

QTL Mapping and KASP Marker Development for Nutritional Quality Traits in Colored Wheat Based on Genome-Wide Association Study

Liang Wang

wangzi651@163.com

Peking University Institute of Advanced Agricultural Sciences <https://orcid.org/0000-0002-7517-4852>

Yanyan Pu

Runfang Li

Dezhou Cui

Xiaojian Fang

Xuan Guo

Xin Jing

Yongchao Hao

Xiaofeng Cao

Ru-Mei Tian

Song Hou

Shanshan Liu

shasha Wang

Nana Li

Yongchao Gong

Chengzhi Jiao

Article

Keywords: Colored wheat, GWAS, Nutritional quality, QTL, KASP

Posted Date: September 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10652140/v1>

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

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

**QTL Mapping and KASP Marker Development for Nutritional Quality Traits in
Colored Wheat Based on Genome-Wide Association Study**

Yanyan Pu¹, Runfang Li¹, Dezhou Cui¹, Xiaojian Fang², Xuan Guo³, Xin Jing⁴,
Yongchao Hao², Xiaofeng Cao², Rumei Tian¹, Song Hou¹, Shanshan Liu⁵, Shasha
Wang⁶, Nana Li^{1*}, Yongchao Gong^{1*}, Chengzhi Jiao^{2*}, Liang Wang^{2*}

¹ Shandong Academy of Agricultural Sciences, Shandong 250100, China

² Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory
of Advanced Agricultural Sciences in Weifang, Shandong 261325, China

³ Shandong Provincial Seed Management Station, Shandong 250100, China

⁴ Smartgenomics Technology Institute, Qingdao, China.

⁵ Qingdao Agricultural Technology Extension Center, Qingdao, Shandong Province
266000, China

⁶ Qingdao Animal Husbandry and Veterinary Institute, Qingdao, Shandong Province
266000, China

29 **Abstract**

30 Colored wheat is rich in nutritional components such as anthocyanins (ANs), crude fat
31 (CF), free amino acids (FAA), and total nitrogen (TN). However, the genetic basis
32 underlying these quality traits remains largely unexplored. This study aimed to identify
33 stable quantitative trait loci (QTL) for four nutritional quality traits in colored wheat
34 and to develop kompetitive allele-specific PCR (KASP) markers for marker-assisted
35 selection. A panel of 57 colored wheat germplasm resources was genotyped by whole-
36 genome resequencing and phenotyped for four traits. Genome-wide association study
37 (GWAS) was performed using Chinese Spring (CS-IAAS) genome as reference. A total
38 of 176 stable QTLs were detected, including 26 for ANs with phenotypic variance
39 explained (PVE) values ranging from 0.23% to 46.41%, 26 for CF with PVE values
40 from 0.19% to 50.70%, 19 for FAA with PVE values from 0.96% to 49.01%, and 105
41 for TN with PVE values from 0.12% to 55.66%. Three KASP markers, *KASP-ANs-7B*,
42 *KASP-CF-2B*, and *KASP-TN-6B*, were successfully developed and validated. Single-
43 marker analysis revealed significant associations with corresponding metabolic traits,
44 while effects on agronomic traits varied among markers. Extending 3 Mb on both sides
45 of each QTL intervals, 15 of 713 genes were prioritized as candidates based on
46 functional relevance. This study provides valuable QTL information and validated
47 KASP markers for nutritional quality improvement in colored wheat breeding programs,
48 and the pleiotropic effects of these markers on agronomic traits offer insights for
49 balancing quality and yield in wheat breeding.

50 **Keywords:** Colored wheat, GWAS, Nutritional quality, QTL, KASP

51 **1 Introduction**

52 Common wheat (*Triticum aestivum* L.) is one of the most widely cultivated and highest-
53 yielding food crops globally, providing approximately 20% of daily caloric and protein
54 intake for humans (Shewry *et al.*, 2015). Rising living standards and growing health
55 awareness have driven a shift in consumer preference from grain yield to grain quality,

rendering nutritional quality improvement a major objective in wheat breeding (Padhy *et al.*, 2024). Colored wheat refers to germplasm with purple, blue, black, or red kernels, with coloration primarily derived from ANs accumulated in the seed coat or aleurone layer (Wu *et al.*, 2025). As natural water-soluble pigments, ANs exhibit potent antioxidant activity, effectively scavenging free radicals, delaying aging, and reducing the risk of cardiovascular diseases and certain cancers (Marcotuli *et al.*, 2025). Moreover, colored wheat grains generally contain higher levels of protein, iron, zinc, dietary fiber, and unsaturated fatty acids than conventional wheat (Padhy *et al.*, 2024). These properties confer broad prospects for colored wheat in specialty agriculture and the health industry. However, its promotion faces the challenge of balancing nutritional quality enhancement with yield maintenance, as most colored wheat varieties exhibit lower yield potential than high-yielding cultivars, limiting large-scale adoption (Nigro *et al.*, 2019). The fundamental cause is insufficient understanding of the genetic basis underlying nutritional quality traits, hampering coordinated improvement through molecular breeding (Wu *et al.*, 2025). Therefore, systematic dissection of the genetic regulatory mechanisms of key quality traits, identification of superior alleles, and development of marker-assisted selection tools are of great theoretical and practical value.

GWAS is an efficient strategy for dissecting the genetic basis of complex quantitative traits (Flint-Garcia *et al.*, 2003; Muqaddasi *et al.*, 2020; Yang *et al.*, 2020). Compared with traditional linkage mapping, GWAS directly utilizes natural populations without constructing segregating populations, substantially shortening research timelines, while achieving higher resolution through denser markers and richer allelic diversity (Cheng *et al.*, 2024; Schulthess *et al.*, 2022). Since GWAS has been widely adopted in crop genetics and breeding, it has been successfully applied in wheat to dissect yield and agronomic traits, grain protein content (GPC), starch quality, gluten strength, and micronutrients, demonstrating its power in uncovering loci associated with quality traits (Hao *et al.*, 2022; Jiao *et al.*, 2023; Jiao *et al.*, 2024; Kartseva *et al.*, 2023; Muqaddasi *et al.*, 2020). Recent advances in wheat reference genomics have greatly facilitated GWAS studies. The release of IWGSC RefSeq v1.0 and the

86 completion of the CS-IAAS genome assembly achieved complete assembly of all 21
87 chromosomes, providing an unprecedented reference for fine-mapping QTLs and
88 candidate gene annotation (IWGSC *et al.*, 2018; Liu *et al.*, 2025). Wheat pan-genome
89 studies have also revealed the importance of structural variations in quality regulation
90 (Cheng *et al.*, 2024; Jiao *et al.*, 2024; Schulthess *et al.*, 2022; Walkowiak *et al.*, 2020).
91 However, the vast majority of existing GWAS studies have been conducted in common
92 wheat, whereas systematic GWAS investigations in colored wheat remain scarce,
93 particularly for signature traits such as ANs, CF, and FAA.

94 Most existing studies on colored wheat have focused on anthocyanin biosynthesis
95 pathways. Key enzyme-encoding genes and regulatory genes involved in anthocyanin
96 biosynthesis have been cloned and functionally validated (Garg *et al.*, 2022; Xu *et al.*,
97 2023). However, these studies largely relied on homology-based candidate gene
98 approaches confined to known pathway genes, failing to capture the full complexity of
99 regulatory networks. Genetic studies on CF and FAA are even more limited (Mandrioli
100 *et al.*, 2024; Peng *et al.*, 2018). Furthermore, complex correlations exist among quality
101 traits, such as the observed negative association between nitrogen level and anthocyanin
102 accumulation, though its genetic basis remains unclear (Nigro *et al.*, 2019).

103 At the methodological level, most GWAS studies employ a single association
104 model, which often suffers from inflated false-positive rates or inadequate statistical
105 power (Peng *et al.*, 2018; Yang *et al.*, 2020). The generalized linear model (GLM) tends
106 to produce false positives due to insufficient control of population structure, while
107 mixed linear model (MLM) may miss genuine loci through overcorrection (Peng *et al.*,
108 2018). EMMAX, as an efficient mixed linear model, controls false positives while
109 maintaining high detection power (Kang *et al.*, 2010). The PCA+GLM strategy
110 incorporates principal components to correct population structure and has been widely
111 adopted (Huang *et al.*, 2017; Lin *et al.*, 2016). The co-localization strategy combining
112 two or multiple models, whereby only QTLs detected consistently across models are
113 considered reliable, has been proven to enhance detection robustness (Bashir *et al.*,
114 2025; Hong *et al.*, 2025). Meanwhile, translating GWAS results into KASP markers is
115 a critical step connecting genetic dissection to breeding application. KASP technology,

with its low cost, high throughput, and accurate genotyping, has become one of the most widely adopted platforms for marker-assisted selection in crops (Chen *et al.*, 2025). In wheat, numerous KASP markers developed from GWAS results have been successfully applied to breeding for disease resistance, quality, and yield traits (Bashir *et al.*, 2025; Chen *et al.*, 2025; Wang *et al.*, 2025). However, KASP markers specifically developed for nutritional quality traits in colored wheat remain extremely limited. This study identified 176 stable QTLs for four nutritional quality traits in 57 colored wheat accessions via GWAS, successfully developed three KASP markers, and screened 15 candidate genes, providing directly applicable molecular tools for quality improvement in colored wheat.

2 Materials and Methods

2.1 Plant Materials

The experiment was conducted at the experimental station in Licheng District, Jinan, Shandong Province, China, during the 2024 to 2025 cropping season. A total of 57 colored wheat accessions were used as plant materials, following the preceding crop of maize on soils with moderate and uniform fertility. The experimental layout followed a randomized complete block design, with 57 plots corresponding to the 57 tested accessions. Each plot covered an area of 15 m², with a length of 10 m and a width of 1.5 m. Six rows were planted in each plot with a row spacing of 30 cm. The seeding rate was 375 kg ha⁻¹. Field management practices were consistent with those applied to local high-yielding wheat fields.

2.2 Phenotypic Measurements

Plant height (PH), spike length (SL), spikelet number per spike (SN), kernel number per spike (KNS), and grain number per spikelet (GNS) were measured in the field in May 2025. Ten representative plants with uniform growth were selected from each plot for agronomic trait evaluation and yield component analysis. Each trait was measured

with ten biological replicates. Thousand-kernel weight (TKW) was determined after harvest and measured with three biological replicates. Results are presented as mean values.

The contents of ANs, CF, FAA, and TN were determined by Wuhan ProNets Testing Technology Co. Ltd. (Wuhan, China). Each sample was analyzed with three technical replicates, and the results are presented as mean values. The determination methods for the four traits are described as follows.

ANs were determined using a spectrophotometric method. Approximately 0.1 g of sample powder was extracted with 0.5 mL of extraction solution under ultrasonication for 30 min, followed by centrifugation at 10,000 g for 10 min. The absorbance of the supernatant was measured at 530 nm, 620 nm, and 650 nm. The anthocyanin content ($\mu\text{g/g}$) was calculated as follows: $\text{ANs } (\mu\text{g/g}) = [\Delta A \times V \times 10^{-3} / (\epsilon \times d) \times M_1 \times F / 10^6] / M$, where $\Delta A = (A_{\text{max}} - A_{620}) - 0.1(A_{650} - A_{620})$, V is the extraction volume (mL), ϵ is the molar extinction coefficient of ANs, d is the light path length of the 96-well plate (cm), M_1 is the molecular weight of ANs (g/mol), F is the dilution factor, and M is the sample mass (g).

CF was determined using the Soxhlet extraction method. An appropriate amount of sample powder was extracted with petroleum ether under reflux for 6–10 h. After solvent recovery, the extract was dried at $(100 \pm 5)^\circ\text{C}$ to constant weight. The CF content was calculated as follows: $\text{CF (g/100g)} = (m_1 - m_0) / m_2 \times 100$, where m_1 is the constant weight of the receiving flask with fat (g), m_0 is the constant weight of the empty receiving flask (g), and m_2 is the sample mass (g).

FAA were determined using the ninhydrin colorimetric method. Approximately 0.1 g of sample powder was extracted with 10% acetic acid solution, and the supernatant was reacted with ninhydrin reagent. The absorbance was measured at 570 nm, and quantification was performed using a standard curve prepared with L-leucine. The free amino acid content was calculated as follows: $\text{FAA (mg/g)} = C \times V / M \times F$, where C is the amino acid concentration determined from the standard curve (mg/mL), V is the extraction volume (mL), M is the sample mass (g), and F is the dilution factor.

TN was determined using the Dumas combustion method. An appropriate amount

of sample was combusted at high temperature using a Dumas nitrogen analyzer, and the TN content was directly quantified by a thermal conductivity detector after conversion to molecular nitrogen. Phenotypic data organization and basic statistical analyses were performed using Microsoft Excel v14.0.7263.5000 and GraphPad Prism v10.1.2.

2.3 Correlation Analysis

Pearson correlation coefficients were calculated for pair-wise comparisons among the four nutritional quality traits, which included ANs, CF, FAA, and TN. Correlation analysis was performed using R v4.5.3 software (<https://www.r-project.org/>) based on the phenotypic data of 57 colored wheat accessions. The following R packages and versions were employed: reshape2 v1.4.4 for data reshaping, ggplot2 v4.0.0 and GGally v2.4.0 for generating scatterplot matrices, and ggpubr v0.6.3 for statistical visualization. Significance testing was conducted using Student's *t*-test, and the significance level was set at $P < 0.05$.

2.4 Genotyping and Quality Control

The wheat accessions were germinated in a growth chamber set at 20 °C with a photoperiod of 12 h of light and 12 h of darkness in 2025. Seedling leaves were collected after two weeks of growth. Genomic DNA was extracted using the CTAB method and subjected to whole-genome resequencing with 5× depth on the DNBSEQ sequencing platform by Smartgenomics Technology Institute. Quality control of the raw genotype data was performed using bcftools v1.18 with the following criteria: $QUAL \geq 30$, missing rate $< 20\%$, and minor allele frequency (MAF) ≥ 0.05 (Gao *et al.*, 2023).

2.5 GWAS Analysis

Genome-wide association analysis was conducted using two models: EMMAX and PCA+GLM (Huang *et al.*, 2017; Kang *et al.*, 2010). The CS-IAAS genome was used as the reference genome. The EMMAX model analysis was performed using EMMAX-

beta software, while the PCA+GLM model analysis and calculation of PVE were performed using PLINK v1.9 (Chang *et al.*, 2015). Manhattan plots were generated using the base R packages utils v4.5.1, graphics v4.5.1, and stats v4.5.1 in R software v4.5.3, with a significance threshold set at $-\log_{10}(P) \geq 5$. SNP density plots were generated using the R package CMplot v4.5.1 (<https://github.com/YinLiLin/CMplot>) (Yin *et al.*, 2021).

2.6 Genetic diversity analysis

Nucleotide diversity was estimated from the whole-genome SNP data of the 57 colored wheat accessions using vcftools v0.1.13 software. The π values were calculated using a sliding-window approach with a window size of 50 kb and a step size of 25 kb. Based on the SNP information within each window, π values were computed for the whole genome and for individual chromosomes, and the average π values for the A, B, and D sub-genomes were subsequently derived. Only SNP markers with MAF greater than 0.05 and missing rate below 20% were retained for the analysis to ensure the reliability of the results (Hao *et al.* 2020).

2.7 KASP Primer Design

Significant SNPs associated with the four traits in the GWAS were selected for KASP marker development. Flanking sequences extending 100 bp upstream and 100 bp downstream of each significant SNP were retrieved from the WheatOmics website (<http://202.194.139.32/>) (Ma *et al.*, 2021), and submitted to the PolyMarker website (<https://www.polymarker.info/>) for KASP primer design (Ramirez-Gonzalez *et al.*, 2015). Primers were synthesized by BGI TECH SOLUTIONS (BEIJING LIUHE) CO., LIMITED.

3 Result

3.1 Phenotypic variation and correlation analysis of nutritional quality traits

The four nutritional quality traits exhibited extensive phenotypic variation across the 57 colored wheat accessions, with all traits showing normal distributions (Fig. 1A). ANs content ranged from 6.34 to 92.93 µg/g, CF content from 0.50 to 2.47 g/100g, FAA content from 3.39 to 7.92 mg/g, and TN content from 2.20% to 2.94% (Table S1).

Pairwise Pearson correlation analysis in Figure 1A revealed that ANs was significantly positively correlated with CF, with a correlation coefficient of 0.366 and $P < 0.01$. The upward-sloping regression line suggested a potential synergistic relationship between anthocyanin accumulation and lipid metabolism. ANs showed a significant negative correlation with TN, with a correlation coefficient of -0.463 and $P < 0.001$. Similarly, CF was also significantly negatively correlated with TN, with a correlation coefficient of -0.355 and $P < 0.01$, indicating potential trade-offs or competition between nitrogen metabolism and both anthocyanin accumulation and lipid accumulation. No significant correlations were detected between ANs and FAA, with a correlation coefficient of 0.164 and $P > 0.05$. Likewise, no significant correlation was found between CF and FAA, with a correlation coefficient of 0.013 and $P > 0.05$, nor between FAA and TN, with a correlation coefficient of -0.199 and $P > 0.05$. In each of these non-significant cases, the regression lines approximated horizontal, suggesting that FAA content may be regulated independently of the other three traits.

3.2 The whole-genome variation map and genetic diversity of colored wheat

Whole-genome resequencing of the 57 colored wheat accessions generated 49,254,023 raw SNPs, of which 39,840,202 high-quality SNPs were retained after quality control. These SNPs were unevenly distributed across the three sub-genomes of wheat, with the A, B, and D subgenomes accounting for 36.32%, 47.93%, and 15.74% of the total SNPs, respectively. At the chromosome level, chromosome 3B harbored the largest number of SNPs with 3,875,162 SNPs, while chromosome 6D contained the fewest with only 718,777 SNPs. In terms of marker density, chromosome 3B exhibited the highest density at 4,435 SNPs per megabase, whereas chromosome 3D showed the lowest density at 1,274 SNPs per megabase (Fig. 1B; Table S2). These results indicate that the

B subgenome possesses a marked advantage in both SNP number and density, which may be attributable to its larger physical length.

Nucleotide diversity was estimated for the 57 colored wheat accessions based on high-quality SNP markers obtained from whole-genome resequencing. The average π value across the whole genome was 0.296×10^{-3} . The mean π values for the A, B, and D sub-genomes were 0.319×10^{-3} , 0.318×10^{-3} , and 0.250×10^{-3} , respectively. Among individual chromosomes, chromosome 7A exhibited the highest π value at 0.346×10^{-3} , while chromosome 7D showed the lowest at 0.236×10^{-3} . These results indicate that the colored wheat germplasm displays a similar pattern of genetic diversity distribution at the sub-genome level to that observed in common wheat, with higher diversity in the A and B sub-genomes compared with the D sub-genome, consistent with the asymmetric sub-genome selection signature reported in Chinese wheat breeding programs.

3.3 Genome-wide identification of QTLs related to nutritional quality

Using the 39,840,202 high-quality SNPs, genome-wide association studies were performed for the four nutritional quality traits employing two models, EMMAX and PCA+GLM, with a significance threshold of $-\log_{10}(P) \geq 5$ (Fig. 2). By consolidating the QTLs identified by both models, a total of 176 stable QTLs and 1,507 significant SNPs were detected.

The number of QTLs varied considerably across traits. For ANs, 26 QTLs and 127 significant SNPs were identified, distributed across 14 chromosomes, with PVE values spanning from 11.42% to 39.50%. Among them, *ANs.pkuias.7B.3* explained 26.28% to 43.21% of the phenotypic variation, representing the largest-effect QTL for this trait. CF also yielded 26 QTLs and 278 significant SNPs, distributed on 12 chromosomes, with PVE values ranging from 18.42% to 42.60%. The locus *CF.pkuias.2B.4* exhibited PVE values from 0.85% to 40.92% and was the most prominent QTL for CF content. FAA yielded the fewest QTLs among the four traits, with 19 QTLs and 79 significant SNPs distributed across 15 chromosomes, and PVE values ranging from 14.29% to 38.84%. TN produced the largest number of QTLs, with 105 QTLs and 1,023

significant SNPs, distributed across all chromosomes except 3D, and PVE values ranging from 9.63% to 47.21%. Notably, *TN.pkuias.6B.4* harbored 303 significant SNPs with PVE values from 30.95% to 53.18%, representing the strongest major QTL among all four traits (Table 1; Table S3).

Comparison of the QTLs identified in this study with previously reported loci revealed that approximately 15.34% of the 176 QTLs coincided with known positions, while the remaining 84.66% were novel discoveries. For ANs, 7 of the 26 QTLs matched previously reported loci, and the 19 newly identified QTLs contributed PVE values ranging from 0.23% to 46.41%. For CF, all 26 QTLs were reported for the first time in this study, with PVE values ranging from 0.39% to 50.70%. For FAA, 3 of the 19 QTLs were consistent with known loci, and the 16 newly identified QTLs exhibited PVE values from 0.96% to 49.01%. For TN, 17 of the 105 QTLs matched previously reported loci, and the 88 newly identified QTLs contributed PVE values ranging from 0.12% to 55.66%. These results demonstrate that a substantial number of novel QTLs were detected across all four nutritional quality traits, with particularly rich discoveries in CF and TN, significantly expanding the genetic resources available for quality improvement in colored wheat.

3.4 KASP marker development and single-marker analysis for nutritional traits

The QTLs harboring the largest number of significant SNPs for each trait were selected for KASP marker development. These included *ANs.pkuias.7B.3* for ANs, *CF.pkuias.2B.4* for CF, *FAA.pkuias.7B* for FAA, and *TN.pkuias.6B.4* for TN, with primer sequences provided in Table S4. Among these, the KASP markers for *ANs.pkuias.7B.3*, *CF.pkuias.2B.4*, and *TN.pkuias.6B.4* were successfully genotyped in the 57 accessions (Fig. 3A-3C), whereas the marker for *FAA.pkuias.7B* failed to develop, and single-marker analysis revealed no significant association with FAA content (Fig. S1).

The three successfully developed KASP markers were further evaluated for their associations with the four nutritional traits. Single-marker analysis of *KASP-ANs-7B*

revealed that accessions carrying the locus, designated as genotype B, had significantly higher ANs content than those without the locus, designated as genotype A, with $P < 0.01$. No significant differences were observed for CF, FAA, or TN between the two groups, indicating that this locus specifically and positively regulates anthocyanin accumulation (Fig. 3D).

For *KASP-CF-2B*, accessions carrying the locus exhibited significantly higher CF content than those without the locus, with $P < 0.05$, while no significant effects were detected for ANs, FAA, or TN. This finding suggests a specific positive effect of this locus on CF content (Fig. 3E). For *KASP-TN-6B*, accessions carrying the locus showed significantly higher TN content with $P < 0.0001$, but significantly lower ANs content with $P < 0.05$, compared with non-carriers. No significant differences were detected for CF or FAA (Fig. 3F). This result indicates that the *TN.pkuias.6B.4* locus may have a negative pleiotropic effect on anthocyanin accumulation while simultaneously enhancing TN content.

3.5 Verification of KASP markers with agronomic traits

To evaluate the potential pleiotropic effects of the three KASP markers on yield-related agronomic traits, we further analyzed the associations of the three marker loci with six agronomic traits, namely GNS, KNS, PH, SL, SN, and TKW (Fig. 4).

The results showed that no significant differences were detected between carrier and non-carrier groups for any of the six agronomic traits at the *KASP-ANs-7B* and *KASP-CF-2B* loci, indicating that these two loci have no apparent pleiotropic effects on yield-related traits. In contrast, the *KASP-TN-6B* locus exhibited significant pleiotropy: carriers showed significantly reduced GNS ($P < 0.05$) and significantly increased PH ($P < 0.001$), whereas no significant effects were observed for KNS, SL, SN, or TKW (Fig. 4).

3.6 Candidate gene prediction in QTL interval

To identify candidate genes within the four target QTL intervals, each original physical

interval was extended by 3 Mb both upstream and downstream. Based on the CS-IAAS genome annotation, a total of 713 candidate genes were identified. Among these, 52 candidate genes were located in the *ANs.pkuias.7B.3* interval, 197 in the *CF.pkuias.2B.4* interval, 175 in the *FAA.pkuias.7B* interval, and 289 in the *TN.pkuias.6B.4* interval (Table S5).

For the 15 representative candidate genes identified within the four major QTL intervals, GO functional annotations and KEGG pathway information were further examined (Table 2). Among the seven candidate genes identified in the *ANs.pkuias.7B.3* interval, six zinc finger family proteins were all annotated with DNA-binding transcription factor activity. Among them, *CSIAAS7BG1255000HC* was annotated with zinc ion binding and protein-folding chaperone binding, with a KEGG annotation of mitochondrial protein import protein ZIM17. The other candidate, *CSIAAS7BG1256000HC*, encoding a bHLH transcription factor, was annotated with DNA-binding transcription factor activity, sequence-specific DNA binding, and protein dimerization activity. Among the five candidate genes identified in the *CF.pkuias.2B.4* interval, the WRKY transcription factor *CSIAAS2BG0786400HC* was annotated with DNA-binding transcription factor activity and sequence-specific DNA binding, with a KEGG annotation of WRKY transcription factor group IIc. The three bHLH transcription factors were all annotated with DNA-binding transcription factor activity and protein dimerization activity, among which *CSIAAS2BG0807200HC* and *CSIAAS2BG0807800HC* were additionally annotated with sequence-specific DNA binding. *CSIAAS2BG0824000HC* was annotated as a 3-oxoacyl-reductase but lacked GO or KEGG annotations. For the *FAA.pkuias.7B* interval, only one candidate gene, *CSIAAS7BG1012200HC*, was identified, which encodes an allantate amidohydrolase with hydrolase activity acting on carbon-nitrogen bonds in linear amidines and a KEGG annotation of allantate deiminase. For the *TN.pkuias.6B.4* interval, *CSIAAS6BG1231200HC*, encoding a major facilitator superfamily protein, was annotated with protein binding, while *CSIAAS6BG1240000HC*, encoding a NAC domain protein, was annotated with DNA binding activity (Table 2).

4 Discussion

4.1 Genetic diversity characteristics of colored wheat and comparison with common wheat

To evaluate the level of genetic diversity in colored wheat germplasm, genome-wide and sub-genome nucleotide diversity values were calculated for the 57 accessions. The results showed that the average genome-wide π value of colored wheat was 0.296×10^{-3} , with the A and B sub-genomes exhibiting significantly higher π values of 0.319×10^{-3} and 0.318×10^{-3} , respectively, compared with 0.250×10^{-3} for the D sub-genome. This distribution pattern is consistent with the resequencing results of 145 representative Chinese wheat cultivars reported by Hao *et al.* (2020).

Notably, the genome-wide π value of colored wheat in this study was 0.296, which is higher than that of modern cultivars at 1.31×10^{-3} and landraces at 0.99×10^{-3} reported by Hao *et al.* (2020). This finding suggests that colored wheat germplasm has retained relatively high genetic diversity, likely due to less intense selection pressure during breeding as a special type of germplasm. It may also reflect that the abundant genetic variation present in colored wheat germplasm has not yet been fully exploited in cultivar improvement. From the perspective of genetic diversity, this result supports the unique value of colored wheat as a specialized germplasm resource for quality improvement and functional variety development.

4.2 Summary and comparison of QTL mapping results

In this study, a total of 176 stable QTLs associated with four nutritional quality traits were identified from 57 colored wheat accessions using two association models, EMMAX and PCA+GLM. These included 26 QTLs for ANs, 26 for CF, 19 for FAA, and 105 for TN (Table S3). This result highlights the complex genetic architecture underlying these four traits. Notably, the number of QTLs detected for TN substantially exceeded that for the other three traits, suggesting that TN content is governed by a larger number of genetic loci, which may be attributed to the involvement of multiple

genes in the intricate physiological processes of nitrogen absorption, transport, assimilation, and remobilization (Kasemsap *et al.*, 2023; Shi *et al.*, 2022). In contrast, only 19 QTLs were identified for FAA. This comparatively low number may reflect a more genetically concentrated regulation of free amino acid content, or alternatively, may be related to the highly dynamic nature of FAA as a transient metabolic intermediate, whose content is susceptible to sampling time and measurement conditions, thereby reducing the statistical power for QTL detection (Peng *et al.*, 2018; Zhong *et al.*, 2018).

Comparison with previously published GWAS studies on wheat nutritional quality revealed that a considerable proportion of the TN-related QTLs identified in this study were distributed on chromosomes 2B, 4A, and 4B, which is highly consistent with the major distribution intervals of GPC QTLs reported in common wheat. For instance, Li *et al.* (2023) mapped a stable QTL, *MQTL2B.3*, on wheat chromosome 2B, which coincided with *TN.pkuias.2B.11* identified in this study at the physical position of 752.47-755.44 Mb. Guo *et al.* (2023) and Li *et al.* (2023) independently identified loci *MQTL4A.3* and *MQTL4A.4* significantly associated with grain nitrogen content on chromosome 4A, corresponding to *TN.pkuias.4A.7* and *TN.pkuias.4A.8* in our study, with physical positions of 571.31-599.00 Mb and 718.67-718.67 Mb, respectively. Jadon *et al.* (2023) reported a stable protein content-related QTL, *QGpc.iari_4B*, on chromosome 4B with a relatively large interval that encompassed three QTLs identified in our study, namely *TN.pkuias.4B.17*, *TN.pkuias.4B.18*, and *TN.pkuias.4B.19* (Table 1; Table S3). Notably, the 105 TN QTLs identified in this study were distributed across all 20 chromosomes except 3D, suggesting that nitrogen use efficiency in colored wheat may be regulated by numerous minor-effect loci throughout the genome.

For the ANs trait, among the 26 QTLs identified, *ANs.pkuias.7B.3* on chromosome 7B harbored the largest number of significant SNPs and represented the most reliable QTL for this trait. Previous studies on wheat anthocyanin content have predominantly focused on chromosomes 1A, 3A, 3B, 6A, 7A, and 7D, whereas the major locus identified on chromosome 7B in this study provides a novel candidate region for anthocyanin regulation, suggesting that chromosome 7B may also harbor as-yet-

undiscovered anthocyanin regulatory genes. Zhou *et al.* (2017) mapped four loci, *GCR-QTL1*, *GCR-QTL20*, *GCR-QTL27*, and *GCR-QTL30*, on chromosomes 1A, 3B, 6A, and 7A, respectively, which shared identical physical positions with *ANs.pkuias.1A.1*, *ANs.pkuias.3B.1*, *ANs.pkuias.6A.1*, and *ANs.pkuias.7A* identified in our study on the corresponding chromosomes. Rasheed *et al.* (2016) cloned the gene *TaPHS1* on chromosome 3A, which fell within the interval of *ANs.pkuias.3A* identified in our study. Meanwhile, Bi *et al.* (2014) cloned the gene *TaDFR-B* on chromosome 3B, which was located within the interval of *ANs.pkuias.3B.2* identified in our study. Furthermore, the distribution of ANs QTLs across 14 chromosomes indicates that anthocyanin content is a complex trait regulated by a multi-gene network.

For the CF trait, the 26 QTLs identified in this study were distributed across 12 chromosomes, among which *CF.pkuias.2B.4* on chromosome 2B contained the largest number of significant SNPs. Previous QTL studies on wheat grain fat content have been relatively scarce. For example, Yu *et al.* (2025) mapped a gene *TaFAD8-D* encoding fatty acid desaturase 8 on chromosome 2D, which was negatively correlated with heat tolerance. Additionally, Yang *et al.* (2022) cloned *TaWRIIL2*, a homologue of the *Arabidopsis AtWRII* gene in wheat, and demonstrated its important regulatory roles in fatty acid synthesis, storage lipid accumulation, and seed development. The identification of CF-related QTLs in our study provides new locus information for the genetic regulation of oil accumulation in wheat. GWAS studies on the FAA trait have also been rarely reported in crops, with only a few publications available. Peng *et al.* (2018) identified a gene *TraesCS1A01G390300* encoding glutamic acid on chromosome 1A, which shared an identical physical position with *FAA.pkuias.1A* identified in our study on chromosome 1A. Gao *et al.* (2016) identified a gene *TaASN3.1* associated with amino acid content on chromosome 1B, whose physical position was close to *FAA.pkuias.1B* identified in our study, with a difference of approximately 5 Mb. Lavoignat *et al.* (2024) mapped a QTL on chromosome 2B that was close to *FAA.pkuias.2B* in our study, with a difference of approximately 8 Mb. The 19 QTLs identified in this study provide valuable initial information for the genetic dissection of free amino acid content in wheat. These QTLs were relatively scattered

across chromosomes without obvious hotspot regions, which may reflect the regulation of free amino acid metabolism by multiple independent metabolic pathways.

From a methodological perspective, the co-localization strategy combining two models, EMMAX and PCA+GLM, effectively enhanced the reliability of QTL detection in this study. EMMAX reduces false-positive rates by controlling population structure and relatedness (Kang *et al.*, 2010), while PCA+GLM performs association tests after correcting for population stratification using principal components. Loci co-localized by both models are considered to have higher confidence (Yang *et al.*, 2020).

4.3 Breeding value of KASP markers

In this study, KASP markers, namely *KASP-ANs-7B*, *KASP-CF-2B*, and *KASP-TN-6B*, were successfully developed for three major QTLs, *ANs.pkuias.7B.3*, *CF.pkuias.2B.4*, and *TN.pkuias.6B.4*, and their significant associations with the target traits were validated in the 57 colored wheat accessions. The genotyping results of the three markers revealed clear clustering patterns and good polymorphism in the test population (Fig. 3A-3C), indicating that these markers can be directly applied as stable molecular tools for marker-assisted selection in colored wheat breeding.

Single-marker analysis further confirmed the specific effects of the three markers on their target traits. *KASP-ANs-7B* and *KASP-CF-2B* exhibited significant positive effects only on their respective target traits, with no notable influence on other nutritional quality traits, indicating that these two loci possess good specificity in regulating metabolic pathways. *KASP-TN-6B*, however, displayed a more complex effect pattern: while significantly increasing TN content, it simultaneously exerted a significant negative effect on ANs content. This result is consistent with the highly significant negative correlation observed between ANs and TN in the correlation analysis, suggesting that anthocyanin biosynthesis and nitrogen assimilation may share some regulatory nodes or compete for carbon metabolic flux. Such negative associations are not uncommon in the relationship between primary and secondary metabolism in plants, as plants tend to allocate more carbon skeletons and energy to

primary metabolism such as amino acid synthesis when nitrogen is abundant, while reducing resource investment in secondary metabolites such as ANs (Kasemsap *et al.*, 2023; Zhong *et al.*, 2023). This finding suggests that when utilizing *KASP-TN-6B* to increase TN content in practical breeding, concurrent attention should be paid to changes in anthocyanin content, and if necessary, this negative effect should be compensated by pyramiding other ANs-enhancing loci.

In terms of agronomic traits, the three KASP markers exhibited differentiated pleiotropic patterns. *KASP-ANs-7B* and *KASP-CF-2B* showed no significant effects on any of the six major agronomic traits, indicating that these two quality trait loci do not have obvious negative effects on yield and can be directly applied to quality improvement breeding without concerns about compromising yield performance. This finding is consistent with previously reported studies showing that some GPC QTLs do not significantly reduce yield (Liu *et al.*, 2024; Nigro *et al.*, 2019), providing a basis for the coordinated improvement of quality and yield in colored wheat. In contrast, *KASP-TN-6B* significantly reduced GNS and increased PH while enhancing TN content, exhibiting pronounced pleiotropy. Both GNS and PH are important yield components, and the negative effect of this locus on GNS may adversely affect final yield. However, the lack of significant effects on KNS, SL, SN, and TKW suggests that the negative yield effect may be partially mitigated through compensation by other yield components. In terms of breeding strategy, if this TN-enhancing locus is to be utilized, marker-assisted selection can be employed to identify recombinant individuals carrying the locus with reduced impact on GNS, or those with a genetic background harboring GNS-enhancing loci, to minimize or eliminate the negative effect of this locus.

It is noteworthy that the *KASP-FAA-7B* marker developed for FAA failed to genotype successfully, and single-marker analysis did not detect any significant association between this locus and FAA content. This result may be attributed to the following reasons: first, the SNP loci within the QTL interval used for KASP primer design lacked sufficient polymorphism to enable effective allele genotyping in the test population; second, the genetic architecture of FAA may be inherently complex, and this locus might be a minor-effect QTL whose effect is insufficient to be detected by

single-marker analysis with the current population size; third, FAA content is highly susceptible to environmental factors, and the single-measurement phenotypic data might introduce substantial environmental noise, thereby reducing the sensitivity of QTL detection. This result suggests that the effectiveness of a single-marker strategy for traits such as FAA, which are controlled by multiple genes and are highly environmentally sensitive, may be limited. Future studies could consider employing multi-marker pyramiding or genomic selection strategies for genetic improvement.

4.4 Biological functions of candidate genes

To explore the potential molecular basis of the four major QTLs, we extended the physical interval of each QTL by 3 Mb on both sides and obtained 713 candidate genes in total. Among these, 15 representative candidate genes were further prioritized, and their potential functions were examined through GO and KEGG annotations (Table 2).

The *ANs.pkuias.7B.3* locus harbored seven candidate genes related to transcriptional regulation, among which six belonged to the zinc finger family. Zinc finger proteins are well-recognized transcription regulators, and accumulating evidence has implicated them in the activation of anthocyanin biosynthetic pathways (Fang *et al.*, 2019; Ren *et al.*, 2025; Wang *et al.*, 2022). Notably, *CSIAAS7BG1255000HC* was annotated by KEGG as mitochondrial protein import protein ZIM17, suggesting a possible indirect effect on anthocyanin biosynthesis through mitochondrial function. In addition, *CSIAAS7BG1256000HC* encodes a bHLH transcription factor, which commonly forms MBW ternary complexes with MYB and WD40 proteins to synergistically activate downstream genes in anthocyanin biosynthesis (Jiang *et al.*, 2018; Liu *et al.*, 2021). The co-occurrence of zinc finger proteins and a bHLH transcription factor at this locus implies that *ANs.pkuias.7B.3* may influence anthocyanin accumulation through a transcriptional regulatory network, which is consistent with its identification as a major QTL for ANs content.

At the *CF.pkuias.2B.4* locus, five candidate genes were identified, covering multiple regulatory levels of lipid biosynthesis. The WRKY transcription factor

CSIAAS2BG0786400HC was annotated by KEGG as WRKY transcription factor group IIc, and WRKY family members have been reported to participate in the transcriptional regulation of lipid biosynthesis (Yang *et al.*, 2022). The enrichment of three bHLH transcription factors further suggests the presence of a complex transcriptional regulatory network at this locus, and bHLH transcription factors have also been associated with lipid metabolism regulation (Liu *et al.*, 2025; Zhao *et al.*, 2026). Although *CSIAAS2BG0824000HC* lacked GO or KEGG annotations, its protein description as a 3-oxoacyl-reductase indicates its catalytic role in fatty acid carbon chain elongation (Kim *et al.*, 2026). From transcriptional regulation to fatty acid biosynthesis, this locus covers multiple critical steps of lipid accumulation, providing a reasonable biochemical basis for *CF.pkuias.2B.4* as a major QTL for CF content.

Only one candidate gene, *CSIAAS7BG1012200HC*, was identified at the *FAA.pkuias.7B* locus. This gene encodes allantoate amidohydrolase, which participates in purine metabolism. This enzyme functions in the nitrogen cycle by catalyzing allantoate degradation, thereby influencing nitrogen remobilization (Islam *et al.*, 2021; Yazdanpanah *et al.*, 2022). Although FAAs are not directly linked to purine metabolism, this gene may indirectly modulate FAA homeostasis by affecting overall nitrogen metabolism. The limited number of candidate genes at this locus is consistent with the failure of KASP marker development and the lack of significant association in single-marker analysis, suggesting that the genetic basis of FAA content may involve more minor-effect loci or non-coding regulatory elements.

At the *TN.pkuias.6B.4* locus, two candidate genes were identified. *CSIAAS6BG1231200HC* encodes a major facilitator superfamily protein with protein binding activity, potentially involved in transmembrane transport of nitrogen-related compounds (Niño-González *et al.*, 2019). *CSIAAS6BG1240000HC* encodes a NAC domain protein with DNA binding activity. Notably, NAM/GPC-B1, a well-characterized NAC transcription factor, has been demonstrated to regulate grain protein and TN content (Andleeb *et al.*, 2023; Kartseva *et al.*, 2023). Therefore, *CSIAAS6BG1240000HC* may represent a key candidate gene through which this locus regulates TN content. However, no genes directly involved in anthocyanin biosynthesis

were identified within this interval. This suggests that the negative effect of *TN.pkuia.6B.4* on anthocyanin content may not arise from direct interference with anthocyanin pathway genes, but rather from altered nitrogen assimilation efficiency affecting the overall carbon-nitrogen metabolic balance, which in turn diminishes anthocyanin accumulation. This interpretation aligns with the pleiotropic effect pattern observed in single-marker analysis.

In summary, the enrichment of candidate genes with defined functional annotations within the four major QTL intervals provides molecular support for the authenticity of each QTL, as the concentration of genes with recognizable physiological functions generally indicates genuine involvement in trait regulation. Nevertheless, these candidate gene predictions remain at the bioinformatics level, and further validation through gene expression profiling, mutant phenotyping, or transgenic complementation experiments is still required.

4.5 Strategies and perspectives for quality improvement in colored wheat

This study provides abundant QTL resources and directly applicable KASP marker tools for nutritional quality improvement in colored wheat, carrying significant theoretical and practical implications for advancing molecular breeding in colored wheat.

In terms of marker-assisted selection, *KASP-ANs-7B* and *KASP-CF-2B*, which have no significant negative effects on agronomic traits, can be directly incorporated into quality breeding programs for directional selection of ANs and CF content. For *KASP-TN-6B*, given its negative effect on GNS and positive effect on PH, careful evaluation is warranted in breeding applications. A feasible strategy would be to combine the selection of the TN-enhancing allele with background selection using other independently inherited GNS-enhancing markers to offset the negative effect of this locus on GNS. Additionally, screening for superior recombinants that retain the TN-enhancing effect but have eliminated the negative effect on GNS through large-scale recombination screening represents a worthwhile breeding strategy to pursue.

In terms of coordinated improvement of multiple traits, the QTL information generated in this study provides direction for trait pyramiding. Although we did not systematically test for QTL co-localization, the positive correlation between ANs and CF and the negative correlation between ANs and TN observed in the correlation analysis suggest that some QTLs may simultaneously regulate multiple traits. Pleiotropic QTLs of this nature can be leveraged to achieve coordinated improvement across multiple traits in breeding programs. Meanwhile, for trait combinations with negative correlations, such as ANs and TN, marker-assisted pyramiding of multiple independent loci to achieve trait balance through additive effects would be necessary.

In recent years, genomic selection has been increasingly widely applied in crop breeding (Juliana *et al.*, 2019; Schulthess *et al.*, 2022). The GWAS information obtained in this study can provide important prior information for constructing genomic selection prediction models, and incorporating significantly associated marker loci as fixed effects into genomic selection models is expected to improve prediction accuracy. Furthermore, with the continuous improvement of reference genome and pan-genome resources for colored wheat (Cheng *et al.*, 2024; Walkowiak *et al.*, 2020), more diverse types of genetic variation, such as structural variations, could be incorporated into GWAS analyses in the future to further elucidate the genetic regulatory networks of quality traits. Colored wheat has broad market prospects as a functional food ingredient (Garg *et al.*, 2022; Padhy *et al.*, 2024). The ANs content-related QTLs and candidate genes identified in this study provide an effective pathway for directionally increasing anthocyanin content through molecular breeding, which will facilitate the development of new functional colored wheat varieties rich in natural antioxidant-active compounds to meet the growing demand for specialty agricultural products in the health industry.

This study has certain limitations. Phenotypic data were derived from only a single measurement, which precludes the assessment of genotype-by-environment interaction effects on QTL detection. Wheat quality traits are greatly influenced by environmental conditions such as planting location, annual climate, and nitrogen fertilizer application level (Kartseva *et al.*, 2023; Muqaddasi *et al.*, 2020), which may have resulted in the failure to detect certain environment-specific QTLs or an underestimation of QTL effect

sizes. In addition, the population size of 57 accessions is relatively modest for GWAS analysis, which may limit the power to detect rare alleles or minor-effect QTLs. Future efforts to expand population size or integrate bulked segregant analysis for complementary validation would help refine the QTL map. It should also be noted that the candidate genes identified in this study are based on reference genome annotation and bioinformatics prediction, and their biological functions require validation through molecular biology experiments such as qRT-PCR expression analysis, virus-induced gene silencing, gene editing, or transgenic overexpression. The causal relationships between candidate genes and traits remain to be further confirmed through genetic transformation or mutant analysis.

5 Conclusion

This study performed whole-genome resequencing and phenotypic evaluation of 57 colored wheat accessions for four nutritional quality traits, identifying 176 stable QTLs through combined EMMAX and PCA plus GLM models, with 84.66 percent being novel discoveries. Three KASP markers were successfully developed for the major QTLs of ANs, CF, and TN, and validated through single-marker analyses across metabolic and agronomic traits, providing practical tools for marker-assisted selection. Notably, the total nitrogen-associated locus exhibited negative pleiotropic effects on anthocyanin accumulation, revealing a genetic trade-off between primary and secondary metabolism. Candidate gene analysis within extended QTL intervals prioritized 15 functional genes involved in anthocyanin biosynthesis, lipid metabolism, and nitrogen assimilation. These findings substantially enrich the genetic resources and molecular tools for nutritional quality improvement in colored wheat breeding.

CRedit authorship contribution statement

Yanyan Pu: Resources, Conceptualization, Writing - original draft. Runfang Li: Visualization. Dezhou Cui: Visualization. Xiaojian Fang: Software. Xuan Guo: Data curation. Xin Jing: Formal analysis. Yongchao Hao: Software. Xiaofeng Cao:

Methodology. Rumei Tian: Investigation. Song Hou: Formal analysis. Shanshan Liu: Software. Shasha Wang: Software. Nana Li: Resources, Supervision, Validation. Yongchao Gong: Project administration, Funding acquisition. Chengzhi Jiao: Writing - review and editing, Project administration, Conceptualization, Funding acquisition. Liang Wang: Writing - original draft, Writing - review and editing, Conceptualization.

Acknowledgments

This work was supported by the Technology Innovation Project of the Shandong Academy of Agricultural Sciences (CXGC2026B21), Shandong Rural Revitalization Science and Technology Innovation Promotion Action Plan (2025TZXD006), the Agricultural Variety Improvement Project of Shandong Province (2025LZGC009), the Key R&D Program of Shandong Province, China (2025CXPT008) and Shandong Provincial Natural Science Foundation (ZR2025QB03 and SYS202206), the Key R&D Program of Shandong Province, China (Innovative Capacity Improvement Project for Science and Technology-based SMEs, No. 2025TSGCCZZB0252), the Qingdao Special Program for Science and Technology Benefiting People Demonstration (No. 26-1-5-xdny-3-nsh), and the Qingdao Key Technology Research and Industrialization Demonstration Project (No. 26-1-1-gjgg-70-nsh).

Conflict of interest

The authors declare that they have no conflict of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used DeepSeek to assist in improving the clarity and readability of the language. All content generated was carefully reviewed and revised by the authors as necessary, and the authors take full responsibility for the final content of the article.

Data

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2025) in National Genomics Data Center (Nucleic Acids Res 2025), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA045555) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.

References

- Andleeb, T., Knight, E. & Borrill, P. Wheat NAM genes regulate the majority of early monocarpic senescence transcriptional changes including nitrogen remobilization genes. *G3 Genes|Genomes|Genetics*. **13**, jkac275 (2023).
- Bashir, L. et al. Identification of quantitative trait nucleotides for grain quality in bread wheat under heat stress. *Scientific Reports*. **15**, 6641 (2025).
- Bi, H. H., Sun, Y. W., Xiao, Y. G. & Xia, L. Q. 2014. Characterization of DFR allelic variations and their associations with pre-harvest sprouting resistance in a set of red-grained Chinese wheat germplasm. *Euphytica*. **195**, 197-207 (2014).
- Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. **4**, s13742-13015-10047-13748 (2015).
- Chen, L. et al. Genome-wide association study of pre-harvest sprouting resistance and grain color in common wheat (*Triticum aestivum* L.). *BMC Plant Biology*. **25**, 973 (2025).
- Cheng, S. F. et al. Harnessing landrace diversity empowers wheat breeding. *Nature*. **632**, 823-831 (2024).
- Fang, H. C. et al. The B-box zinc finger protein MdBBX20 integrates anthocyanin accumulation in response to ultraviolet radiation and low temperature. *Plant, Cell & Environment*. **42**, 2090-2104 (2019).
- Flint-Garcia, S. A., Thornsberry, J. M. & Buckler, E. S. Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*. **54**, 357-374 (2003).
- Gao, J. et al. Deciphering genetic basis of developmental and agronomic traits by integrating high-throughput optical phenotyping and genome-wide association

704 studies in wheat. *Plant Biotechnology Journal*. **21**, 1966-1977 (2023).

705 Gao, R. H. et al. Food safety: Structure and expression of the asparagine synthetase
706 gene family of wheat. *Journal of Cereal Science*. **68**, 122-131 (2016).

707 Garg, M. et al. Rising Demand for Healthy Foods-Anthocyanin Biofortified Colored
708 Wheat Is a New Research Trend. *Frontiers in Nutrition*. **9**, 878221 (2022).

709 Guo, Y. X. et al. Genetic dissection of protein and starch during wheat grain
710 development using QTL mapping and GWAS. *Frontiers in Plant Science*. **14**,
711 1189887 (2023).

712 Hao, C. Y. et al. Resequencing of 145 landmark cultivars reveals asymmetric sub-
713 genome selection and strong founder genotype effects on wheat breeding in
714 China. *Molecular Plant*. **13**, 1733-1751 (2020).

715 Hao, S. Y. et al. Genome-Wide Association Study Reveals the Genetic Basis of Five
716 Quality Traits in Chinese Wheat. *Frontiers in Plant Science*. **13**, 835306 (2022).

717 Hong, M. J., Ko, C. S. & Kim, D. Y. Genome-Wide Association Study of Agricultural
718 and Biochemical Traits in Radiation-Induced Colored Wheat. *Agronomy*. **158**,
719 1933 (2025).

720 Huang, C. et al. Population structure and genetic basis of the agronomic traits of upland
721 cotton in China revealed by a genome-wide association study using high-density
722 SNPs. *Plant Biotechnology Journal*. **15**, 1374-1386 (2017).

723 Islam, S., Zhang, J., Zhao, Y., She, M. & Ma, W. Genetic regulation of the traits
724 contributing to wheat nitrogen use efficiency. *Plant Science*. **303**, 110759 (2021).

725 IWGSC et al. Shifting the limits in wheat research and breeding using a fully annotated
726 reference genome. *Science*. **361**, eaar7191 (2018).

727 Jadon, V. et al. Molecular Mapping of Biofortification Traits in Bread Wheat (*Triticum*
728 *aestivum* L.) Using a High-Density SNP Based Linkage Map. *Genes*. **14**, 221
729 (2023).

730 Jiang, W. H. et al. Two transcription factors *TaPpm1* and *TaPpb1* co-regulate
731 anthocyanin biosynthesis in purple pericarps of wheat. *Journal of Experimental*
732 *Botany*. **69**, 2555-2567 (2018).

733 Jiao, C. Z. et al. Fast integration and accumulation of beneficial breeding alleles through

734 an AB-NAMIC strategy in wheat. *Plant Communications*. **4**, 100549 (2023).
 735 Jiao, C. Z. et al. Accumulation of beneficial haplotypes in Huang-Huai-Hai wheat
 736 region and its application in molecular breeding. *Journal of Integrative*
 737 *Agriculture*. 10.1016/j.jia.2024.12.003 (2024).
 738 Jiao, C. Z. et al. Pan-genome bridges wheat structural variations with habitat and
 739 breeding. *Nature*. **637**, 384-393 (2025).
 740 Juliana, P. et al. Improving grain yield, stress resilience and quality of bread wheat using
 741 large-scale genomics. *Nature Genetics*. **51**, 1530-1539 (2019).
 742 Kang, H. M. et al. Variance component model to account for sample structure in
 743 genome-wide association studies. *Nature Genetics*. **42**, 348-354 (2010).
 744 Kartseva, T. et al. Nutritional Genomic Approach for Improving Grain Protein Content
 745 in Wheat. *Foods*. **12**, 1399 (2023).
 746 Kasemsap, P. & Bloom, A. J. Breeding for Higher Yields of Wheat and Rice through
 747 Modifying Nitrogen Metabolism. *Plants*. **12**, 85 (2023).
 748 Kim, H. et al. Polyketide synthase-like functionality acquired by plant fatty acid
 749 elongase. *Science Advances*. **12**, eaed2892 (2026).
 750 Lavoignat, M. et al. Different wheat loci are associated to heritable free asparagine
 751 content in grain grown under different water and nitrogen availability.
 752 *Theoretical and Applied Genetics*. **137**, 46 (2024).
 753 Li, N. et al. Consensus genomic regions for grain quality traits in wheat revealed by
 754 Meta-QTL analysis and in silico transcriptome integration. *The Plant Genome*.
 755 **16**, e20336 (2023).
 756 Lin, M. et al. Genome-wide association analysis on pre-harvest sprouting resistance
 757 and grain color in U.S. winter wheat. *BMC genomics*. **17**, 794 (2016).
 758 Liu, S. C. et al. A telomere-to-telomere genome assembly coupled with multi-omic data
 759 provides insights into the evolution of hexaploid bread wheat. *Nature Genetics*.
 760 **57**, 1008-1020 (2025).
 761 Liu, X. et al. *TbMYC4A* is a candidate gene controlling the blue aleurone trait in a
 762 wheat-triticum boeoticum substitution line. *Frontiers in Plant Science*. **12**,
 763 762265 (2021).

Liu, Z. H. et al. Selection and application of four QTLs for grain protein content in modern wheat cultivars. *Journal of Integrative Agriculture*. **23**, 2557-2570 (2024).

Liu, Z. J. et al. A bHLH7-PDF2 module mediates transcriptional regulation to repress seed oil accumulation in *Arabidopsis thaliana*. *Plant Communications*. **6**, 101459 (2025).

Ma, S. W. et al. WheatOmics: A platform combining multiple omics data to accelerate functional genomics studies in wheat. *Molecular Plant*. **14**, 1965-1968 (2021).

Mandrioli, M. et al. Lipids and Fatty Acid Composition Reveal Differences between Durum Wheat Landraces and Modern Cultivars. *Plants*. **13**, 1817 (2024).

Marcotuli, I. et al. Genetic Study of Total Phenolic Content and Antioxidant Activity Traits in Tetraploid Wheat via Genome-Wide Association Mapping. *Antioxidants*. **14**, 1048 (2025).

Muqaddasi, Q. H. et al. Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. *Scientific Reports*. **10**, 12541 (2020).

Nigro, D. et al. Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat. *Planta*. **249**, 1157-1175 (2019).

Niño-González, M., Novo-Uzal, E., Richardson, D. N., Barros, P. M. & Duque, P. More transporters, more substrates: The *Arabidopsis* major facilitator superfamily revisited. *Molecular Plant*. **12**, 1182-1202 (2019).

Padhy, A. K., Kaur, P., Singh, S., Kashyap, L. & Sharma, A. Colored wheat and derived products: key to global nutritional security. *Critical Reviews in Food Science and Nutrition*. **64**, 1894-1910 (2024).

Peng, Y. C. et al. Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Frontiers in Plant Science*. **9**, 1196 (2018).

Ramirez-Gonzalez, R. H., Uauy, C. & Caccamo, M. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics*. **31**, 2038-2039 (2015).

Rasheed, A. et al. Development and validation of KASP assays for genes underpinning key economic traits in bread wheat. *Theoretical and Applied Genetics*. **129**,

1843-1860 (2016).

Ren, Y. X. et al. C2H2-type zinc finger protein transcription factor MdZAT1 plays a negative role in anthocyanin biosynthesis in apple. *Molecular Horticulture*. **5**, 28 (2025).

Schulthess, A. W. et al. Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. *Nature Genetics*. **54**, 1544-1552 (2022).

Shewry, P. R. & Hey, S. J. The contribution of wheat to human diet and health. *Food and Energy Security*. **4**, 178-202 (2015).

Shi, H. W. et al. Genome-wide association study of agronomic traits related to nitrogen use efficiency in wheat. *Theoretical and Applied Genetics*. **135**, 4289-4302 (2022).

Walkowiak, S. et al. Multiple wheat genomes reveal global variation in modern breeding. *Nature*. **588**, 277-283 (2020).

Wang, D. R. et al. Overexpression of *MdZAT5*, an C2H2-type zinc finger protein, regulates anthocyanin accumulation and salt stress response in apple calli and *Arabidopsis*. *International Journal of Molecular Sciences*. **23**, 1897 (2022).

Wang, L. et al. Characterization of adult plant leaf rust resistance loci in an international wheat collection using a re-sequencing genotyping system. *The Crop Journal*. **13**, 1563-1573 (2025).

Wu, Z. et al. Advances in understanding wheat grain color: Genetic, nutritional, and agronomic perspectives. *Agronomy*. **15**, 1108 (2025).

Xu, D. D. et al. *HvGST* plays a key role in anthocyanin accumulation in colored barley. *The Plant Journal*. **113**, 47-59 (2023).

Yang, F. P. et al. Cloning and functional analysis of *TaWRILs*, the key genes for grain fatty acid synthesis in bread wheat. *International Journal of Molecular Sciences*. **23**, 5293 (2022).

Yang, Y. et al. Multi-locus GWAS of quality traits in bread wheat: Mining more candidate genes and possible regulatory network. *Frontiers in Plant Science*. **11**, 1091 (2020).

Yang, Y. Z. et al. Transcriptional regulation of oil biosynthesis in seed plants: Current

understanding, applications, and perspectives. *Plant Communications*. **3**, 100328 (2022).

Yazdanpanah, F., Willems, L. A. J., He, H., Hilhorst, H. W. M. & Bentsink, L. A role for allantoate amidohydrolase (*AtAAH*) in the germination of *Arabidopsis thaliana* seeds. *Plant Cell Physiol.* **63**, 1298-1308 (2022).

Yin, L. L. et al. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*. **19**, 619-628 (2021).

Yu, H. J. et al. Natural variation in *TaFAD8-D* promoter enhances thermotolerance in wheat through fatty acid and lipid remodelling. *Plant Biotechnology Journal*. **24**, 1251-1268 (2025).

Zhao, Y. et al. The bHLH112-ATML1 transcriptional module represses oil accumulation in seeds. *The Plant Journal*. **125**, e70832 (2026).

Zhong, C. et al. Metabolomic analysis reveals patterns of whole wheat and pearling fraction flour quality response to nitrogen in two wheat lines with contrasting protein content. *Journal of Agricultural and Food Chemistry*. **71**, 2290-2300 (2023).

Zhong, Y. X. et al. Nitrogen topdressing timing modifies free amino acids profiles and storage protein gene expression in wheat grain. *BMC Plant Biology*. **18**, 353 (2018).

Zhou, Y. et al. Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of chinese wheat landraces. *Frontiers in Plant Science*. **8**, 401 (2017).

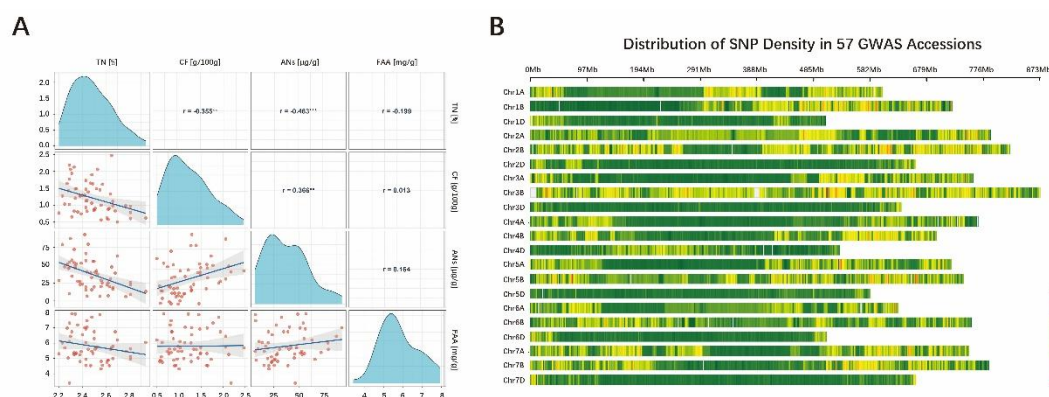


Figure 1. Correlation analysis of nutritional quality traits and genome-wide SNP density in colored wheat.

A. Scatterplot matrix of four nutritional quality traits. Upper panel: Pearson correlation coefficients with significance levels (** $P < 0.01$, *** $P < 0.001$). Diagonal: Kernel density distributions of each trait. Lower panel: Scatter plots with linear regression lines (blue solid line) and 95% confidence intervals (gray shading). TN, total nitrogen (%); CF, crude fat (g/100g); ANs, anthocyanins (μg/g); FAA, free amino acids (mg/g). **B. Genome-wide SNP density distribution.** X-axis, physical position (Mb); Y-axis, chromosome.

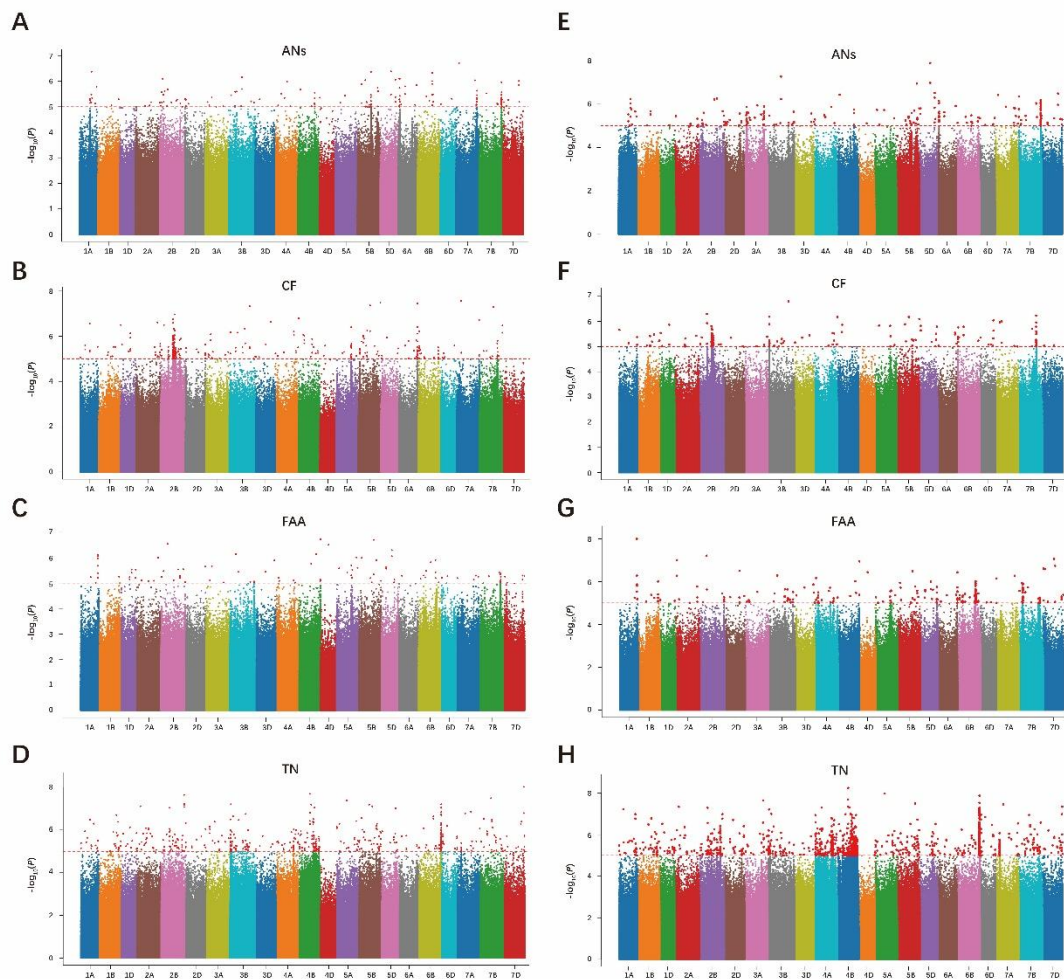


Figure 2. Manhattan plots of genome-wide association studies for four nutritional quality traits.

A-D. Association results using the EMMAX model. A. Anthocyanins (ANs); B. Crude fat (CF); C. Free amino acids (FAA); D. Total nitrogen (TN). **E-H. Association results using the PCA+GLM model.** E. Anthocyanins (ANs); F. Crude fat (CF); G. Free amino acids (FAA); H. Total nitrogen (TN). X-axis, chromosome (Chromosome 1A-7D); Y-axis, $-\log_{10}(P)$. The horizontal red dashed line indicates the significance threshold ($-\log_{10}(P)=5$, corresponding to $P=1\times 10^{-5}$). Each dot represents an individual SNP marker, and markers exceeding the threshold are considered significantly associated with the respective trait.

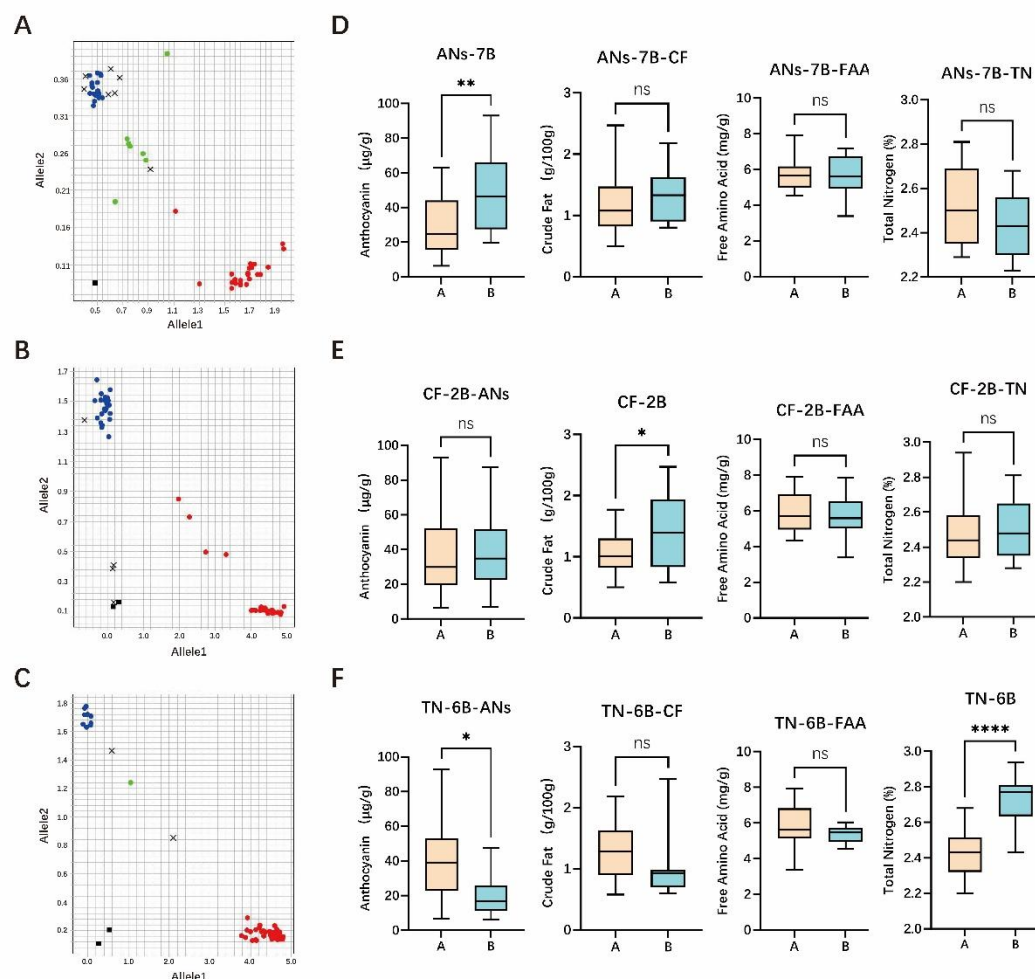


Figure 3. KASP marker genotyping and single-marker analysis for metabolic traits.

A-C. KASP genotyping results. X-axis, Allele 1; Y-axis, Allele 2. Red dots: homozygous recessive genotype (A); Blue dots: homozygous positive genotype (B); Green dots: heterozygous genotype; Crosses (×): missing data; Black dots: negative control (NTC). A. Marker *KASP-ANs-7B*; B. Marker *KASP-CF-2B*; C. Marker *KASP-TN-6B*. **D-F. Single-marker analysis of three KASP markers for four metabolic traits.** Orange boxplots: genotype A group; Blue boxplots: genotype B group. Traits: ANs (anthocyanins, $\mu\text{g/g}$), CF (crude fat, $\text{g}/100\text{g}$), FAA (free amino acids, mg/g), TN (total nitrogen, %). Significance levels: * $P<0.05$, ** $P<0.01$, **** $P<0.0001$, ns, not significant. D. Marker *KASP-ANs-7B*; E. Marker *KASP-CF-2B*; F. Marker *KASP-TN-6B*.

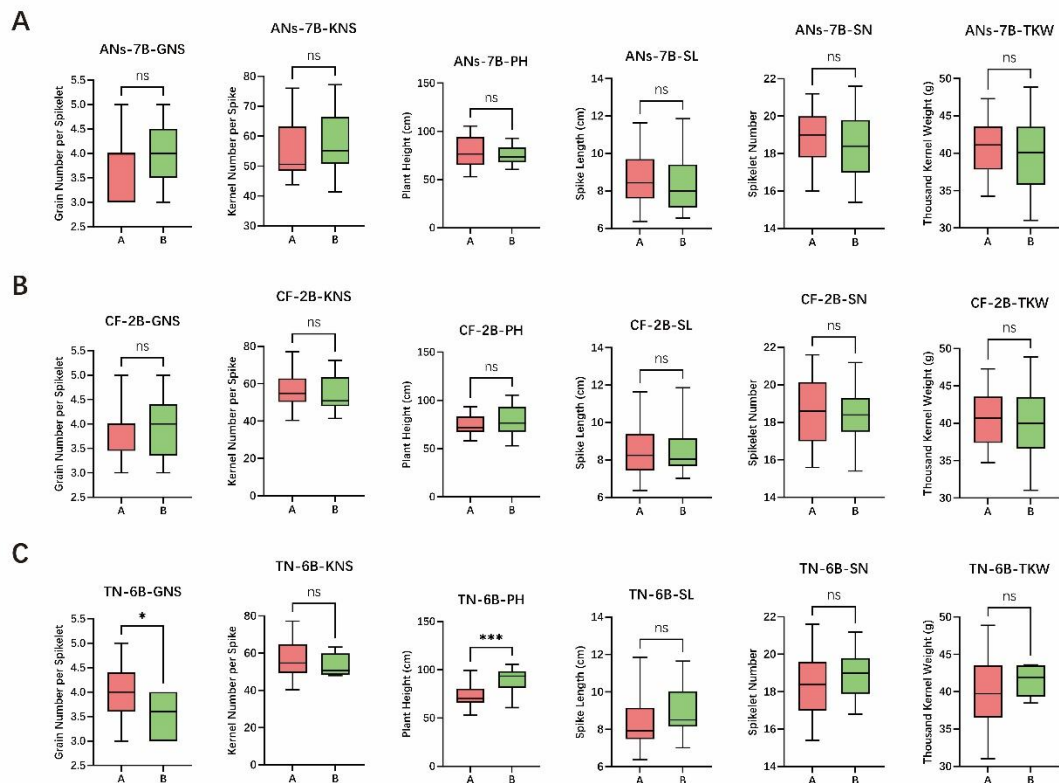


Figure 4. Single-marker analysis of three KASP markers for agronomic traits.

A-C. Single-marker analysis of three KASP markers for six agronomic traits. Red boxplots: genotype A group; Green boxplots: genotype B group. Traits: GNS (grain number per spikelet), KNS (kernel number per spike), PH (plant height, cm), SL (spike length, cm), SN (spikelet number per spike), TKW (thousand kernel weight, g). Significance levels: * $P < 0.05$, *** $P < 0.001$, ns, not significant. A. Marker *KASP-ANs-7B*; B. Marker *KASP-CF-2B*; C. Marker *KASP-TN-6B*.

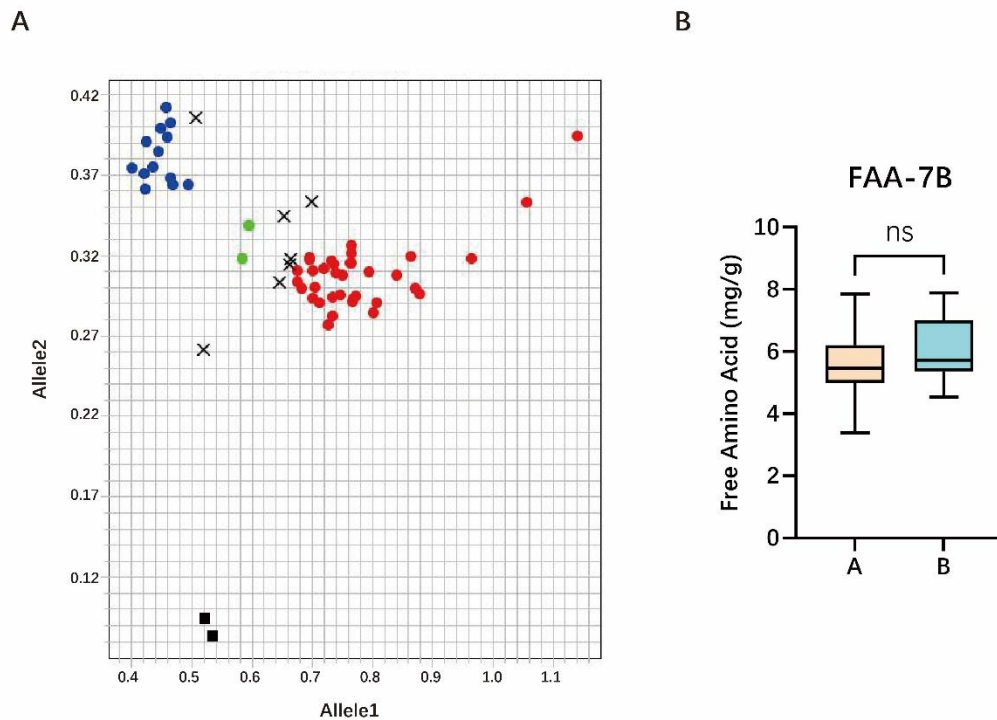


Figure S1. Genotyping and single-marker analysis of the KASP-FAA-7B marker.

A. KASP genotyping results. X-axis, Allele 1; Y-axis, Allele 2. Red dots: homozygous recessive genotype (A); Blue dots: homozygous positive genotype (B); Green dots: heterozygous genotype; Crosses (×): missing data; Black dots: negative control (NTC).

B. Single-marker analysis of *KASP-FAA-7B* for free amino acids (FAA). Orange boxplot: genotype A group; Blue boxplot: genotype B group. The Y-axis represents FAA content (mg/g). ns indicates no significant difference between the two genotype groups ($P>0.05$).

Supplementary Files

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

- [TableS2.pdf](#)
- [TableS3.pdf](#)
- [Table2.pdf](#)
- [TableS5.pdf](#)
- [Table1.pdf](#)
- [TableS4.pdf](#)
- [TableS1.pdf](#)