Article

Not peer-reviewed version

# Rice (Oryza sativa L.) Grain Size, Shape, and Weight-Related QTLs Identified using GWAS with Multiple GAPIT Models and High-Density SNP Chip DNA Markers

Nkulu Rolly Kabange *, Gamenyah Daniel Dzorkpe, Dong-Soo Park, Youngho Kwon, Sais-Beul Lee, So-Myeong Lee, Ju-Won Kang, Seong-Gyu Jang, Ki-Won Oh, Jong-Hee Lee *

Posted Date: 20 September 2023

doi: 10.20944/preprints202309.1280.v1

Keywords: SNP Chip DNA Marker; GAPIT; GWAS; Genomic Selection; Grain traits; Rice

*Article*

# Rice (*Oryza sativa* L.) Grain Size, Shape, and Weight-Related QTLs Identified using GWAS with Multiple GAPIT Models and High-Density SNP Chip DNA Markers

**Nkulu Rolly Kabange [1,*], Gamenyah Daniel Dzorkpe [2], Dong-Soo Park [1], Youngho Kwon [1], Sais-Beul Lee [1], So-Myeong Lee [1], Ju-Won Kang [1], Seong-Gyu Jang [1], Ki-Won Oh [1] and Jong-Hee Lee [1,*]**

[1] Department of Southern Area Crop Science, National Institute of Crop Science, RDA, Miryang 50424, Korea; parkds9709@korea.kr (D.-S.P.); kwon6344@korea.kr (Y.K.); pappler@korea.kr (S-B.L.); olivetti90@korea.kr (S.-M.L.); kangjw81@korea.kr (J.-W.K); sgjang0136@korea.kr (S.-G.J.); ohkw1004@korea.kr (K.-W.O)

[2] Council for Scientific and Industrial Research (CSRI), Crops Research Institute, Kumasi, Ghana; gamenyahdaniel@gmail.com

* Correspondence: rollykabange@korea.kr (N.R.K.), Tel.: +82-10-5413-4881; ccriljh@korea.kr; Tel: +82-53-350-1169; F.: +82-55-352-3059.

**Abstract:** This study investigated novel quantitative traits loci (QTLs) associated with the control of grain shape and size as well as grain weight in rice. We employed a joint strategy multiple GAPIT (Genome Association and Prediction Integrated Tool) models [(Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK)), Fixed and random model Circulating Probability Uniform (FarmCPU), Settlement of MLM Under Progressive Exclusive Relationship (SUPER), and General Linear Model (GLM)]–High Density SNP Chip DNA Markers (60,461) to conduct a Genome-Wide Association Study (GWAS). GWAS was performed using genotype and grain-related phenotypes of 143 recombinant inbred lines (RILs). Data show that parental lines (Ilpum and Tung Tin Wan Hein 1, TTWH1, *Oryza sativa* L., ssp. *japonica* and *indica*, respectively) exhibited divergent phenotypes for all analyzed grain traits), which was reflected in their derived population. GWAS results revealed the association between seven SNP Chip makers and quantitative trait loci (QTLs) for grain length, co-detected by all GAPIT models on (Chr) 1–3, 5, 7, and 11), were *qGL1-1*$^{BFSG}$ (AX-95918134, Chr1: 3820526 bp) explains 65.2%–72.5% of the phenotypic variance explained (PVE). In addition, *qGW1-1*$^{BFSG}$ (AX-273945773, Chr1: 5623288 bp) for grain width explains 15.5%–18.9% of PVE. Furthermore, BLINK or FarmCPU identified three QTLs for grain thickness independently, and explain 74.9% (*qGT1*$^{Blink}$, AX-279261704, Chr1: 18023142 bp) and 54.9% (*qGT2-1*$^{Farm}$, AX-154787777, Chr2: AX-154787777 bp) of the observed PVE. For t length-to-width ratio, the *qLWR2*$^{BFSG}$ (AX-274833045, Chr2: 10000097 bp) explains nearly 15.2%–32% of PVE for LWR. Likewise, the major QTL for thousand-grain weight (TGW) was detected on Chr6 (*qTGW6*$^{BFSG}$, AX-115737727, 28484619 bp) and explains 32.8%–54% of PVE. The *qTGW6*$^{BFSG}$ QTL coincides with *qGW6-1*$^{Blink}$ for grain width and explained 32.8%–54% of PVE. Putative Candidate genes pooled from major QTLs for each grain traits have interesting annotated functions that require functional studies to elucidate their function in the control of grain size, shape, or weight in rice. Genome selection analysis proposed makers useful for downstream marker-assisted selection based on genetic merit of RILs.

**Keywords:** SNP Chip DNA marker; GAPIT; GWAS; genomic selection; grain traits; rice

 doi:10.20944/preprints202309.1280.v1



## 1. Introduction

Rice (*Oryza sativa L.*) remains a staple cereal crop for more than half of the world's population, [1,2], and serves as an important source of calories for human health and fitness [3]. Its consumption is increasing faster than any other cereals [4]. Despite the increase in population growth estimated to about 9.8 billion by 2050 [2,5] coupled with food insecurity and climate change, the production of rice must increase to over 852 million tons by 2035 [6] to meet the growing food demands. Generally, rice is consumed as whole grain and can be processed into different forms of food. Rice grain size, shape, appearance, and quality of the grain directly influence the market value [7,8]. Based on grain size, preferences for its qualities vary across the world. Rice grain size and shape determined the milling efficiency and grain recovery, which influence its price.

Many rice-breeding programs have long been oriented to develop rice varieties that are high yielding and disease-resistant [9–11]. As part of the diversification process to address the rising food demands in terms of quantity and quality, the trend of rice breeding has shown a keen interest in the quality of grains coupled with productivity [2,12–14].

The phenotype of rice appearance is determined by grain shape (length, width, and thickness), translucency, and thousand-grain weight [15,16]. Studies have identified genes controlling grain size, shape, and weight of rice, which happens to be the result of a complex interaction between major and minor quantitative trait loci (QTLs) [17]. For instance, grain size has a high heritability, and a number of the major genes [15] and minor QTLs linked with grain size have been proposed [18].

Several of these grain traits are controlled by quantitative trait loci (QTLs) that are regulated by environmental influences and genetic variation in natural populations, intervarietal lines, mutant populations [6], doubled haploid lines [19], near-isogenic lines (NILs) [20] and recombinant inbred lines (RILs) [21]. Many genetic studies on rice inbred lines have effectively used molecular markers that include amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), or microsatellite (RM) [6,22], single nucleotide polymorphism (SNP) [23], restriction fragment length polymorphism (RFLP) [24], to improve grain appearance and other quality parameters [2].

Likewise, genome-wide association study (GWAS), linkage mapping, and omics tools were used to investigate genetic loci controlling the complex grain-quality traits such as grain shape (appearance), milling quality, nutritional quality, and eating and cooking qualities have been elucidated [25]. To date, several QTLs associated with the control of grain traits-related phenotypes are reported, and mapped to all chromosomes of rice. Among them, we could mention grain size 3 (GS3, controlling both grain length and weight [26,27], grain width 2 (GW2) [28], wide and thick grain (OsOTUB1/WTG1) [29], GS5 regulates a putative serine carboxypeptidase (SCP) that specifically affects grain width and filling [30], *GS2* encode growth-regulating factor 4 (GRF4) that regulates grain length and width [31], grain length 3 (*GL3.1*/*qGL3*) acts on a putative protein phosphatase and influences grain length, width and weight [32,33], *GW5*/*qSW5* a calmodulin-binding protein responsible for grain width and weight [34,35], thousand-grain weight 3 (*TGW3*) and *TGW6* encodes the auxin signaling pathway which regulated grain weight [36,37], *GW6* encodes a gibberellin-regulated GAST family protein that control grain width and weight [38]. The major QTLs (*GLW7*) and *GW8* not only encode the *OsSPL13* and *OsSPL16* transcription factors but also contribute to grain size formation in rice [39–41]. All these QTLs function as either positive or negative regulators in a number of signaling pathways, including the G-protein signaling, ubiquitin-proteasome, phytohormone, and transcriptional regulation pathways that influence cell division, endosperm development expressing the grain size, shape, and overall grain appearance [42–44].

This study aimed at investigating novel QTLs controlling grain size and shape of rice using a RIL population consisting of 143 lines derived from a cross between *indica* and *japonica* cultivars. To achieve that, a joint strategy employing GWAS with multiple GAPIT (Genome Association and Prediction Tool) models coupled with high-density SNP Chip markers was used.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



## 2. Results

### *2.1. Diverging Grain Phenotypes between Parental Lines and RILs*

Considering that they belong to the most cultivated subspecies of rice (*Oryza sativa* L. ssp. *indica* and *japonica*), parental lines (Ilpum and Tung Tin Wan Hein 1, TTWH1) exhibited differential phenotypes for all analyzed rice grain size and shape-related traits, as expected (Figure 1A,B). The *japonica* parent (Ilpum) showed relatively shorter grain and lower LWR, while having larger grains (grain width, GW) and higher thousand-grain weight (TGW). In contrast, the *indica* TTHW1 had longer grains and higher LWR, but thinner grains and lower TGW. However, although we recorded an arithmetic or numerical difference between the grain thickness of Ilpum (thinner grains) compared to that of TTWH1 (thicker grains), a non-significant statistical difference was observed.

In addition, we observed a normal distribution for grain length (Figure 2A,K). However, Grain Width, GT, and thousand grain weight exhibited a left skewness (Figures 2C,G, I,K), while LWR showed a right skewness-like pattern (Figure 2E,K). As displayed in panel D, F, and J of Figure 2, a total shift (Ilpum-like pattern) in grain width, thickness (, and thousand-grain weight of the RIL population was observed. Meanwhile, panels B and H of Figure 2 show that nearly 82.5% and 97.9% of the RIL population exhibited relatively short grains (Ilpum-like) and LWR value, respectively, against 17.5% and 2.1% having long grains and LWR (TTWH1-like phenotype).

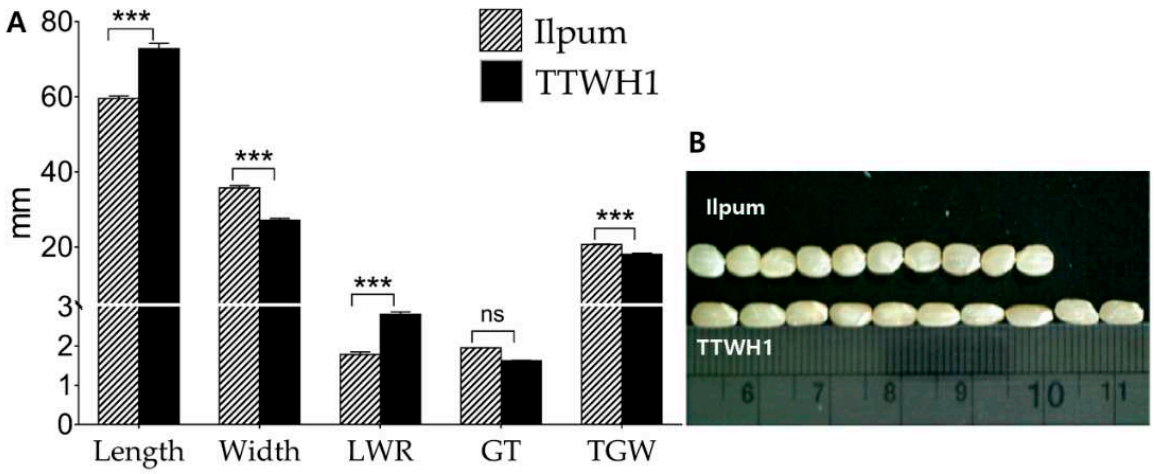



**Figure 1.** Differential phenotypic difference between parental lines. (A) Comparison of rice grain trait values of Ilpum (japonica) and Tung Tin Wan 1 (indica) and (B) grain phenotypes of parental lines.

### *2.2. Relatedness, Correlation, Heritability, and Genomic Selection*

We constructed a kinship matrix to assess the relatedness of the RIL population. As indicated in Figure 3A, based on the color pattern in the heat map, the genotypes of RILs used in this are diverse and not closely related. The density of SNP Chip DNA markers across all 12 chromosomes of rice is provided in panel B in Figure 3. Figure 3C shows that RILs were grouped into three clusters based on their recorded grain size and shape phenotypes. Principal Component Analysis (PCA) Results suggest that PC1 (55.6%), PC2 (30.8%), and PC3 (9.8%) explain 96.2% of the proportion of phenotypic variance of the RILs population.

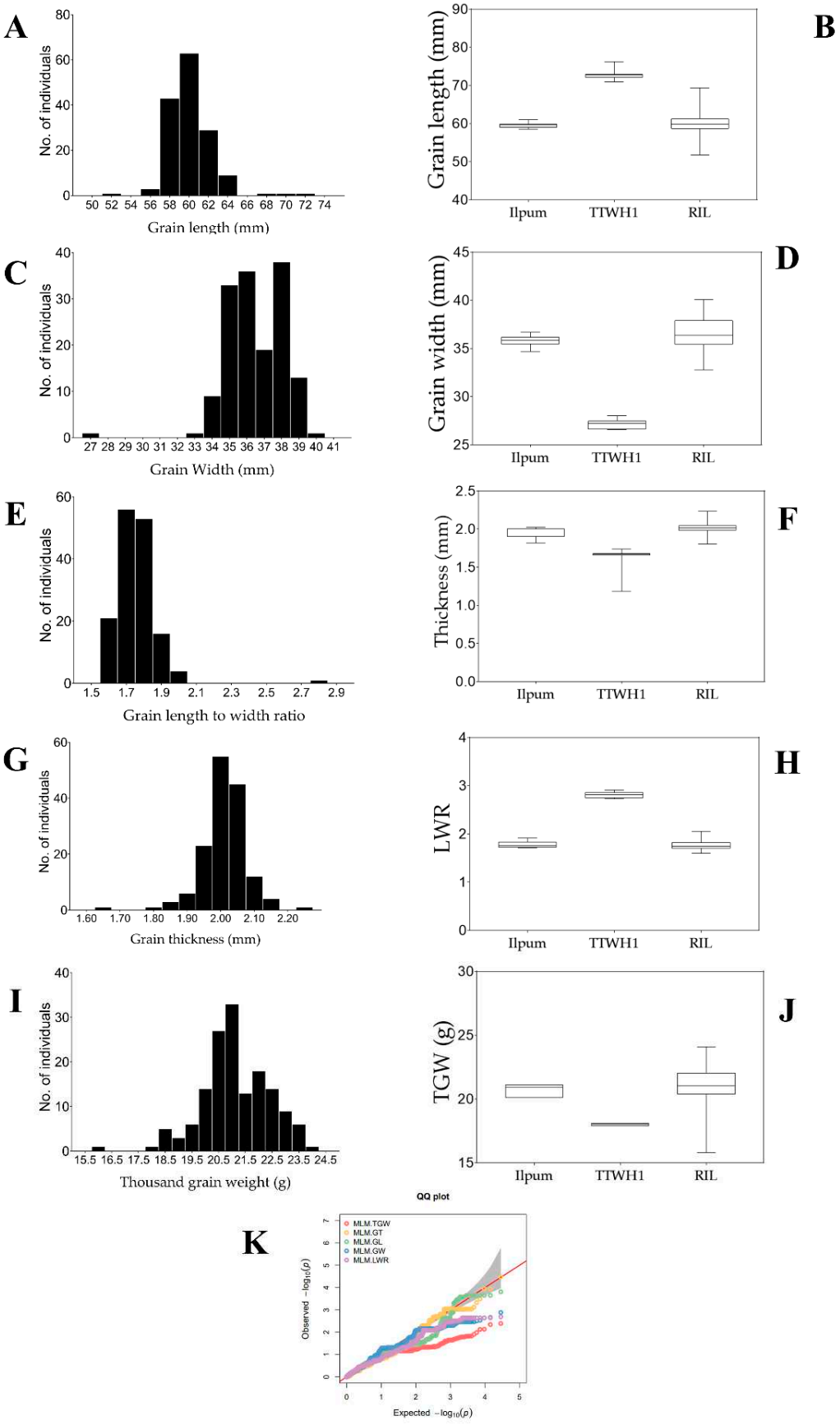
Furthermore, to understand the proportion of variation explained by the individuals' breeding values for the target traits, we estimated the narrow sense heritability ($h^2$) of traits. Data in panel D indicate that grain length had an $h^2$ of 0.915, while grain width, grain thickness, grain width, and thousand-grain weight showed an $h^2$ of 0.885, 0.454, 0.852, and 0.831, respectively (Figures 3E–H).

To further gain insights and assess the genetic merits of individuals in the RIL population for the target traits, we performed a genomic selection analysis based on the MLM (gBLUP) method known to have a high prediction accuracy for genomic estimated breeding value (GEBV) for traits

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



controlled by a large number of genes. The resulting output of the genomic selection analysis shows the predicted and observed GEBV of individuals in the RIL population for thousand-grain weight in the reference (Figure 3I) and inference (Figure 3J) groups.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



**Figure 2.** Frequency distribution of traits, box plots, and parental phenotypes. (**A–E**) frequency distribution of traits, (**F–J**) box plots showing the shift in grain trait values in the doubled haploid lines relative to their parental lines, and (**K**) Quantile–Quantile (Q–Q) plot.

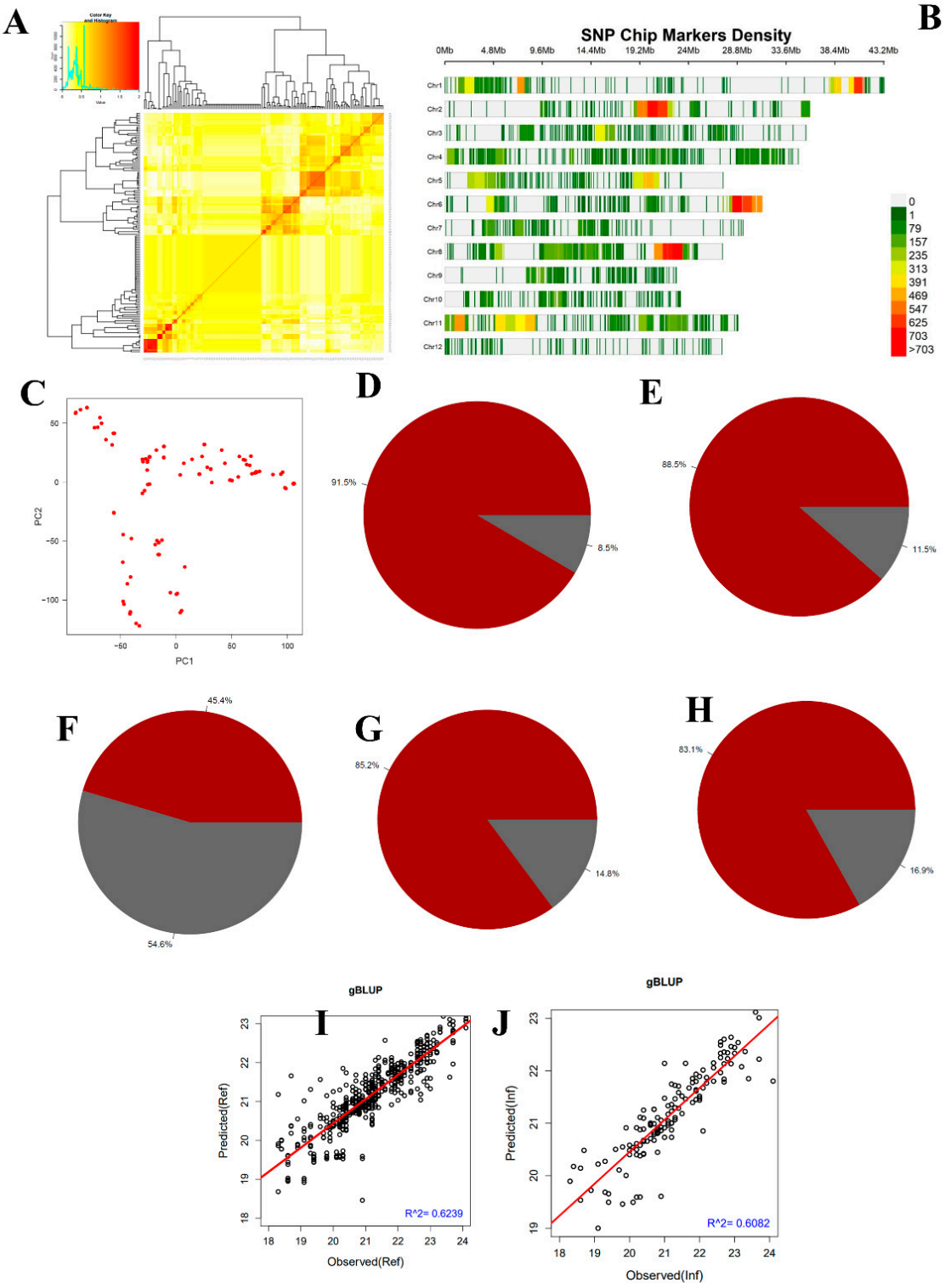



**Figure 3.** Kinship matrix, marker density, PCA, heritability, and Genome selection results. (**A**) heat map showing the relatedness or the level of co-ancestry of the population, (**B**) Density map of SNP Chip DNA markers, (**C**) principal component analysis( PCA), (**D–H**) narrow sense heritability of traits, and (**I,J**) genome prediction of genomic estimated breeding value (GEBV) of individuals.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



Correlation analysis is useful for understanding the relationship between two variables or identifying possible inputs for testing changes in a dependent variable while holding other variables constant. To explore the relationship between traits, we conducted a correlation analysis using the Pearson correlation method. Results in panels A and D in Figure 4 reveal a weak positive correlation between grain length ($R^2$=0.153***) or grain thickness ($R^2$=0.205***) and thousand-grain weight. In contrast, panel B in Figure 4 suggests the existence of a strong positive correlation between grain width and thousand-grain weight ($R^2$=0.392***).

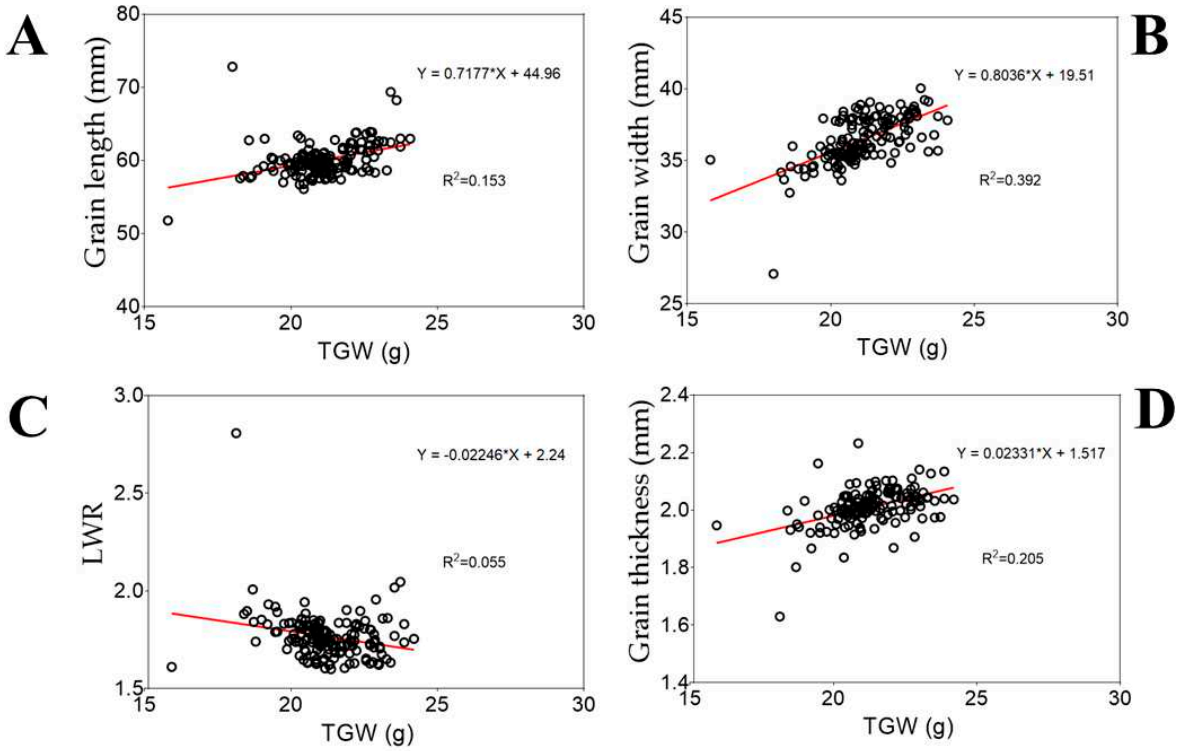



**Figure 4.** Pearson Correlation analysis results between traits. (**A**) Correlation results between grain length and thousand-grain weight, (**B**) grain width and TGW, (**C**) length-to-width and TGW, and (**D**) grain thickness and TGW.

*2.3. Identified QTLs for Grain Size and Shape in Rice through GWAS with Multiple GAPIT Models*

We conducted a GWAS employing multiple GAPIT models, with enhanced power and accuracy for genome association, to investigate novel genetic loci for grain size and shape, and thousand-grain weight (TGW), which are essential rice yield's components. GWAS results identified 43 QTLs (all grain traits considered and GAPIT models cumulated), of which number 10 QTLs were associated with grain length (GL), 14 with grain width (GW), 3 with grain thickness (GT), 8 with length-to-width ratio (LWR), and 8 with TGW (Table 1). From these results, we were interested to see co-detected QTLs with the highest contribution to the trait values. In this regard, we found that among the detected GL-related QTLs, seven out of ten were co-detected by both BLINK and FarmCPU GAPIT models), and mapped on chromosomes 1–3, 5, 7, and 11 (Figure 5A–N). Among them, four QTLs were co-detected by all GAPIT models, and two QTLs by three GAPIT models, of which AX-95918134 (*qGL1-1*$^{BFSG}$, Chr1, 3820526 bp, allelic effect: TTWH1) explains 72.5% –65.5% of the observed phenotypic variance (PVE; Figure 5A, Tables 1 and S1).

Likewise, among the fourteen QTLs associated with grain width (GW) identified here, six QTLs were co-detected 2–4 GAPIT models; were *qGW1-1*$^{BFSG}$ (AX-273945773, Chr1:5623288 bp) explains a PVE of about 15.5%–18.9%, and the allele from Ilpum contributed to the trait value. Besides, the GW-related QTL *qGW6-1*$^{Blink}$ coincides with the *qTGW6*$^{BFG}$ locus for TGW, which can be noted as *qGW6-1*$^{Blink}$/*TGW6*$^{BFSG}$ (AX-115737727, Chr6: 28484619 bp) (Figure 5B, Table 1). Other QTLs for GW are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



located on Chr1–3, 6, 8, and 12. The *qGW6-2*$^{FSG}$, *qGW8*$^{FSG}$ and *qGW12*$^{FSG}$ were co-detected by FarmCPU, SUPER, and GLM.

Concerning grain thickness (GT), three QTLs (AX-279261704, Chr1: 18023142 bp, PVE 74.9%; AX-154787777, Chr2: 2118477 bp, PVE 54.9%, and AX-154913392, Chr2: 25105471 bp, PVE 5.3%) were detected by BLINK (*qGT1*$^{Blink}$) and FarmCPU (*qGT2-1*$^{Farm}$ and *qGT2-2*$^{Farm}$) (Figure 5C, Table 1). Meanwhile, four out of eight QTLs associated with length–to–width ratio (LWR), were co-detected by 2–3 GAPIT models, with *qLWR2*$^{BFSG}$ (AX-274833045, Chr2: 10000097 bp, allelic effect: TTHW1) being the only one co-detected by all four GAPIT models. However, *qLWR6-1*$^{FSG}$ (AX-115851421, Chr6: 10178858 bp, recorded the highest PVE value (PVE 30.5%) (Figure 5D, Table 1).

Thousand-grain weight (TGW) is an important component of rice yield, and is determined by several factors, including GL, GW, and GT, among others. Our data in Table 1 shows that two out of eight QTLs associated with the control of TGW were co-detected by all four GAPIT models; meanwhile, BLINK and FarmCPU co-detected two other. The SNP Chip marker AX-115737727, linked to the *qTGW6*$^{BFSG}$ QTL (Chr6: 28484619 bp), which coincides with GW QTL *qGW6-1*$^{Blink}$ as indicated earlier, is here regarded as the major QTL for TGW identified by this study, considering its co-detection by all GAPIT models used and its high PVE value for TGW. The latter is followed by *qTGW2-1*$^{BF}$ (AX-279699609, Chr2 (10805604 bp, PVE 18.7%–27.9%) and *qTGW3-2*$^{Farm}$ (AX-123153600, Chr3: 7887961 bp, PVE 13.9% (Figure 5E, Tables 1 and S1).

### *2.4. Putative Candidate Genes Harbored by Grain Traits-Related QTLs*

Following the detection of major QTLs associated with the control of target grain size or shape-related traits, we were interested in unraveling the identity of genes harbored by these QTLs. To achieve that, we used the known physical positions of associated SNP Chip DNA markers co-detected by both BLINK and FarmCPU, in the rice genome database (http://rice.uga.edu/cgi-bin/gbrowse/rice/#search, accessed on September 1, 2023). From data in Table 2, we can see that genes harbored by *qGL1-1*$^{BFSG}$ are proposed to be involved in post-embryonic development, reproduction, and/or signal transduction, secondary metabolic (Os01g07880 and Os01g07930, encoding a HY5 and Zinc finger transcription factors, respectively). In the same region, genes associated with transport events (Os01g07870, encoding an ATP binding cassette (ABC) transporter), protein modification process (Os01g07920) or cellular homeostasis (Os01g07950, encoding a glutaredoxin subunit II), protein binding activities (Os01g07980, encoding an Ankyrin repeat domain), or response to abiotic stimuli (Os01g07910, encoding NADH-cytochrome b5 reductase) are found.

The *qGW1-1*$^{BFSG}$ region (associated with the control of grain width in rice) harbors genes with similar annotated functions to those found in *qGL1-1*$^{BFSG}$. The latter includes the Os01g10580 (Encoding a B-box (BBX) zinc finger transcription factor protein) proposed to be involved in post-embryonic development, cellular component organization, or secondary metabolic process; the Os01g10590 (*OsFTL8*, encoding an FT-like 8 homologous to flowering locus T gene) involved in flower development and reproduction; the Os01g10550 (*OsDEFL35*, encoding Defensin-like DEFL protein), Os01g10600 (Aquaporin), or Os01g10610 (encoding a Brassinosteroids-regulated transcription factor BES1/BZR1 protein) involved in protein binding and transport activity, respectively.

Likewise, in the *qGT1-1*$^{Blink}$ (for grain thickness on Chr1), the Os01g32930 gene (encoding an SGS domain-containing protein) are proposed to be involved in embryo development, reproduction, or post-embryonic development, among others. The *qGT1-1*$^{Blink}$ also harbors genes encoding transcription factors (Os01g32920, ZOS1-08, a C2H2 zinc finger TF), transport-related proteins (Os01g32880, AP-3 complex protein DnaJ), or protein metabolic process (Os01g32800, a proteasome subunit, PINT motif (Proteasome, Int-6, Nip-1 and TRIP15)). In the same way, *qGT2-1*$^{Farm}$ carries genes associated with lipid metabolic process, multicellular organization, or flower development (Os02g04690, Os02g04725, OsSPL3 TF (Os02g04680), or transcriptional regulatory event (Os02g04640, myeloblastosis (MYB)-like DNA binding domain), etc.

Like in the case of other genetic loci, putative candidate genes were pooled from QTLs co-detected by at least two GAPIT models. Otherwise, independent QTLs, detected by either BLINK,

 doi:10.20944/preprints202309.1280.v1



FarmCPU, or GLM, with the highest PVE value were considered. Thus, in the case LWR, *qLWR2-1*$^{BFSG}$ (AX-274833045, Chr2: 10000097 bp (PVE 15.2% (BLINK) or 32.9% (FarmCPU)), was retained to uncover the identity of putative candidate genes. In this region (*qLWR2-1*$^{BFSG}$), a set of genes encoding interesting annotated predicted functions are found. We could mention the VHS (VPS-27, Hrs, and STAM) and GAT (GGA and Tom1, Os02g17350 responsible for ubiquitin binding and ubiquitination), the Os02g17380, encoding pentatricopeptide (PPR) repeat domain-containing protein associated with the restoration of fertility (Cytoplasmic male sterility, CMS), the restoration of fertility 2 (*Rf2*, Os02g17380, encoding a mitochondrial glycine-rich protein) in LD-CMS, the Os02g17390 (encoding 3-hydroxyacyl-CoA dehydrogenase [45]), involved in flower development or multicellular organismal development, and the Tesmin/TSO1-like transcription factor (Os02g17460).

Concerning TGW, the major QTL (AX-115737727 *qGW6-1*$^{Blink}$/*qTGW6*$^{BFG}$), co-detected by all GAPIT models, harbors genes such as Os06g46910 coding for a ZOS6-07 C2H2 zinc finger transcription factor, Os06g46920 (encoding dihydroflavonol-4-reductase, associated with fatty acid catabolism, gibberellin biosynthesis and signaling, or seed dormancy). In the same QTL region (*qTGW6*$^{BFG}$), the *Os6bglu25* gene (Os06g4930, encoding a beta-glucosidase homologue proposed to be involved in carbohydrate metabolism), Os06g46950 (encoding an EF hand protein) associated with anatomical structure morphogenesis, cell differentiation or cellular component organization, are found.

Considering the genetic variability between the *japonica* and *indica* rice subspecies, we were interested to see the degree of similarity of genes found in major QTLs for grain traits identified in Table 2. To achieve that, the coding sequence (CDS) of each genes in the *japonica* group were aligned with their orthologues in the *indica* group. Results in Table 2 (Column 7, CDS *japonica* vs *indica*) reveal mutations sites (deletion or substitution) in a set of genes, while others showed a 100% similarity between the two subspecies.

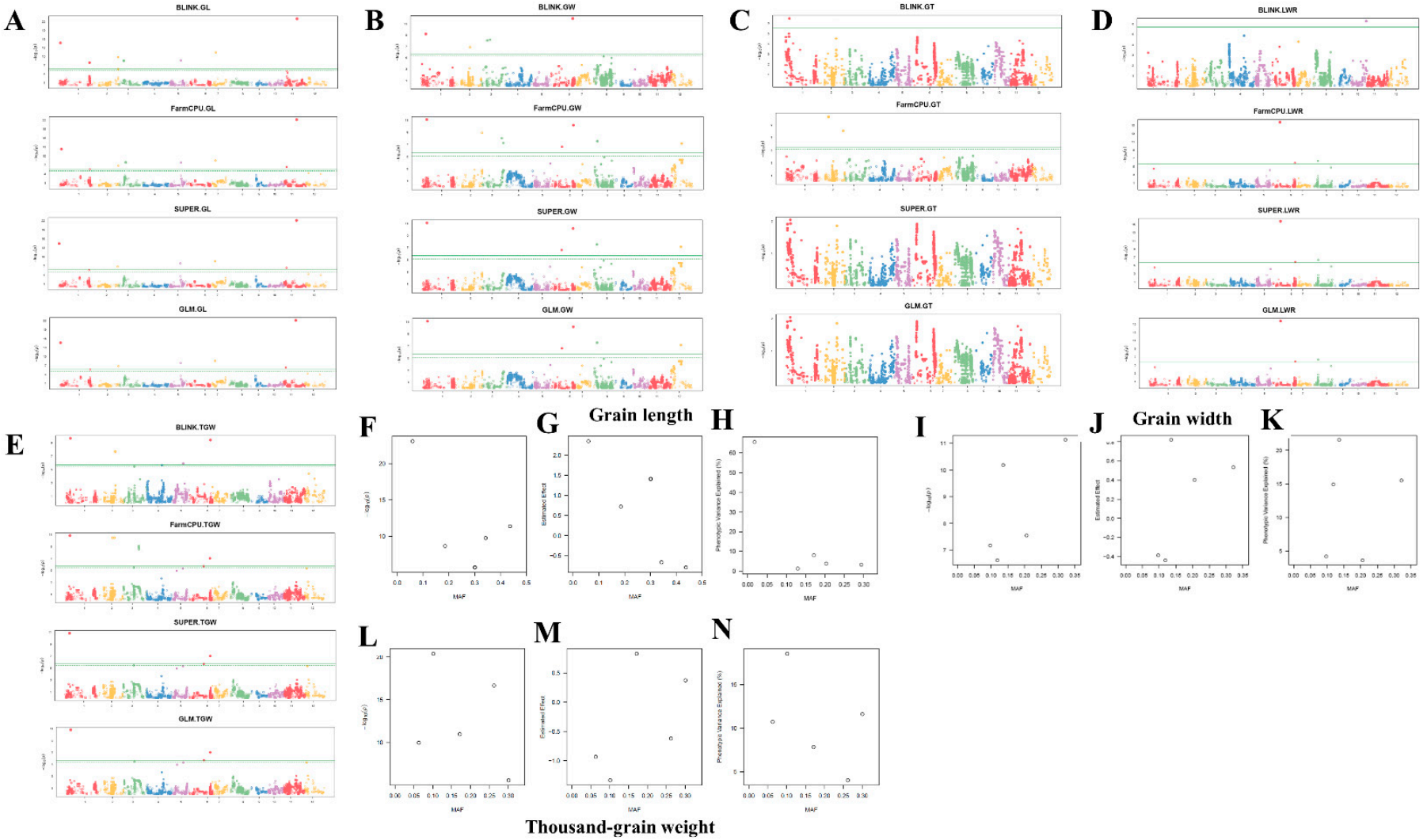



**Figure 5.** Manhattan plots, QTL estimated effects, and Phenotypic variance explained. (A–E) Manhattan plots showing significant SNP Chip DNA markers with their associated traits, detected by BLINK, FarmCPU, SUPER, and/or GLM GAPIT models. (F,I,L) logarithm of the odds (LOD) scores for significant SNP Chip DNA markers linked to grain traits loci, (G,J,M) estimated effects of QTLs, and (H,K,N) phenotypic variance explained (PVE) values of QTLs.



doi:10.20944/preprints202309.1280.v1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



**Table 1.** Detection grain traits-related QTLs by GAPIT model.

| Traits/QTLs | SNP markers | Chr | Position (bp) | PVE (%) | GWAS-GAPIT Models | | | | Allele |
|---|---|---|---|---|---|---|---|---|---|
| **Grain length** | | | | | | | | | |
| *qGL1-1*$^{BFSG}$ | AX-95918134 | 1 | 3820526 | 72.5 | BLINK | FarmCPU | SUPER | GLM | TTWH1 |
| *qGL11-1* $^{BFSG}$ | AX-274862201 | 11 | 16356105 | 31.9 | BLINK | FarmCPU | SUPER | GLM | Ilpum |
| *qGL2-1*$^{BFSG}$ | AX-115751685 | 2 | 35100558 | 22.9 | BLINK | FarmCPU | SUPER | GLM | TTWH1 |
| *qGL3-1*$^{BF}$ | AX-154437636 | 3 | 2253428 | 5.5 | BLINK | FarmCPU | - | - | Ilpum |
| *qGL2-2*$^{BF}$ | AX-154880023 | 2 | 35133528 | 3.1 | BLINK | FarmCPU | - | - | TTWH1 |
| *qGL7-1*$^{BFSG}$ | AX-153903748 | 7 | 6690089 | 2.6 | BLINK | FarmCPU | SUPER | GLM | TTWH1 |
| *qGL1-2*$^{BF}$ | AX-279584700 | 1 | 41489588 | 1.7 | BLINK | FarmCPU | - | - | TTWH1 |
| *qGL5*$^{BFSG}$ | AX-282746698 | 5 | 16169537 | 1.4 | BLINK | FarmCPU | - | - | Ilpum |
| *qGL11-2*$^{FSG}$ | AX-115751092 | 11 | 2823622 | 14.0 | - | FarmCPU | SUPER | GLM | Ilpum |
| *qGL3-2*$^{Farm}$ | AX-154551783 | 3 | 8242439 | 8.0 | - | FarmCPU | - | - | TTWH1 |
| **Grain Width** | | | | | | | | | |
| *qGW1-1*$^{BFSG}$ | AX-273945773 | 1 | 5623288 | 18.9 | BLINK | FarmCPU | SUPER | GLM | Ilpum |
| *qGW2-1*$^{Blink}$ | AX-279699609 | 2 | 10805604 | 14.9 | BLINK | - | - | - | TTWH1 |
| *qGW1-2*$^{BF}$ | AX-115791785 | 1 | 43103625 | 8.8 | BLINK | FarmCPU | - | - | TTWH1 |
| *qGW1-3*$^{Blink}$ | AX-281116133 | 1 | 20864932 | 6.9 | BLINK | - | - | - | Ilpum |
| *qGW6-1*$^{BF}$ | AX-115737727 | 6 | 28484619 | 6.2 | BLINK | FarmCPU | - | - | Ilpum |
| *qGW3-1*$^{Blink}$ | AX-154073979 | 3 | 7895651 | 4.8 | BLINK | - | - | - | Ilpum |
| *qGW3-2*$^{Blink}$ | AX-115811160 | 3 | 14888685 | 4.1 | BLINK | - | - | - | Ilpum |
| *qGW6-2*$^{FSG}$ | AX-273990782 | 6 | 13986482 | 14.9 | - | FarmCPU | SUPER | GLM | TTWH1 |
| *qGW1-4*$^{Farm}$ | AX-280898927 | 1 | 2483022 | 9.4 | - | FarmCPU | - | - | TTWH1 |
| *qGW12*$^{FSG}$ | AX-284265976 | 12 | 13013702 | 4.2 | - | FarmCPU | SUPER | GLM | TTWH1 |
| *qGW8* $^{FSG}$ | AX-115796459 | 8 | 3875546 | 3.7 | - | FarmCPU | SUPER | GLM | Ilpum |
| *qGW2-2*$^{Farm}$ | AX-279994820 | 2 | 35461009 | 3.5 | BLINK | FarmCPU | - | - | TTWH1 |
| *qGW2-3* $^{Farm}$ | AX-154042022 | 2 | 24704256 | 3.4 | BLINK | FarmCPU | - | - | Ilpum |
| *qGW3-3* $^{Farm}$ | AX-154797543 | 3 | 2973374 | 1.2 | BLINK | FarmCPU | - | - | Ilpum |
| **Grain thickness** | | | | | | | | | |
| *qGT1*$^{Blink}$ | AX-279261704 | 1 | 18023142 | 74.9 | BLINK | - | - | - | TTWH1 |
| *qGT2-1*$^{Farm}$ | AX-154787777 | 2 | 2118477 | 54.9 | - | FarmCPU | - | - | TTWH1 |
| *qGT2-2*$^{Farm}$ | AX-154913392 | 2 | 25105471 | 5.3 | - | FarmCPU | - | - | Ilpum |
| **Length-to-Width Ratio** | | | | | | | | | |
| *qLWR10*$^{Blink}$ | AX-115835839 | 10 | 22038978 | 26.5 | BLINK | - | - | - | Ilpum |
| *qLWR2*$^{BFSG}$ | AX-274833045 | 2 | 10000097 | 15.2 | BLINK | FarmCPU | SUPER | GLM | TTWH1 |
| *qLWR1-1*$^{BF}$ | AX-154960834 | 1 | 1595394 | 13.5 | BLINK | FarmCPU | - | - | TTWH1 |
| *qLWR1-2*$^{Blink}$ | AX-115737888 | 1 | 600441 | 10.7 | BLINK | - | - | - | TTWH1 |
| *qLWR3*$^{Blink}$ | AX-154834762 | 3 | 8098398 | 6.9 | BLINK | - | - | - | TTWH1 |
| *qLWR6-1*$^{FSG}$ | AX-115851421 | 6 | 10178858 | 30.5 | - | FarmCPU | SUPER | GLM | Ilpum |
| *qLWR6-2*$^{FSG}$ | AX-155522120 | 6 | 30842264 | 9.4 | - | FarmCPU | SUPER | GLM | TTWH1 |
| *qLWR8*$^{FSG}$ | AX-154176130 | 8 | 5398451 | 5.9 | - | FarmCPU | SUPER | GLM | TTWH1 |
| **Thousand grain weight** | | | | | | | | | |
| *qTGW6*$^{BFSG}$ | AX-115737727 | 6 | 28484619 | 32.8 | BLINK | FarmCPU | SUPER | GLM | Ilpum |
| *qTGW2-1*$^{BF}$ | AX-279699609 | 2 | 10805604 | 18. 6 | BLINK | FarmCPU | - | - | TTWH1 |
| *qTGW3-2*$^{Farm}$ | AX-123153600 | 3 | 7887961 | 13.9 | - | FarmCPU | - | - | Ilpum |
| *qTGW1-1*$^{Blink}$ | AX-154298059 | 1 | 5644298 | 11.6 | BLINK | - | - | - | Ilpum |
| *qTGW3-1*$^{BF}$ | AX-154471576 | 3 | 15332432 | 10.7 | BLINK | FarmCPU | - | - | TTWH1 |
| *qTGW2-2*$^{BF}$ | AX-154096541 | 2 | 10773042 | 7.8 | BLINK | FarmCPU | - | - | Ilpum |
| *qTGW1-2*$^{BFSG}$ | AX-154333920 | 1 | 5860250 | 4.9 | BLINK | FarmCPU | SUPER | GLM | Ilpum |
| *qTGW1-3*$^{BF}$ | AX-154810092 | 1 | 42931550 | 4.03 | BLINK | FarmCPU | - | - | TTWH1 |

GL: grain length, GW: grain width, GT: grain thickness, LWR: Length-to-width ratio, and TGW: thousand-grain weight. Chr: chromosome, MAF: minor allelic frequency, nobs: number of observations, PVE: phenotype variance explained. qTrait$^{Blink}$: QTL detected by BLINK only, qTrait$^{Farm}$: QTL detected by FarmCPU only, qTrait$^{FSG}$: QTL co-detected by FarmCPU, SUPER, and GLM, qTrait$^{BFSG}$: QTL co-detected by BLINK, FarmCPU, SUPER, and GLM.

**Table 2.** Candidate Genes harbored by qGL1-1$^{BFSG}$, qGW1-1$^{BFSG}$, qGT1$^{Blink}$ and qGT2-1$^{Farm}$, qLWR2-1$^{BFSG}$, and qTGW6$^{BFSG}$ loci.

| No. | japonica/indica | Description | Biological process | Molecular function | Cellular component | CDS *japonica* vs *indica*/*Similar Report* |
|---|---|---|---|---|---|---|
| | ***qGL1-1$^{BFSG}$*** | **Chr1:3804000..3883000** | | | | |
| 1 | Os01g07870 | ATP-bindinc cassette (ABC) transporter family protein, Peroxidase 56 | Transport | Hydrolase activity, transporter activity | Extracellular region, integral component of membrane, vacuole | -; [46] |
| 2 | Os01g07880 | OsbZIP01/OsRE1, Transcription factor HY5, putative, expressed | Post-embryonic development, signal transduction, secondary metabolism | Sequence-specific DNA binding transcription factor activity | Nucleus | -; [47] |
| 4 | Os01g07910/ BGIOSGA002284 | NADH-cytochrome b5 reductase, putative | Response to stress, response to abiotic stimulus | Binding, catalytic activity | Cell wall, mitochondrion | 100% similar |
| 5 | Os01g07920 | Prolyl 4-hydroxylase, putative | Protein modification process | Binding, catalytic activity | Golgi apparatus, vacuole, membrane | -; - |
| 6 | Os01g07930/ BGIOSGA002287 | Zinc finger C-x8-C-x5-C-x3-H (CCCH)-domain containing protein family, transcription factor | Biosynthetic process | Sequence-specific DNA binding transcription factor activity | - | 100% similar |
| 7 | Os01g07940/ BGIOSGA002282 | AGC_PVPK_like_kin82y.3 - ACG kinases include homologs to PKA, PKG and PKC | Reproduction, post-embryonic development, embryo development, protein modification process | Nucleotide binding, kinase activity | - | Deletion in *japonica* (126–131 bp); indica (924–932, 1201–3 bp), and SNPs |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| 8 | Os01g07950 | OsGrx_S15.2 - glutaredoxin subgroup II | Cellular homeostasis | Binding | Mitochondrion | -; - |
| 9 | Os01g07960 | Acyl-protein thioesterase, Similar to Biostress-resistance-related protein | - | Hydrolase activity | - | - |
| 10 | Os01g07980 | Ankyrin, putative, expressed. SGT1, suppressor of G2 allele of SKP1; Provisional | - | Binding | - | -; [48] |
| 11 | Os01g08000 | Fibronectin type 3 and ankyrin repeat domains 1 protein | - | Protein binding | - | -; - |
| 12 | Os01g08020/ BGIOSGA002278 | Boron transporter protein, Bicarbonate transporter, eukaryotic domain containing protein. | Anion transport, | Borate efflux transmembrane transporter activity; inorganic anion exchanger activity | Integral component of membrane | Deletion in *japonica* (1–174 bp) |
| | ***qGW1-1*$^{BFSG}$** | **Chr1:5623500..5684500** | | | | |
| 13 | Os01g10550 | DEFL35 - Defensin-like DEFL family | | | | - |
| 14 | Os01g10580/ BGIOSGA002958 | B-box (BBx) zinc finger family protein, transcription factor | Post-embryonic development, cellular component organization, secondary metabolic process, response to abiotic stimulus | Sequence-specific DNA binding transcription factor activity | Nucleoplasm | Deletion in *indica* (1–184; 197; 241; 841–846) |
| 15 | Os01g10590/ BGIOSGA002959 | OsFTL8 FT-Like8 homologous to Flowering Locus T gene | Flower development, reproduction, post-embryonic development, response to abiotic stimulus | Protein binding, lipid binding | Nucleus, cytoplasm | Deletion in *japonica* (161–194 bp) |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| 16 | Os01g10600 | OsNIP1;2 encoding Aquaporin protein, putative, expressed | Transport | Transporter activity | Membrane, plasma membrane | -; [49] |
| 17 | Os01g10610/ BGIOSGA002172 | BRI1-EMS-SUPPRESSOR1/ BRASSINOZANOL RESISTANT 1 (BES1/BZR1); transcriptional repressor family protein. | Brassinosteroids signaling- | - | - | Deletion in *indica* (1–93 bp) and SNPs (831:G/T, 836: T/C, 879: C/T); [50,51] |
| | ***qGT1-1$^{Blink}$*** | **Chr1:17993000..18054000** | | | | |
| 18 | Os01g32780/ BGIOSGA001545 | Universal stress protein domain-containing protein, UspA domain containing protein | Response to stress, response to molecule of fungal origin | - | - | 100% similar; [52] |
| 19 | Os01g32800/ BGIOSGA001543 | Proteasome subunit, putative, expressed. PCI domain, also known as PINT motif (Proteasome, Int-6, Nip-1, and TRIP-15). | Protein metabolic process | Protein binding | Nucleus, intracellular, cytosol, proteasome complex | Deletion in *indica* (972–1004 bp) |
| 20 | Os01g32870 | Heat shock protein DnaJ, Similar to Chaperone protein dnaJ 15 (Protein ALTERED RESPONSE TO GRAVITY) (AtARG1) (AtJ15) (AtDjB15). | Protein metabolic process, response to abiotic stimulus, protein binding tropism | - | - | -; - |
| 21 | Os01g32880 | AP-3 complex subunit delta, Armadillo-type fold domain containing protein | Intra-Golgi vesicle-mediated transport, protein storage vacuole organization | Transporter activity, protein binding | Membrane, cytoplasm, Golgi apparatus | -; - |
| 22 | Os01g32920/ BGIOSGA003627 | ZOS1-08 - C2H2 zinc finger protein, expressed, Transcription factor | Biosynthetic process | Sequence-specific DNA binding transcription factor activity | Intracellular | SNP689: T/C |
| 23 | Os01g32930/ BGIOSGA003628 | SGT1-specific (SGS) domain-containing protein | Embryo development, reproduction, post- | - | Cytosol | Deletion in *indica* (1–12, 270, 275–294 bp), *japoica* (479–484), SNPs (272: T/C, 319: C/A, 447: C/T, 504–505: GG/AC) |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | embryonic development, protein binding, signal transduction, protein metabolic process, response to biotic stimulus | | | |
| | ***qGT2-1*** $^{Farm}$ | **Chr2:2088000..2151000** | | | | |
| 24 | Os02g04630 | Sodium/calcium exchanger protein, putative, expressed. The Ca2+:Cation Antiporter (CaCA) Family (TC 2.A.19) proteins | Transport | Transporter activity | Cell, vacuole, membrane | -; - |
| 25 | Os02g04640/ BGIOSGA007162 | PHOSPHATE STARVATION RESPONSE 3 (OsPHR3), Myb-like DNA-binding domain containing protein, , transcription factor | Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process | Sequence-specific DNA binding transcription factor activity | - | 100% similar; [53] |
| 26 | Os02g04650/ BGIOSGA007161 | Activator of 90 kDa heat shock protein ATPase homolog | Catabolic process | Enzyme regulator activity, protein binding | - | 100% similar; - |
| 27 | Os02g04660 | Arginine N-methyltransferase 5 | Response to abiotic stimulus, protein modification process | Transferase activity | Cytosol | -; - |
| 28 | Os02g04670/ BGIOSGA007498 | Glucan endo-1,3-beta-glucosidase precursor | Carbohydrate metabolic process | Binding, hydrolase activity | Plasma membrane, membrane | Deletion in *japonica* (44–52 bp) |
| 29 | Os02g04680/ BGIOSGA007499 | Squamosa promoter-binding-like protein 3 (OsSPL3) - SBP-box gene family member, Transcription factor | Flower development, multicellular organismal development | Sequence-specific DNA binding transcription factor activity | Nucleus | 100% similar; [54] |
| 30 | Os02g04690 | Cycloartenol synthase | Multicellular organismal | Catalytic activity | Vacuole | -; - |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | development, cellular component organization, lipid metabolic process | | | |
| 31 | Os02g04700 | tRNA synthetases class II domain-containing protein | Translation | Catalytic activity, nucleic acid binding | Cytosol, cytoplasm | -; - |
| 32 | Os02g04710 | Cycloartenol synthase | Multicellular organismal development, cellular component organization, lipid metabolic process | Catalytic activity | Vacuole | -; - |
| 33 | Os02g04725 | Dolichol phosphate-mannose biosynthesis regulatory protein | Macromolecule biosynthetic process | - | Cell, integral component of endoplasmic reticulum membrane | -; - |
| | ***qLWR2*$^{BFSG}$** | **Chr2:9970000..10030000** | | | | |
| 34 | Os02g17350/ BGIOSGA007951 | VPS-27, Hrs, and STAM (VHS) and GGA and Tom1 (GAT) domain-containing protein | Transport | Transporter activity | Golgi apparatus, plasma membrane | 100% similar |
| 35 | Os02g17360/ BGIOSGA006711 | Restorer of fertility gene, Rf, pentatricopeptide repeat (PPR) repeat domain-containing protein | Mitochondrial cytoplasmic male sterility (CMS) | Nuclease activity | Plastid, mitochondrion | deletion in *indica* (1–84 bp); [55] |
| 36 | Os02g17380 | Fertility restorer 2 (Rf2), Mitochondrial glycine-rich protein, Fertility restoration in LD-CMS | - | - | - | -; [56] |
| 37 | Os02g17390/ BGIOSGA007953 | ABNORMAL INFLORENSCENCE MERISTEM 1(MFP/AIM1); 3-hydroxyacyl-CoA dehydrogenase | Flower development, multicellular organismal development, post-embryonic | Catalytic activity | Plastid, cell wall, peroxisome | 100% similar; [45] |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | development, lipid metabolic process | | | |
| 38 | Os02g17400/ BGIOSGA006709 | Leucine-rich repeat protein | Signal transduction, response to biotic stimulus, response to stress | - | Cell wall | Deletion in *indica* (96–101 bp) |
| 39 | Os02g17460 | Tesmin/TSO1-like CXC domain-containing protein; transcrption factor | Biosynthetic process | Sequence-specific DNA binding transcription factor activity | - | -; - |
| | ***qTGW6$^{BFSG}$*** | **Chr6:28484608..28484625** | | | | |
| 40 | Os06g46910/ BGIOSGA023481 | ZOS6-07 - C2H2 zinc finger transcription factor, expressed | Biosynthetic process | Sequence-specific DNA binding transcription factor activity | Intracellular | (SNP329: A/C; SNP445: T/C; SNP676: A/G; SNP1318: G/A); [57] |
| 41 | Os06g46920 | Dihydroflavonol-4-reductase, NAD(P)-binding domain containing protein | Fatty acid catabolic process, gibberellin (GA) biosynthesis process, Seed dormancy process, GA-mediated signaling pathway | Cinnamyl-alcohol dehydrogenase activity, coenzyme binding, nucleotide binding, catalytic activity | - | -; - |
| 42 | Os06g46930/ BGIOSGA020659 | 50S ribosomal protein L24, chloroplast precursor (CL24) | Pastid translation | Structural constituent of ribosome | Ribosome, plastid, large ribosomal subunit, chloroplast stroma | (SNP51: T/G; SNP282: A/G) |
| 43 | Os06g46940 | *Os6bglu25* - beta-glucosidase homologue, similar to *Os3bglu6*, expressed | Carbohydrate metabolic process | Hydrolase activity, binding | Cell wall | -; [58] |
| 44 | Os06g46950/ BGIOSGA023482 | carotenoid cleavage dioxygenase 1(OsCCD1), EF-hand calcium | Anatomical structure morphogenesis, | Calcium ion binding | - | 100% similar; [59,60] |



doi:10.20944/preprints202309.1280.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| | | ($Ca^{2+}$)-binding protein familyexpressed | cellular component organization, cell differentiation, multicellular organismal development | | | |
| 45 | Os06g46995 | Armadillo/beta-catenin repeat family protein, putative, expressed | - | Protein binding | - | -; - |
| 46 | Os06g47000/ BGIOSGA020655 | External NADH-ubiquinone oxidoreductase 1, mitochondrial precursor, putative, expressed | Metabolic process | Catalytic activity | Membrane, mitochondrion | 100% similar; - |



doi:10.20944/preprints202309.1280.v1

 doi:10.20944/preprints202309.1280.v1



## 3. Discussion

### *3.1. Grain Length, Width, and Thickness are Closely Related to Thousand Grain Weight but Not Length-to-Width Ratio*

Understanding the correlation between factors helps quantify the strength of the direct relationship between them and figure out their affiliation [61]. Thousand-grain weight (TGW) is a determinant component of rice yield, and is influenced by several factors, including grain length (GL), width (GW), and thickness (GT) [62–64]. In addition to the grain-filling [65], it has been established that grain weight is determined by factors such as grain length and width, which contribute to enhancing the yield of rice [66,67]. Hence, the observed strong positive correlation between grain width and thousand-grain weight ($R^2$=0.392***), and that between grain length ($R^2$=0.153***) or grain thickness ($R^2$=0.205***) and thousand-grain weight would partially explain the shift in the thousand-grain weight of the RIL population as shown in panels I and J in Figure 2.

### *3.2. Genomic Estimated Breeding Value of RILs Population and Traits Heritability*

The use of genomic selection (GS) in plant breeding has proven essential to increase the genetic gain of complex traits per unit time and cost by enhancing the genomic estimated breeding value (GEBV) accuracies, through employing dense markers, and traits heritability [68]. GS also estimates the genetic merit of individuals (in this case the RILs) based on a large set of dense markers (here SNP Chip DNA makers) across the whole genome. GS then derives the GEBVs of all individuals in the breeding population based on their genotype and phenotype profiles and predicts those are suitable for downstream breeding programs, relying on their actual performance [69]. Here, data obtained from GS analysis revealed the GEBV profile of RILs for thousand-grain weight, which is useful for downstream breeding using best-performing RILs and associated SNP Chip DNA markers. It was interesting to see that grain length ($h^2$=0.915) and width ($h^2$=0.885), that were earlier shown to be closely related to thousand-grain weight ($h^2$=0.852), recorded high heritability scores. A study by Chen, *et al.* [70] observed a high heritability for grain shape and weight have a high heritability rate, but environmental factors, including temperature, largely influence the phenotypic values of these traits.

### *3.3. The qGL1-1$^{BFSG}$ QTL Harbors Genes Involved in Post-Embryonic Development and Reproduction*

Grain length-related QTLs have been reported on Chr1 (*qGL1*), Chr2 (*qGL2.1, qGL2.2*), Chr3 (*GS3*), Chr4 (*qGL4*), Chr6 (*qGL6*), Chr7 (*qGS7, qGL7*), Chr8 (*qGL8.1*), Chr10 (*qGL10*), Chr11 [70,71]. Here, we noted that the major QTL *qGL1-1$^{BFSG}$* associated with the control of grain length in rice harbors genes proposed to be involved in reproduction, post-embryonic development, and embryo-development or protein modification events (Os01g07880 (HY5: elongated hypocotyl 5) and Os01g07940 (AGC-PVPK)). The HY5 encodes a bZIP (basic leucine zipper) transcription factor highly conserved across plant species, and it is described as central regulator of light signaling, acting as a pivotal regulator of light-dependent development [72]. The HY5 also functions in the regulation of nutrient uptake and utilization by controlling the expression of a large set of genes involved in nitrogen uptake and transport [73–75]. Other reports suggest the role of HY5 in light-mediated root growth [76], sucrose efflux events (by inducing the expression of *SWEET11* and *SWEET12* (*SUCROSE TRANSPORTER*) [75,77]. Likewise, HY5 physically interacts with a group of B-box proteins (BBXs) [78–81] and other proteins [82] to regulate the expression of several target genes as well as multiple molecular and biological events.

In the same region (*qGL1-1$^{BFSG}$*), a gene encoding a Zinc finger (CCCH) encoding a TF and two others encoding Ankyrin repeat domain-containing protein. Genes encoding the CCCH Zinc-finger protein have been proposed to regulate the adaptation of plants to abiotic stress [83–85]. Likewise, Ankyrin repeat domain-containing protein-encoding genes are thought to exclusively function to meditate protein-protein interactions and disease response [86].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



### *3.4. The Grain Width, Thickness, and LWR-Associated QTLs qGW1-1$^{BSFG}$, qGT1$^{Blink}$, qGT2-1$^{Farm}$ and qLWR2$^{BSFG}$ Carry Genes Involved in Flower Development, Post-Embryonic Development and Reproduction*

Several loci controlling grain thickness (GT), width (GW), and length–to–width ratio (LWR) have been reported under various growth conditions, and mapped to almost all chromosomes of rice (Chr1, 2, 3, 6–9, 11, 12) [2,70,71]. In the *qGW1$^{BF}$* region, we noticed the presence of a gene encoding the *flowering time-like 8* locus (*OsFTL8*, Os01g10590), associated with flower development and reproduction. A previous report proposed that a member of the FTL family, *OsFTL4* (Os09g33850) regulates flowering time in rice in response to changing environmental conditions [87]. Likewise, a set of genes encoding a B-box (BBX) zinc finger protein (Os01g40580) or *OsNIP1* (Os01g10600, Aquaporin) are located within the *qGW1-1$^{BFSG}$* region. Members of the BBXs family are a class of zinc finger proteins that encode transcription factors, and are mapped across the rice genome [88–90]. Among them, the *OsBBX14* (Os05g11510) was proposed to promote photomorphogenesis in rice [88]. In the same way, aquaporin is mainly associated with water movement in- and outside the cell. A study conducted by He, *et al.* [91] revealed that *OsPIP1* encoding aquaporin interacts with other proteins to promote water uptake and seed germination. Furthermore, BES1/BZR1, a family of Brassinosteroids transcriptional regulator, were recently proposed to regulate plant development [92], kernel size in rice [50] and maize [51] through interaction with several proteins [93].

It was also interesting to see that genes located within the *qGT1$^{Blink}$* locus or *qGT2-1$^{Farm}$*, based on their predicted annotated functions, are associated with growth-related biological processes, including embryo development, reproduction, flower development (*OsSGT1*, Os01g322890; *OsSPL3*, Os02g04660), or transport, as well as transcriptional regulation (*ZOS1-08*, Os01g32920; PHR3, Os02g04640) [94].

As for the *qLWR2$^{BFSG}$*, this QTL harbors genes with interesting annotated functions, including two genes (Os02g17350 and Os02g17380, *OsRf2*) described as being involved in the restoration of fertility (cytoplasmic male sterility, CMS). The *Rf2* gene was earlier suggested to be involved in the mechanism for the restoration of fertility in CMS lines in rice [56].

### *3.5. The Grain Weight-Related QTL qTGW6$^{BFSG}$ Harbors Genes Associated with Anatomical Structure Morphogenesis, Cell Differentiation, and Carbohydrate Metabolism*

Thousand-grain weight (TGW) is controlled by several genetic loci. To date, many quantitative trait loci (QTLs) proposed to control TGW in rice have been identified, and mapped on all 12 chromosomes of rice, and a few genes have been functionally characterized. Multiple genetic and molecular aspects of plants affects grain weight, leading to dynamic changes in cell division, expansion, and differentiation [95].

The marker AX-115737727 is linked to the major QTL for TGW (*qTGW6$^{BFSG}$*, Chr6: 28484619 bp) that coincides with the *qGW6-1$^{Blink}$* QTL identified by the present study. We could mention the Os06g46950 encoding the carotenoid cleavage dioxygenase 1 (CCD1) protein, the ZOS6-07 C2H2 Zinc finger TF (Os06g46920) or the *Os6bglu25* (Os06g46940, encoding the β-glucosidase homologue). A study by Ren, *et al.* [58] suggested that a member of the β-glucosidase protein family, *Os06gGlu24* plays a role in seed germination and root elongation, while interacting with indole-3-acetic acid (IAA) and abscisic acid (ABA) signaling. Likewise, Ilg, *et al.* [59] proposed the *CCD1* gene as being involved in the control of endosperm color in rice.

Although grain length, width, thickness, or thousand grain weight are known to be controlled by multiple loci, genes harbored by *qGL1-1$^{BFSG}$*, *qGT2-1$^{Farm}$*, *qGW1-1$^{BFSG}$*, or *qTGW6$^{BFSG}$* share commonalities such as being involved in multicellular organismal development, flower development or reproduction, cell division or differentiation, among other annotations. It has been evidenced that TGW largely depends on GL, GW and GT [96], in addition to grain filling ratio.

 doi:10.20944/preprints202309.1280.v1



## 4. Materials and Methods

*Plant Materials, Growth Conditions, and Phenotypic Measurements*

A hundred and forty-three recombinant inbred lines (RILs), derived from a cross between Ilpum (*Oryza sativa* L. ssp. *japonica*) and Tung Tin Wan Hein1 (TTWH1, *Oryza sativa* L. ssp. *indica*) were used to conduct the experiments. Initially, pre-germinated seeds of RILs were sown and grown in 50-well trays until transplanting time. Then, healthy and vigorous four-week-old seedlings were transplanted (Cropping season May to October 2022) in the experimental field (altitude: 11 m, 35°29′31.4′′ N, and 128°44′30.0″ E), located at the National Institute of Crop Science (NICS), Department of Southern Area Crop Science, Paddy Crop Division, Rural Development Administration, Miryang, Republic of Korea.

Soon after harvesting and postharvest processing, the grain size and shape-related phenotypes, including grain length (GL), grain width (GW), grain thickness (GT), grain length-to-width ratio (LWR, calculated as the GL divided by GW), and thousand-grain weight (TGW) were measured. The GL, GT, GW, and LWR were measured or calculated using the SmartGrain v.1.2 (copyright© 2010-2012, Takanari TANABTA, Japan; http://phenotyping.image.coocan.jp). Before analysis, 100 rice seeds, with a label that helps identify the RIL under analysis, were placed on the Canon scanner 5600F model using a typical rectangular rice seed dispenser (Figure S1A–C), and the dispenser was removed thereafter. Seeds were scanned and the image saved in an appropriate folder, for further processing (Figure S1D,E). Prior to analyzing the phenotype of grains, basic settings are performed, such as the selection of seed detection sensitivity strength, picking seed and background colors by right-clicking inside the imported image, determining the scale bar, etc. To analyze, click on "Analyze" in the title bar, select "Analyze area" in the drop window, and select the target region on the open image to analyze. Final quality control was performed to ensure the accuracy of the measurement as follows: Set [Disable/Enable] (right click on the mouse) to unselect or select seeds on the image, followed by exporting as Excel "csv." Format (Figure S1F).

The GT was measured manually using a digital Vernier Caliper (CD-20CP, Mitutoyo Corp, Tokyo, Japan). However, the TGW was calculated as the [(average grain weight of 100 dehulled seeds/the number of samples (n)) × 10].

*Frequency Distribution, Correlation Analysis, Quantile–Quantile Plots, Kinship Matrix*

To assess the frequency distribution of traits, generate the box plots, and investigate the Pearson correlation between the target traits, GraphPad Prism 7.0 (© 1992–2016 GraphPad Software, Inc., ODESA) was used. The Quantile–Quantile (Q–Q) plots and the pairwise kinship matrix, also known as the co-ancestry or half relatedness, as well as the principal component analysis (PCA) plot were generated from the GAPIT package using R software. The SNP density plot was generated using filtered SNP Chip DNA markers with their relative *p*-values (GWAS results in .csv file) using the below script:

```
install.packages('CMplot')
library(CMplot)
head(my_data)
CMplot(my_data,type="p",plot.type="d",bin.size=1e6,chr.den.col=c("darkgreen", "yellow", "red"),file="jpg",file.name="",dpi=300, main="SNP Chip Markers Density",file.output=TRUE,verbose=TRUE,width=9,height=6)
```

Genomic Selection or Prediction Analysis

To investigate the genetic merit of the RILs for specific target traits, a genomic prediction or selection analysis was conducted as described by Zhang, *et al.* [97]. The genomic best linear unbiased prediction (gBLUP), commonly used for the genomic selection based on mixed model (MLM), and having a higher prediction accuracy for traits controlled by a large number of genes was used perform the genomic selection [98]. The genotype data was converted from the Haplotype Map (HapMap) format to numerical (see R script below) prior to performing the analysis.

To convert HapMap to numerical format:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2023 doi:10.20944/preprints202309.1280.v1



```
myG <-fread("file:///D:genotype data location.txt", head = FALSE)
myGAPIT <- GAPIT(G=myG, output.numerical=TRUE)
myGD= myGAPIT$GD
myGM= myGAPIT$GM
To conduct a genomic prediction:
myY<-read.csv("phenotype file location pathway.csv", sep = ",")
myGD=read.csv("numerical genotype file location pathway.csv", sep = ",")
myGM=read.csv("markers file location pathway.csv", sep = ",")
set.seed(99163)
GAPIT.Validation(
    Y=myY[,1:2],
    model=c("gBLUP"),
    GD=myGD,
    GM=myGM,
    PCA.total=3,
    file.output=T,
    nfold=5
```

The GS/GP of the inference groups (based on the ties with corresponding groups in the reference panel) was derived from Henderson's formula as follows:

$$u_I = K_{IR} K_{RR}^{-1} u_R,$$

where $K_{RR}$ is the variance-covariance matrix for all groups in the reference panel, $K_{RI}$ is the covariance matrix between the groups in the reference and inference panels, $K_{IR}$ is the covariance matrix between the groups inference and reference panels, $u_R$ is the predicted genomic values of the individuals in the inference group. To assess the reliability of the genomic prediction, the following formula is used:

$$\text{Reliability} = 1 - \text{PEV}/\sigma^2_a,$$

where PEV is the prediction error variance, representing the diagonal element in the inverse left-hand side of the mixed model equation, and $\sigma^2_a$ is the genetic variance.

*Genome-Wide Association Study (GWAS) Analysis*

To assess the association between potential genetic loci and the traits of interest at the whole genome level, we performed a Genome-Association Study (GWAS) employing the Genome Association and Prediction Integrated Tool (GAPIT) version 3 [99] with multiple models with enhanced power and accuracy for genome association. The GAPIT models used in this study include the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) [100], the Fixed and random model Circulating Probability Uniform (FarmCPU) [101], Settlement of MLM Under Progressively Exclusive Relationship (SUPER), and the General Linear Model (GLM) [102]. FarmCPU and SUPER supports genomic selection, while BLINK and GLM are commonly used for breeding through marker-assisted selection (MAS).

To perform a GWAS analysis, the below R script was used, after setting the results directory (*setwd()*), installing (install.packages ("package name") and launching all necessary packages and their libraries (package name)), installing the GAPIT source code, importing the genotype (*geno.raw <- fread("file:///D:/... .csv or .txt*) and phenotype (*myY <- fread("file:///D:/... .csv or .txt*) files, and performing initial data quality control:

my_GAPIT <- GAPIT(Y=myY, G=myG, model=c("SUPER", "FarmCPU", "BLINK"), PCA.total=3, SNP.MAF = 0.05, Multiple_analysis=TRUE)

*In Silico Analysis for Gene Ontology Search*

GWAS results provided useful information on novel genetic loci for grain size and shape in rice. The physical positions of associated significant SNP Chip markers were utilized to uncover the identity of genes harbored by the target genetic loci for more insights. To achieve that, we conducted

a search using the browser of the Rice Genome Annotation Project database (http://rice.uga.edu/cgi-bin/gbrowse/rice/#search, accessed on September 7, 2023) and PlantPAN 3.0 (http://plantpan.itps.ncku.edu.tw/plantpan3/search.php?#results, accessed on September 7, 2023) for each specific gene locus ID. Genes encoding similar domain-containing proteins were searched in the literature (https://funricegenes.github.io/geneKeyword.table.txt, accessed on September 7, 2023).

To assess the degree of sequence similarity of genes found in major QTLs for grain traits, the coding sequence (CDS) of each genes in the *japonica* group were aligned with that of their orthologues genes in the *indica* group. The respective CDS of target gene locus IDs (LOC_Osxxgxxxxx: Nipponbare database (http://rice.uga.edu/analyses_search_locus.shtml, accessed on September 11, 2023), and BGIOSGAxxxxxx: *indica* database (https://plants.ensembl.org/Oryza_indica/Info/Index, accessed on September 11, 2023) were obtained, and aligned using the ClustalW multiple alignment feature in Bioedit sequence Alignment Editor Software (Copyright © 1997-2013 Tom Hall) [103].

## 5. Conclusions

Rice grain-related traits are controlled by multiple genetic loci in plants. Grain length, width, and thickness determine the thousand-grain weight, thus influencing rice yield. A total of 43 QTLs associated with grain size, shape, or weight in rice, distributed across almost all rice chromosomes. GWAS results show seven SNP Chip makers (co-detected by both BLINK and FarmCPU) with strong association with grain length on Chr1–3, 5, 7, and 11, with $qGL1\text{-}1^{BFSG}$ explaining 65.2%–72.5% of the observed phenotypic variance for grain length. In addition, one ($qGW1\text{-}1^{BFSG}$) out of fourteen QTLs for grain width was co-detected by all four GAPIT models on Chr1. The $qGW1\text{-}1^{BFSG}$ explains 15.5%–18.9% of PVE. Likewise, either BLINK or FarmCPU identified three QTLs for grain thickness. Two of them explain 74.9% ($qGT1^{Blink}$) and 54.9% ($qGT2\text{-}1^{Farm}$) of the observed PVE. Regarding length-to-width ratio, the $qLWR2^{BFSG}$, detected by all GAPIT models, explains about 15.2%–32%) for LWR. As for thousand-grain weight, the $qTGW6^{BFSG}$ QTL coincided with $qGW6\text{-}1^{Blink}$ for grain width and explained 32.8%–54% of PVE. Putative candidate genes pooled from co-detected regions by all four GAPIT models have interesting annotated functions, and either associated with flower development, reproduction, post-embryonic development, carbohydrate metabolisms, or transcription regulation. Downstream functional studies, through the use of genetic engineering approaches or mutagenesis, would help elucidate the molecular functions of the candidate genes. The major QTLs for each grain trait can serve for downstream marker-assisted selection based on genome selection results.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Author Contributions:** "Conceptualization, J-H.L and N.R.K.; methodology, J-H.L. and N.R.K.; software, N.R.K.; formal analysis and investigation, N.R.K., G.D.D., and S.-B.L.; resources, D.-S.P. and S.-B.L.; data curation and visualization, N.R.K.; writing—original draft preparation, N.R.K and G.D.D.; writing—review and editing, J.-H.L., D.-S.P., J.-W.K., S.-G.J., and K.-W.O.; supervision, project administration, and funding acquisition, J.-H.L. and K.-W.O. All authors have read and agreed to the published version of the manuscript.".

**Funding:** This work was supported by the Young Scientist Research Program (Project No. RS-2022-RD010353) of the Korea-Africa Food and Agriculture Cooperation Initiative and the Rural Development Administration (KAFACI/RDA), Republic of Korea.

**Data Availability Statement:** Not Applicable.

**Acknowledgments:** Not Applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Campbell, M.T.; Bandillo, N.; Al Shiblawi, F.R.A.; Sharma, S.; Liu, K.; Du, Q.; Schmitz, A.J.; Zhang, C.; Véry, A.-A.; Lorenz, A.J.J.P.g. Allelic variants of OsHKT1; 1 underlie the divergence between indica and japonica subspecies of rice (Oryza sativa) for root sodium content. **2017**, *13*, e1006823.

2. Kang, J.-W.; Kabange, N.R.; Phyo, Z.; Park, S.-Y.; Lee, S.-M.; Lee, J.-Y.; Shin, D.; Cho, J.H.; Park, D.-S.; Ko, J.-M.J.A. Combined linkage mapping and genome-wide association study identified QTLs associated with grain shape and weight in rice (Oryza sativa L.). **2020**, *10*, 1532.
3. Jasim Aljumaili, S.; Rafii, M.; Latif, M.; Sakimin, S.Z.; Arolu, I.W.; Miah, G.J.B.r.i. Genetic diversity of aromatic rice germplasm revealed by SSR markers. **2018**, *2018*.
4. Adjah, K.L.; Asante, M.D.; Toure, A.; Aziadekey, M.; Amoako-Andoh, F.O.; Frei, M.; Diallo, Y.; Agboka, K.J.R.S. Improvement of rice production under drought conditions in West Africa: application of QTLs in breeding for drought resistance. **2022**, *29*, 512-521.
5. Desa, U.J.U.D.; economic, U.N.D.o.; org, s.a.u. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. **2017**.
6. Bazrkar-Khatibani, L.; Fakheri, B.-A.; Hosseini-Chaleshtori, M.; Mahender, A.; Mahdinejad, N.; Ali, J.J.I.J.o.G. Genetic mapping and validation of quantitative trait loci (QTL) for the grain appearance and quality traits in rice (Oryza sativa L.) by using recombinant inbred line (RIL) population. **2019**, *2019*.
7. Sharma, N.; Khanna, R.J.R.a.i.g.c.r. Rice grain quality: current developments and future prospects. **2019**, *5772*, 89367.
8. Feng, F.; Li, Y.; Qin, X.; Liao, Y.; Siddique, K.H.J.F.i.P.S. Changes in rice grain quality of indica and japonica type varieties released in China from 2000 to 2014. **2017**, *8*, 1863.
9. Yuan, L.; Denning, G.; Mew, T.J.D.G., Mew TW, editors. Hybrid rice breeding for super high yield. **1998**, 10-12.
10. Dixit, S.; Singh, A.; Kumar, A.J.I.J.o.A. Rice breeding for high grain yield under drought: a strategic solution to a complex problem. **2014**, *2014*, 1-15.
11. Goto, A.; Sasahara, H.; Shigemune, A.; Miura, K.J.J.A.R.Q.J. Hokuriku 193: A new high-yielding indica rice cultivar bred in Japan. **2009**, *43*, 13-18.
12. He, P.; Li, S.; Qian, Q.; Ma, Y.; Li, J.; Wang, W.; Chen, Y.; Zhu, L.J.T.; Genetics, A. Genetic analysis of rice grain quality. **1999**, *98*, 502-508.
13. Juliano, B.; Duff, B.J.r.g.m.; quality issues. Los Banos, L., IRRI. Rice grain quality as an emerging priority in National rice breeding programmes. **1991**, 55-64.
14. Hori, K.; Yamamoto, T.; Yano, M.J.B.s. Genetic dissection of agronomically important traits in closely related temperate japonica rice cultivars. **2017**, *67*, 427-434.
15. Li, N.; Xu, R.; Duan, P.; Li, Y.J.P.R. Control of grain size in rice. **2018**, *31*, 237-251.
16. Zhao, D.-S.; Li, Q.-F.; Zhang, C.-Q.; Zhang, C.; Yang, Q.-Q.; Pan, L.-X.; Ren, X.-Y.; Lu, J.; Gu, M.-H.; Liu, Q.-Q.J.N.C. GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. **2018**, *9*, 1240.
17. Mahender, A.; Anandan, A.; Pradhan, S.K.; Pandit, E.J.S. Rice grain nutritional traits and their enhancement using relevant genes and QTLs through advanced approaches. **2016**, *5*, 1-18.
18. Nagata, K.; Ando, T.; Nonoue, Y.; Mizubayashi, T.; Kitazawa, N.; Shomura, A.; Matsubara, K.; Ono, N.; Mizobuchi, R.; Shibaya, T.J.B.s. Advanced backcross QTL analysis reveals complicated genetic control of rice grain shape in a japonica× indica cross. **2015**, *65*, 308-318.
19. Yoshida, S.; Ikegami, M.; Kuze, J.; Sawada, K.; Hashimoto, Z.; Ishii, T.; Nakamura, C.; Kamijima, O.J.B.S. QTL analysis for plant and grain characters of sake-brewing rice using a doubled haploid population. **2002**, *52*, 309-317.
20. Yao, G.; Li, J.; Zhang, Q.; Hu, G.; Chen, C.; Tang, B.; Zhang, H.; Li, Z.J.A.A.S. Mapping QTLs for grain weight and shape using four sister near isogenic lines in rice (Oryza sativa L.). **2010**, *36*, 1310-1317.
21. Yong-Zhong, X.; Yi-Fang, T.; Cai-Guo, X.; Jin-Ping, H.; Xin-Li, S.J.J.o.I.P.B. Mapping quantitative trait loci for grain appearance traits of rice using a recombinant inbred line population. **2001**, *43*, 840.
22. Rabiei, B.; Valizadeh, M.; Ghareyazie, B.; Moghaddam, M.; Ali, A.J.E. Identification of QTLs for rice grain size and shape of Iranian cultivars using SSR markers. **2004**, *137*, 325-332.
23. Chen, J.; Zhou, H.; Gu, Y.; Xia, D.; Wu, B.; Gao, G.; Zhang, Q.; He, Y.J.M.B. Mapping and verification of grain shape QTLs based on high-throughput SNP markers in rice. **2019**, *39*, 1-10.
24. Jian-long, X.; Qing-zhong, X.; Li-jun, L.; Zhi-kang, L.J.C.J.o.R.S. Genetic dissection of grain weight and its related traits in rice (Oryza sativa L.). **2002**, *16*, 6.
25. Li, P.; Chen, Y.-H.; Lu, J.; Zhang, C.-Q.; Liu, Q.-Q.; Li, Q.-F.J.R. Genes and their molecular functions determining seed structure, components, and quality of rice. **2022**, *15*, 18.
26. Fan, C.; Xing, Y.; Mao, H.; Lu, T.; Han, B.; Xu, C.; Li, X.; Zhang, Q.J.T.; genetics, a. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. **2006**, *112*, 1164-1171.
27. Mao, H.; Sun, S.; Yao, J.; Wang, C.; Yu, S.; Xu, C.; Li, X.; Zhang, Q.J.P.o.t.N.A.o.S. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. **2010**, *107*, 19579-19584.
28. Song, X.-J.; Huang, W.; Shi, M.; Zhu, M.-Z.; Lin, H.-X.J.N.g. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. **2007**, *39*, 623-630.

29. Huang, K.; Wang, D.; Duan, P.; Zhang, B.; Xu, R.; Li, N.; Li, Y.J.T.P.J. WIDE AND THICK GRAIN 1, which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice. **2017**, *91*, 849-860.
30. Li, Y.; Fan, C.; Xing, Y.; Jiang, Y.; Luo, L.; Sun, L.; Shao, D.; Xu, C.; Li, X.; Xiao, J.J.N.g. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. **2011**, *43*, 1266-1269.
31. Duan, P.; Ni, S.; Wang, J.; Zhang, B.; Xu, R.; Wang, Y.; Chen, H.; Zhu, X.; Li, Y.J.N.p. Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. **2015**, *2*, 1-5.
32. Qi, P.; Lin, Y.-S.; Song, X.-J.; Shen, J.-B.; Huang, W.; Shan, J.-X.; Zhu, M.-Z.; Jiang, L.; Gao, J.-P.; Lin, H.-X.J.C.r. The novel quantitative trait locus GL3. 1 controls rice grain size and yield by regulating Cyclin-T1; 3. **2012**, *22*, 1666-1680.
33. Zhang, X.; Wang, J.; Huang, J.; Lan, H.; Wang, C.; Yin, C.; Wu, Y.; Tang, H.; Qian, Q.; Li, J.J.P.o.t.N.A.o.S. Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. **2012**, *109*, 21534-21539.
34. Weng, J.; Gu, S.; Wan, X.; Gao, H.; Guo, T.; Su, N.; Lei, C.; Zhang, X.; Cheng, Z.; Guo, X.J.C.r. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. **2008**, *18*, 1199-1209.
35. Liu, J.; Chen, J.; Zheng, X.; Wu, F.; Lin, Q.; Heng, Y.; Tian, P.; Cheng, Z.; Yu, X.; Zhou, K.J.N.P. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. **2017**, *3*, 1-7.
36. Hu, Z.; Lu, S.-J.; Wang, M.-J.; He, H.; Sun, L.; Wang, H.; Liu, X.-H.; Jiang, L.; Sun, J.-L.; Xin, X.J.M.P. A novel QTL qTGW3 encodes the GSK3/SHAGGY-like kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. **2018**, *11*, 736-749.
37. Ishimaru, K.; Hirotsu, N.; Madoka, Y.; Murakami, N.; Hara, N.; Onodera, H.; Kashiwagi, T.; Ujiie, K.; Shimizu, B.-i.; Onishi, A.J.N.g. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. **2013**, *45*, 707-711.
38. Shi, C.L.; Dong, N.Q.; Guo, T.; Ye, W.W.; Shan, J.X.; Lin, H.X.J.T.P.J. A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. **2020**, *103*, 1174-1188.
39. Wang, S.; Li, S.; Liu, Q.; Wu, K.; Zhang, J.; Wang, S.; Wang, Y.; Chen, X.; Zhang, Y.; Gao, C.J.N.g. The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. **2015**, *47*, 949-954.
40. Si, L.; Chen, J.; Huang, X.; Gong, H.; Luo, J.; Hou, Q.; Zhou, T.; Lu, T.; Zhu, J.; Shangguan, Y.J.N.g. OsSPL13 controls grain size in cultivated rice. **2016**, *48*, 447-456.
41. Wang, S.; Wu, K.; Yuan, Q.; Liu, X.; Liu, Z.; Lin, X.; Zeng, R.; Zhu, H.; Dong, G.; Qian, Q.J.N.g. Control of grain size, shape and quality by OsSPL16 in rice. **2012**, *44*, 950-954.
42. Qian, Q.J.R.G., Genetics; Breeding. Gene Network of Grain Size and Number in Rice. **2018**, 191-206.
43. Li, N.; Xu, R.; Li, Y.J.A.r.o.p.b. Molecular networks of seed size control in plants. **2019**, *70*, 435-463.
44. Zuo, J.; Li, J.J.A.r.o.g. Molecular genetic dissection of quantitative trait loci regulating rice grain size. **2014**, *48*, 99-118.
45. Xu, L.; Zhao, H.; Ruan, W.; Deng, M.; Wang, F.; Peng, J.; Luo, J.; Chen, Z.; Yi, K.J.T.P.C. ABNORMAL INFLORESCENCE MERISTEM1 functions in salicylic acid biosynthesis to maintain proper reactive oxygen species levels for root meristem activity in rice. **2017**, *29*, 560-574.
46. Li, H.; Liu, Y.; Qin, H.; Lin, X.; Tang, D.; Wu, Z.; Luo, W.; Shen, Y.; Dong, F.; Wang, Y.J.N.P. A rice chloroplast-localized ABC transporter ARG1 modulates cobalt and nickel homeostasis and contributes to photosynthetic capacity. **2020**, *228*, 163-178.
47. Chai, J.; Zhu, S.; Li, C.; Wang, C.; Cai, M.; Zheng, X.; Zhou, L.; Zhang, H.; Sheng, P.; Wu, M.J.P.B.J. OsRE1 interacts with OsRIP1 to regulate rice heading date by finely modulating Ehd1 expression. **2021**, *19*, 300-310.
48. Zhang, X.; Li, D.; Zhang, H.; Wang, X.; Zheng, Z.; Song, F.J.M.b.r. Molecular characterization of rice OsBIANK1, encoding a plasma membrane-anchored ankyrin repeat protein, and its inducible expression in defense responses. **2010**, *37*, 653-660.
49. Wang, Y.; Yang, S.; Li, C.; Hu, T.; Hou, S.; Bai, Q.; Ji, X.; Xu, F.; Guo, C.; Huang, M.J.F.i.P.S. The plasma membrane-localized OsNIP1; 2 mediates internal aluminum detoxification in rice. **2022**, *13*, 970270.
50. Yang, S.; Yuan, D.; Zhang, Y.; Sun, Q.; Xuan, Y.H.J.F.i.p.s. BZR1 regulates brassinosteroid-mediated activation of AMT1; 2 in rice. **2021**, *12*, 665883.
51. Sun, Z.; Liu, X.; Zhu, W.; Lin, H.; Chen, X.; Li, Y.; Ye, W.; Yin, Z.J.I.J.o.M.S. Molecular Traits and Functional Exploration of BES1 Gene Family in Plants. **2022**, *23*, 4242.
52. Sauter, M.; Rzewuski, G.; Marwedel, T.; Lorbiecke, R.J.J.o.E.B. The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. **2002**, *53*, 2325-2331.
53. Guo, M.; Ruan, W.; Li, C.; Huang, F.; Zeng, M.; Liu, Y.; Yu, Y.; Ding, X.; Wu, Y.; Wu, Z.J.P.P. Integrative comparison of the role of the PHOSPHATE RESPONSE1 subfamily in phosphate signaling and homeostasis in rice. **2015**, *168*, 1762-1776.

54. Shao, Y.; Zhou, H.-Z.; Wu, Y.; Zhang, H.; Lin, J.; Jiang, X.; He, Q.; Zhu, J.; Li, Y.; Yu, H.J.T.P.C. OsSPL3, an SBP-domain protein, regulates crown root development in rice. **2019**, *31*, 1257-1275.
55. Jiang, H.; Lu, Q.; Qiu, S.; Yu, H.; Wang, Z.; Yu, Z.; Lu, Y.; Wang, L.; Xia, F.; Wu, Y.J.P.o.t.N.A.o.S. Fujian cytoplasmic male sterility and the fertility restorer gene OsRf19 provide a promising breeding system for hybrid rice. **2022**, *119*, e2208759119.
56. Itabashi, E.; Iwata, N.; Fujii, S.; Kazama, T.; Toriyama, K.J.T.p.j. The fertility restorer gene, Rf2, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. **2011**, *65*, 359-367.
57. Agarwal, P.; Arora, R.; Ray, S.; Singh, A.K.; Singh, V.P.; Takatsuji, H.; Kapoor, S.; Tyagi, A.K.J.P.m.b. Genome-wide identification of C 2 H 2 zinc-finger gene family in rice and their phylogeny and expression analysis. **2007**, *65*, 467-485.
58. Ren, R.; Li, D.; Zhen, C.; Chen, D.; Chen, X.J.P. Specific roles of Os4BGlu10, Os6BGlu24, and Os9BGlu33 in seed germination, root elongation, and drought tolerance in rice. **2019**, *249*, 1851-1861.
59. Ilg, A.; Yu, Q.; Schaub, P.; Beyer, P.; Al-Babili, S.J.P. Overexpression of the rice carotenoid cleavage dioxygenase 1 gene in Golden Rice endosperm suggests apocarotenoids as substrates in planta. **2010**, *232*, 691-699.
60. Jing, P.; Zou, J.; Kong, L.; Hu, S.; Wang, B.; Yang, J.; Xie, G.J.P.S. OsCCD1, a novel small calcium-binding protein with one EF-hand motif, positively regulates osmotic and salt tolerance in rice. **2016**, *247*, 104-114.
61. Senthilnathan, S.J.A.a.S. Usefulness of correlation analysis. **2019**.
62. Li, R.; Li, M.; Ashraf, U.; Liu, S.; Zhang, J.J.F.i.p.s. Exploring the relationships between yield and yield-related traits for rice varieties released in China from 1978 to 2017. **2019**, *10*, 543.
63. Xie, Q.; Mayes, S.; Sparkes, D.L.J.J.o.e.b. Carpel size, grain filling, and morphology determine individual grain weight in wheat. **2015**, *66*, 6715-6730.
64. Huang, R.; Jiang, L.; Zheng, J.; Wang, T.; Wang, H.; Huang, Y.; Hong, Z.J.T.i.p.s. Genetic bases of rice grain shape: so many genes, so little known. **2013**, *18*, 218-226.
65. Yang, W.; Peng, S.; Dionisio-Sese, M.L.; Laza, R.C.; Visperas, R.M.J.F.C.R. Grain filling duration, a crucial determinant of genotypic variation of grain yield in field-grown tropical irrigated rice. **2008**, *105*, 221-227.
66. Du, Z.; Huang, Z.; Li, J.; Bao, J.; Tu, H.; Zeng, C.; Wu, Z.; Fu, H.; Xu, J.; Zhou, D.J.T.; et al. qTGW12 a, a naturally varying QTL, regulates grain weight in rice. **2021**, *134*, 2767-2776.
67. Ahmed, N.; Ahamed, K.U.; Rahman, M.S.J.A.-A.J.o.B.; Biotechnology. Yield and yield contributing attributes of rice (Oryza sativa L.) under different planting dates in boro season. **2018**, *3*, 162-169.
68. Bhat, J.A.; Ali, S.; Salgotra, R.K.; Mir, Z.A.; Dutta, S.; Jadon, V.; Tyagi, A.; Mushtaq, M.; Jain, N.; Singh, P.K.J.F.i.g. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. **2016**, *7*, 221.
69. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.J.g. Prediction of total genetic value using genome-wide dense marker maps. **2001**, *157*, 1819-1829.
70. Chen, J.; Liu, K.; Zha, W.; Zhou, L.; Li, M.; Xu, H.; Li, P.; Chen, Z.; Yang, G.; Chen, P.J.P.o. Identification and verification of grain shape QTLs by SNP array in rice. **2021**, *16*, e0260133.
71. Aslam, K.; Naveed, S.A.; Sabar, M.; Shabir, G.; Shah, S.M.; Khan, A.R.; Shah, M.M.; Fiaz, S.; Xu, J.; Arif, M.J.F.i.G. Identification of QTLs for rice grain size and weight by high-throughput SNP markers in the IR64 x Sadri population. **2022**, 1900.
72. Xiao, Y.; Chu, L.; Zhang, Y.; Bian, Y.; Xiao, J.; Xu, D.J.F.i.p.s. HY5: A pivotal regulator of light-dependent development in higher plants. **2022**, *12*, 800989.
73. Jonassen, E.M.; Sévin, D.C.; Lillo, C.J.J.o.p.p. The bZIP transcription factors HY5 and HYH are positive regulators of the main nitrate reductase gene in Arabidopsis leaves, NIA2, but negative regulators of the nitrate uptake gene NRT1. 1. **2009**, *166*, 2071-2076.
74. Huang, L.; Zhang, H.; Zhang, H.; Deng, X.W.; Wei, N.J.P.S. HY5 regulates nitrite reductase 1 (NIR1) and ammonium transporter1; 2 (AMT1; 2) in Arabidopsis seedlings. **2015**, *238*, 330-339.
75. Sakuraba, Y.; Yanagisawa, S. Light signalling-induced regulation of nutrient acquisition and utilisation in plants. In Proceedings of the Seminars in cell & developmental biology, 2018; pp. 123-132.
76. Yang, Y.; Liang, T.; Zhang, L.; Shao, K.; Gu, X.; Shang, R.; Shi, N.; Li, X.; Zhang, P.; Liu, H.J.N.P. UVR8 interacts with WRKY36 to regulate HY5 transcription and hypocotyl elongation in Arabidopsis. **2018**, *4*, 98-107.
77. Chen, X.; Yao, Q.; Gao, X.; Jiang, C.; Harberd, N.P.; Fu, X.J.C.b. Shoot-to-root mobile transcription factor HY5 coordinates plant carbon and nitrogen acquisition. **2016**, *26*, 640-646.
78. Zhang, X.; Huai, J.; Shang, F.; Xu, G.; Tang, W.; Jing, Y.; Lin, R.J.P.P. A PIF1/PIF3-HY5-BBX23 transcription factor cascade affects photomorphogenesis. **2017**, *174*, 2487-2500.
79. Bursch, K.; Toledo-Ortiz, G.; Pireyre, M.; Lohr, M.; Braatz, C.; Johansson, H.J.N.P. Identification of BBX proteins as rate-limiting cofactors of HY5. **2020**, *6*, 921-928.
80. Gangappa, S.N.; Crocco, C.D.; Johansson, H.; Datta, S.; Hettiarachchi, C.; Holm, M.; Botto, J.F.J.T.P.C. The Arabidopsis B-BOX protein BBX25 interacts with HY5, negatively regulating BBX22 expression to suppress seedling photomorphogenesis. **2013**, *25*, 1243-1257.

81. Heng, Y.; Lin, F.; Jiang, Y.; Ding, M.; Yan, T.; Lan, H.; Zhou, H.; Zhao, X.; Xu, D.; Deng, X.W.J.P.P. B-Box containing proteins BBX30 and BBX31, acting downstream of HY5, negatively regulate photomorphogenesis in Arabidopsis. **2019**, *180*, 497-508.
82. Li, C.; Qi, L.; Zhang, S.; Dong, X.; Jing, Y.; Cheng, J.; Feng, Z.; Peng, J.; Li, H.; Zhou, Y.J.T.P.C. Mutual upregulation of HY5 and TZP in mediating phytochrome A signaling. **2022**, *34*, 633-654.
83. Wang, D.; Guo, Y.; Wu, C.; Yang, G.; Li, Y.; Zheng, C.J.B.g. Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. **2008**, *9*, 1-20.
84. Han, G.; Qiao, Z.; Li, Y.; Wang, C.; Wang, B.J.I.j.o.m.s. The roles of CCCH zinc-finger proteins in plant abiotic stress tolerance. **2021**, *22*, 8327.
85. Chen, F.; Liu, H.-L.; Wang, K.; Gao, Y.-M.; Wu, M.; Xiang, Y.J.F.i.p.s. Identification of CCCH zinc finger proteins family in Moso bamboo (Phyllostachys edulis), and PeC3H74 confers drought tolerance to transgenic plants. **2020**, *11*, 579255.
86. Li, J.; Mahajan, A.; Tsai, M.-D.J.B. Ankyrin repeat: a unique motif mediating protein− protein interactions. **2006**, *45*, 15168-15178.
87. Gu, H.; Zhang, K.; Chen, J.; Gull, S.; Chen, C.; Hou, Y.; Li, X.; Miao, J.; Zhou, Y.; Liang, G.J.R. OsFTL4, an FT-like gene, regulates flowering time and drought tolerance in rice (Oryza sativa L.). **2022**, *15*, 47.
88. Bai, B.; Lu, N.; Li, Y.; Guo, S.; Yin, H.; He, Y.; Sun, W.; Li, W.; Xie, X.J.P.S. OsBBX14 promotes photomorphogenesis in rice by activating OsHY5L1 expression under blue light conditions. **2019**, *284*, 192-202.
89. Lei, L.; Cao, L.; Ding, G.; Zhou, J.; Luo, Y.; Bai, L.; Xia, T.; Chen, L.; Wang, J.; Liu, K.J.F.i.P.S. OsBBX11 on qSTS4 links to salt tolerance at the seeding stage in Oryza sativa L. ssp. Japonica. **2023**, *14*, 1139961.
90. Huang, J.; Zhao, X.; Weng, X.; Wang, L.; Xie, W.J.P.o. The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. **2012**, *7*, e48242.
91. He, Y.; Chen, S.; Liu, K.; Chen, Y.; Cheng, Y.; Zeng, P.; Zhu, P.; Xie, T.; Chen, S.; Zhang, H.J.P.B.J. OsHIPL1, a hedgehog-interacting protein-like 1 protein, increases seed vigour in rice. **2022**, *20*, 1346-1362.
92. Shi, H.; Li, X.; Lv, M.; Li, J.J.I.J.o.M.S. BES1/BZR1 family transcription factors regulate plant development via brassinosteroid-dependent and independent pathways. **2022**, *23*, 10149.
93. Tong, H.; Liu, L.; Jin, Y.; Du, L.; Yin, Y.; Qian, Q.; Zhu, L.; Chu, C.J.T.P.C. DWARF AND LOW-TILLERING acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate brassinosteroid responses in rice. **2012**, *24*, 2562-2577.
94. Verma, L.; Bhadouria, J.; Bhunia, R.K.; Singh, S.; Panchal, P.; Bhatia, C.; Eastmond, P.J.; Giri, J.J.J.o.E.B. Monogalactosyl diacylglycerol synthase 3 affects phosphate utilization and acquisition in rice. **2022**, *73*, 5033-5051.
95. Chen, K.; Łyskowski, A.; Jaremko, Ł.; Jaremko, M.J.F.i.P.S. Genetic and molecular factors determining grain weight in rice. **2021**, *12*, 605799.
96. Li, G.; Tang, J.; Zheng, J.; Chu, C.J.T.C.J. Exploration of rice yield potential: Decoding agronomic and physiological traits. **2021**, *9*, 577-589.
97. Zhang, Z.; Todhunter, R.; Buckler, E.; Van Vleck, L.D.J.J.o.a.s. Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. **2007**, *85*, 881-885.
98. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.J.N.g. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **2006**, *38*, 203-208.
99. Wang, J.; Zhang, Z.J.G., proteomics; bioinformatics. GAPIT version 3: boosting power and accuracy for genomic association and prediction. **2021**, *19*, 629-640.
100. Zhang, Z.; Liu, X.; Zhou, Y.; Summers, R.J.G. BLINK: a package for the next level of genome-wide association studies with both individuals and markers Meng Huang. **2018**, *8*, 1-12.
101. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z.J.P.g. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. **2016**, *12*, e1005767.
102. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D.J.N.g. Principal components analysis corrects for stratification in genome-wide association studies. **2006**, *38*, 904-909.
103. Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Proceedings of the Nucleic acids symposium series, 1999; pp. 95-98.