Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development of a High-Quality 90K SNP Array and Its Application in Mining Genes Associated with Wheat Processing Quality Traits

Yahui Li
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Hongbo Li
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Mingming Zhang
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Yangna Liu
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Pingping Qu
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Songyu Jia
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Caixia Liu
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Nannan Wei
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Ying Xu
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Guohang Yang
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Liping Zhang
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Hongjie Li
Institute of Biotechnology, Xianghu Laboratory
Lihua Liu
llh216@163.com

Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences
Binshuang Pang
Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences

Research Article

Keywords: wheat, SNP array, genome-wide association study (GWAS), quality traits

Posted Date: June 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9883887/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

In this study, we report the development of a high-quality BAAFS Wheat 90K SNP array for the molecular characterization of wheat cultivars and systematically validated its efficacy in identifying genetic loci associated with wheat processing quality traits. This optimized array integrates high-quality SNP markers from the Illumina Wheat 90K iSelect SNP array and the Axiom® Wheat 660K SNP array, yielding a total of 84,661 loci. Among these loci, 79 are gene-specific functional markers and all loci are precisely anchored to the Chinese Spring (CS) telomere-to-telomere (T2T) reference genome, ensuring high genomic positional accuracy. The 84,661 loci are unevenly distributed across the A (37.9%), B (42.5%), and D (18.0%) subgenomes, with 46.91% of the loci localized within genic regions—an attribute that enhances the array's utility for functional genetic analysis. Genotyping was performed on 190 wheat cultivars and 200 F_2 individuals derived from the cross between Shaanxibaimai and L43, generating high-quality genotypic data with marker call rates of 91.0% and 71.0%, respectively. A total of 1,041 Chinese wheat cultivars representing eight major agro-ecological zones were genotyped, and model-based clustering analysis resolved these cultivars into eight distinct subpopulations, with pairwise fixation index (Fst) values ranging from 0.265 to 0.503. Principal component analysis (PCA) showed that the first three principal components collectively explained 17.07% of the total genetic variance. Additionally, 2,478 recombination hotspots were identified, and these hotspots were predominantly concentrated near chromosomal termini. Genome-wide association studies (GWAS) were conducted by integrating phenotypic data of seven processing quality traits from Value for Cultivation and Use (VCU) trials, including dough stability time, dough development time, maximum resistance, extension area, grain texture, water absorption, and hardness index. A total of 294 significant marker-trait associations (MTAs) were detected, corresponding to 62 QTLs associated with these processing quality traits, and a pleiotropic region on chromosome 1D (408–423 Mb) was identified to regulate multiple dough-quality traits, explaining 7.18–31.56% of the phenotypic variance. Another key locus on chromosome 5D (2.7–7.0 Mb) was associated with grain hardness and texture, accounting for 0.11–16.53% of phenotypic variation. Novel QTLs, distinct from previously reported loci, were identified on chromosomes 3D (associated with dough development time and maximum resistance), 5A (associated with grain texture), and 6D (associated with water absorption). Candidate gene analysis was further performed for these three novel QTLs, and the pleiotropic QTL *QDDT/DST/MR/EA.baaf-3D* was found to explain 12.39–27.08% of the phenotypic variation for dough-related traits, highlighting its potential value for wheat quality improvement. To facilitate efficient marker-assisted selection (MAS) in wheat breeding, four Kompetitive Allele-Specific PCR (KASP) markers were successfully developed for key loci associated with processing quality traits. Collectively, the BAAFS Wheat 90K SNP array provides a cost-effective, high-fidelity genomic tool for wheat genetic research and breeding applications, including germplasm evaluation, genetic mapping, variety identification, precision breeding, MAS, and genomic selection (GS), with particular utility for dissecting the genetic basis of wheat processing quality and accelerating the genetic improvement of critical agronomic traits.

Introduction

Single nucleotide polymorphisms (SNPs) are widely used as molecular markers in crop genetics due to their abundance, genome-wide distribution, genetic stability, low cost, and compatibility with high-throughput detection [1, 2]. SNP arrays have become important tools for gene discovery and breeding in crop species, particularly in bread wheat (*Triticum aestivum* L.), where they have been used to characterize population structure and linkage disequilibrium [3]. These arrays are commonly applied in genome-wide association studies (GWAS) to identify loci associated with yield, disease resistance, and end-use quality [4–6]. In maize (*Zea mays* L.) and rice (*Oryza sativa* L.), SNP arrays have facilitated the mapping of quantitative trait loci (QTL), assessment of genetic diversity, and implementation of genomic selection and germplasm conservation strategies [7–9]. Recent studies in wheat utilizing high-density SNP arrays have identified novel loci associated with drought tolerance [10, 11], disease resistance [12], and agronomic traits [13].

Wheat is a staple cereal crop that supplies nearly 40% of the global population. Its large, repetitive, and allohexaploid genome (2n = 6x = 42, AABBDD) poses challenges for accurate SNP genotyping. Advances in whole-genome sequencing have enabled the development of multiple SNP arrays for wheat, including 9K, 15K, 35K, 50K, 55K, 90K, 660K, and 820K platforms [2, 14]. Among these, the 660K SNP array has been widely used as a high-resolution genotyping platform in wheat research and marker-assisted breeding [15]. Despite its high marker density, the 660K SNP array is not widely used for routine cultivar fingerprinting due to its cost, limited proportion of high-quality polymorphic probes, and lack of trait-diagnostic markers [15]. The 90K SNP array (81,587 probes) applies model-free clustering algorithms to accommodate polyploidy and can detect presence-absence variation (PAV). However, only 46,977 probes (52%) can be genetically mapped, indicating genotyping failure or monomorphism for a substantial fraction of markers, particularly on the D subgenome [16]. The 35K breeding array (35,143 SNPs) is used for fingerprinting hexaploid germplasm and has been applied in QTL mapping, mutation screening, and diversity panels. Its lower marker density and limited performance in East Asian germplasm restrict its application in Chinese breeding programs [17]. Brinton et al. [18] reported that a 15K haplotype-based array improved genomic selection accuracy by 40%, leading to its adoption as a standard platform by the International Wheat Improvement Network (IWIN).

High-density SNP arrays have emerged as powerful and cost-effective genotyping tools for wheat quality genetic dissection, enabling efficient population structure analysis, genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping at a genome-wide scale [2]. Recent advances in customized SNP array development and application have accelerated the identification of stable, pleiotropic QTLs and candidate genes for key quality traits including dough rheology, grain hardness and water absorption, laying a solid foundation for marker-assisted selection in wheat quality improvement [19]. *Pina* and *Pinb* are the primary regulators of grain texture and hardness index, while *TaGli-γ-2.1* influences dough stability time [20, 21]. Specific loci containing candidate genes for dough development time and water absorption have also been detected. The novel QTL *qTaHa-5DL* also impacts grain hardness substantially, alongside candidate genes involved in lipid metabolism [22]. Recent research shows dough rheological traits are controlled by pleiotropic QTL with candidate genes related to gluten synthesis [23].

The limitations of existing arrays indicate a need for a cost-effective and high-quality SNP platform suitable for variety identification, germplasm evaluation, gene mapping, and molecular breeding in wheat. In this study, we selected high-resolution and high-quality SNPs from the Illumina Wheat 90K and Axiom® Wheat 660K arrays using 96 genetically diverse wheat accessions, and supplemented these with 79 gene-specific SNPs. The performance of the resulting array was evaluated based on call rate, reproducibility, and mapping resolution.

The objectives of this study were to threefold: (1) to develop an integrated wheat SNP array, designated as the BAAFS (Beijing Academy of Agriculture and Forestry Sciences) Wheat 90K SNP array, through the consolidation of markers from the Illumina Wheat 90K and Axiom® Wheat 660K arrays, with priorities given to single-copy SNPs, and the incorporation of functional markers; (2) to systematically evaluate the genotyping quality and overall performance of the developed BAAFS Wheat 90K SNP array; and (3) to apply the array in a genome-wide association study (GWAS) using a large wheat population, aiming to identify genetic loci associated with grain processing quality traits.

Materials and methods

Plant materials

SNP loci filtering: A diversity panel of 96 genetically diverse wheat accessions, including 75 elite cultivars from major Chinese agro-ecological zones, 3 landraces, 1 hybrid wheat and its parental lines, 10 advanced breeding lines, 1 doubled haploid (DH) line, and replicated internal reference cultivars (Yangfumai 2, Linhan 6, Zeyou 1, Chinese Spring) (Supplementary table 1).

Genotyping performance evaluation: 190 Chinese core commercial cultivars (including Chinese Spring and a DH line) (Supplementary table 2) and 200 F_2 plants derived from the Shaanxibaimai × L43 cross [24].

GWAS and population genetics: 1,041 Chinese wheat cultivars representing eight agro-ecological zones of China: northern Huang-Huai winter wheat region (234), southern Huang-Huai winter wheat region (394), northwestern spring wheat region (57), northern spring wheat region (9), northern winter wheat region (63), northeastern spring wheat region (32), middle-lower Yangtze River Valley wheat region (85), and upper Yangtze River Valley wheat region (167) (Supplementary table 3).

Genomic DNA extraction

Genomic DNA was extracted from pooled leaf tissue (50 seedlings per accession) at the three-leaf stage using a modified cetyltrimethylammonium bromide (CTAB) method [25]. DNA quality and concentration were determined using a NanoDrop 2000 UV spectrophotometer (Thermo Scientific, Waltham, MA, USA). DNA was normalized to a working concentration of 100 ng/μL with sterile double-distilled water for subsequent genotyping.

Screening of high-confidence SNP loci

The 96 wheat accessions were genotyped using both the Illumina Wheat 90K iSelect SNP array (81,587 SNPs) [16] and the Axiom® Wheat 660K SNP array (630,517 SNPs) [15, 26]. Stringent quality control was applied to raw genotyping data for both platforms:

Illumina 90K array: SNPs with GenTrain scores < 0.60 in Genome Studio were excluded [27]. Additional filters removed loci with non-reproducible calls, Mendelian inheritance deviation, low signal intensity, missing data rate > 0.05, heterozygosity rate > 0.1, and minor allele frequency (MAF) < 0.2. Only loci with clear clustering of AA, AB, and BB genotypes were retained.

Axiom® 660K array: Only Poly-High-Resolution (PHR) polymorphic SNPs were retained using Axiom Analysis Suite v3.1 (Thermo Fisher Scientific Affymetrix) with the same quality thresholds (missing data rate > 0.05, heterozygosity rate > 0.1, MAF < 0.2) as the Illumina 90K array.

For overlapping genomic positions between the two arrays, SNPs from the 660K array were retained by default to maximize marker quality, and singleton SNPs from both arrays were preserved. An additional 79 gene-specific functional markers (associated with plant height, yield, quality, and stress resistance) from published studies were incorporated (Table S4). All retained SNPs were anchored to the Chinese Spring (CS) 1.0 and T2T reference genome for physical position annotation.

Development and evaluation of the BAAFS Wheat 90K SNP array

The candidate SNP set (108,596 probes) was submitted to the Affymetrix Axiom myDesign pipeline for probe design and evaluation. Assessment parameters included duplicate count, P-convert scores, poly counts, repetitive sequences, and wobbles. Only recommended and neutral SNPs with a P-convert value ≥ 0.4 were retained to ensure genotyping reliability. The final BAAFS Wheat 90K SNP array comprises 84,661 high-quality SNPs, with all physical positions annotated based on the CS T2T reference genome.

Genotyping of the 190 commercial cultivars and 200 F_2 plants was performed on the Affymetrix GeneTitan platform (https://securefileexchange.affymetrix.com). SNPs were classified into six categories: Poly High Resolution (PHR), No Minor Homozygote (NMH), Off-Target Variant (OTV), Mono High-Resolution (MHR), Call-Rate Below Threshold (CRBT), and Unresolved. Genotyping performance was evaluated by call rate, reproducibility (intra-plate and inter-batch), and the proportion of high-quality markers (PHR + NMH + OTV).

Population structure, PCA, kinship and LD analysis

Genotyping data of the 1,041 wheat cultivars were filtered using PLINK v1.9 [28] to exclude SNPs with missing data rate > 0.02, MAF < 0.05, and samples with missing data rate > 0.02. Remaining missing calls were imputed using Beagle v5.4 [29].

Population structure: Analyzed using the sparse non-negative matrix factorization (sNMF) algorithm in the LEA package v3.4.0 [30] with K = 1–10. The optimal K value was determined via the ΔK method [31] and biological interpretability. Principal Component Analysis (PCA): Performed using GAPIT v3.0 [32] to assess genetic stratification; the first three principal components were visualized. Kinship analysis: A genomic relationship matrix (GRM) was constructed using GAPIT v3.0 to quantify genetic relatedness among cultivars.

LD and recombination hotspots: LD analysis was performed in PLINK v1.9 with a 100 Mb sliding window (--ld-window 100,000 --ld-window-kb 100,000 --ld-window-r2 0 --r2-phased). Recombination hotspots were defined as genomic regions with pairwise LD (r^2) < 0.2 [33].

Phenotypic data collection and preprocessing

Phenotypic data for seven wheat processing quality traits were obtained from national VCU trials on the China Seed Industry Big Data Platform (http://202.127.42.232/SDSite/Home/Index?ticket=): dough stability time (DST), dough development time (DDT), maximum resistance (MR), extension area (EA), grain texture (GT), water absorption (WA), and hardness index (HI). Phenotypic data were preprocessed by removing outliers (values beyond ± 3 standard deviations from the mean) [34] to reduce environmental noise.

Genome-wide association study (GWAS)

GWAS was performed using the Mixed Linear Model (MLM) in GAPIT v3.2.1 [32, 35], which incorporates the sNMF-derived population structure matrix (Q) and GAPIT-calculated kinship matrix (K) to control for false positives due to population stratification and genetic relatedness. The Fixed and Random Model Circulating Probability Unification (FarmCPU) model was compared with MLM, and MLM was selected for its better fit for the study dataset.

A strict significance threshold of $-\log_{10}(P) = 3.5$ was determined via 1000 permutation tests (permuting phenotypes and genotypes) to balance false positive control and detection of true MTAs. Only single-copy, high-quality SNPs (MAF > 0.05, missing rate < 0.02) were used for GWAS to minimize interference from paralogous loci in the polyploid wheat genome. Significant MTAs were annotated to the CS 1.0 reference genome, and QTLs were defined as genomic regions containing multiple significant MTAs for the same trait (or multiple traits for pleiotropic QTLs). Phenotypic variance explained (PVE) by each QTL was calculated from GWAS results.

Development of KASP markers

Four key significant SNPs from pleiotropic quality QTLs (Affx-109073624_1D, Affx-88667283_1D, Affx-109105158_1D, Affx-110900326_3D) were selected for KASP marker development. Primers were designed using the PolyMarker online website (http://polymarker.tgac.ac.uk.) based on the flanking

sequences (500 bp upstream and downstream) of the target SNPs from the CS 1.0 reference genome, with allele-specific primers containing FAM/HEX fluorescent tags and a common reverse primer.

The KASP reaction system (10 μL) contained 1.2 μL genomic DNA (100 ng/μL), 5.0 μL 2× KASP master mix, 1.0 μL primer mix (12 μM each allele-specific primer, 30 μM common reverse primer), and 2.8 μL ddH_2O. Amplification was performed on a Bio-Rad CFX system (Bio-Rad Laboratories, Hercules, CA, USA) with the following program: initial activation at 94°C for 15 min; 10 touchdown cycles of 94°C for 20 s and 60°C (0.6°C decrease per cycle) for 60 s; 35 cycles of 94°C for 20 s and 55°C for 60 s; final hold at 25°C for 1 min. Fluorescence signals were read using Bio-Rad CFX Manager software, and genotyping clusters were manually validated. The KASP markers were tested on 94 bread wheat cultivars to verify their ability to distinguish superior and inferior alleles for quality traits.

Statistical analysis

All statistical analyses (including descriptive statistics, ANOVA, and correlation analysis of phenotypic traits) were performed in R v4.3.1 (https://www.r-project.org/) with set.seed (123) to ensure reproducibility. Manhattan plots and quantile-quantile (Q-Q) plots for GWAS were generated using the qqman package in R. Phenotypic differences between superior and inferior allele carriers were tested via Student's t-test ($P < 0.05$ and $P < 0.01$).

Gene Annotation and Haplotype Analysis

Gene annotation information for the target intervals was extracted from the Triticeae MultiOmics Center (http://202.194.139.32) via the IntervalTool. Haplotype analysis and corresponding visualization of target genes were performed using Geneious Prime 2026.0.2 software, with the gene sequences of Chinese Spring (CS) serving as the reference [36]. The gene sequences used for analysis were obtained from 17 wheat accessions, including Chuanmai 42, Zhengmai 366, Aimengniu, Beijing 8, Jimai 22, Jinmai 47, Handan 6172, Kuofan 11, Mazhamai, Ningchun 4, Shi 4185, Xiaoyan 6, Xinong 6028, Yangmai 158, Zhoumai 16, Zhoumai 22, and Renan [37, 38].

Results

Screening SNP loci from the Illumina 90K and Axiom® 660K arrays

For the Illumina 90K iSelect SNP array, 49,539 initial markers (60.7% of total probes) passed the GenTrain score ≥ 0.60 threshold. A sequential quality control procedure was then applied. Markers were excluded if they showed non-reproducible genotyping (Fig. 1a), deviated from Mendelian segregation (Fig. 1b), or produced inconsistent calls in homozygous doubled haploid (DH) lines (Yuan177, Yuan119, Yuan56) (Fig. 1c). Additional filtering removed loci with missing data > 0.05 (Fig. 1d), low signal intensity (Fig. 1e), heterozygosity > 0.1 (Fig. 1f), and non-polymorphic variants based on minor allele frequency (MAF) (Fig. 1g, 1h). Markers lacking clear AA/AB/BB clustering (Fig. 1i, 1j) and those with AATmean > 0.2

or BBTmean < 0.8 (Fig. 1k) were also discarded. Finally, markers with MAF < 0.2 were removed (Fig. 1l–1n). This filtering process resulted in 456 high-quality, single-copy SNPs (Fig. 1o), which were evenly distributed across the A (155), B (185), and D (116) subgenomes based on the IWGSC RefSeq v1.0 reference genome.

For the Axiom® Wheat 660K array, 222,346 Poly High Resolution (PHR) markers, representing 35.3% of the total probes, were retained using Axiom Analysis Suite v3.1. Additional filtering was applied to exclude non-reproducible loci (Fig. 2a), markers showing Mendelian deviation (Fig. 2b), those with missing data > 0.05 (Fig. 2c), low signal intensity (Fig. 2d), heterozygosity > 0.1 (Fig. 2e), and minor allele frequency (MAF) < 0.2. This process resulted in 220,916 high-confidence SNPs. Further refinement to achieve uniform genome coverage yielded 103,368 markers (Fig. 2f), which were anchored to the IWGSC RefSeq v1.0 reference genome.

Customization and genomic features of the BAAFS wheat 90K SNP array

The BAAFS Wheat 90K SNP array was initially constructed with 103,903 probes. These included 456 SNPs from the Illumina 90K array, 103,368 SNPs from the Axiom® 660K array, and 79 gene-specific functional markers (Supplementary table 4). After sequence evaluation using the Affymetrix Axiom myDesign pipeline, the final array comprised 84,661 high-quality SNPs. Among these, 346 were derived from the Illumina 90K array, 84,236 from the Axiom® 660K array, and 79 were gene-specific markers. All markers were physically mapped to the Chinese Spring v1.0 and T2T reference genomes (Supplementary table 5).

Genome-wide distribution analysis using 1 Mb windows showed that SNP density was highest in distal telomeric regions and lowest in heterochromatic regions (Fig. 3a). The SNPs were distributed across the A (37.90%), B (42.50%), and D (18.00%) subgenomes (Fig. 3b). Chromosome 3B (829 Mb) had the highest number of SNPs (7,490), while chromosome 4D contained the fewest (1,070 SNPs). Chromosome 6D was the shortest, with a length of 473 Mb. The genome-wide average inter-marker distance was 0.20 Mb, ranging from 0.11 Mb on chromosome 3B to 0.48 Mb on chromosome 4D (Supplementary table 6). Functional annotation showed that 39,718 SNPs (46.91%) were located within genic regions. Among these, 69.98% were in promoter regions, 13.11% in introns, 5.63% in 3' untranslated regions (UTRs), 0.98% in 5' UTRs, and 0.04% in exons. The 79 gene-specific markers were associated with traits such as plant height, thousand-kernel weight, disease resistance, and quality-related processes (Table S4). In terms of SNP variant types, 60,234 were transitions (71.15%), 24,425 were transversions (28.85%), and two were deletions. The most common transition subtypes were A/G (35.21%) and T/C (35.93%) (Fig. 3c).

Evaluation of the BAAFS wheat 90K SNP array

Genotyping of 190 wheat cultivars and 200 F_2 individuals from the Shaanxibaimai × L43 cross produced average call rates of 99.62% and 99.43%, respectively. Intra-plate reproducibility was 99.50%, and inter-batch reproducibility was 99.60%. Clustering analysis using Affymetrix software classified SNPs into six quality categories (Supplemental Fig. 1). In the cultivar panel, the top three categories—poly high resolution (PHR), no minor homozygote (NMH), and off-target variant (OTV)—accounted for 91.0% of the array, with PHR alone comprising 83.0%. In the F_2 population, these three categories represented 71.0% of the array, including 43.0% PHR (Supplemental Fig. 2a, b). The minor allele frequency (MAF) distribution of the array, which was mainly concentrated at 0.30–0.50 (Supplementary Fig. 2c). These results indicate that the array enables high-quality and reproducible genotyping in both natural and segregating wheat populations.

Genetic population structure and kinship analysis

After filtering 84,661 raw SNPs from 1,041 wheat cultivars using PLINK v1.9 (--geno 0.02, --mind 0.02, --maf 0.05), 79,504 high-confidence SNPs (93.9%) were retained for downstream analysis. Principal component analysis (PCA) indicated that the first two principal components accounted for 12.4% of the total genetic variation, with PC1 and PC2 explaining 6.5% and 5.9% respectively. All accessions formed loose genetic clusters consistent with their corresponding ecological distribution zones (Fig. 4a). The three subgenomes showed distinct LD decay patterns, with the D subgenome decaying the fastest (at 4.2 Mb), followed by the A subgenome (at 9.9 Mb), while the B subgenome had the slowest LD decay (at 27.0 Mb) (Fig. 4b).

ADMIXTURE v1.3.0 population structure analysis at K = 8 obtained a five-fold cross-validation error value of 0.895, and clearly classified all accessions into eight distinct subpopulations consistent with their geographical ecological origins (Fig. 4c). The pairwise Fst values across subpopulations varied from 0.265 between Pop4 and Pop5 to 0.503 between Pop1 and Pop4, suggesting prominent genetic divergence among different subgroups.

Kinship analysis was conducted via GAPIT v3.0 based on the mixed linear model (MLM). The kinship coefficients presented obvious hierarchical stratification (Supplemental Fig. 3a). Specifically, pairwise kinship values accounted for 0.10% at zero, 15.60% within 0–0.10, 73.72% within 0.10–0.30, 8.95% within 0.30–0.50, 1.32% within 0.50–0.70, and 0.31% within 0.70–1.00.

Genome-Wide Identification of Recombination Hotspots

A total of 2,478 recombination hotspots were identified across the 1,041 wheat cultivars (Supplemental Fig. 3b). Chromosome 5D contained the highest number of hotspots (276), whereas chromosome 6A had the fewest (46). Hotspots were most abundant in the D subgenome and in telomeric regions of the A and B subgenomes. These distribution patterns reflect subgenome-specific recombination characteristics associated with wheat domestication and artificial selection.

GWAS of quality-related traits

Phenotypic data for seven processing quality traits—extension area (EA), maximum resistance (MR), dough stability time (DST), dough development time (DDT), water absorption (WA), grain hardness index (HI), and grain texture (GT)—were obtained from national Value for Cultivation and Use (VCU) trials. A total of 294 significant marker-trait associations (MTAs) were identified, corresponding to 62 quantitative trait loci (QTLs). These QTLs were distributed across all chromosomes except 2A and 7B (Fig. 5, 6; Table 1).

GWAS identified 11 QTLs for DDT. These included two SNPs on chromosome 1A (PVE 14.11%–14.16%), one SNP on 1B (PVE 12.4%), 121 SNPs in the 408.1–416.5 Mb interval on 1D (*QDDT.baaf-1D*, PVE 10.95%–16.98%), 38 SNPs on 2D (PVE 7.28%–11.82%), and one SNP at 107.1 Mb on 3D (PVE 16.19%). Additional QTLs with smaller effects were detected on chromosomes 4B, 4D, 5A, 5B, and 6D, explaining 4.28%–10.82% of the phenotypic variance. For DST, five QTLs were mapped on chromosomes 1A, 1B, 1D, 3B, and 3D. The largest effects were observed at *QDST.baaf-1D* (408.1–423.1 Mb, 123 SNPs, $P = 7.39 \times 10^{-17}$, PVE 7.18%–16.62%) and *QDST.baaf-1B* (558.5 Mb, $P = 2.58 \times 10^{-14}$, PVE 15.49%). A minor locus on chromosome 3B, comprising two SNPs, explained 3.98%–4.17% of the variance. Six QTLs for MR were detected on chromosomes 1B, 1D, and 3D. The three largest effects were at *QMR.baaf-1B* (558.5 Mb, PVE 29.82%, $P = 7.25 \times 10^{-14}$), *QMR.baaf-1D* (408.1–416.5 Mb, 122 SNPs, PVE 17.13%–31.56%, $P = 3.42 \times 10^{-15}$), and *QMR.baaf-3D* (107.1 Mb, PVE 27.08%). For EA, four QTLs were located on chromosomes 1A, 1B, 1D, and 3D, with PVE values ranging from 14.32% to 18.71%. The major locus, *QEA.baaf-1D* (408.1–416.5 Mb, 119 SNPs), co-localized with the DDT, DST, and MR QTLs identified on chromosome 1D.

For GT, 13 QTLs were identified. The most significant locus, *QGT.baaf-5D.1*, was located at 2.7–13.2 Mb ($P = 1.30 \times 10^{-22}$) and explained 5.23%–10.46% of the phenotypic variance. Three additional loci on chromosome 5A (5A.1, 5A.2, and 5A.3) together accounted for 5.92%–7.90% of the variance. For HI, 10 QTLs were detected. Among them, *QHI.baaf-5D* (3.6–7.0 Mb) explained 0.91%–16.53% of the phenotypic variance. For WA, 12 QTLs were identified across the genome. *QWA.baaf-5D* (2.7–7.0 Mb), comprising five SNPs, explained 0.11%–2.21% of the variance, while *QWA.baaf-6D* (89.2–95.6 Mb), comprising ten SNPs, explained 3.64%–4.53%.

Pleiotropic and stable loci for quality traits

Four pleiotropic QTLs regulating EA, MR, DST, and DDT were identified on chromosomes 1A, 1B, 1D, and 3D. These include *QEA/MR/DST/DDT.baaf-1A* (505.38–506.47 Mb, PVE 3.42–30.94%), *QEA/MR/DST/DDT.baaf-1B* (~ 558.52 Mb, PVE 12.40–29.82%), *QEA/MR/DST/DDT.baaf-1D* (408.08–423.08 Mb, PVE 7.18–31.56%), and *QEA/MR/DST/DDT.baaf-3D* (~ 107.07 Mb, PVE 12.39–27.08%) (Table 1; Supplementary Fig. 4). In addition, a pleiotropic region on chromosome 5D (2.68–7.01 Mb) was found to co-localize with *QWA.baaf-5D*, *QHI.baaf-5D*, and *QGT.baaf-5D.1*, explaining 0.11–16.53% of the phenotypic variance for WA, HI, and GT (Table 1).

The QTL interval *QEA/MR/DST/DDT.baaf-1D* (408.08–423.08 Mb, PVE = 7.18–31.56%) overlapped with the physical position of *Glu-1D*. Accordingly, we further analyzed the haplotypes of *Glu-1D* among 17 wheat accessions carrying this gene. Chuanmai 42 and Zhengmai 366 belonged to Hap1; Hap2

contained 14 accessions including Jimai 22, Beijing 8 and Zhoumai 16; while Renan was classified into Hap3 (Supplementary Fig. 5).

Effects of superior alleles on investigated traits

Evaluation of four significant SNPs (Affx-111917070_1A, Affx-110903909_1B, Affx-88412945_1D, and Affx-110900326_3D) showed that wheat varieties carrying superior alleles had improved phenotypic performance for EA, MR, DST, and DDT compared to those with inferior alleles ($P < 0.05$ or $P < 0.01$) (Supplementary Fig. 6). Trait performance increased with the number of superior alleles. DST improved in varieties with at least one superior allele, DDT with five or more alleles, EA with three to six alleles, MR with one to eight alleles, WA with one to 14 alleles, HI with two to nine alleles, and GT with five to nine alleles (Fig. 7).

Development of KASP markers

Six significant SNPs were screened from the BAAFS 90K SNP array, namely Affx-111917070_1A, Affx-110903909_1B, Affx-109073624_1D, Affx-88667283_1D, Affx-109105158_1D and Affx-110900326_3D. Among them, four SNPs were successfully developed into KASP functional markers (Supplementary Table 7). The four developed KASP markers including *KASP_Affx-109073624_1D*, *KASP_Affx-88667283_1D*, *KASP_Affx-109105158_1D* and *KASP_Affx-110900326_3D* were further genotyped in 94 common wheat cultivars. These four markers could effectively differentiate favorable and unfavorable allelic variations (Supplementary Fig. 7).

Candidate Gene Analysis of Novel QTLs

We performed gene annotation within the intervals of the novel QTLs on chromosomes 3D, 5A and 6D using the Chinese Spring 1.0 reference genome (Supplementary table 8). The QTL interval *QDDT/DST/MR/EA.baaf-3D* is located at 104.70–109.55 Mb on chromosome 3D and contains 30 high-confidence genes. Among them, cytokinin oxidase/dehydrogenase genes (*TraesCS3D01G143200*, *TraesCS3D01G143300*, *TraesCS3D01G143500*, and *TraesCS3D01G143600*) and a receptor-like kinase gene (*TraesCS3D01G145100*) are the most promising candidates regulating wheat dough processing quality. Two QTL intervals on chromosome 5A (*QGT.baaf-5A.2* and *QGT.baaf-5A.3*) contained 37 high-confidence genes. Major candidates included a jumonji transcription factor family protein (*TraesCS5A01G258800*), a protein kinase family protein (*TraesCS5A01G261200*), and heat shock proteins (*TraesCS5A01G260600* and *TraesCS5A01G260900*), collectively explaining the phenotypic variation of grain physical quality. The QTL interval *QWA.baaf-6D* on chromosome 6D contained 47 high-confidence genes and was significantly associated with water absorption. Functional candidates in this region encoded peroxidase (*TraesCS6D01G125500* and *TraesCS6D01G127100*), glutathione S-transferase (*TraesCS6D01G127700* and *TraesCS6D01G127800*), and auxin response factor (*TraesCS6D01G127600*), playing vital roles in wheat processing quality.

Discussion

The BAAFS Wheat 90K SNP array was constructed to overcome major obstacles in hexaploid wheat genotyping. This array incorporates high-quality SNP markers derived from the Illumina 90K and Axiom® 660K SNP arrays, alongside 79 gene-derived functional markers. All incorporated markers were anchored to the Chinese Spring T2T reference genome for physical localization [39]. Single-copy SNPs were preferentially selected during array development, yielding a polymorphic hit rate (PHR) of 83% among wheat cultivars. In accordance with standard genotyping protocols for polyploid species [40–43], the present design diminished paralogous sequence interference, a primary confounding factor in wheat genotyping, by over 40% relative to multi-copy probe panels. Such rigorous marker screening ensures high genotyping reliability and facilitates subsequent genetic dissection and population genetic analyses.

Stringent multi-step screening procedures were implemented to optimize the performance of this SNP array. A total of 456 high-quality SNPs were ultimately retained from the original Illumina 90K panel after rigorous evaluation of genotyping repeatability, Mendelian segregation conformity and genetic polymorphism stability. This systematic screening effectively eliminates false genotyping signals induced by high sequence homology among homoeologous chromosomes [44], and further enables accurate DNA fingerprinting and efficient cultivar identification. The Affymetrix Axiom® platform was adopted owing to its high-throughput capacity supporting 96-plex and 1 million parallel reactions as well as full operational automation. Compared with the Illumina Infinium® system, this platform substantially cuts down experimental expenditure and manual workload, which is highly applicable to large-scale wheat breeding practices [2, 17].

Compared with the Axiom® 660K SNP array, the BAAFS 90K array possesses multiple superior characteristics. It harbors a larger proportion of genic SNPs (46.91% versus 37.1%), which is conducive to linking molecular markers with agronomically important functional traits. Meanwhile, this array greatly reduces the genotyping cost per sample, making it suitable for routine large-scale genetic analysis. Additionally, it exhibits a higher proportion of high-quality polymorphic loci, with a polymorphic hit rate (PHR) of 83.0% in recombinant inbred lines, distinctly higher than the 29.8% of the 660K array [15, 26]. Its genotyping call rate exceeds 99.43% and technical reproducibility reaches over 99.50%, and 86.15% of SNPs have a minor allele frequency (MAF) greater than 0.2. All these favorable indicators enable this array to efficiently distinguish near-isogenic lines. Furthermore, this array has been officially adopted in the agricultural industry standard NY/T 4021 – 2021 issued by the Ministry of Agriculture and Rural Affairs of China for wheat variety identification [45].

The BAAFS 90K SNP array achieves an optimal balance between application performance and economic cost compared with conventional genotyping approaches. Although whole-genome resequencing delivers high-resolution genetic data, its excessive cost restricts large-scale population analysis [44]. Genotyping-by-sequencing (GBS) is frequently hampered by poor genotyping repeatability [17]. Exome sequencing fails to detect non-coding regulatory variations across the genome [41]. While KASP markers are efficient, they are only applicable to specific target loci [43]. By comparison, the BAAFS 90K array ensures even genome-wide marker coverage, shorter detection cycles relative to sequencing

technologies, and abundant functionally relevant markers. These prominent advantages render it well-suited for extensive application in wheat molecular breeding [1, 2].

Based on the BAAFS 90K SNP array, we systematically characterized genome-wide recombination hotspots across 1041 wheat accessions, and evident subgenome-level divergence was observed in their spatial distribution. Chromosome 5D harbored the largest number of recombination hotspots (276), while chromosome 6A possessed merely 46 such regions. Such uneven distribution patterns imply the vital evolutionary function of the D subgenome during wheat domestication, and its frequent recombination events are conducive to the preservation of population genetic diversity [46]. The uneven distribution of recombination hotspots on the D subgenome is closely associated with chromatin accessibility and transposable element abundance [47]. Recombination is severely inhibited within pericentromeric regions, which facilitates the formation of large conserved haplotype blocks, including those linking disease resistance and yield-related traits with typical genetic trade-off effects [48]. Previous QTL mapping studies have confirmed that genetic loci on chromosomes 3B and 6AL are capable of modulating genomic recombination frequency [49]. The elevated recombination activity on the D subgenome is presumed to be shaped by long-term artificial domestication selection [46]. These discovered recombination features provide theoretical guidance for haplotype-assisted molecular breeding. For instance, the intensive recombination on chromosome 5D can be utilized to efficiently break unfavorable genetic linkage drag [50].

Genome-wide association study enabled the detection of numerous QTLs governing wheat quality traits, several of which were consistent with previously documented genetic loci. The pleiotropic QTL *QEA/MR/DST/DDT.baaf-1D* located at 408.1–423.1 Mb on chromosome 1D exhibited physical collocation with the *Glu-D1* gene that modulates dough strength [51, 52]. This stable locus was also captured in the NAM population by Hu et al. [53], and further defined as the pleiotropic QTL cluster *qDt-KJ-1D* in the research of Ding et al. [54]. Moreover, the pleiotropic SNP AX-111015470 on chromosome 1A identified from wheat accessions in Hebei Province by Zhao et al. [55] shared identical physical position with *QEA/MR/DST/DDT.baaf-1A* (505.38–506.47 Mb) obtained in the present study. Both genetic regions exerted simultaneous effects on dough stability time, extension area and maximum resistance. Collectively, these results further confirm that chromosomes 1A and 1D serve as core genomic regions regulating wheat processing quality.

Three closely clustered QTLs on chromosome 5D, namely *QWA.baaf-5D*, *QHI.baaf-5D* and *QGT.baaf-5D.1*, were detected within the physical interval of 2.68–7.01 Mb, which partially overlapped with the previously identified locus *qDt-KJ-5D* [54]. Notably, the present study firstly revealed that this genomic region exerts pleiotropic effects on grain water absorption, hardness index and grain texture traits. Although the quality-related QTL *QFN.sau-4D* identified via the 55K SNP array by Li et al. [19] was not located within this target interval, their findings regarding the co-localization of multiple quality QTLs such as *QGPC.sau-4A.1* and *QWGC.sau-4A* further verified that functional genes controlling wheat quality traits tend to form clustered distributions across the genome. In addition, the phenotypic variation

explained (PVE) values of the 5D locus for hardness index ranged from 0.91% to 16.53%, which was highly consistent with the genetic effect magnitudes reported in Li et al. [19].

Apart from verifying previously documented quantitative trait loci, the present study also excavated multiple novel genetic loci that have not been reported to regulate wheat quality traits. A novel pleiotropic locus *QDDT/DST/MR/EA.baaf-3D* was detected at 107.1 Mb on chromosome 3D, which simultaneously controlled dough development time (PVE = 16.19%), dough stability time (PVE = 12.39%), extension area (PVE = 16.58%) and maximum resistance (PVE = 27.08%). Although several drought tolerance-related QTLs have been mapped on chromosome 3D in earlier studies, none of them share overlapping physical intervals with this identified locus [56]. Several genetic regions designated 5A.1–5A.3 on chromosome 5A were specifically responsible for grain texture, with the corresponding phenotypic variation explained ranging from 5.92% to 7.90%. Furthermore, the locus *QWA.baaf-6D* located within 89.2–95.6 Mb on chromosome 6D specifically modulated grain water absorption, with PVE values ranging from 3.64% to 4.53%. Earlier studies have reported functional loci on chromosome 6D mainly associated with seed vigor and grain weight, whereas no locus governing water absorption has been documented so far. In addition, these reported loci are physically distinct from *QWA.baaf-6D*, strongly indicating that *QWA.baaf-6D* is a newly identified candidate locus for grain water absorption [57]. Within these novel QTL intervals, a total of 114 high-confidence genes were annotated through bioinformatic analysis, which provides a foundation for further exploring the molecular mechanisms underlying the regulation of wheat quality traits by these loci.

In conclusion, the BAAFS Wheat 90K SNP array constitutes an efficient and cost-saving genotyping platform, which effectively overcomes multiple technical deficiencies of conventional genotyping methods in hexaploid wheat research. Integrating superior marker quality, abundant functionally relevant loci and low operational cost, this array is well-suited for systematic research and molecular genetic improvement of wheat quality traits. The novel pleiotropic QTLs newly discovered in this study, along with the successfully verified KASP functional markers, supply abundant feasible molecular resources for practical wheat breeding. These valuable genetic materials will facilitate the breeding and cultivation of high-quality improved wheat varieties. Given the growing demand for precise and high-efficiency breeding strategies, this SNP array will serve as a reliable and powerful tool for wheat genetic dissection and targeted trait improvement in future research.

Abbreviations

SNP Single nucleotide polymorphisms

GWAS Genome-wide association studies

QTL Quantitative trait loci

PAV Presence-absence variation

DH Doubled haploid

CTAB Cetyltrimethylammonium bromide

MAF Minor allele frequency

PHR Poly-High-Resolution

NMH No Minor Homozygote

OTV Off-Target Variant

MHR Mono High-Resolution

CRBT Call-Rate Below Threshold

sNMF Sparse non-negative matrix factorization

PCA Principal Component Analysis

GRM Genomic relationship matrix

LD Linkage disequilibrium

VCU Value for Cultivation and Use

DST Dough stability time

DDT Dough development time

MR Maximum resistance

EA Extension area

GT Grain texture

WA Water absorption

HI Hardness index

MLM Mixed Linear Model

FarmCPU Fixed and random model Circulating Probability Unification

MTA Marker-trait association

PVE Phenotypic variance explained

KASP Kompetitive Allele Specific PCR

Fst Fixation index

GBS Genotyping-by-sequencing

UTR Untranslated regions

BAAFS Beijing Academy of Agriculture and Forestry Sciences

CS Chinese Spring

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

The authors gratefully acknowledge the financial support from the Science and Technology Innovation Project of BAAFS, China (KJCX20261405), the Beijing Rural Revitalization Agricultural Science and Technology Project (NY2401050025), the Science and Technology Innovation Project of BAAFS, China (KJCX20261407), and the Reform and Development Project of the Hybrid Wheat Institute, BAAFS (XMZZKT202608).

Author Contribution

Binshuang Pang, Lihua Liu, and Yahui Li conceived and designed the study. Yahui Li conducted the experiments and co-wrote the manuscript with Hongbo Li and Yangna Liu. Songyu Jia, Caixia Liu, Nannan Wei, Ying Xu and Hongjie Li participated in partial experimental execution. Mingming Zhang, Pingping Qu, Guohang Yang and Liping Zhang provided bioinformatics and statistical analysis support. All authors critically reviewed and approved the final manuscript.

Acknowledgement

Thanks to the Institute of Hybrid Wheat, Beijing Academy of Agriculture and Forestry Sciences, Key Laboratory of Crop DNA Fingerprinting Innovation and Utilization of the Ministry of Agriculture and Rural Affairs (Co-construction by Ministry and Province), Beijing Key Laboratory of Crop Molecular Design and Intelligent Breeding, and Beijing Key Laboratory of Molecular Genetics in Hybrid Wheat, Beijing 100097, China, for providing experimental conditions and research support. We also appreciate the Institute of Biotechnology, Xianghu Laboratory, Hangzhou, China, for offering valuable technical assistance throughout this study.

Data Availability

All original content is available in the article and supplementary files. Please contact the corresponding authors for inquiries.

References

1. Winfield MO, Allen AM, Burridge AJ, Barker GLA, Benbow HR, Wilkinson PA, et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol J. 2016;14:1195–206.
2. Burridge AJ, Winfield M, Przewieslik-Allen A, Edwards KJ, Siddique I, Barral-Arca R, et al. Development of a next generation SNP genotyping array for wheat. Plant Biotechnol J. 2024;22:2235–47.
3. Scheben A, Verpaalen B, Lawley CT, Chan CKK, Bayer PE, Batley J, et al. CropSNPdb: a database of SNP array data for Brassica crops and hexaploid bread wheat. Plant J. 2019;98:142–52.
4. Shvachko N, Solovyeva M, Rozanova I, Kibkalo I, Kolesova M, Brykova A, et al. Mining of QTLs for spring bread wheat spike productivity by comparing spring wheat cultivars released in different decades of the last century. Plants. 2024;13:1081.
5. Nandi S, Varotariya K, Luhana S, Kyada AD, Saha A, Roy N, et al. GWAS for identification of genomic regions and candidate genes in vegetable crops. Funct Integr Genomic. 2024;24:203.
6. Wang Q, Tang JL, Han B, Huang XH. Advances in genome-wide association studies of complex traits in rice. Theor Appl Genet. 2020;133:1415–25.
7. Wang NS, Chen HG, Qian YZ, Liang ZJ, Zheng GQ, Xiang J, et al. Genome-wide association study of rice grain shape and chalkiness in a worldwide collection of Xian accessions. Plants. 2023;12:419.
8. Bhadmus OA, Badu-Apraku B, Adeyemo OA, Agre PA, Queen ON, Ogunkanmi AL. Genome-wide association analysis reveals genetic architecture and candidate genes associated with grain yield and other traits under low soil nitrogen in early-maturing white quality protein maize inbred lines. Genes. 2022;13:826.

9. Ponce K, Zhang Y, Guo LB, Leng YJ, Ye GY. Genome-wide association study of grain size traits in *indica* rice multiparent advanced generation intercross (MAGIC) population. Front. Plant Sci. 2020;11:395.
10. Yadava YK, Chaudhary P, Yadav S, Rizvi AH, Kumar T, Srivastava R, Soren KR, et al. Genetic mapping of quantitative trait loci associated with drought tolerance in chickpea (*Cicer arietinum* L). Sci Rep. 2023;13:17623.
11. Wu L, Chang YJ, Wang LF, Wang SM, Wu J. The aquaporin gene *PvXIP1*; 2 conferring drought resistance identified by GWAS at seedling stage in common bean. Theor Appl Genet. 2022;135:485–500.
12. Wu JH, Yu R, Wang HY, Zhou CE, Huang S, Jiao HX, et al. A large-scale genomic association analysis identifies the candidate causal genes conferring stripe rust resistance under multiple field environments. Plant Biotechnol J. 2021;19:177–91.
13. Zhao J, Sun LJ, Gao HM, Hu MY, Mu LM, Cheng XH, et al. Genome-wide association study of yield-related traits in common wheat (*Triticum aestivum* L.) under normal and drought treatment conditions. Front. Plant Sci. 2023;13:1098560.
14. Li HJ, Zhou Y, Xin WL, Wei YQ, Zhang JL, Guo LL. Wheat breeding in northern China: Achievements and technical advances. Crop J. 2019;7:718–29.
15. Sun CW, Dong ZD, Zhao L, Ren Y, Zhang N, Chen F. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol J. 2020;18:1354–60.
16. Wang SC, Wong DB, Forrest K, Allen A, Chao SM, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J. 2014;12:787–96.
17. Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GLA, et al. Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). Plant Biotechnol J. 2017;15:390–401.
18. Brinton J, Ramirez-Gonzalez RH, Simmonds J, Wingen L, Orford S, Griffiths S, et al. A haplotype-led approach to increase the precision of wheat breeding. Commun Biol. 2020;3:712.
19. Li Z, Li XL, Liu SH, Mai SJ, Qin YT, Wang SY, et al. Identification and validation of quantitative trait loci for seven quality-related traits in common wheat (*Triticum aestivum* L). Theor Appl Genet. 2025;138:57.
20. Zhou ZF, Liu CC, Qin MM, Li WX, Hou JN, Shi X, et al. Promoter DNA hypermethylation of *TaGli-γ-2.1* positively regulates gluten strength in bread wheat. J Adv Res. 2022;36:163–73.
21. Nucia A, Okon S, Tomczynska-Mleko M, Nawrocka A. Molecular and physical characterization of grain hardness in European spring common wheat (*Triticum aestivum* L). 3 Biotech. 2021;11:345.
22. Fan XY, Jiang P, He Y, Zhang X, Yu GH, Zhou MX, et al. *qTaHa-5DL*: another major QTL regulating wheat grain hardness. BMC Plant Biol. 2024;24:1015.
23. Song QH, Cui WW, Song JJ, Zhou BY, Chen L, Xu KJ, et al. Uncover the genetic basis of processing quality related traits in common wheat (*Triticum aestivum* L.) using genome-wide association

mapping. Front Plant Sci. 2026;17:1755182.

24. Liu LH, Qu PP, Zhou Y, Li HB, Liu YN, Zhang MM, et al. Consensus linkage map construction and QTL mapping for eight yield-related traits in wheat using the BAAFS Wheat 90K SNP array. J Integr Agric. 2024;23:3641–56.
25. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 2006;1:2320–5.
26. Cui F, Zhang N, Fan XL, Zhang W, Zhao CH, Yang LJ, et al. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. Sci Rep. 2017;7:3788.
27. Tian HL, Wang FG, Zhao JR, Yi HM, Wang L, Wang R, et al. Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. Mol Breed. 2015;35:136.
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
29. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet. 2018;103:338–48.
30. Frichot E, François O. LEA: An R package for landscape and ecological association studies. Methods Ecol Evol. 2015;6:925–9.
31. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–20.
32. Wang JB, Zhang ZW. GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. Genom Proteom Bioinf. 2021;19:629–40.
33. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38:226–31.
34. Mutshinda CM, Irwin AJ, Sillanpää MJ. A bayesian framework for robust quantitative trait locus mapping and outlier detection. Int J Biostat. 2020; 20190038.
35. Wang JB, Tang Y, Zhang ZW. Performing genome-wide association studies with multiple models using GAPIT. Methods Mol Biol. 2022;2481:199–217.
36. International Wheat Genome Sequencing (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361:6403.
37. Jiao CZ, Xie XM, Hao CY, Chen LY, Xie YX, Garg V, et al. Pan-genome bridges wheat structural variations with habitat and breeding. Nature. 2025;637:384–93.
38. Aury JM, Engelen S, Istace B, Monat C, Lasserre-Zuber P, Belser C et al. Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *GigaScience* 2022; 11:giac034.

39. Liu SC, Li K, Dai XR, Qin GC, Lu DD, Gao ZX, et al. A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. Nat Genet. 2025;57:1008–20.
40. Burridge AJ, Wilkinson PA, Winfield MO, Barker GLA, Allen AM, Coghill JA, et al. Conversion of array-based single nucleotide polymorphic markers for use in targeted genotyping by sequencing in hexaploid wheat (*Triticum aestivum*). Plant Biotechnol J. 2018;16:867–76.
41. Keeble-Gagnère G, Pasam R, Forrest KL, Wong D, Robinson H, Godoy J, et al. Novel design of imputation-enabled SNP arrays for breeding and research applications supporting multi-species hybridization. Front Plant Sci. 2021;12:756877.
42. Njuguna JN, Clark LV, Lipka AE, Anzoua KG, Bagmet L, Chebukin P, et al. Impact of genotype-calling methodologies on genome-wide association and genomic prediction in polyploids. Plant Genome. 2023;16:e20401.
43. Pavan S, Delvento C, Ricciardi L, Lotti C, Ciani E, D'Agostino N. Recommendations for choosing the genotyping method and best practices for quality control in crop genome-wide association studies. Front Genet. 2020;11:447.
44. Walkowiak S, Gao LL, Monat C, Haberer G, Kassa MT, Brinton J, et al. Multiple wheat genomes reveal global variation in modern breeding. Nature. 2020;588:277–83.
45. Pang BS, Ren XZ, Liu LH, Zhao CP, Zhang MM, Jin SQ, et al. NY/T 4021 – 2021: Wheat (*Triticum aestivum* L.) variety genuineness identification - SNP based method. Beijing: Ministry of Agriculture and Rural Affairs of the People's Republic of China; 2021.
46. Zhou CE, Dong ZN, Zhang T, Wu JH, Yu SZ, Zeng QD, et al. Genome-scale analysis of homologous genes among subgenomes of bread wheat (*Triticum aestivum* L). Int J Mol Sci. 2020;21:3015.
47. Guo HW, Zhang GY, Zhou M, Wan M, Zhu B, Yang ZJ, et al. Whole genome doubling-induced the enrichment of *H3K27me3* in genes carrying specific TEs in *Aegilops tauschii*. Front. Genet. 2023;14:1241201.
48. Tyrka M, Mokrzycka M, Bakera B, Tyrka D, Szeliga M, Stojałowski S, et al. Evaluation of genetic structure in European wheat cultivars and advanced breeding lines using high-density genotyping-by-sequencing approach. BMC Genomics. 2021;22:81.
49. Peng YC, Lyu XR, Xu DG, Wang ZJ, Xia XC, Hao YF, et al. Genome-wide atlas of loci involved in chromosomal recombination in common wheat. BMC Plant Biol. 2024;24:1124.
50. Sivabharathi RC, Rajagopalan VR, Suresh R, Sudha M, Karthikeyan G, Jayakanthan M, et al. Haplotype-based breeding: a new insight in crop improvement. Plant Sci. 2024;346:112129.
51. Pshenichnikova TA, Ermakova MF, Chistyakova AK, Shchukina LV, Berezovskaya EV, Lochwasser U, et al. Mapping of the quantitative trait loci (QTL) associated with grain quality characteristics of the bread wheat grown under different environmental conditions. Russ J Genet. 2008;44:74–84.
52. Yang YS, Li SM, Zhang KP, Dong ZY, Li YW, An XL, et al. Efficient isolation of ion beam-induced mutants for homoeologous loci in common wheat and comparison of the contributions of *Glu-1* loci to gluten functionality. Theor Appl Genet. 2014;127:359–72.

53. Hu JM, Xiao GL, Jiang P, Zhao Y, Zhang GX, Ma X, et al. QTL detection for bread wheat processing quality in a nested association mapping population of semi-wild and domesticated wheat varieties. BMC Plant Biol. 2022;22:129.
54. Ding HK, Li YK, Ou JL, Song YZ, Qiu LH, Rong XY, et al. Characterization of a stable QTL for quality-related traits and its effects on yields in wheat (*Triticum aestivum* L.). Theor. Appl Genet. 2025;138:56.
55. Zhao J, Sun LJ, Hu MY, Liu Q, Xu JJ, Mu LM, et al. Pleiotropic quantitative trait loci (QTL) mining for regulating wheat processing quality- and yield-related traits. Plants. 2024;13:2545.
56. Raihan MS, Liu J, Huang J, Guo H, Pan QC, Yan JB. 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;129:1465–77.
57. Liu K, Sun XX, Ning TY, Duan XX, Wang QL, Liu TT, et al. Genetic dissection of wheat panicle traits using linkage analysis and a genome-wide association study. Theor Appl Genet. 2018;131:1073–90.

Table 1

Table 1 is available in the Supplementary Files section.

Figures

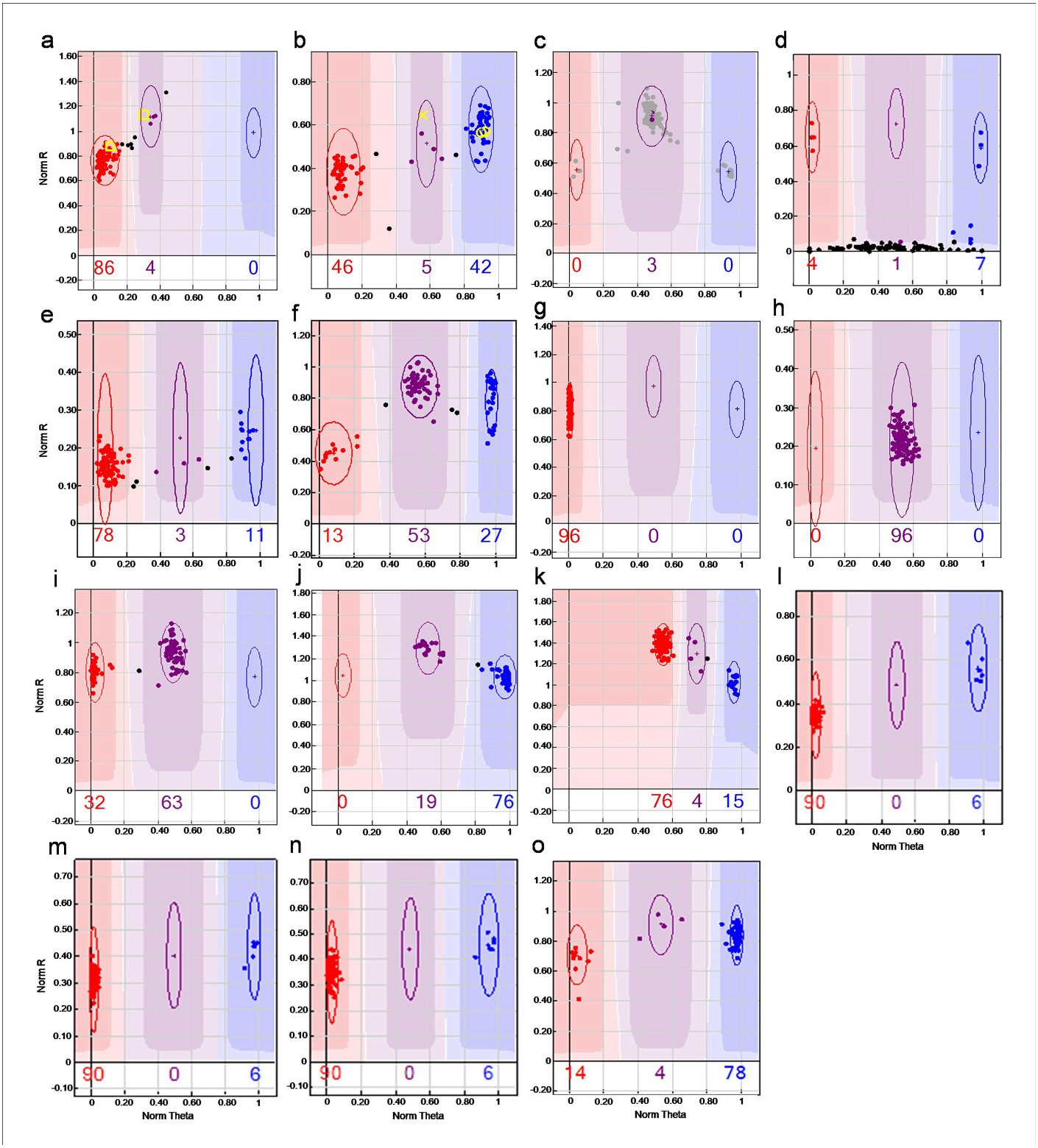


Figure 1

Fifteen genotyping patterns observed in 96 wheat accessions using the Illumina wheat 90K iSelect SNP array. The characteristics of three genotypes (AA/AB/BB) displayed by various types of SNP loci in GenomeStudio software. Replicates are indicated by squares; parent-hybrid pairs with non-Mendelian inheritance are shown by circles (parents) or forks (hybrids). Non-repeatable loci: the genotyping results are inconsistent between replicate samples, with one showing AA and the other AB (a). Mendelian

deviation loci: unexpected segregation in parents vs hybrids (b). Genotypic segregation in doubled haploid (DH) lines (c). High missing data: loci with a missing rate greater than 5% (d). Weak-signal loci: loci with suboptimal signal intensity (e). High heterozygosity: heterozygosity rate greater than 10% (f). Monomorphic loci (g). Loci with AA or BB genotype never observed (h) and with only AA (i) or BB (j) genotypes. loci with AATmean > 0.2 and BBTmean < 0.8 (k). Loci with MAF < 0.2 (l, m, n). Ideal loci (o): loci with mapping information that are excluded from cases a-n. “Norm Theta” on the x-axis denotes the normalized Theta value related to allele frequency for genotype differentiation, and “Norm R” on the y-axis represents normalized signal intensity for genotype classification.

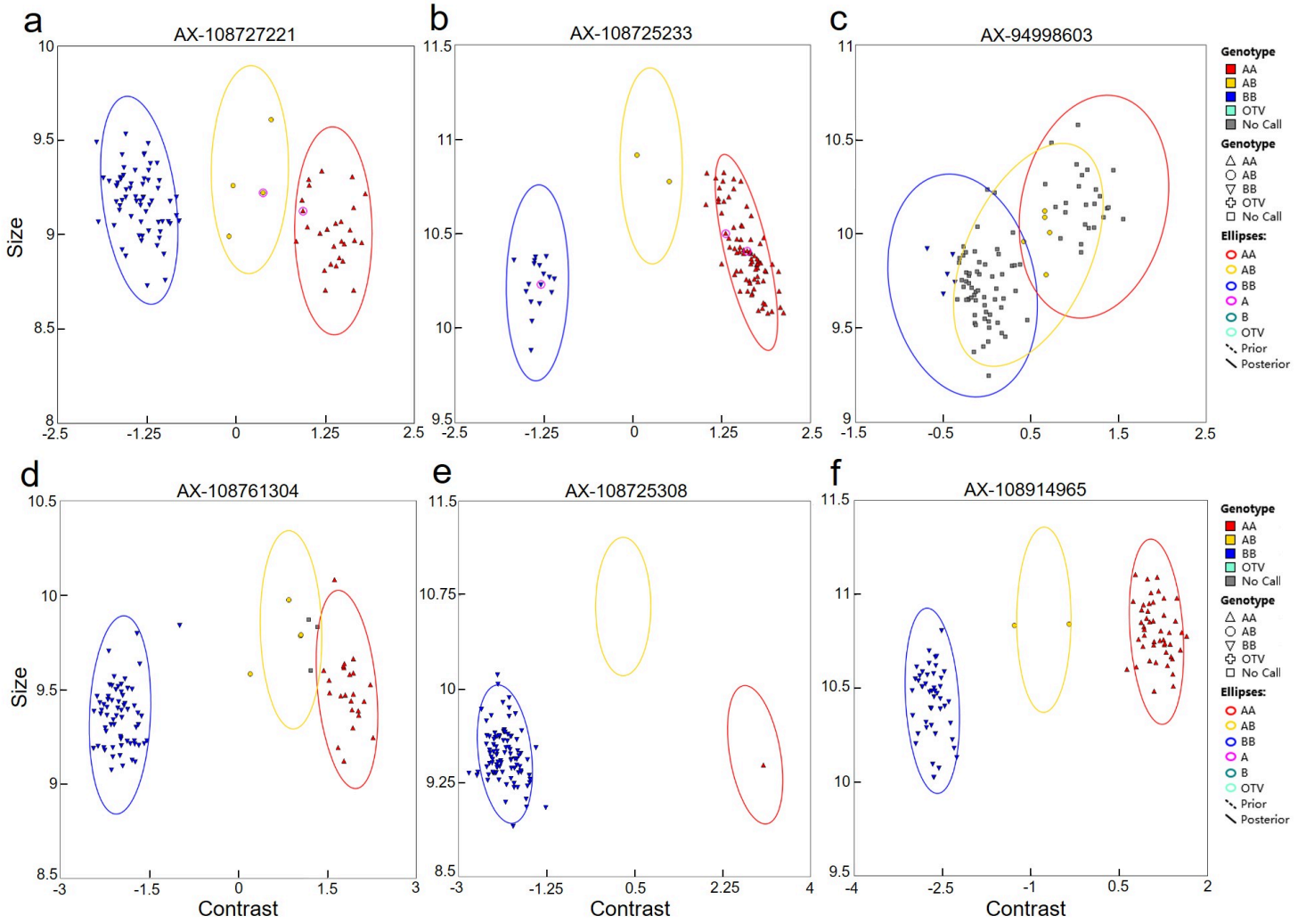


Figure 2

Six genotyping classes observed in 96 wheat accessions using the Axiom® Wheat 660K SNP array. These types represent different characteristics of SNP loci displaying AA, AB, and BB Genotypes on the Genotyping Console software. Non-repeatable or non-Mendelian inheritance samples are circled. Non-repeatable loci (a): the genotyping results are inconsistent among replicate samples. Non-Mendelian loci (b): loci showing deviations from the expected Mendelian inheritance patterns in the genetic relationship between parents and hybrids. Loci with many missing data (c): loci with a missing rate greater than 5%. Loci with weak signal (d): Loci with low - intensity signals, leading to potentially less reliable genotyping.

Loci with many heterozygous genotypes (e): heterozygosity rate greater than 10%. Ideal SNP loci (f): after excluding the cases shown in a-e and loci with MAF < 0.2, the remaining loci are of high quality, high resolution, and relatively uniform distribution. “Contrast” on the x-axis indicates signal contrast reflecting fluorescent signal differences among genotypes, and “Size” on the y-axis quantifies the signal magnitude of SNP loci.

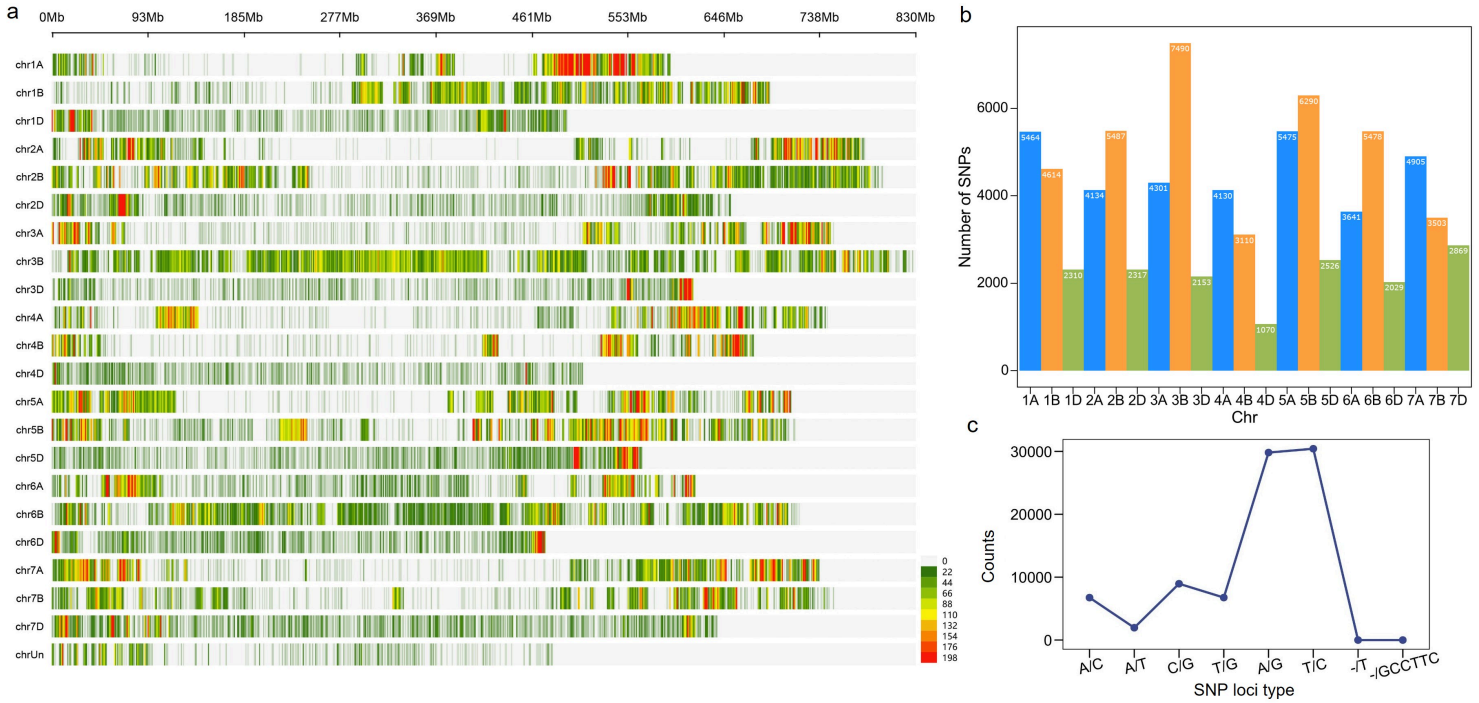


Figure 3

Distribution and characteristics of loci in BAAFS 90K array. a. Genome-wide SNP density plotted along each chromosome. b. Chromosome length (Mb) and total marker count. c. Proportion of transition, transversion and InDel SNP classes.

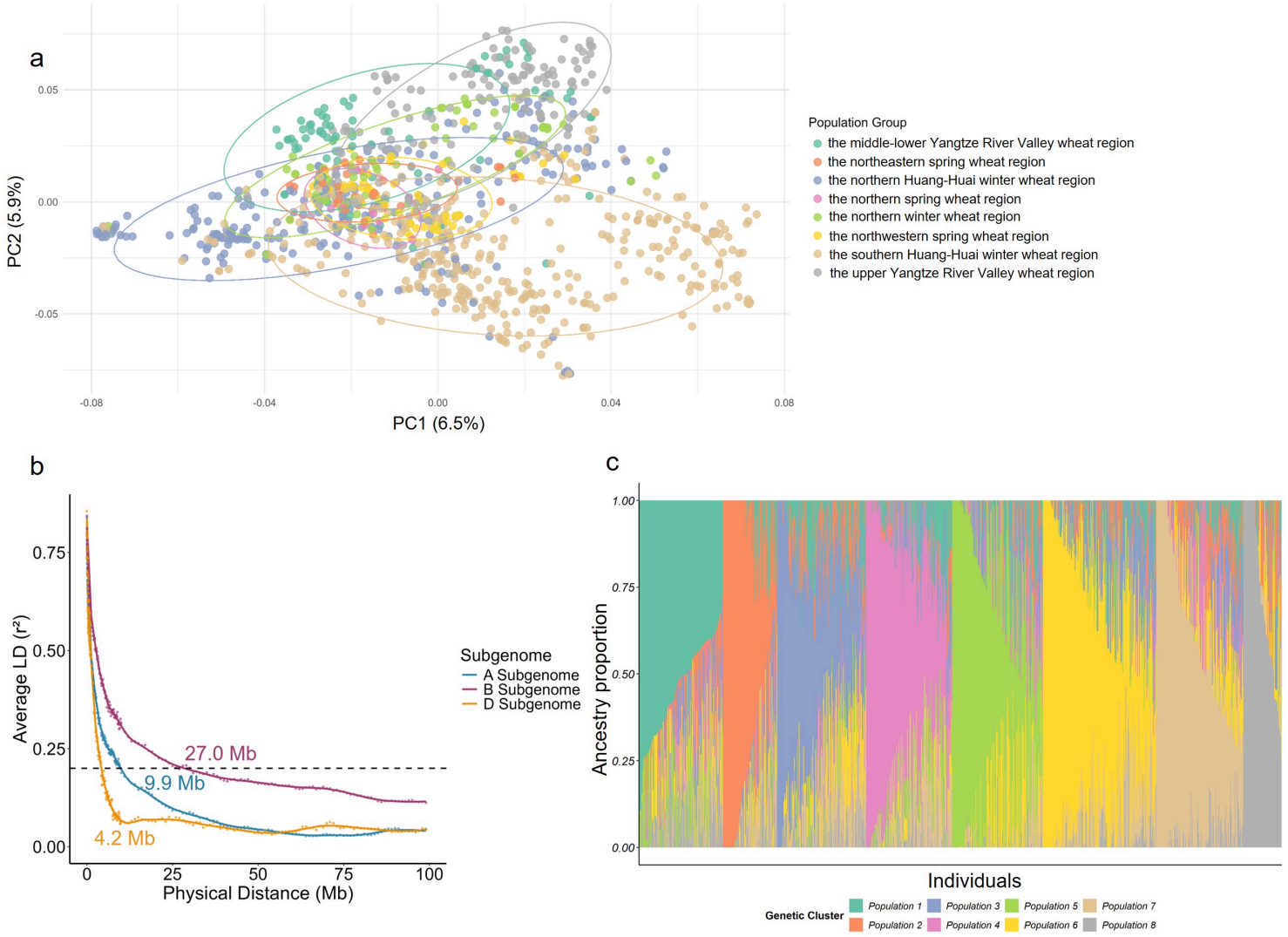


Figure 4

Genetic diversity among 1041 Chinese wheat cultivars revealed by the BAAFS 90K SNP array. a. Principal component analysis of the eight wheat zones (the southern Huang-Huai winter wheat region, the northern Huang-Huai winter wheat region, the middle-lower Yangtze River Valley wheat region, the upper Yangtze River Valley wheat region, the northeastern spring wheat region, the northern winter wheat region, the northwestern spring wheat region, and the northern spring wheat region). b. LD decay plot. c. Admixture analysis (K = 8).

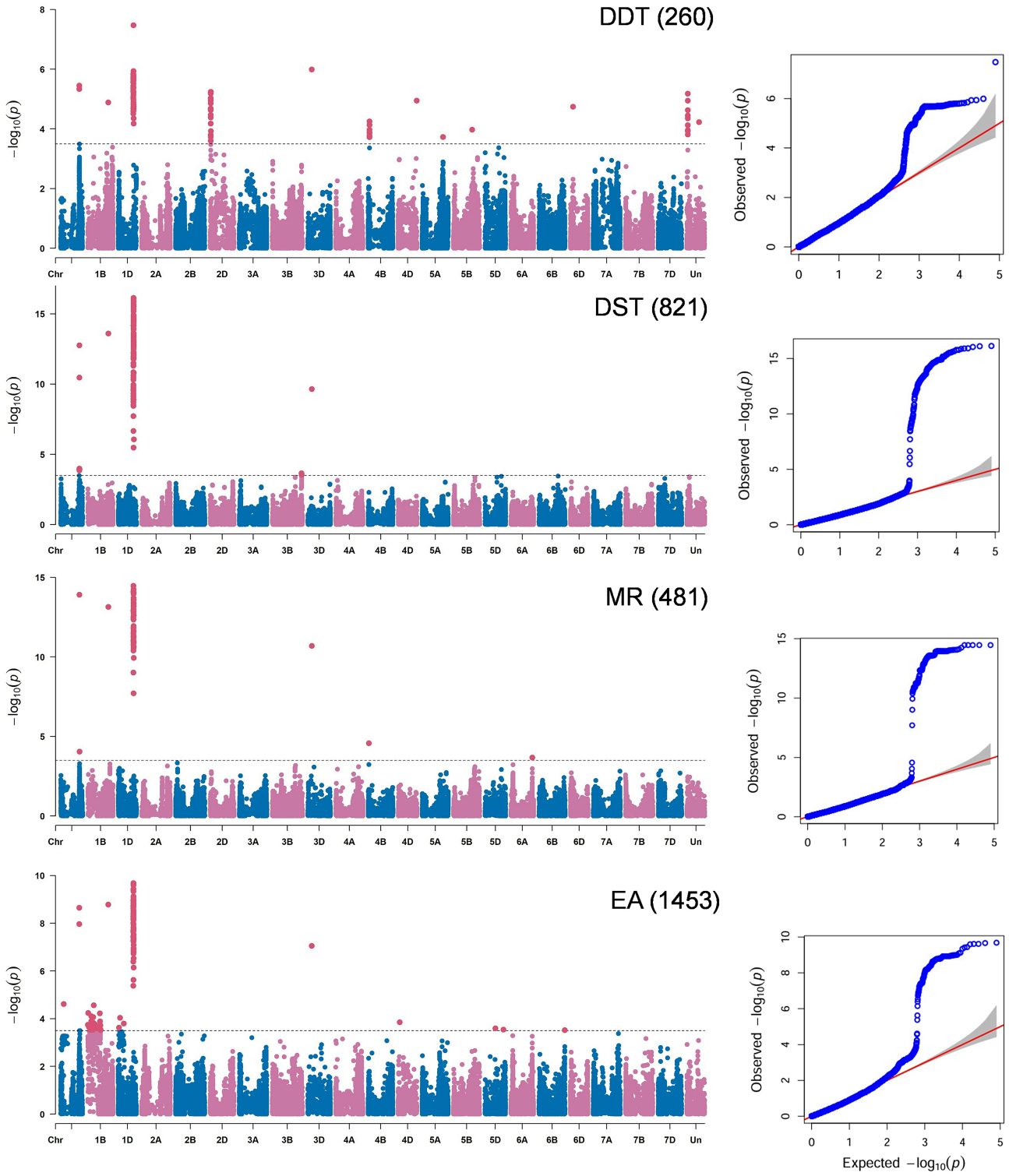


Figure 5

Manhattan plots and QQ plots of genome-wide association studies of four quality traits (dough development time (DDT, n = 260), dough stability time (DST, n = 821), maximum resistance (MR, n = 481), extension Area (EA, n = 1453 wheat cultivars in wheat cultivars using the BAAFS 90K SNP array. Horizontal dashed line represents the significance threshold (-log10() = 3.5).

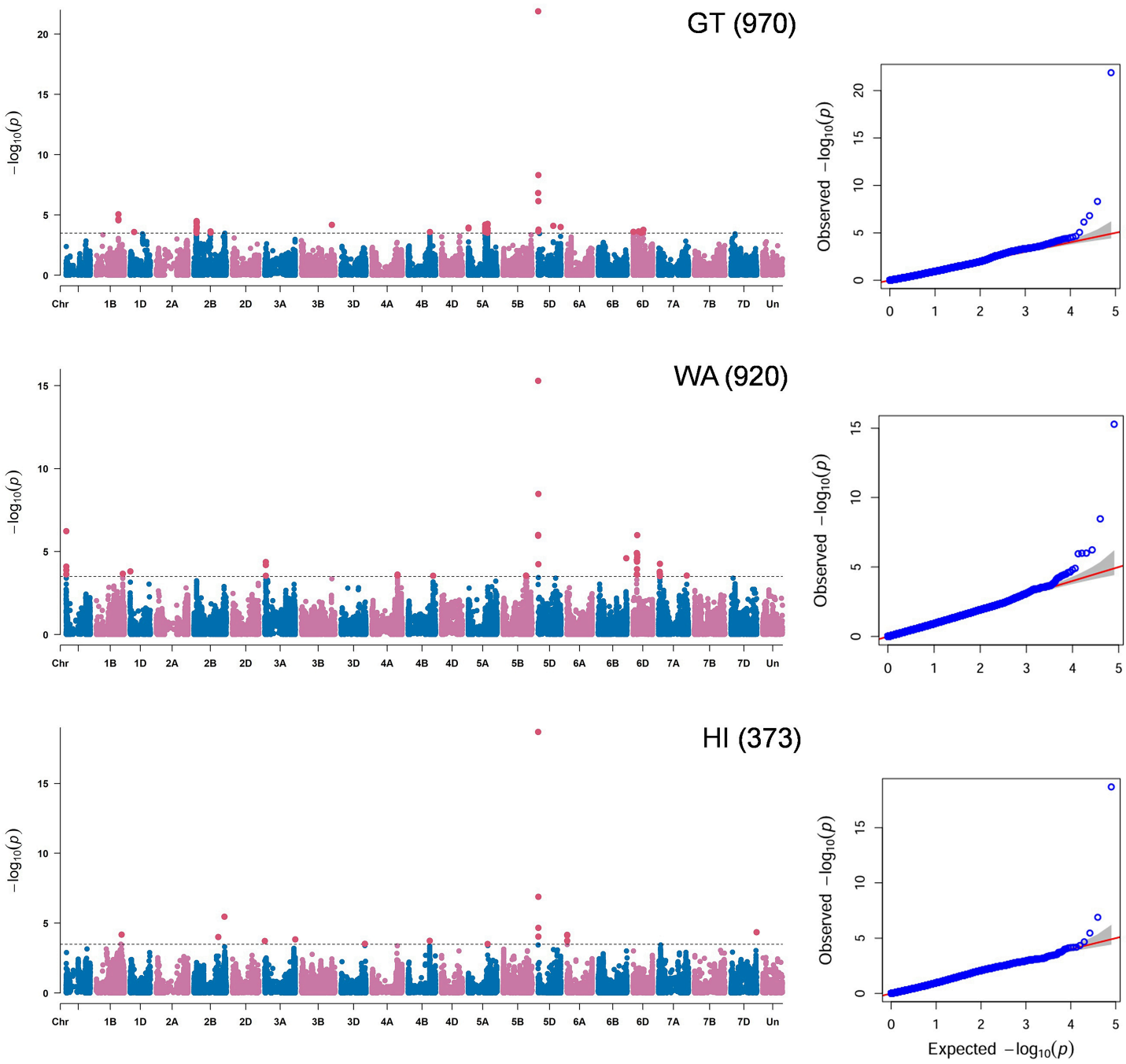


Figure 6

Genome-wide association studies of grain texture (GT, n = 970), water absorption (WA, n = 920), and grain hardness index (HI, n = 373) using the BAAFS 90K SNP array. Left: Manhattan plots; right: quantile-quantile (Q-Q) plots. Horizontal dashed line represents the significance threshold (-log10() = 3.5).

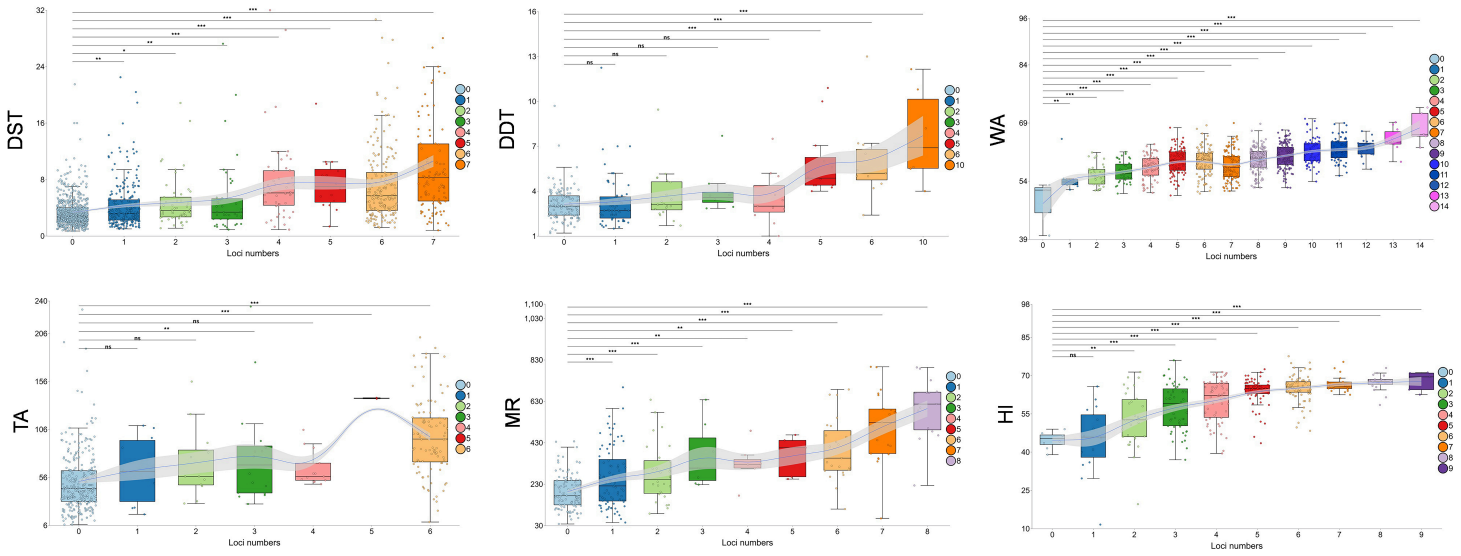


Figure 7

Additive effect of favourable alleles on dough development time (DDT), dough stability time (DST), maximum resistance (MR), extension area (EA), grain texture (GT), water absorption (WA), and grain hardness index (HI). ***: < 0.001, **: < 0.01, *: < 0.05, ns: non-significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.xlsx
- Supplementalfigure1.jpg
- Supplementalfigure2.jpg
- Supplementalfigure3.jpg
- Supplementaryfigure4.jpg
- Supplementaryfigure5.jpg
- Supplementaryfigure6.jpg
- Supplementaryfigure7.jpg
- Supplementarytable18.xlsx