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.

Identification of QTNs, QTN-by-Environment Interactions, and Their Candidate Genes for Rice Grain-Related Traits across Years using 3VmrMLM

Suhong Bu

busuhong@scau.edu.cn

South China Agricultural University

Shuai Tu

South China Agricultural University

Fangping Li

South China Agricultural University

Qiong Zhao

Huazhong Agricultural University

Qingyu Song

South China Agricultural University

Hongxu Huang

South China Agricultural University

Jieying Liu

South China Agricultural University

Weifeng Yang

South China Agricultural University

Shaokui Wang

South China Agricultural University

Junliang Zhao

Guangdong Academy of Agricultural Sciences

Research Article

Keywords: Rice, Grain shape, Grain weight, 3VmrMLM, QTN-by-environment interaction, Candidate genes

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11350363/v1

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

Additional Declarations: No competing interests reported.

Identification of QTNs, QTN-by-Environment Interactions, and Their Candidate Genes for Rice Grain-Related Traits across Years using 3VmrMLM

Suhong Bu[1+*], Shuai Tu[1+], Fangping Li[1*], Qiong Zhao[3], Qingyu Song[1], Hongxu Huang[1], Jieying Liu[1], Weifeng Yang[1], Shaokui Wang[1], Junliang Zhao[1, 2*]

[1]Guangdong Basic Research Center of Excellence for Precise Breeding of Future Crops, Guangdong Provincial Key Laboratory of Plant Molecular Breeding, South China Agricultural University, Guangzhou, China.

[2]Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Key Laboratory of Rice Science and Technology, Guangzhou, China.

[3]College of Plant Science & Technology, Huazhong Agricultural University,Wuhan, China

[*]Correspond: Suhong Bu (busuhong@scau.edu.cn; ORCID: 0009-0006-8791-3145); Fangping Li (Fangping.li@scau.edu.cn); Junliang Zhao (zhao_junliang@gdaas.cn)

[+]These authors contributed equally to this work.

Abstract

Background: Grain shape and grain weight are major determinants of rice yield and quality. Although numerous loci controlling these traits have been identified, their stable and environment-dependent genetic components remain incompletely understood. This study aimed to dissect the genetic architecture of rice grain-related traits across years using a multi-locus 3VmrMLM algorithm.

Results: A globally diverse panel of 413 rice accessions, genotyped with 1,974,377 high-quality single-nucleotide polymorphisms, was evaluated for grain length (GL), grain width (GW), grain length-to-width ratio (GLWR), and thousand-grain weight (TGW) across years. 3VmrMLM, along with MLM and FarmCPU, was used to analyze their phenotypes and best linear unbiased prediction values. A total of 407 nonredundant quantitative trait nucleotides (QTNs) were identified, almost each explaining small proportion of phenotypic variation. Of these, 116 were consistently detected across at least two analytical datasets, while 44 previously reported grain-size genes (e.g., *GS3*, *GW5*, *GS2*, *GW7*, and *GS9*) were located near the QTNs detected by IIIVmrMLM. Multi-year joint analysis further identified 10 significant QTN-by-environment interactions (QEIs) for GL, GW, and GLWR, but none for TGW. Integrating common QTN/QEI positions, developmental transcriptome data, functional annotation, and haplotype analysis prioritized 14 new candidate genes for grain size or weight. Among them, *OsFKII/FRK2*, *OsPCF5*, and *OsSWN3/OsNAC31* were supported by multiple lines of evidence, whereas *PHYA* showed significant haplotype and haplotype-by-year interaction effects on GW.

Conclusions: Multi-locus GWAS identified both stable QTNs and QEIs associated with rice grain traits across years. These QTNs, QEIs, and prioritized candidate genes provide valuable genetic resources for functional validation and for marker-assisted and environment-aware breeding.

Keywords: Rice; Grain shape; Grain weight; 3VmrMLM; QTN-by-environment interaction; Candidate genes

Introduction

Rice (*Oryza sativa* L.) is grown nearly worldwide and provides the staple food for more than half of the global population. Grain size (GS), as an important determinant of grain quality and yield in rice, was subject to deliberate selection and breeding. It is generally characterized by a combination of grain length (GL), grain width (GW), and length-width ratio (GLWR), whereas thousand-grain weight (TGW) is a closely related yield component. These traits have undergone strong natural and artificial selection during rice domestication and breeding [1, 2]. Genetically, rice grain size and shape are complex quantitative traits determined by cell proliferation and expansion in the spikelet hull, grain filling, and the coordinated development of maternal and filial tissues. Numerous genes and quantitative trait loci (QTLs), including *GS3*, *GW5*, *GL3.1*, *GW8/OsSPL16*, *GL7/GW7*, and *GL10*, have been cloned and functionally characterized. These genes participate in interconnected regulatory pathways involving heterotrimeric G-protein signaling, mitogen-activated protein kinase cascades, ubiquitin–proteasome-mediated protein degradation, phytohormone signaling, transcriptional regulation, and epigenetic modification [3–9]. Nevertheless, the continuing discovery of regulators such as *qGW12*/*OsPUB23* and the *GS8–GL10* module indicates that the genetic and molecular architecture of grain shape remains incompletely understood [10, 11]. Although numerous grain-shape-related loci have been identified through linkage mapping and genome-wide association studies (GWAS), several limitations remain. QTLs identified in biparental segregation populations represent only the alleles segregating between two parents, and their

estimated effects may depend strongly on the parental combination and genetic background [12]. GWAS in diverse germplasm panels can exploit historical recombination and a broader range of natural allelic variation, thereby providing higher mapping resolution and greater relevance to breeding populations [13]. However, many previous studies have analyzed phenotypes from individual environments or have focused primarily on marker effects. Consequently, they cannot adequately distinguish loci with stable effects across environments from those with environment-dependent effects. Grain development is sensitive to year-specific variation in temperature, solar radiation, precipitation, and other environmental factors [14, 15]. Joint analysis of multi-year phenotypes is therefore important for distinguishing stable effect QTNs from QTN-by-environment interactions (QEIs). Stable QTNs may be useful for improving broad adaptation, whereas environment-responsive QTNs may facilitate the development of varieties targeted to specific production environments. Recent multi-environment GWAS have identified both stable QTNs and QEIs for grain-size traits; however, QEIs remain much less characterized than QTNs [16–18].Conventional single-locus GWAS methods, such as general linear model and mixed linear model (MLM), test markers individually while accounting for population structure and relatedness. Although these approaches are widely used, stringent multiple-testing corrections may reduce the power to detect the numerous small-effect loci expected for highly polygenic traits [19–21]. FarmCPU improves GWAS by iteratively fitting fixed- and random-effect models [22], whereas multi-locus GWAS methods provide a complementary strategy by evaluating multiple

potentially associated markers within a unified analytical framework [23–25]. The three-variance-component multi-locus random-SNP-effect mixed linear model, 3VmrMLM, was developed to identify QTNs, QEIs, and QTN-by-QTN interactions (QQIs) while accounting for all possible genetic effects and their polygenic backgrounds [26]. Applications of 3VmrMLM to rice yield and grain-size traits have demonstrated its potential to identify both QTNs and QEIs and to facilitate candidate-gene prioritization through the integration of functional annotation, transcriptomic data, and haplotype analysis [27, 28].

In the present study, an internationally diverse panel of 413 rice accessions was resequenced and evaluated for grain-shape-related traits across years. GL, GW, and GLWR were measured in three years and TGW in two, with each year considered a distinct environment. We applied 3VmrMLM to identify QTNs and QEIs for these traits and wvaluated locus stability used single-year, multi-year joint, and BLUP datasets; mixed linear model (MLM) and random model circulating probability unification (FarmCPU) were additionally used to assess the reproducibility of association signals across different GWAS methods. Known genes located near the detected QTNs and QEIs were examined, and panicle transcriptomic data from the P4 and P5 developmental stages, together with Gene Ontology annotation, were integrated to prioritize candidate genes. This study aimed to dissect the stable QTNs and environment-responsive QEIs for rice grain shape traits and to provide candidate loci and genes for the breeding of high-yielding, high-quality, and environmentally adapted rice varieties.

Materials and methods

Plant materials and phenotypic evaluation

A total of 413 rice accessions from 53 countries were selected from a diverse collection of 1,568 accessions, based on geographic origins and genotypes at approximately 700,000 single-nucleotide polymorphism (SNP) loci [29, 30]. Seeds of all accessions were obtained from the International Rice Research Institute (IRRI). Detailed information on the accessions is provided in Table S1. The 413 rice accessions were grown in a paddy field in Guangzhou, China (23°07′ N, 113°15′ E), under standard local cultivation practices, using a randomized complete block design with three replicates and ten plants per accession per replicate for phenotypic evaluation.Grain length (GL), grain width (GW), and grain length-to-width ratio (GLWR) were evaluated in 2014, 2016, and 2018. Thousand-grain weight (TGW) was evaluated in 2016 and 2018. GL and GW were measured using a Microtek scanner with ScanWizard EZ software version 2.140 and a Wanshen grain analyzer. GLWR was calculated as the ratio of GL to GW. Each experimental year was treated as a distinct environment because all field trials were conducted at the same location but under naturally varying annual environmental conditions. Although the panel comprised 413 accessions, the number of accessions with valid phenotypic records varied among traits and years because of missing or incomplete field observations.

Phenotypic data analysis

We calculated descriptive statistics for each trait in each year, including the mean, range, standard deviation, coefficient of variation, skewness, and kurtosis. A

mixed linear model was also analyzed for phenotypic data:

$$y_{ijk} = \mu + E_i + G_j + (GE)_{ij} + R_{k(i)} + \varepsilon_{ijk}$$

Where y_{ijk} is the phenotypic observation of genotype *j* in replicate *k* of year *i*; μ is the overall mean; E_i is the effect of the *i*th year; G_j is the effect of the *j*th genotype; $(GE)_{ij}$ is the genotype-by-year interaction; $R_{k(i)}$ is the effect of replicate *k* nested within year *i*; ε_{ijk} is the residual error. Year (environment) effect was treated as a fixed effect, whereas genotype, genotype-by-year interaction and replicate nested within year were treated as random effects. Variance components and genotypic best linear unbiased predictors (BLUPs) were estimated using the lme4 package in R (http://lme4.r-forge.r-project.org/). The genotypic BLUPs obtained from the random genotype effects were subsequently used as across-year phenotypic values for genome-wide association analyses.

Broad-sense heritability (H_B^2) for each trait was estimated as described as:

$$H_B^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gy}^2 / y + \sigma_e^2 / yr}$$

where σ_g^2, σ_{gy}^2, andσ_e^2 are genetic, genotype-by-environment interaction and residual error variances, respectively, *r* is the number of replications, and *y* is the number of years.

whole-genome resequencing and SNP identification

The 413 accessions were subjected to whole-genome resequencing using an Illumina sequencing platform [30]. Raw sequencing reads were aligned to the *Oryza sativa* reference genome MSU version 7 using BWA-MEM. The resulting alignment files were processed, sorted, and converted into the required formats using SAMtools

158 [31]. SNP calling and quality filtering were performed following the Genome
159 Analysis Toolkit version 4.0 Best Practices workflow. SNPs with a minor allele
160 frequency < 0.01 or a missing genotype rate >0.20 were excluded. After quality
161 control and filtering, a total of 1,974,377 high-quality SNPs were retained for
162 subsequent population genetic analysis and GWAS. The total length of the rice
163 chromosomes is ~ 380 Mbp, hence, on average there was one SNP per 192.47 bp.

164 *Population Structure and Linkage Disequilibrium (LD)*

165 Population structure was evaluated using ADMIXTURE v1.3.0 [32]. The number
166 of ancestral subpopulations, K, was evaluated across a predefined range, and the
167 population structure inferred at K = 3 was incorporated into subsequent GWAS. A
168 genomic kinship matrix among the 413 accessions was calculated using TASSEL 5.
169 Principal component analysis was also performed to characterize the genetic
170 relationships and population stratification within the diversity panel. All 1,974,377
171 SNPs were used for linkage disequilibrium (LD) analysis with PopLDdecay
172 (https://github.com/BGIShenzhen/PopLDdecay). LD decay was determined by
173 plotting the r^2 values against the genetic distance of a pair of loci (kb) for each
174 chromosome.

175 *Genome-Wide Association Study for main-effect QTN and QTN-by-environment*
176 *interactions*

177 GWAS were performed for GL, GW, GLWR, and TGW. The 3VmrMLM method
178 was used to identify QTNs and QEIs by utilizing the IIIVmrMLM R package version
179 1.0 (https://github.com/YuanmingZhang65/IIIVmrMLM) [26, 33]. Single-year

phenotypic values and across-year BLUPs were analyzed separately to identify QTNs. In addition, the original phenotypic observations from multiple years were jointly analyzed using IIIVmrMLM to detect both QTNs and QEIs.

For comparison, MLM and FarmCPU were implemented using the rMVP R package (https://github.com/xiaolei-lab/rMVP). The estimated population structure covariates and genomic kinship matrix were included in GWAS model to control for population stratification and relatedness. The significance thresholds were set to 0.05/*m* for MLM, FarmCPU and 3VmrMLM, while suggested threshold was set to LOD >= 3 for 3VmrMLM, where *m* is the number of markers. The gap size of linked significant QTNs on the same chromosome less than 200 kb was regarded as a same biological QTN.

Identification of known grain-size genes

To evaluate the biological relevance of the detected association signals, previously cloned genes for rice grain-related traits were collected from the China Rice Data Center (https://www.ricedata.cn/) and previous studies. Known grain-size genes located within 200 kb upstream or downstream of each detected QTN/QEI were considered to colocalize with the corresponding association locus. The same genomic window was used for the results obtained from the above three methods to ensure comparability across methods.

Identification and Mining of Novel Candidate Genes

To identify novel candidate genes for grain size-related traits, QTNs repeatedly detected across datasets and located within a 200-kb interval were defined as common

QTNs. Genes located within the 200-kb windows around common QTNs/QEIs were extracted, and previously reported genes for grain-related traits were excluded. The remaining genes were considered putative novel candidate genes for further prioritization.

Panicle expression profiles of the indica rice varieties Minghui 63 (MH63) and Zhenshan 97 (ZS97), which differ significantly in grain morphology, were obtained from the publicly available GSE19024 dataset in the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19024). Because gene expression during the P4 and P5 panicle developmental stages has been reported to be closely associated with rice grain-size determination, differential expression analyses between MH63 and ZS97 were performed separately at the P4 and P5 stages. Genes showing significant differential expression at both developmental stages were identified by taking the intersection of the P4 and P5 differentially expressed gene (DEG) sets. ·Genes with an adjusted $P < 0.05$ and $|\log2FC| > 1$ were considered differentially expressed.

The resulting DEGs located within the candidate QTN/QEI regions were further evaluated by haplotype analysis. SNPs within each candidate gene and its 2-kb upstream promoter region were used to define gene-level haplotypes. Haplotypes with frequencies $\geqslant 5\%$ were retained as major haplotypes, and phenotypic differences across haplotypes were evaluated using the corresponding BLUP values. ANOVA was used to test overall haplotype effects, followed by multiple-comparison tests when having significant differences. Finally, Gene Ontology (GO) functional annotations

were obtained using eggNOG-mapper. Genes supported by replicated QTN/QEI localization, differential expression at both P4 and P5 stages, significant haplotype–phenotype associations, and biological functions potentially related to grain development were prioritized as candidate genes.

Results

Phenotypic Variation

Phenotypic variation for four grain-related traits (GL, GW, GLWR, and TGW) was evaluated in 413 rice accessions across three (GL, GW, GLWR) or two (TGW) environments. All traits exihibited broad phenotypic variation among accessions (Figure 1; Table 1). Mean values ranged from 8.22 to 8.93 mm (GL), 2.77 to 2.93 mm (GW), 3.02 to 3.12 (GLWR), and 23.97 to 24.98 g (TGW). Coefficients of variation ranged from 10.34% to 19.40%, with GLWR and TGW showing greater dispersion than GL and GW. Histogram and year-wise boxplot indicated continuous distributions and moderate year-to-year shifts, particularly for GL and TGW. Broad-sense heritability was high for all traits (95.07%–98.29%), suggesting that phenotypic variation was largely genetically controlled and suitable for genetic dissection. Trait distributions also varied among years: kurtosis ranged from 0.07 to 0.96 for GLWR and 0.28 to 0.65 for TGW, indicating that grain-related traits were influenced by annual environmental variation.

Population Structure and Linkage Disequilibrium Analysis

Population structure analysis indicated the inflection point for parameter estimation erroroccurred at K (number of preset groups) = 3 to 4 (Figure 2 A). PCA

246 further grouped the 413 accessions into three distinct clusters in a two-dimensional
247 plot (Figure 2 B). These resulted indicated that the panel could be divided into three
248 subgroups (K=3) for subsequent analysis. The population structure estimated by
249 ADMIXTURE (Figure 2 C) was incoreporated as a covariate in GWAS. Genome-
250 wide linkage disequilibrium was evaluated using pairwise r^2 values between SNPs.
251 The average r^2 decreased with increasing physical distance, and the genome-wide LD
252 decay distance, defined as the distance at which r^2 declined to half of its maximum
253 value, was approximately 122.5 kb (Figure 2 D).

254 *Identification of QTNs for rice grain-related traits*

255 To dissect the genetic basis of rice grain-related traits, we applied IIIVmrMLM
256 to single-year phenotypes, BLUP values and multi-year datasets for GL, GW, GLWR
257 and TGW. After merging signals across datasets, 407 non-redundant QTNs were
258 identified (100, 102, 94, and 111 for GL, GW, GLWR, and TGW, respectively; Table
259 2 and Tables S2–S5). For GL, 19, 25, 26, 17 and 51 significant QTNs were detected in
260 2014, 2016, 2018, BLUP and multi-year joint analysis, respectively (Table S2), with
261 LOD scores of 3.54–234.28 and r^2 values of 0.06%–22.60% (mean 1.77%). For GW,
262 the corresponding datasets yielded 20, 25, 19, 15, and 49 QTNs, giving 102 merged
263 loci (LOD = 3.33–194.82 and r^2 values of 0.09%–19.82% (Table S3). For GLWR, 22,
264 19, 23, 22, and 53 QTNs were identified, resulting in 94 merged loci (LOD = 3.10–
265 282.98 and r^2 = 0.08%–11.21%) (Table S4). For TGW, 26, 34, 27, and 50 QTNs were
266 detected in 2016, 2018, BLUP, and multi-year joint analysis, respectively, with 111
267 merged loci (LOD = 3.34–145.06 and r^2 = 0.18%–5.50%) (Table S5). Overall, most

loci explained only a small proportion of phenotypic variance, consistent with the polygenic nature of rice grain-shape-related traits.

To evaluate whether the 3VmrMLM signals captured established grain-related loci, we compared all detected QTNs with known cloned or functionally characterized grain-related genes. After merging repeated records and grouping alternative names of the same gene, 44 known grain-related genes were located within 200 kb of 3VmrMLM-detected QTNs across all 12 chromosomes (Table S6). These genes were associated with one or more of the four traits, including 13, 14, 21, and 6 genes for GL, GW, GLWR, and TGW, respectively. Several well-characterized loci were detected repeatedly across traits and years. For example, QTNs near *GS3* on chromosome 3 were associated with GL, GW and GLWR across single-year, BLUP and multi-year joint analyses, while QTNs near *GW5* on chromosome 5 were associated with all four grain-related traits. Other known regulators, including *GS2*, *SMG*3, *SSG6*, *GW7* and *WTG1*, were also located near significant QTNs. The average distance between QTNs and their corresponding known genes ranged from 9.35 to 183.45 kb (mean 94.43 kb). These results indicate that 3VmrMLM successfully recovered multiple established grain-size loci, further supporting the reliability of the detected QTN set.

Common QTNs stably detected across environments

To identify robust loci, we screened QTNs detected within 200 kb in at least two datasets (single-year, BLUP, and multi-year). A total of 116 common QTNs were identified, including 31 for GL, 27 for GW, 27 for GLWR, and 31 for TGW (Figure

3A and Table S7). These common QTNs were distributed across most rice chromosomes and represented stable genomic regions for grain-related traits. Of these, 4, 7, 9 and 10 QTNs for GL, GW, GLWR and TGW, respectively, were also detected by MLM and/or FarmCPU. In addition, 22 common QTNs were located near cloned grain-size genes, including *GS3*, *GW5*, *GS2*, *GW7*, *SSG6*, *SMG3*, *GS9* and *GSN1*[4, 7, 34–39], further supporting the reliability of the multi-environment 3VmrMLM signals (Figures 3 B & C).

Comparison of QTNs for rice grain-related traits among three GWAS methods

To evaluate the consistency of association signals across different GWAS frameworks, MLM and FarmCPU were used in parallel with 3VmrMLM based on single-year datasets. After merging nearby significant associations within each method, MLM detected 6, 6, 8, and 24 loci for GL, GW, GLWR, and TGW respectively, whereas FarmCPU detected 9, 10, 15, and 11 loci (Table S8). Thus, 3VmrMLM detected substantially more associations than either comparison method, especially small-effect loci with significant LOD support. Comparison with cloned grain-size genes further supported this pattern: MLM recovered 3, 1, 2 and 1 cloned-gene-associated loci for GL, GW, GLWR, and TGW, respectively, whereas FarmCPU recovered 2, 1, 2 and 2 such loci.

Grain width was used as a representative trait to visualize the correspondence among the three GWAS methods across single-year, BLUP, and multi-environment joint analyses (Figure 4). A prominent region on chromosome 5 containing the well-characterized grain-width gene *GW5* was consistently supported by 3VmrMLM,

MLM, and FarmCPU across multiple datasets, providing strong cross-model support for this locus. In addition, 3VmrMLM recovered several previously reported grain-size genes, including *GS3*, *FUWA*[40], *RAG2*[41], *OsCEP6.1*[42], and *OsBZR1*[43], across different analytical datasets. The integrated analysis also highlighted several putative novel candidate genes near 3VmrMLM signals, such as *OsPCF5*, *OsGA2ox11*, and *CYP75B3*. Notably, the multi-environment joint analysis additionally identified candidate genes associated with QTN-by-environment interaction signals, including PHYA, which were distinguishable from the main-effect loci (Figure 4). Together, these results show that the three GWAS methods captured partially overlapping genetic signals, while integration of 3VmrMLM results with known-gene information, candidate-gene prioritization, and QEI analysis provided additional biological interpretation of the detected loci.

Identification of QTN-by-environment interactions (QEI) for rice grain-related traits

Multi-year joint analysis via IIIVmrMLM identified 10 significant QEIs for GL, GW, and GLWR, but none for TGW (Table 3). Two GL QEIs were located on chromosomes 1 and 9, with LOD scores of 29.45 and 22.61, respectively. Five GW QEIs were distributed on chromosomes 3, 9 and 10; among these, the locus at 14,674,801 bp on chromosome 10 showed both additive-by-environment and dominance-by-environment interaction effects, whereas the remaining four showed additive-by-environment interaction effects only. Three GLWR QEIs were located on chromosomes 3, 5, and 10. The GLWR QEI at 17,058,771 bp on chromosome 3 showed both additive-by-environment and dominance-by-environment interaction

effects and was located near *OsMYB3/SMG3* (cited reference), a cloned gene reported to coordinately regulate rice grain size, grain number per panicle, and grain weight. This colocalization suggests that natural variation at or near *OsMYB3/SMG3* may contribute to environment-dependent variation in grain shape, although further functional validation is required. Across all QEIs, LOD scores ranged from 14.13 to 93.42 and r^2 values ranged from 0.28% to 2.35%, indicating that environmental interactions contributed modest but detectable effects to grain-trait variation.

Prediction of new candidate genes for grain size and weight

To prioritize candidate genes underlying rice grain-size variation, we integrated genes located within ±200 kb of replicatedly detected QTNs and QEIs with differentially expressed genes between the P4 and P5 panicle developmental stages of MH63 and ZS97. After excluding previously cloned grain-size genes, we used differential expression analysis, functional annotations, and haplotype–phenotype associations to prioritize candidate genes. Fourteen genes met these criteria, including three associated with GL, four with GW, four with GLWR, two with TGW, and one located within a GW-QEI region (Table 4). These genes were annotated in biological processes potentially related to grain development, such as carbohydrate metabolism, seed development, phytohormone responses, protein phosphorylation, secondary cell-wall biosynthesis, and light signaling.Among these candidates, *OsFKII/FRK2*, *OsPCF5*, and *OsSWN3/OsNAC31* emerged as particularly promising genes. To further characterize natural allelic variation at these loci, SNPs spanning the gene body and the 2-kb upstream promoter region were used to reconstruct gene-level haplotypes

and mapped to the corresponding gene structures (Figure 5A, C, E). After retaining haplotypes with frequencies ≥5%, three major haplotypes were identified for *OsFKII/FRK2*, three for *OsPCF5*, and six for *OsSWN3/OsNAC31*. The haplotype-defining variants were distributed across the upstream promoter and transcribed regions, revealing substantial gene-level sequence variation at these candidate loci.

OsFKII/FRK2 (*Os08g0113100*), encoding a fructokinase involved in carbohydrate metabolism and starch accumulation, was located near a replicatedly GL-associated QTN and was significantly differentially expressed between MH63 and ZS97 (adjusted $P=2.99\times10^{-5}$, $\log_2FC=1.61$). Its three major haplotypes showed significant differences of GL BLUP values ($P=3.99\times10^{-6}$) (Table 4, Figure 5A, B). *OsPCF5* (*Os01g0213800*), a plant-specific transcription factor and miR319 target located near a GW-associated QTN, also showed significant differential expression (adjusted $P=1.22\times10^{-6}$, $log_2FC=2.63$), and its three major haplotypes differed significantly ($P=3.07\times10^{-4}$) in GW BLUP values (Table 4, Figure 5C, D). *OsSWN3/OsNAC31* (*Os08g0103900*), a NAC transcription factor involved in cellulose synthesis and secondary cell-wall biosynthesis, was located close to a TGW QTN, and its major haplotypes differed significantly ($P=8.22\times10^{-9}$) in TGW BLUP values (Table 4, Figure 5E, F). Together, the concordance among QTN localization, expression variation, gene function, and structure-resolved haplotype effects further supported these three genes as candidates for grain-related trait variation.

In addition, *PHYA* (*Os03g0719800*), encoding the red/far-red-light photoreceptor phytochrome A, was identified near a GW-associated QEI on chromosome 3. *PHYA*

was significantly differentially expressed between MH63 and ZS97(adjusted P=7.02×10^{-5}, log_2FC=-1.32). Gene-structure-based haplotype analysis identified eight major PHYA haplotypes defined by multiple variants across the gene body and 2-kb upstream region (Figure 5G). These haplotypes differed significantly in grain width, and their phenotypic effects varied across 2014, 2016, and 2018 (Figure 5H). A significant haplotype-by-year interaction further indicated that the effects of natural variation at the *PHYA* locus were dependent on year-environment, supporting *PHYA* as a promising candidate underlying environment-responsive variation in rice grain width.

Discussion

Grain shape and weight are key determinants of rice yield and are shaped by both genetic factors and environmental variation. Therefore, the simultaneous detection of main-effect QTNs and QTN-by-environment interactions (QEIs) is important for understanding the genetic basis of grain size and for identifying loci with potential breeding value [18, 44]. In this study, we applied the multi-locus GWAS method IIIVmrMLM to dissect four grain-related traits across multiple environments. Compared with MLM and FarmCPU, IIIVmrMLM detected a larger number of QTNs, while most QTNs identified by MLM and FarmCPU were also recovered by IIIVmrMLM. Notably, a relatively large number of previously reported grain size-related genes were located within or near the QTN regions detected by IIIVmrMLM, further supporting the reliability of this method in identifying trait-associated loci. These results suggest that IIIVmrMLM can provide a useful supplement to

conventional single-locus GWAS models and improve the discovery of loci contributing to complex quantitative traits [45].

Stable loci detected across multiple datasets or environments are generally considered more reliable for candidate gene prediction [46]. In this study, QTNs repeatedly detected within 200-kb intervals were defined as common QTNs and were used for downstream candidate gene mining. Several common QTN regions overlapped with or were located near previously cloned grain size-related genes, including well-characterized regulators such as *GS2*, *GW5*, *GS3*, *GW7*, and *GS9* [35, 47–50]. The recovery of these known genes supports the reliability of the GWAS results and indicates that common QTNs can effectively enrich trait-relevant genomic regions.

To further identify novel candidate genes, we integrated common QTN/QEI localization, differential gene expression, GO functional annotation, and haplotype analysis and finally identified 14 high-confidence candidate genes potentially involved in grain-related traits variation. Among them, *OsFKII/FRK2*, *OsPCF5*, and *OsSWN3/OsNAC31* showed particularly consistent evidence. *OsFKII/FRK2*, located near a recurrent GL-associated locus, encodes a fructokinase associated with carbohydrate metabolism and may affect grain development through sugar utilization and starch accumulation [51, 52]. *OsPCF5*, located near a GW-associated locus, encodes a transcription factor potentially involved in organ growth and developmental regulation [53, 54]. *OsSWN3/OsNAC31* was located particularly close to a TGW-associated QTN and is functionally associated with secondary cell-wall biosynthesis,

422 suggesting a possible role in determining the growth capacity of the developing grain
423 [55, 56]. Importantly, mapping haplotype-defining variants onto the gene structures
424 provided an additional view of natural allelic diversity at these loci. Three major
425 haplotypes were identified for both OsFKII/FRK2 and OsPCF5, whereas six were
426 identified for OsSWN3/OsNAC31, and these haplotypes showed significant
427 differences in their corresponding grain traits. The defining variants occurred across
428 both upstream promoter and transcribed regions, indicating that the observed
429 haplotype effects represent composite allelic backgrounds rather than single marker
430 associations. Together with their differential expression between MH63 and ZS97 and
431 their biological annotations, these results provide mutually supportive evidence for
432 prioritizing the three genes. However, the present analyses do not distinguish causal
433 polymorphisms from linked variants, and the haplotype associations were evaluated in
434 the same diversity panel used for GWAS; they should therefore be regarded as gene-
435 level candidate-prioritization evidence rather than independent functional validation.
436 In contrast to the large number of main-effect loci, only 10 significant QEIs were
437 identified for GL, GW, and GLWR, and no significant QEI was detected for TGW.
438 This difference does not necessarily indicate that environmental effects on grain traits
439 are unimportant. Because all field trials were conducted at a single location, the
440 environmental variation in this study primarily reflected differences among years
441 rather than geographically contrasting conditions. The high heritabilities of the grain-
442 related traits and the lower statistical power for detecting interaction effects may
443 partly explain the limited number of QEIs identified. Multi-location and multi-season

trials incorporating detailed climatic variables may reveal additional environment-dependent loci. A particularly interesting candidate gene associated with environmental responsiveness was *PHYA* (*Os03g0719800*), identified near a GW-related QEI region. It encodes phytochrome A, a red/far-red-light photoreceptor that participates in light perception, photomorphogenesis, gibberellin regulation, internode elongation, and photoperiodic flowering in rice; notably, its effects on flowering differ between short- and long-day conditions, indicating strong environmental responsiveness [57–60]. In the present study, the major *PHYA* haplotypes differed significantly in grain width and showed distinct phenotypic trajectories across years (Figure 5G, H), suggesting that natural variation in or near *PHYA* may contribute to environment-dependent regulation of grain width, potentially through light-responsive developmental or hormonal pathways.

Overall, the present study revealed both recurrent main-effect loci and environment-dependent associations for rice grain-related traits. The recovery of known grain-size genes supports the reliability of the GWAS results, whereas the prioritized candidates provide new hypotheses concerning carbohydrate metabolism, transcriptional regulation, cell-wall development, and environmental signal perception. These loci and haplotypes may provide useful genetic resources for improving rice grain shape and grain weight through marker-assisted selection and environment-aware breeding strategies.

Conclusions

This study used a multi-year, multi-locus GWAS to dissect the genetic basis of

rice grain-related traits in a globally diverse panel of 413 accessions. A total of 407 nonredundant QTNs were identified, including 116 common QTNs supported by at least two datasets, and 44 previously reported grain-size genes were near the 3VmrMLM QTNs, including *GS3*, *GW5*, *GS2*, *GW7*, and *GS9*. Multi-year joint analysis further detected 10 QEIs for GL, GW, and GLWR, revealing a detectable environment-responsive component of grain-trait variation. Integration of replicatedly QTN/QEI regions with transcriptomic data, functional annotation, and haplotype analysis prioritized 14 prioritized candidate genes associated with grain size or weight, among which *OsFKII/FRK2*, *OsPCF5*, *OsSWN3/OsNAC31*, and *PHYA* showed particularly consistent supporting evidence. Together, these findings provide a more comprehensive view of both robust and year-dependent genetic variation underlying rice grain shape and weight, and offer candidate genes for further functional validation and potential use in marker-assisted and environment-aware rice breeding.

Supplemental Information

Supplementary Table S1 Accession information of the 413 international rice accessions used in this study.

Supplementary Table S2 All main-effect QTNs of grain length in rice for 2014,2016,2018, BLUP and multiple years.

Supplementary Table S3 All main-effect QTNs of grain width in rice for 2014,2016,2018, BLUP and multiple years.

Supplementary Table S4 All main-effect QTNs of grain length-width ratio in rice for 2014,2016,2018, BLUP and multiple years.

Supplementary Table S5 All main-effect QTNs of thousand-grain weight in rice for 2016,2018, BLUP and multiple years.

Supplementary Table S6 Known genes around ± 200kb window of QTNs from IIIVmrMLM.

Supplementary Table S7 Common QTN loci associated with grain related traits across environments in rice.

Supplementary Table S8 All QTNs of grain related traits by MLM and FarmCPU.

Declarations

Funding

This work was supported by National Key R&D Program of China (2024YFD1200801) and the National Natural Science Foundation of China (31601375).

Acknowledgements

Not applicable.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Authors' Contributions

S.B., F.L. and J.Z. conceived and designed the study. S.T., Q.S., H.H., and J.L. performed the experiments and collected the data. S.B., S.T. and F.L. conducted data analyses and interpreted the results. S.W. and J.Z. supervised the research and provided resources. S.B., S.T. and F.L. drafted the manuscript. S.B. revised the manuscript. All authors read and approved the final manuscript.

Availability of data and material

The raw read data (FASTQ files) of 413 accessions were uploaded to NCBI's sequence read archive (BioProject accession PRJNA820969) [30]. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang Q. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor Appl Genet. 2000;101:823–9. https://doi.org/10.1007/s001220051549.
2. Lu L, Shao D, Qiu X, Sun L, Yan W, Zhou X, et al. Natural variation and artificial selection in four genes determine grain shape in rice. New Phytologist. 2013;200:1269–80. https://doi.org/10.1111/nph.12430.
3. Abbas W, Xu L, Zhu Y. The molecular regulation of rice grain size: Pathways and prospects for precision breeding. Plant Science. 2026;364:112882. https://doi.org/10.1016/j.plantsci.2025.112882.
4. Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, et al. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA. 2010;107:19579–84. https://doi.org/10.1073/pnas.1014419107.
5. Qi P, Lin Y-S, Song X-J, Shen J-B, Huang W, Shan J-X, et al. The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. Cell Res. 2012;22:1666–80. https://doi.org/10.1038/cr.2012.151.
6. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, et al. Control of grain size, shape and quality by OsSPL16 in rice. Nature Genetics. 2012;44:950–4. https://doi.org/10.1038/ng.2327.
7. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, et al. The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat Genet. 2015;47:949–54. https://doi.org/10.1038/ng.3352.
8. Yaseen M, Tariq N, Kanwal R, Farooq A, Wang H, Yuan H. Rice grain size: current regulatory mechanisms and future perspectives. J Plant Res. 2025;138:403–17. https://doi.org/10.1007/s10265-025-01626-8.
9. Zhan P, Ma S, Xiao Z, Li F, Wei X, Lin S, et al. Natural variations in grain length 10 (GL10) regulate rice grain size. J Genet Genomics. 2022;49:405–13. https://doi.org/10.1016/j.jgg.2022.01.008.
10. Ang Y, Li B, Shen J, Li J, Zhao Y, Li T, et al. The GRAIN SIZE 8- GRAIN LENGTH 10 module controls the grain size in rice by regulating the expression of GA - and CTK -related genes. The Plant Journal. 2025;122:e70247. https://doi.org/10.1111/tpj.70247.
11. Li H, Wang Y, Qiao W, Zhu Z, Wang Z, Tian Y, et al. Identification of a novel locus qGW12/OsPUB23 regulating grain shape and weight in rice (Oryza sativa L.). Theor Appl Genet. 2024;137:267. https://doi.org/10.1007/s00122-024-04776-w.
12. Xing Y, Zhang Q. Genetic and Molecular Bases of Rice Yield. Annu Rev Plant Biol. 2010;61:421–42. https://doi.org/10.1146/annurev-arplant-042809-112209.
13. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet. 2012;44:32–9. https://doi.org/10.1038/ng.1018.
14. Chen Q, Zhu Y, Ruan B, Yu Y. Advances in Genetics and Breeding of Grain Shape in Rice. Agriculture. 2025;15:1944. https://doi.org/10.3390/agriculture15181944.
15. Deng N, Ling X, Sun Y, Zhang C, Fahad S, Peng S, et al. Influence of temperature and solar radiation on grain yield and quality in irrigated rice system. European Journal of Agronomy. 2015;64:37–46. https://doi.org/10.1016/j.eja.2014.12.008.

16. Wang J-T, Han X-L, Zhao M-M, Zhang H-Q, Chen Y, Jiang Q-Y, et al. A fast method for breeding by design via G × E interactions detected in large-scale climatic, phenomic and genomic data. National Science Review. 2026;13:nwag095. https://doi.org/10.1093/nsr/nwag095.
17. Wei M, Luo T, Huang D, Ma Z, Liu C, Qin Y, et al. Construction of High-Density Genetic Map and QTL Mapping for Grain Shape in the Rice RIL Population. Plants. 2023;12:2911. https://doi.org/10.3390/plants12162911.
18. Zhao Q, Shi X-S, Wang T, Chen Y, Yang R, Mi J, et al. Identification of QTNs, QTN-by-environment interactions, and their candidate genes for grain size traits in main crop and ratoon rice. Front Plant Sci. 2023;14:1119218. https://doi.org/10.3389/fpls.2023.1119218.
19. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–8. https://doi.org/10.1038/ng1702.
20. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010;42:348–54. https://doi.org/10.1038/ng.548.
21. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4. https://doi.org/10.1038/ng.2310.
22. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genetics. 2016;12:e1005767. https://doi.org/10.1371/journal.pgen.1005767.
23. Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6:19444. https://doi.org/10.1038/srep19444.
24. Wen Y-J, Zhang H, Ni Y-L, Huang B, Zhang J, Feng J-Y, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Briefings in Bioinformatics. 2018;19:700–12. https://doi.org/10.1093/bib/bbw145.
25. Wang J, Chen Y, Shu G, Zhao M, Zheng A, Chang X, et al. Fast3VmrMLM: A fast algorithm that integrates genome-wide scanning with machine learning to accelerate gene mining and breeding by design for polygenic traits in large-scale GWAS datasets. Plant Communications. 2025;6:101385. https://doi.org/10.1016/j.xplc.2025.101385.
26. Li M, Zhang Y-W, Xiang Y, Liu M-H, Zhang Y-M. IIIVmrMLM: the R and C++ tools associated with 3VmrMLM, a comprehensive GWAS method for dissecting quantitative traits. Molecular Plant. 2022;15:1251–3. https://doi.org/10.1016/j.molp.2022.06.002.
27. Zhang J, Wang S, Wu X, Han L, Wang Y, Wen Y. Identification of QTNs, QTN-by-environment interactions and genes for yield-related traits in rice using 3VmrMLM. Front Plant Sci. 2022;13:995609. https://doi.org/10.3389/fpls.2022.995609.
28. Zhao Q, Shi X-S, Wang T, Chen Y, Yang R, Mi J, et al. Identification of QTNs, QTN-by-environment interactions, and their candidate genes for grain size traits in main crop and ratoon rice. Front Plant Sci. 2023;14:1119218. https://doi.org/10.3389/fpls.2023.1119218.
29. McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, et al. Open access resources for genome-wide association mapping in rice. Nat Commun. 2016;7:10532. https://doi.org/10.1038/ncomms10532.
30. Wang J, Yang W, Zhang S, Hu H, Yuan Y, Dong J, et al. A pangenome analysis pipeline

provides insights into functional gene identification in rice. Genome Biol. 2023;24:19. https://doi.org/10.1186/s13059-023-02861-9.
31. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10:giab008. https://doi.org/10.1093/gigascience/giab008.
32. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64. https://doi.org/10.1101/gr.094052.109.
33. Li M, Zhang Y-W, Zhang Z-C, Xiang Y, Liu M-H, Zhou Y-H, et al. A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. Molecular Plant. 2022;15:630–50. https://doi.org/10.1016/j.molp.2022.02.012.
34. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, et al. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat Plants. 2017;3:17043. https://doi.org/10.1038/nplants.2017.43.
35. Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H, et al. A Rare Allele of GS2 Enhances Grain Size and Grain Yield in Rice. Molecular Plant. 2015;8:1455–65. https://doi.org/10.1016/j.molp.2015.07.002.
36. Matsushima R, Maekawa M, Kusano M, Tomita K, Kondo H, Nishimura H, et al. Amyloplast Membrane Protein SUBSTANDARD STARCH GRAIN6 Controls Starch Grain Size in Rice Endosperm. Plant Physiology. 2016;170:1445–59. https://doi.org/10.1104/pp.15.01811.
37. Li J, Zhang B, Duan P, Yan L, Yu H, Zhang L, et al. An endoplasmic reticulum-associated degradation–related E2–E3 enzyme pair controls grain size and weight through the brassinosteroid signaling pathway in rice. The Plant Cell. 2023;35:1076–91. https://doi.org/10.1093/plcell/koac364.
38. Zhao D-S, Li Q-F, Zhang C-Q, Zhang C, Yang Q-Q, Pan L-X, et al. GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. Nat Commun. 2018;9:1240. https://doi.org/10.1038/s41467-018-03616-y.
39. Wang T, Zou T, He Z, Yuan G, Luo T, Zhu J, et al. GRAIN LENGTH AND AWN 1 negatively regulates grain size in rice. JIPB. 2019;61:1036–42. https://doi.org/10.1111/jipb.12736.
40. Chen J, Gao H, Zheng X, Jin M, Weng J, Ma J, et al. An evolutionarily conserved gene, *FUWA* , plays a role in determining panicle architecture, grain shape and grain weight in rice. The Plant Journal. 2015;83:427–38. https://doi.org/10.1111/tpj.12895.
41. Zhou W, Wang X, Zhou D, Ouyang Y, Yao J. Overexpression of the 16- κD a α-amylase/trypsin inhibitor RAG 2 improves grain yield and quality of rice. Plant Biotechnology Journal. 2017;15:568–80. https://doi.org/10.1111/pbi.12654.
42. Sui Z, Wang T, Li H, Zhang M, Li Y, Xu R, et al. Overexpression of Peptide-Encoding OsCEP6.1 Results in Pleiotropic Effects on Growth in Rice (O. sativa). Front Plant Sci. 2016;7. https://doi.org/10.3389/fpls.2016.00228.
43. Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, et al. Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. The Plant Journal. 2015;82:570–81. https://doi.org/10.1111/tpj.12820.
44. Welch JR, Vincent JR, Auffhammer M, Moya PF, Dobermann A, Dawe D. Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures. Proc Natl Acad Sci U S A. 2010;107:14562–7.

https://doi.org/10.1073/pnas.1001222107.
45. Zhang Y-M, Jia Z, Dunwell JM. Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. Frontiers in Plant Science. 2019;10.
46. Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, et al. GWAS Analysis and QTL Identification of Fiber Quality Traits and Yield Components in Upland Cotton Using Enriched High-Density SNP Markers. Front Plant Sci. 2018;9:1067. https://doi.org/10.3389/fpls.2018.01067.
47. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet. 2006;112:1164–71. https://doi.org/10.1007/s00122-006-0218-1.
48. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, et al. Deletion in a gene associated with grain size increased yields during rice domestication. Nat Genet. 2008;40:1023–8. https://doi.org/10.1038/ng.169.
49. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat Genet. 2015;47:944–8. https://doi.org/10.1038/ng.3346.
50. Zhao D-S, Li Q-F, Zhang C-Q, Zhang C, Yang Q-Q, Pan L-X, et al. GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. Nat Commun. 2018;9:1240. https://doi.org/10.1038/s41467-018-03616-y.
51. Jiang H, Dian W, Liu F, Wu P. Isolation and characterization of two fructokinase cDNA clones from rice. Phytochemistry. 2003;62:47–52. https://doi.org/10.1016/S0031-9422(02)00428-4.
52. Ye X, Dong J, Wei S, Li X, Qian H, Zhou Y, et al. Rice fructokinases orchestrate carbon partitioning to regulate photosynthesis and sugar metabolism. The Crop Journal. 2026;14:710–22. https://doi.org/10.1016/j.cj.2025.12.002.
53. Fang Y, Guo D, Wang Y, Wang N, Fang X, Zhang Y, et al. Rice transcriptional repressor OsTIE1 controls anther dehiscence and male sterility by regulating JA biosynthesis. The Plant Cell. 2024;36:1697–717. https://doi.org/10.1093/plcell/koae028.
54. Yang C, Li D, Mao D, Liu X, Ji C, Li X, et al. Overexpression of micro RNA 319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (*O ryza sativa* L .). Plant Cell & Environment. 2013;36:2207–18. https://doi.org/10.1111/pce.12130.
55. Huang D, Wang S, Zhang B, Shang-Guan K, Shi Y, Zhang D, et al. A Gibberellin-Mediated DELLA-NAC Signaling Cascade Regulates Cellulose Synthesis in Rice. Plant Cell. 2015;27:1681–96. https://doi.org/10.1105/tpc.15.00015.
56. Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye Z-H. Transcriptional Activation of Secondary Wall Biosynthesis by Rice and Maize NAC and MYB Transcription Factors. Plant and Cell Physiology. 2011;52:1856–71. https://doi.org/10.1093/pcp/pcr123.
57. Takano M, Inagaki N, Xie X, Kiyota S, Baba-Kasai A, Tanabata T, et al. Phytochromes are the sole photoreceptors for perceiving red/far-red light in rice. Proc Natl Acad Sci USA. 2009;106:14705–10. https://doi.org/10.1073/pnas.0907378106.
58. Iwamoto M, Kiyota S, Hanada A, Yamaguchi S, Takano M. The Multiple Contributions of Phytochromes to the Control of Internode Elongation in Rice. Plant Physiology. 2011;157:1187–95. https://doi.org/10.1104/pp.111.184861.
59. Hirose F, Inagaki N, Hanada A, Yamaguchi S, Kamiya Y, Miyao A, et al. Cryptochrome and

699 Phytochrome Cooperatively but Independently Reduce Active Gibberellin Content in Rice
700 Seedlings under Light Irradiation. Plant and Cell Physiology. 2012;53:1570–82.
701 https://doi.org/10.1093/pcp/pcs097.
702 60. Lee Y-S, Yi J, An G. OsPhyA modulates rice flowering time mainly through OsGI under short
703 days and Ghd7 under long days in the absence of phytochrome B. Plant Mol Biol.
704 2016;91:413–27. https://doi.org/10.1007/s11103-016-0474-7.

705

706

707 Table 1 Phenotypic characteristics of grain-related traits in 413 rice accessions across years

Traits	Year	Mean	Stdev	Skew	Kurt	Min	Max	CV(%)	H_B^2(%)
GL (mm)	2014	8.93	0.97	0.31	0.64	5.81	11.71	10.85	97.23
	2016	8.49	0.91	0.19	0.69	5.58	11.38	10.77	
	2018	8.22	0.85	0.17	0.77	5.22	10.74	10.34	
GW (mm)	2014	2.93	0.34	-0.16	-0.09	1.93	3.85	11.54	96.21
	2016	2.77	0.32	-0.02	-0.14	1.91	3.69	11.41	
	2018	2.78	0.32	0	-0.13	1.84	3.65	11.35	
GLWR	2014	3.11	0.60	0.92	0.96	1.82	5.70	19.40	98.29
	2016	3.12	0.58	0.73	0.23	1.82	5.45	18.64	
	2018	3.02	0.56	0.73	0.07	1.78	5.11	18.45	
TGW (g)	2016	23.97	4.35	0.26	0.65	11.30	37.40	18.17	95.07
	2018	24.98	4.38	0.19	0.28	13.62	38.01	17.52	

708

709

710 Table 2 The number of QTNs identified by 3VmrMLM for grain-related traits across years

Trait	2014	2016	2018	BLUP	Multiple years	Total[a]
GL	19	25	26	17	51	100
GW	20	25	19	15	49	102
GLWR	22	19	23	22	53	94
TGW	-	26	34	27	50	111

711 a: the total number is after merging the same significant associated SNPs.

Table 3 QTN-by-environment interactions (QEIs) identified by 3VmrMLM for rice grain-related traits

Trait	Chr	Position (bp)	LOD (QE)	QTN-by-environment interaction effect						variance	r^2(%)	P-value
				a*e1	d*e1	a*e2	d*e2	a*e3	d*e3			
GL	1	23437725	29.4475	0.1319		-0.0481		-0.0838		0.0089	0.9719	3.58E-30
	9	7160268	22.6117	-0.1141		0.0369		0.0772		0.0068	0.7402	2.45E-23
GW	3	29019277	24.7692	-0.043		0.0145		0.0285		0.001	0.8742	1.71E-25
	9	13174533	21.4598	0.0323		0.0056		-0.0379		0.0008	0.7639	3.48E-22
	9	22395152	14.9688	0.0315		-0.0033		-0.0283		0.0006	0.5502	1.08E-15
	10	14674801	59.9493	0.0637	-0.1117	-0.0445	-0.3338	-0.0192	0.4455	0.0025	2.3182	1.57E-58
	10	17755647	31.0921	0.049		-0.0287		-0.0203		0.0012	1.1056	8.11E-32
GLWR	3	17058771	14.1315	-0.027	-0.3146	0.0112	0.2025	0.0158	0.1121	0.0009	0.2788	2.48E-13
	5	16805885	21.2216	-0.052		0.0138		0.0382		0.0014	0.428	6.01E-22
	10	9057210	93.4224	-0.1226		0.0869		0.0357		0.0079	2.3473	3.81E-94

Chr, chromosome; LOD (QE), logarithm of odds score for the QTN-by-environment interaction; a*e1, a*e2, and a*e3, additive-by-environment interaction effects in 2014, 2016, and 2018, respectively; d*e1, d*e2, and d*e3, dominance-by-environment interaction effects in the corresponding year-environments; variance, estimated variance component of the QEI; r^2 (%), proportion of phenotypic variance explained by the QEI. GL, grain length; GW, grain width; GLWR, grain length-to-width ratio.

717 Table 4 Candidate genes for rice grain size-related traits

Gene	gene_name	Trait	Chr	QTN position (bp)	Gene_start	Gene_end	Differential expression analysis		Haplotype analysis P-value	GO_ID	Gene_annotation
							Adjust P-value	log2FC			
Os08g0113100	*OsFKII, FRK2*	GL	8	582071	701252	703998	0.0000299	1.61	3.98991E-06	GO:0005975	Fructokinase, Starch storage in rice grains ;Similar to cDNA clone:J023041C07, full insert sequence.
Os10g0438100	*OSK2*	GL	10	15619785	15695559	15696486	0.000215	1.09	2.17454E-05	GO:0016567	Similar to Fimbriata-associated protein (Fragment).
Os12g0562500		GL	12	23082248	23061642	23065235	0.0000197	1.53	0.004510344	GO:0010105	Similar to Protein kinase-like protein (Fragment).
Os01g0213800	*OsPCF5, PCF5*	GW	1	6256729	6221640	6224242	0.00000122	2.63	0.000307352	GO:0048316	Plant-specific transcription factor, miR319 target gene
Os04g0407800	*OsGA2ox11, 2ODD31, Os2ODD31*	GW	4	20123813	20200632	20201883	0.000722	-1.06	0.000754586	GO:0009685	Similar to H0321H01.3 protein.
Os04g0619300	*OsFbox232, Os_F0576, Ctb1, OsCtb1*	GW	4	31316711	31458951	31462013	0.000299	1.21	0.000180263	GO:0043161	Similar to H0114G12.4 protein;Similar to L1332.8 protein.
Os10g0320100	*CYP75B3, F3'H*	GW	10	8517195	8679311	8681266	0.000876	1.58	7.98342E-06	GO:0009733	Flavonoid 3'-hydroxylase, Cytochrome P450 75B3, Flavone C-glycosides biosynthesis

718 Table 4 Candidate genes of rice grain size-related traits (continued)

Gene	gene_name	Trait	Chr	QTN position (bp)	Gene_start	Gene_end	Differential expression analysis		Haplotype analysis P-value	GO_ID	Gene_annotation
							Adjust P-value	log2FC			
Os04g0461000	*Os-MYB39, OsMYB39, MYB39*	GLWR	4	23085234	23024212	23026644	0.000428	1.13	4.47335E-08	GO:0009739	Similar to Myb-related transcription factor LBM1.
Os04g0528300	*OsABCG9*	GLWR	4	26298845	26395912	26401438	0.000473	1.05	2.18968E-06	GO:0009737	Similar to ABC transporter;Similar to OSIGBa0115K01-H0319F09.18 protein;Similar to OSIGBa0115K01-H0319F09.18 protein
Os08g0526400		GLWR	8	26151439	26192694	26196096	0.000428	-1.18	0.000380824	GO:0006468	Defence response, Rin4 domain containing protein.
Os09g0379900		GLWR	9	12746524	12802546	12805244	0.00247	1.19	5.17027E-05	GO:0044275	Similar to Beta-glucan binding protein.
Os08g0103900	*NAC31, OsSWN3*	TGW	8	217525	210422	212411	0.0211	-1.47	8.21882E-09	GO:0071365	NAC transcription factor, Regulation of cellulose synthesis, Regulation of secondary wall biosynthesis
Os08g0157000	*OsMAPK2, MPK2*	TGW	8	3233702	3307597	3310513	0.00108	1.04	2.33068E-08	GO:0000165	Similar to Mitogen-activated protein kinase 4.
Os03g0719800	*PHYA, PHY18, OsPHYA*	GW-QEI	3	29019277	29168197	29176006	0.0000702	-1.32	3.8282E-05	GO:0009642	Phytochrome A, Photoreceptor

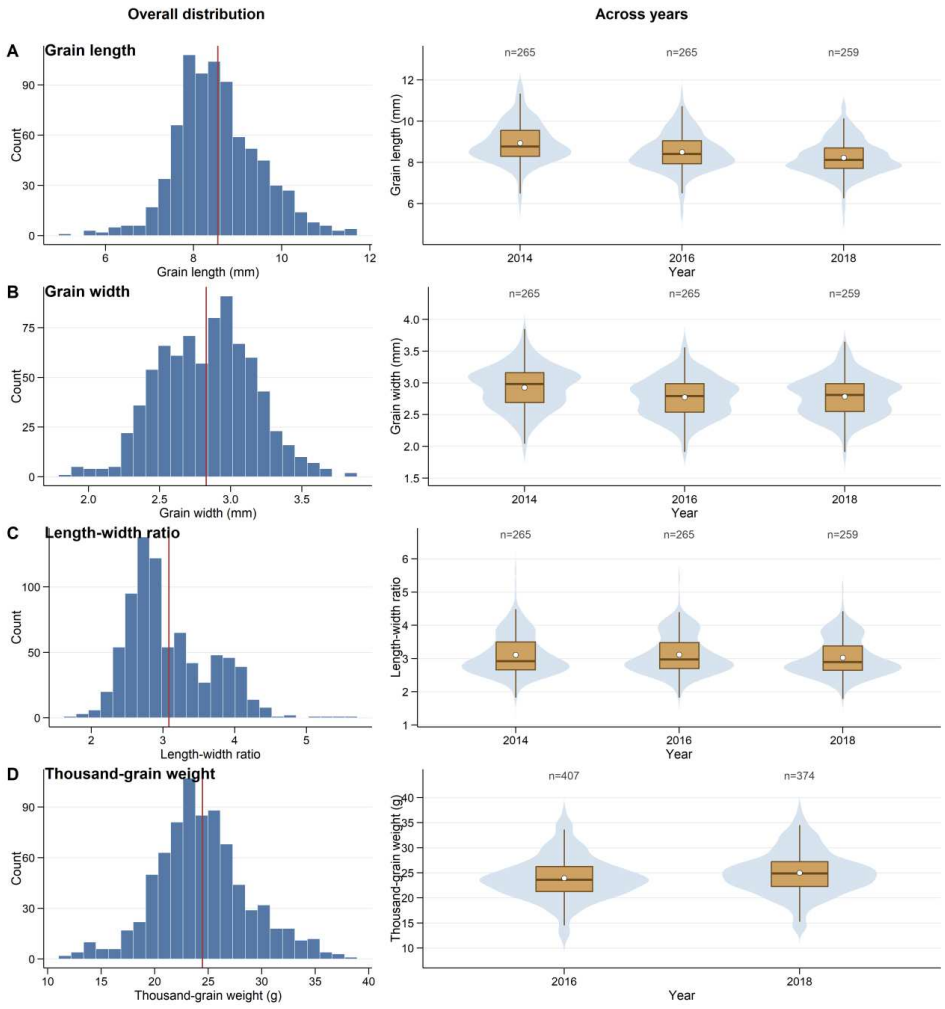


Figure 1 Phenotypic variation of grain-related traits across years

Histograms show the overall distributions of grain length (A), grain width (B), grain length-to-width ratio (C), and thousand-grain weight (D) across all evaluated accessions and years. Boxplots show year-wise phenotypic variation for each trait, with boxes indicating the interquartile range, centre lines indicating medians, whiskers representing 1.5 times the interquartile range, and points denoting outliers. The broad and continuous distributions indicate substantial phenotypic diversity for grain-related traits in the population.

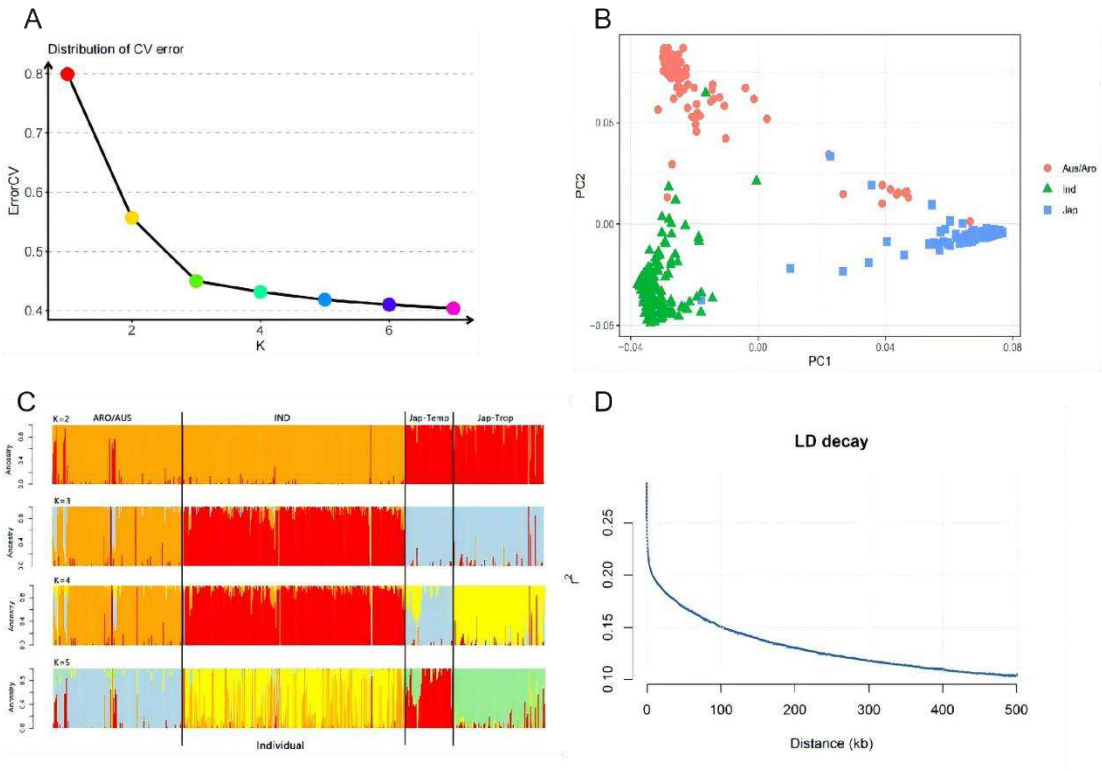


Figure 2 Population structure and LD decay in the rice accession panel

A, Cross-validation error across different K values, supporting the selection of the optimal population structure model. B, Principal component analysis (PCA) showing the genetic differentiation among Aus/Aro, Ind and Jap accessions. C, Population structure inferred from ancestry coefficients across K = 2 to K = 5, with vertical lines separating major genetic groups. D, LD decay estimated by plotting pairwise r^2 values against physical distance.

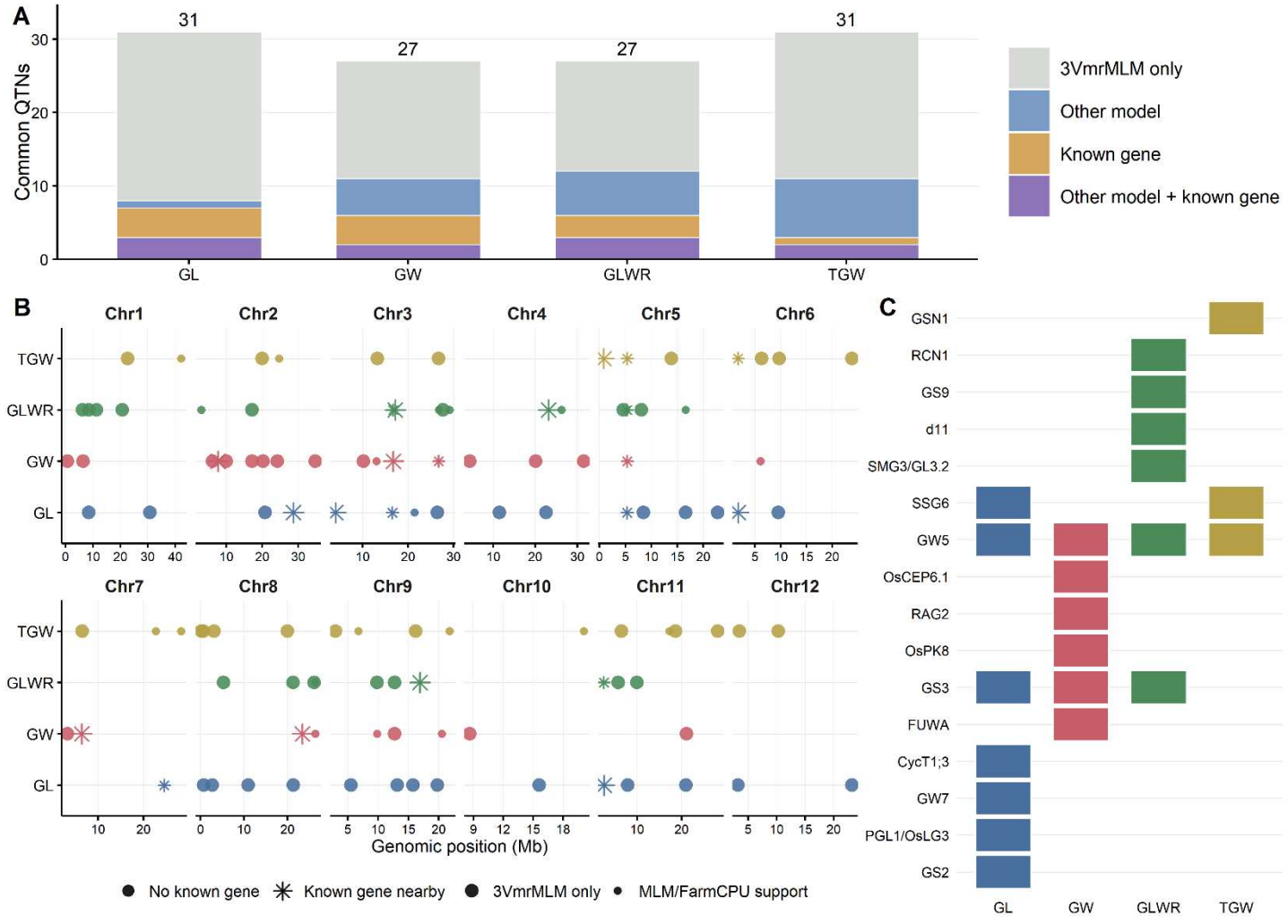
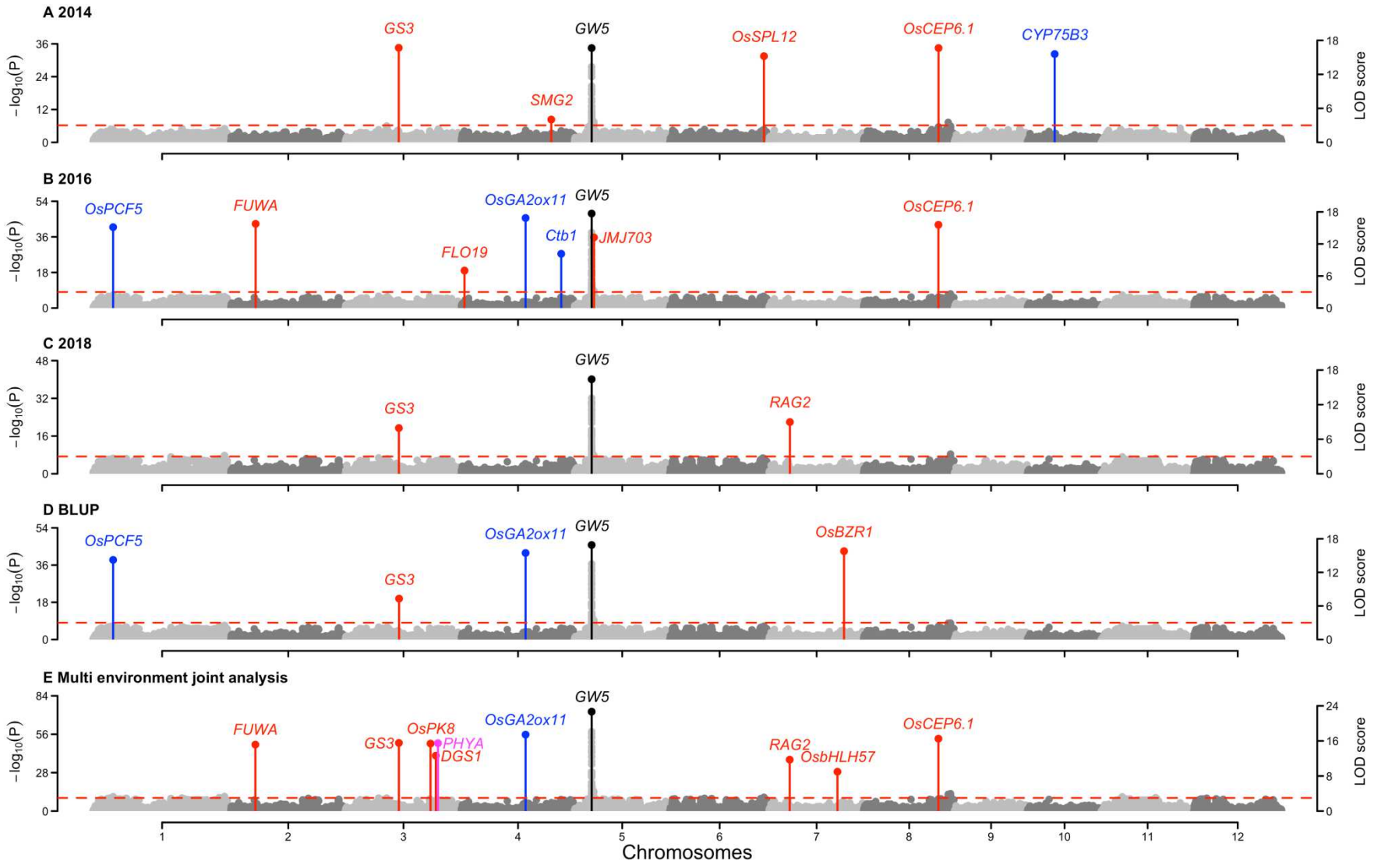


Figure 3 Common QTNs for rice grain-related traits across environments

738 A, Support categories of common QTNs for GL, GW, GLWR and TGW. Numbers above the bars indicate the total number of common QTNs for each

739 trait. B, Genomic distribution of common QTNs across the 12 rice chromosomes. C, Known grain-size genes located near common QTNs and their associated traits.

740


A 2014
GS3
GW5
OsSPL12
OsCEP6.1
CYP75B3
SMG2
B 2016
OsPCF5
FUWA
OsGA2ox11
GW5
Ctb1
JMJ703
FLO19
OsCEP6.1
C 2018
GW5
GS3
RAG2
D BLUP
OsPCF5
OsGA2ox11
GW5
OsBZR1
GS3
E Multi environment joint analysis
GW5
FUWA
OsPK8
OsGA2ox11
GS3
PHYA
DGS1
RAG2
OsbHLH57
OsCEP6.1
−log10(P)
LOD score
Chromosomes

741

Figure 4 Integrated genome-wide association signals for rice grain width detected by 3VmrMLM, MLM, and FarmCPU

Genome-wide association results are shown for grain width in 2014 (a), 2016 (b), 2018 (c), across-year BLUP values (d), and the multi-environment joint analysis (e). Chromosomes are displayed along the x-axis. The left y-axis represents $-\log_{10}(P)$ values for MLM and FarmCPU, whereas the right y-axis represents LOD scores from 3VmrMLM. The red dashed horizontal line indicates the 3VmrMLM significance threshold of LOD = 3. Black labels and vertical lines indicate grain-width-related genes supported by all three GWAS methods. Red labels indicate previously reported grain-size-related genes located near QTNs detected by 3VmrMLM, blue labels indicate novel candidate genes prioritized from 3VmrMLM signals, and magenta labels indicate candidate genes associated with QTN-by-environment interaction signals identified by 3VmrMLM.

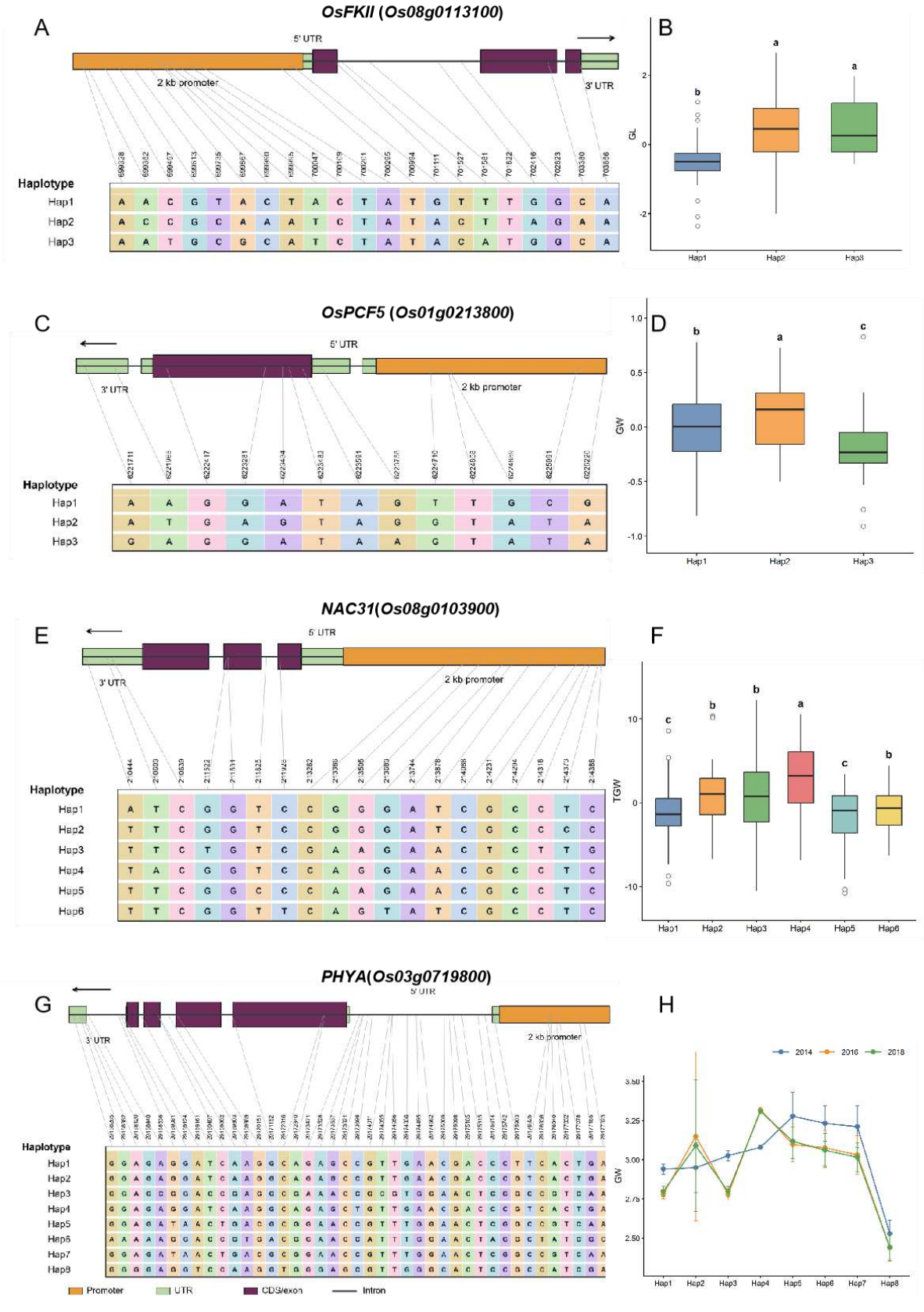


Figure 5 Gene structures, major haplotypes, and phenotypic effects of four prioritized candidate

753 genes

754 (A, B) Gene structure and major haplotypes of OsFKII/FRK2 and their effects on grain length

755 (GL). (C, D) Gene structure and major haplotypes of OsPCF5 and their effects on grain width

756 (GW). (E, F) Gene structure and major haplotypes of OsSWN3/OsNAC31 and their effects on

757 thousand-grain weight (TGW). (G, H) Gene structure and major haplotypes of PHYA and their

758 year-specific effects on GW. Haplotype-defining SNPs are connected to their corresponding

759 physical positions within the gene body or 2-kb upstream promoter region. Only major haplotypes

760 with frequencies ⩾5% were retained. In B, D, and F, different lowercase letters indicate

761 significant differences among haplotypes according to Duncan's multiple range test at $P<0.05$. In

762 H, points represent mean GW values for each haplotype across 2014, 2016, and 2018, and error

763 bars indicate SE.

764

Supplementary Files

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

- TableS1S8.xlsx