Superior haplotypes towards development of low glycemic index rice with preferred grain and cooking quality

Ramchander Selvaraj1,4, Arun Kumar Singh1,4, Vikas Kumar Singh1,4, Ragavendran Abbai3, Sonali Vijay Habde2, Uma Maheshwar Singh2 & Arvind Kumar1,2*

Increasing trends in the occurrence of diabetes underline the need to develop low glycemic index (GI) rice with preferred grain quality. In the current study, a diverse set of 3 K sub-panel of rice consisting of 150 accessions was evaluated for resistant starch and predicted glycemic index, including nine other quality traits under transplanted situation. Significant variations were noticed among the accessions for the traits evaluated. Trait associations had shown that amylose content possess significant positive and negative association with resistant starch and predicted glycemic index. Genome-wide association studies with 500 K SNPs based on MLM model resulted in a total of 41 marker-trait associations (MTAs), which were further confirmed and validated with mrMLM multi-locus model. We have also determined the allelic effect of identified MTAs for 11 targeted traits and found favorable SNPs for 8 traits. A total of 11 genes were selected for haplo-pheno analysis to identify the superior haplotypes for the target traits where haplotypes ranges from 2 (Os10g0469000-GC) to 15 (Os06g18720-AC). Superior haplotypes for RS and PGI, the candidate gene Os06g11100 (H4-3.28% for high RS) and Os08g12590 (H13-62.52 as intermediate PGI). The identified superior donors possessing superior haplotype combinations may be utilized in Haplotype-based breeding to developing next-generation tailor-made high quality healthier rice varieties suiting consumer preference and market demand.

Abbreviations
AC  Amylose content
KL  Kernel length
KB  Kernel breadth
LBR  Length/ Breadth ratio
ASV  Alkali spreading value
KLAC  Kernel length after cooking
KBAC  Kernel breadth after cooking
LER  Linear elongation ratio
GC  Gel consistency
RS  Resistant starch
PGI  Predicted glycemic index
TPR  Transplanted rice
MTAs  Marker trait associations
GWAS  Genome wide association studies

Rice (*Oryza sativa* L.) is a major source for carbohydrate of more than 50% of the global population. In recent times, an increase in living standards has created awareness among the peoples all over the world for consumption of superior quality rice to improve their health status1. Thus, the presence of premium grain quality in rice

---

1IRRI South Asia Hub (IRRI-SAHI), ICRISAT Campus, Patancheru, Hyderabad, India. 2South-Asia Regional Centre (SARC), International Rice Research Institute (IRRI), Varanasi, India. 3Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. *These authors contributed equally: Ramchander Selvaraj, Arun Kumar Singh and Vikas Kumar Singh.*

*email: arvind.kumar@cgiar.org
Starch is made up of two components namely amylose and amylopectin. Amylose, an important parameter and soft to medium gel consistency are preferred by consumers in South Asia. Apart from quality, these days people have more concerns about their health, and prefer low glycaemic index rice varieties which could be safe for diabetics and obesity. Hence, grain quality improvement of rice together with the lower glycaemic index are of vital importance for the rice breeders in the current scenario of increasing diabetic population all over the world, especially in the Asia. Low GI diets effectively prevent type II diabetes and consumption of high amylose class rice varieties with soft-textured cooking might be an alternative for intermediate amylose rice varieties especially for those who are suffering from type II diabetes. The most critical factor responsible for breeding low glycaemic index (GI) rice is resistant starch (RS). It is the starch portion resistant to enzyme hydrolysis that escapes the small intestine and enters the large intestine where it gets fermented and slowly releases glucose to the bloodstream. Generally, it has been reported that increased level of RS content in rice grains lowers the GI value ensuring a negative correlation with amylose content. Recently low-to-intermediate RS phenotypic variations in rice panel were identified, resulting novel RS associations to numerous genes associated with amylopectin biosynthesis and degradation.

In any crop species, genetic diversity plays important role in breeding elite varieties. Identification of favorable alleles and its superior haplotype of the various genes associated with traits linking to cooking and eating quality of rice are pre-requisite of breeding to develop healthier rice. Nowadays, with the improvement of high-throughput sequencing technologies with reduced cost makes genome wide association studies (GWAS) as one of the prominent techniques to identify marker-trait associations (MTAs). Research on the glycaemic index identified a novel association of candidate loci Os05g03600 reporting intermediate to high GI variations. Another hotspot on chromosome 6 was found to include GBSSI, hydrolase genes and genes involved in signaling and chromatin modification with differential methylation patterns in GI region. Alternative splicing of GBSSI promoter region resulted in intermediate to high GI variations. Novel SNP associations on chromosomes namely 2, 5, 6 and 11 has been reported and these SNPs influence the final viscosity variations but no significant association with GI. Besides, the predicted glycaemic index (PGI) is estimated by in-vitro enzymatic action of starch digestion leads to hydrolysis and the glycaemic index (GI) estimated in-vivo, which requires human clinical evaluation of the two to three hours of blood glucose response after food intake. Thus, it is more time consuming and resource demanding. The in-vitro digestion methods have been developed to measure the starch hydrolysis index and it is used to calculate PGI using the formula developed. Significant positive association (r = 0.946) between pGI and GI of rice samples by using bread as the reference.

Recently, 11 candidate loci controlling grain quality traits has been identified that are involved in the starch biosynthesis. Nine MTAs were identified for seven QTLs namely GS3, TUD, qGR1.1, qGOWC7, qG3L3.4, qGW1.1 and qGW10.2 controlling quality traits. Numerous genes/QTLs were identified for grain number and grain length viz., GW2, GFI1, qSW5, GS3, G5S, qGL3, GW8, G6S, GS2, GL7/GW7, OsMA PK6, GLW7 and GAD114–26. QTLs controlling grain size have also been detected via genetic mapping and association studies24,27–32. Starch is made up of two components namely amylose and amylopectin. Amylose, an important parameter affecting cooking and eating quality of rice directly involved in changing the grain texture by absorbing water on cooking. However, the difference in the amylopectin structure of similar AC rice explained variations in the gelatinization temperature (GT) or alkali spreading value (ASV). Generally, rice with intermediate amylose, GT and soft to medium gel consistency are preferred by consumers in South Asia. Apart from quality, these days people have more concerns about their health, and prefer low glycaemic index rice varieties which could be safe for diabetics and obesity. Hence, grain quality improvement of rice together with the lower glycaemic index are of vital importance for the rice breeders in the current scenario of increasing diabetic population all over the world, especially in the Asia. Low GI diets effectively prevent type II diabetes and consumption of high amylose class rice varieties with soft-textured cooking might be an alternative for intermediate amylose rice varieties especially for those who are suffering from type II diabetes. The most critical factor responsible for breeding low glycaemic index (GI) rice is resistant starch (RS). It is the starch portion resistant to enzyme hydrolysis that escapes the small intestine and enters the large intestine where it gets fermented and slowly releases glucose to the bloodstream. Generally, it has been reported that increased level of RS content in rice grains lowers the GI value ensuring a negative correlation with amylose content. Recently low-to-intermediate RS phenotypic variations in rice panel were identified, resulting novel RS associations to numerous genes associated with amylopectin biosynthesis and degradation.

In any crop species, genetic diversity plays important role in breeding elite varieties. Identification of favorable alleles and its superior haplotype of the various genes associated with traits linking to cooking and eating quality of rice are pre-requisite of breeding to develop healthier rice. Nowadays, with the improvement of high-throughput sequencing technologies with reduced cost makes genome wide association studies (GWAS) as one of the prominent techniques to identify marker-trait associations (MTAs). Research on the glycaemic index identified a novel association of candidate loci Os05g03600 reporting intermediate to high GI variations. Another hotspot on chromosome 6 was found to include GBSSI, hydrolase genes and genes involved in signaling and chromatin modification with differential methylation patterns in GI region. Alternative splicing of GBSSI promoter region resulted in intermediate to high GI variations. Novel SNP associations on chromosomes namely 2, 5, 6 and 11 has been reported and these SNPs influence the final viscosity variations but no significant association with GI. Besides, the predicted glycaemic index (PGI) is estimated by in-vitro enzymatic action of starch digestion leads to hydrolysis and the glycaemic index (GI) estimated in-vivo, which requires human clinical evaluation of the two to three hours of blood glucose response after food intake. Thus, it is more time consuming and resource demanding. The in-vitro digestion methods have been developed to measure the starch hydrolysis index and it is used to calculate PGI using the formula developed. Significant positive association (r = 0.946) between pGI and GI of rice samples by using bread as the reference.

Recently, 11 candidate loci controlling grain quality traits has been identified that are involved in the starch biosynthesis. Nine MTAs were identified for seven QTLs namely GS3, TUD, qGR1.1, qGOWC7, qG3L3.4, qGW1.1 and qGW10.2 controlling quality traits. Numerous genes/QTLs were identified for grain number and grain length viz., GW2, GFI1, qSW5, GS3, G5S, qGL3, GW8, G6S, GS2, GL7/GW7, OsMA PK6, GLW7 and GAD114–26. QTLs controlling grain size have also been detected via genetic mapping and association studies24,27–32. Starch is made up of two components namely amylose and amylopectin. Amylose, an important parameter affecting cooking and eating quality of rice directly involved in changing the grain texture by absorbing water on cooking. However, the difference in the amylopectin structure of similar AC rice explained variations in the gelatinization temperature (GT) or alkali spreading value (ASV). Generally, rice with intermediate amylose, GT and soft to medium gel consistency are preferred by consumers in South Asia. Apart from quality, these days people have more concerns about their health, and prefer low glycaemic index rice varieties which could be safe for diabetics and obesity. Hence, grain quality improvement of rice together with the lower glycaemic index are of vital importance for the rice breeders in the current scenario of increasing diabetic population all over the world, especially in the Asia. Low GI diets effectively prevent type II diabetes and consumption of high amylose class rice varieties with soft-textured cooking might be an alternative for intermediate amylose rice varieties especially for those who are suffering from type II diabetes. The most critical factor responsible for breeding low glycaemic index (GI) rice is resistant starch (RS). It is the starch portion resistant to enzyme hydrolysis that escapes the small intestine and enters the large intestine where it gets fermented and slowly releases glucose to the bloodstream. Generally, it has been reported that increased level of RS content in rice grains lowers the GI value ensuring a negative correlation with amylose content. Recently low-to-intermediate RS phenotypic variations in rice panel were identified, resulting novel RS associations to numerous genes associated with amylopectin biosynthesis and degradation.

In any crop species, genetic diversity plays important role in breeding elite varieties. Identification of favorable alleles and its superior haplotype of the various genes associated with traits linking to cooking and eating quality of rice are pre-requisite of breeding to develop healthier rice. Nowadays, with the improvement of high-throughput sequencing technologies with reduced cost makes genome wide association studies (GWAS) as one of the prominent techniques to identify marker-trait associations (MTAs). Research on the glycaemic index identified a novel association of candidate loci Os05g03600 reporting intermediate to high GI variations. Another hotspot on chromosome 6 was found to include GBSSI, hydrolase genes and genes involved in signaling and chromatin modification with differential methylation patterns in GI region. Alternative splicing of GBSSI promoter region resulted in intermediate to high GI variations. Novel SNP associations on chromosomes namely 2, 5, 6 and 11 has been reported and these SNPs influence the final viscosity variations but no significant association with GI. Besides, the predicted glycaemic index (PGI) is estimated by in-vitro enzymatic action of starch digestion leads to hydrolysis and the glycaemic index (GI) estimated in-vivo, which requires human clinical evaluation of the two to three hours of blood glucose response after food intake. Thus, it is more time consuming and resource demanding. The in-vitro digestion methods have been developed to measure the starch hydrolysis index and it is used to calculate PGI using the formula developed. Significant positive association (r = 0.946) between pGI and GI of rice samples by using bread as the reference.

Recently, 11 candidate loci controlling grain quality traits has been identified that are involved in the starch biosynthesis. Nine MTAs were identified for seven QTLs namely GS3, TUD, qGR1.1, qGOWC7, qG3L3.4, qGW1.1 and qGW10.2 controlling quality traits. Numerous genes/QTLs were identified for grain number and grain length viz., GW2, GFI1, qSW5, GS3, G5S, qGL3, GW8, G6S, GS2, GL7/GW7, OsMA PK6, GLW7 and GAD114–26. QTLs controlling grain size have also been detected via genetic mapping and association studies24,27–32. Starch is made up of two components namely amylose and amylopectin. Amylose, an important parameter affecting cooking and eating quality of rice directly involved in changing the grain texture by absorbing water on cooking. However, the difference in the amylopectin structure of similar AC rice explained variations in the gelatinization temperature (GT) or alkali spreading value (ASV). Generally, rice with intermediate amylose, GT and soft to medium gel consistency are preferred by consumers in South Asia. Apart from quality, these days people have more concerns about their health, and prefer low glycaemic index rice varieties which could be safe for diabetics and obesity. Hence, grain quality improvement of rice together with the lower glycaemic index are of vital importance for the rice breeders in the current scenario of increasing diabetic population all over the world, especially in the Asia. Low GI diets effectively prevent type II diabetes and consumption of high amylose class rice varieties with soft-textured cooking might be an alternative for intermediate amylose rice varieties especially for those who are suffering from type II diabetes. The most critical factor responsible for breeding low glycaemic index (GI) rice is resistant starch (RS). It is the starch portion resistant to enzyme hydrolysis that escapes the small intestine and enters the large intestine where it gets fermented and slowly releases glucose to the bloodstream. Generally, it has been reported that increased level of RS content in rice grains lowers the GI value ensuring a negative correlation with amylose content. Recently low-to-intermediate RS phenotypic variations in rice panel were identified, resulting novel RS associations to numerous genes associated with amylopectin biosynthesis and degradation.
from 0.57 to 10.00%, whereas, the PGI ranged from 52.91 to 99.94. Significant variation was also observed for the other quality traits like AC, LBR, LER, ASV and GC that play an important role in the selection of desirable quality of rice varieties in the view of consumer preference. In our study, AC ranged from 12 to 33%, KL in the range of 4.60 to 7.7 mm, KB ranged from 1.80 to 3.00 mm, LBR in the range of 1.58 to 3.60, ASV from 1 to 7 score, LER ranged between 1.02 to 1.96 and GC in the range of 14.50 to 100 mm.

**Correlation among grain quality traits.** Association among the traits was observed. RS and PGI showed significant negative correlations. The important quality trait AC showed a significant positive ($r = 0.19^*$, $p$-value=0.048) and negative correlation ($r = -0.16^*$, $p$-value=0.049) with RS and PGI, respectively whereas; it showed a negative significant relationship with ASV and GC. KL had a significant positive correlation with LBR and KLAC whereas it was negatively correlated with KB (Fig. 2).
Genome-wide association study (GWAS) of targeted traits. Single locus GWAS (MLM) was performed using the phenotypic data with approximately more than 0.5 million (500 K) SNPs of concerned accessions and the peak associations were predicted based on considering the value of PVE (phenotypic variance explained) with suggestive significant $p$-value < 0.00001 ($-\log_{10}(p) > 5$). In general, modification of bonferroni correction significant threshold value to control the false positive rate in a single-locus GWAS are so conservative that some associated SNPs may be eliminated. To avoid this problem multi-locus GWAS of mrMLM was also conducted to confirm the MTAs with the LOD score of greater than 3. A set of 41 causatives MTAs were detected for all the investigated quality traits in MLM (Table 1). In MLM, single MTA was identified for KL, KBAC, L/B ratio and LER, and up to nine MTAs for AC, RS and PGI. We identified nine MTAs associated with RS located on chromosomes 1, 2, 6, 8 and 11 with PVE ranging between 15.92 to 19.31% (Fig. 3). For PGI, nine MTAs were identified on chromosomes 1, 3, 7, 8, 9 and 11 with PVE ranging between 23.47 to 17.77% (Fig. 3). For AC, nine MTAs were detected on chromosomes 2, 6, 8, 10 and 11, respectively with the PVE ranging between 14.38 to 17.96% (Fig. S1). For KL, one MTA was identified on chromosome 11 with the PVE of 11.87% (Fig. S2). For KB, 4 MTAs were identified on chromosome 2, 3 and 5 with a PVE of 14.48, 14.79, 15.35 and 21.48%, respectively (Fig. S3). LBR recorded an MTA on chromosome 5 with the PVE of 16.99% (Fig. S4). For ASV, an indicator for GT, one peak was detected on chromosome 6 with a PVE of 25.53% (Fig. S5). Three MTAs on chromosome 1, 5 and 12 were detected for KLAC with PVE of 23.55, 24.17 and 18.05%, respectively (Fig. S6). For KBAC only one association was detected on chromosome 8 with the PVE of 18.35 (Fig. S7). For LER, one MTA was identified on chromosome 9 explaining PVE of 16.69%. Two MTAs for GC on chromosome 10 and 7 with the PVE of 18.39 and 18.11% were detected in TPR (Fig. S8).

To confirm the real associations, multi-locus GWAS (mrMLM) was also conducted with three models mrMLM, FastmrMLM and FASTmrEMMA by adopting the critical threshold of significance for SNP-trait association was set at a LOD score of greater than 3. On comparing both MLM and mrMLM, we found that all the MTAs identified by MLM were also detected by any one method of mrMLM namely mrMLM, FastmrMLM and FASTmrEMMA with significant LOD score (Table 1, Table S2a). In mrMLM GWAS, S6_6896749 and S11_28809145 exhibited with AC with a high PVE range of 11.52 to 21.83 and 9.26 to 11.50%, S5_5361877 for KB and LBR with PVE of 1.01 to 5.29% and 1.01 to 12.26%, S6_6711302 for ASV with the PVE of 10.55 to 21.94%, S11_28809145 for KLAC with 16.62–17.19% of PVE, S9_17873147 for LER with 1.00–16.07% PVE, S7_24356119 for GC with 37.55% PVE and S11_5737145 for PGI with PVE of 5.71 to 23.55%. As per the model of mrMLM GWAS, the common MTAs identified shown higher values of PVE in the range of 5.29% (S5_5361877
| Trait | SNP ID | Gene ID (RAP) | Gene function |
|-------|--------|--------------|---------------|
| AC    | S8_10622940 | Os08g0276000 | transmembrane 9 superfamily member, putative, expressed |
|       | S6_6896749  | Os06g0232700 | transposon protein, putative, unclassified, expressed |
|       | S8_10166593 | Os08g0266700 | Rad21 / Rec8 like protein, putative, expressed |
|       | S11_28809145 | Os11g0704000  | selT/selW/selH seloprotein domain containing protein, expressed |
|       | S2_1904604  | Os02g0135800  | WD domain, G-beta repeat domain containing protein, expressed |
|       | S10_5085990 | Os08g09430   | retrotransposon protein, putative, Ty3-gypsy subclass, expressed |
|       | S6_10612482 | Os06g18770   | retrotransposon protein, putative, unclassified |
|       | S12_1286113 | Os08g0299700 | OsFBO20—F-box and other domain containing protein, expressed |
|       | S2_2474176  | Os02g05180   | retrotransposon protein, putative, Ty3-gypsy subclass, expressed |
|       | S11_28909306 | Os11g0706100 | alpha-hemolysin, putative, expressed |
|       |          | Os11g0706500 | expressed protein |
| KB    | S5_5361877 | Os05g09510   | hypothetical protein |
|       | S3_13543674 | Os05g0187500 | IQ calmodulin-binding motif family member, expressed |
|       | S3_18993628 | Os03g35870   | retrotransposon protein, putative, Ty3-gypsy subclass, expressed |
|       | S13_1386763 | Os02g0326500 | expressed protein |
|       | S5_5361877 | Os05g09510   | hypothetical protein |
|       | S6_6711302  | Os06g0229000 | OsFtsH6 FtsH protease, homologue of AtFtsH6, expressed |
|       |          | Os06g12380   | expressed protein |
| LBR   | S5_28169331 | Os05g49110   | retrotransposon protein, putative, Ty1-copia subclass, expressed |
|       | S1_9126931 | Os05g056200  | IQ calmodulin-binding motif family protein, expressed |
|       | S12_17888167 | Os12g29920 | retrotransposon protein, putative, Ty3-gypsy subclass, expressed |
|       | S6_5819355  | Os12g484375  | conserved hypothetical protein |
| ASV   | S8_681     | Os08g01010   | hypothetical protein |
|       | S9_17873147 | Os09g0469400 | glycosylase operon protein gGx, putative, expressed |
|       |          | Os09g0469900 | queuine tRNA-ribosyltransferase, putative, expressed |
| KLAC  | S10_17333120 | Os10g0469000 | leucine-rich repeat receptor protein kinase EXS precursor, putative, expressed |
|       | S7_24356119 | Os07g0597000 | expressed protein |
|       | S2_2299187 | Os02g0594100 | protein kinase domain containing protein, expressed |
|       | S10_1500020 | Os08g0335500 | protein kinase, putative, expressed |
|       | S10_4533502 | Os10g08370   | transposon protein, putative, CACTA, En/Spm sub-class, expressed |
|       | S2_5060167 | Os02g0191000 | disease resistance protein RPM1, putative, expressed |
|       | S1_8485858 | Os01g15720   | tobamovirus movement protein containing protein, expressed |
|       | S6_5819355 | Os06g11100   | retrotransposon protein, putative, unclassified, expressed |
|       | S6_9155098 | Os06g16080   | transposon protein, putative, CACTA, En/Spm sub-class, expressed |
|       | S11_11627020 | Os11g20120 | retrotransposon protein, putative, unclassified, expressed |
|       |          | Os11g20130   | retrotransposon protein, putative, unclassified, expressed |
|       | S2_9683531 | Os02g16970   | retrotransposon protein, putative, unclassified, expressed |
|       | S11_5737145 | Os11g10500   | retrotransposon protein, putative, Ty1-copia subclass, expressed |
|       | S11_12661155 | Os11g22020 | retrotransposon protein, putative, unclassified, expressed |
|       | S1_20415780 | Os11g22030   | retrotransposon protein, putative, unclassified, expressed |
|       | S9_17671895 | Os09g29090   | retrotransposon protein, putative, unclassified, expressed |
|       | S7_28567910 | Os07g0675200 | expressed protein |
|       | S8_25864590 | Os07g0673000 | expressed protein |
|       | S9_7528951  | Os08g0250300 | RNA recognition motif containing protein, putative, expressed |
|       | S3_21132448 | Os08g0520400 | spermidine synthase-related, putative, expressed |
|       | S8_7447826  | Os09g12580   | retrotransposon protein, putative, unclassified, expressed |
|       |          | Os09g12590   | retrotransposon, putative, centromere-specific |
Table 1. GWAS for grain quality traits, RS content and predicted glycemic index by single locus (MLM) and multi-locus model (mrMLM). AC—Amylose content, KL—Kernel length, KB—Kernel breadth, LBR—Length/Breadth ratio, ASV—Alkali spreading value, KLAC—Kernel length after cooking, KBAC—Kernel breadth after cooking, LER—Linear elongation ratio, GC—Gel consistency, RS—Resistant starch, PGI—Predicted glycemic index, TPR—Transplanted rice and significant associations were detected with the P value < 0.00001 in MLM and LOD > 3 in mrMLM. Bolded MTAs shown strong association confirmed by MLM and mrMLM with high value of PVE.

Detection of favorable SNP alleles associated with a trait of interest. Associated MTAs of RS, PGI and nine quality traits were subjected for the identification of favorable SNP alleles (Table S2b). In this investigation, the positive effect of candidate SNP alleles that led to increase in AC, KL, L/B ratio, ASV, KLAC, LER, GC, and RS or decrease in KB, KBAC and PGI were defined as “favorable alleles”, and those that resulted in decrease of AC, KL, L/B ratio, ASV, KLAC, LER, GC, and RS or an increase in KB, KBAC and PGI were defined as “unfavorable alleles”. As a result, one MTA for each trait like AC, ASV, GC and RS had an increased phenotypic effect whereas KB, KBAC and PGI revealed decreased effect in their phenotype are designated as favorable alleles. S6_6896749 possessing 'G' allele had strong phenotypic effect (0.88%) on AC, S5_5361877 possessing 'A' allele with 0.18 of L/B ratio, S6_6711302 with 'T' allele had 1.96 phenotypic effects on ASV, S6_5819355 with 'A' allele had 1.74% increase in phenotypic effect in RS and S8_7447826 with 'A' allele had resulted in decrease of −2.91 in PGI (Table S2b). Findings of this study indicated that the favorable SNP significantly increased or decreased the phenotypic effect of the trait in the genotypes they were present. Identification of superior haplotypes of these mined favorable alleles would be beneficial to develop elite entries with preferred grain quality and lower GI value.

Haplotype analysis of identified candidate loci. All the MTAs associated with the quality-related traits were subjected to the identification of candidate loci using RAP database (Table S2b). Out of the 41 MTAs detected in this investigation, we found 27 MTAs were within the candidate loci and 14 MTAs were found to be flanked by two candidate loci. Of these, nine loci for PGI, nine loci for AC, nine loci for RS, four loci each for KB, KLAC, two loci for GC, one loci each for KL, LBR, ASV, LER and KBAC were identified in this study. The identified candidate loci were subjected to haplotype analysis by rice SNP seek database to estimate the number of haplotypes present in the sub-set of 3 K-RGP for all the investigated quality-related traits and to use superior haplotypes to breed quality rice varieties. Haplotype analysis reported a minimum of one haplotype to the maximum of 15 haplotypes for the identified loci. The loci Os08g0276000, Os02g0135800 and Os10g09430

Figure 3. Genome wide association studies—Manhattan plot showing peak associations for resistant starch and predicted glycemic index in (a) Peak association for resistant starch on chromosome 2, 6 and 8 in transplanted rice, (b) Peak association for PGI on chromosome 1, 8 and 11 in transplanted rice.
for AC, Os05g09510 for KB and LBR, Os05g0566200 and Os12g0484375 for KLAC, Os09g0469400 for LER, Os02g0594100, Os02g0191000, Os11g20120 and Os11g20130 for RS and Os08g0520400 for PGI had registered one haplotype, respectively. The loci namely, Os06g18710 and Os06g18720 for AC, Os06g0187500 for KB and LBR, Os06g0187500 for ASV, Os08g0180100 for KBAC, Os06g16080 and Os02g16970 for RS and Os07g0675300 for PGI were recorded to be harbored 15 haplotypes explaining maximum diversity. On the other hand, 14 loci had 2 haplotypes, 7 loci had 3 haplotypes, 4 loci had 4 haplotypes, 3 loci had 5, 1 locus had 8, 10 and 14 haplotypes, respectively (Table S2b).

**Haplotype frequency of identified causative genes.** The major loci Os06g0232700 associated with AC, had five haplotypes where H5 representing the higher frequency of 57.33%. The other loci on chromosome 6, Os06g18710 representing maximum haplotype diversity of 15 haplotypes where H4 recorded a higher frequency of 52.66% (Table S3). The candidate SNP for KL centered between two loci namely Os11g0706100 and Os11g0706500 exhibited a higher frequency of H5 with 58.66% and H3 with 42.66%, respectively. S5_5361877 is the major allele for KB and LBR falls within candidate genes namely Os05g09510 and Os05g0187500, where Os05g0187500 possessing 15 haplotypes with the higher haplotype frequency of 48.66% in H14. The trait ASV and LER for AC, Os09g0469400 and Os11g0706100, respectively (Table 2). For grain size (LBR), H9-2.81 of mean score and H5 (4.48 of mean score) as anticipated one since intermediate ASV is desirable to breed useful varieties. For LER, Os12g0232700 and Os06g18710 (H4-26.50% for high amylose; H5-24.00% for intermediate amylose), Os06g18710 (H3-25.50%; H6-23.30% for intermediate amylose) and Os06g18720 (H4-26.58% for high amylose; H14-23.36% for intermediate amylose) had shown positive allelic effect with the superior haplotype of increasing in amylose content. For RS and PGI, the candidate loci Os06g0232700 (H4-3.32% for high RS; H2-2.42% for intermediate RS) and Os08g12590 (H13-62.52 as intermediate PGI) possesses positive and negative allelic effect of 1.74 and -2.91, respectively. For grain size (LBR), H9-2.81 of (Os05g0187500 had the positive allelic effect of 0.18. For ASV and GC, H1-3.43 and H1-56.06 as the desirable haplotype for the candidates of Os06g0229000 and Os11g046900 (Table 2).

For KL, Os11g0706100 had H8 (6.71 mm) and H7 (5.67 mm), Os11g0706500 had H2 (6.41 mm) as a desirable haplotype for development of long and medium slender grain rice varieties. Os05g0187500 and Os03g3587000 were associated with KB had H7 (2.37 mm) and H1 (2.42 mm) as superior haplotypes for the corresponding loci. For LBR, Os05g0187500 had H9 as desirable haplotype having genotypes recorded higher length-breadth ratio (2.89) possessing slender grain rice lines. Os06g0229000 and Os06g12380 associated with ASV showed H4 (3.67 of mean score) and H5 (4.48 of mean score) as anticipated one since intermediate ASV is desirable to breed good quality rice varieties. The other loci for RS Os06g16080 registered H11 (5.37%) for high and H7 (2.95%) as desirable haplotypes for intermediate RS content, respectively. Other loci, Os02g16970 linked with RS reported H14 (4.49%) as a superior haplotype for improving RS content (Table S4).

**Discussion**

Rice, one of the essential staple food crop species is consumed by half of the global population especially in Asian continents. The ever-increasing population of the world demands more than 50% additional rice production to be attained by 2030 to meet future needs. Dissecting the complexity of grain quality is possible by means of genome wide association mapping using a diverse set of population with a higher precision approach to map the causative alleles than the bi-parental mapping population since GW AS population consist higher evolutionary diversity.

The causative alleles than the bi-parental mapping population since GW AS population consist higher evolutionary diversity. The causative alleles than the bi-parental mapping population since GW AS population consist higher evolutionary diversity. Starchy rice grains and these are complex traits governed by multiple genes with various gene expression and regulatory network pathways and several QTLs associated with grain quality were identified and cloned in rice. Starch biosynthesis in rice grains during grain formation and development is one of the critical factors in improving the quality aspects of rice varieties and regulation of this complex metabolic pathway controlled by a network of various genes and gene combinations are poorly understood. Waxy gene (GRSII) plays a key role in amylopectin biosynthesis and the formation of long-chain amylopectin. Another gene Starch synthase IIa (SSIa) modifies amylopectin structure and improving grain quality. Likewise, several other genes namely SSI, SSIla, isoamylase1 (ISA1), pullulanase (PUL), branching enzyme (BEI and BEIIb) have also been reported to control starch structure formation and modifies the physicochemical properties of rice grains. Therefore, amylopectin content and structure alteration have a significant impression on the modification of starch granules and affect cooking properties in rice.

In recent times, non-infectious diseases cause several health-related issues like cardiovascular problems, obesity and diabetics due to the high consumption of calories and improper physical exercises. Recent study reported that high amylose starch-rich RS decreases the glycemic value of starch, insulin response in human metabolism and reduce the risk of type II diabetes mellitus and other health-related issues. Currently, breeders and geneticists are making efforts to develop high RS rice varieties. Cultivars rich in amylose content of rice, maize and barley developed by either mutational breeding or biotechnological approaches have been found to possess high RS content.
Table 2. Haplo-phenotype analysis of favourable candidate loci for grain quality traits, RS content and predicted glycemic index in rice. AC—Amylose content, KL—Kernel length, KB—Kernel breadth, LBR—Length/Breadth ratio, ASV—Alkali spreading value, KLAC—Kernel length after cooking, KBAC—Kernel breadth after cooking, LER—Linear elongation ratio, GC—Gel consistency, RS—Resistant starch, PGI—Predicted glycemic index, TPR—Transplanted rice.

| Trait  | Gene ID (RAP) | Allelic effect | Gene function | Average performance of superior haplotype | Average performance of other haplotype | Lines with superior haplotype |
|--------|---------------|----------------|---------------|------------------------------------------|----------------------------------------|-------------------------------|
|        |               |                | Haplotype     |                                          |                                        |                               |
| AC     | Os06g0232700  | 0.88           | transposon protein, putative, unclassified, expressed | H4-26.50                        | H1-24.23<sup>a</sup>, H2-24.44<sup>a</sup>, H3-24.03<sup>a</sup>, H5-24.00<sup>a</sup> | 4                             |
|        | Os06g18710    | 0.28           | retrotransposon protein, putative, unclassified | H3-25.50<sup>a</sup> | H1-24.50<sup>a</sup>, H4-24.12<sup>a</sup>, H6-23.30<sup>a</sup>, H7-24.28<sup>a</sup>, H9-22.11<sup>a</sup> and H15-24.37<sup>a</sup> | 2                             |
|        | Os06g18720    | 0.28           | retrotransposon protein, putative, 5'3' gypsy subclass, expressed | H4-26.58<sup>a</sup> | H1-20.00<sup>a</sup>, H2-24.45<sup>a</sup>, H3-25.00<sup>a</sup>, H5-23.80<sup>a</sup>, H6-24.64<sup>a</sup>, H7-24.94<sup>a</sup>, H8-25.73<sup>a</sup>, H10-23.11<sup>e</sup>, H12-19.53<sup>e</sup>, H13-24.75<sup>a</sup>, H14-23.36<sup>a</sup> | 3                             |
| KB     | Os05g0187500  | -0.1           | IQ calmodulin-binding motif family protein, expressed | H7-2.37<sup>a</sup> | H1-2.40<sup>a</sup>, H2-2.42<sup>a</sup>, H3-2.59<sup>a</sup>, H4-2.40<sup>a</sup>, H5-2.56<sup>e</sup>, H6-2.46<sup>e</sup>, H9-2.40<sup>a</sup>, H11-2.47<sup>bc</sup>, H12-2.58<sup>bc</sup>, H13-2.45<sup>e</sup>, H14-2.46<sup>e</sup>, H15-2.46<sup>e</sup> | 6                             |
| LBR    | Os05g0187500  | 0.18           | IQ calmodulin-binding motif family protein, expressed | H9-2.81<sup>a</sup> | H1-2.71<sup>a</sup>, H2-2.69<sup>a</sup>, H3-2.14<sup>bc</sup>, H4-2.66<sup>e</sup>, H5-2.38<sup>e</sup>, H6-2.68<sup>e</sup>, H7-2.66<sup>e</sup>, H11-2.59<sup>e</sup>, H12-2.33<sup>abc</sup>, H13-2.61<sup>e</sup>, H14-2.57<sup>e</sup>, H15-2.51<sup>e</sup> | 73                            |
| ASV    | Os06g0229000  | 1.96           | OsFtsH6 FtsH protease, homologue of AtFtsH6, expressed | H1-3.43<sup>ab</sup> | H2-3.47<sup>ab</sup>, H3-3.00<sup>ab</sup>, H4-3.57<sup>ab</sup> | 9                             |
|        | Os06g12380    | 1.96           | expressed protein | H11-2.75<sup>ef</sup> | H5-4.48<sup>ab</sup>, H6-2.92<sup>ab</sup>, H8-3.67<sup>ef</sup>, H9-3.25<sup>ef</sup>, H10-3.00<sup>ef</sup>, H13-2.58<sup>ef</sup>, H15-2.60<sup>ef</sup> | 8                             |
| KBAC   | Os08g01010    | -1.16          | hypothetical protein | H11-3.17<sup>ef</sup> | H1-3.40<sup>ef</sup>, H3-3.52<sup>ef</sup>, H4-3.31<sup>ef</sup>, H6-3.18<sup>ef</sup>, H10-3.29<sup>ef</sup>, H11-3.35<sup>ef</sup>, H14-3.53<sup>ef</sup>, H15-3.60<sup>ef</sup> | 8                             |
| GC     | Os10g0469000  | 5.01           | leucine-rich repeat receptor protein kinase EXS precursor, putative, expressed | H1-56.06<sup>ef</sup> | H2-47.73<sup>ef</sup> | 6                             |
| RS     | Os06g11100    | 1.74           | retrotransposon protein, putative, unclassified, expressed | H4-3.28<sup>ef</sup> | H1-3.06<sup>ef</sup>, H2-2.42<sup>ef</sup>, H3-2.85<sup>ef</sup> | 48                            |
| PGI    | Os08g12590    | -2.91          | retrotransposon, putative, centromere-specific | H13-62.52<sup>ab</sup> | H1-69.34<sup>ef</sup>, H2-67.64<sup>ef</sup>, H5-63.30<sup>ef</sup>, H6-66.68<sup>ef</sup>, H9-65.41<sup>ef</sup>, H11-65.53<sup>ef</sup>, H14-71.67<sup>ef</sup> | 5                             |

In the present investigation, diverse panel of rice accessions (3 K-RGP sub-set) were not only phenotyped for grain quality-related traits but also phenotyped for the most important health/well-being related traits such as RS and PGI to map candidate SNPs, functional annotations and to unravel the haplotype diversity of identified candidate loci across the 3 K-RGP. Then, the haplo-phenotype analysis was carried out to ascertain superior or appropriate haplotype for the improvement of grain quality in rice breeding programs.

RS, a valuable trait was found to have a significant positive correlation with amylose content and was negatively correlated with glycemic index as suggested by several previous reports and thus a suitable parameter for food consumption to improve the health status of those who are suffering from diabetics, colon cancer, and obesity. We identified nine strong MTAs for RS with significant PVE on chromosomes 1, 2, 6, 8 and 11 (Table 1). Primarily, previous reports found that sbe3-rs on chromosome 2<sup>26,27</sup> and SSIIIa (Os08g0191433) on chromosome 8 involved in the production of RS content and it may be associated with expression of Waxy gene as well. In our study, we report two MTAs in proximity to the candidate regions mentioned above of chromosome 2 (Os02g05943100) and chromosome 8 (Os08g0335500) with not much significant haplotype diversity but significant variations in the trait value was observed among the accessions.<sup>27</sup> We also identified a locus on chromosome 6 (Os06g11100) near (1 Mb away) to the position of starch synthase IIa (SSIIIa: Os06g0229800, 6,748,398–6,753,302) gene, significantly affecting GC, degree of starch crystallinity by increasing the fractional
amount of A chains in amylopectin. All the identified MTAs and their corresponding candidate loci exhibit a significant number of variable haplotypes for the traits studied. In recent past, the discovery of mutation/variants each in GBSSI, SSI, SSIIa, and SSIIIa with a potential to increase RS content and hydrolysis index in rice has been reported. For PGI, we identified a total of 9 associations on chromosome 1 (Os01g0548000), 3 (Os03g38070), 7 (Os07g0523000 and Os07g03060 and Os09g29090) and 11 (Os11g10500 and Os11g22020) (Table 1). Similar kind of results were reported and we also found significant MTAs influencing PGI in close proximity on chromosome 1 (snp_01_30302075 and snp_01_36980087), Chr 3 (snp_03_22422723), Chr 9 (snp_09_21456056 and snp_09_21523496) and Chr 11 (snp_11_334055 and snp_11_28758733). Similar is the case for GC; our results match with qAC7 associated with RM8261 in the physical position Chr7: 25,866,581. Here, we identified a candidate Os07g0597400 on chromosome 7 associated with GC. A QTL qRS7-2 on chromosome 7 associated with RS content between RM3404 (chr7:20,105,832) and RM478 (chr7:25,949,521) was identified and reported.

Over several decades, quality traits like AC, GC, ASV, LER, etc., of rice grains have been extensively studied and it has been reported that AC is governed by waxy gene, by several other loci and even by unidentified non-waxy genes. Our study reported nine associations for AC (Fig. S1) and the major SNP S6_6896749 located with the candidate loci of Os06g0232700 found adjacent to ALK gene. Several other loci were identified on Chromosome 2, Os02g0135800 encoding for WD domain, G-beta repeat domain-containing protein, expressed and Os02g05180 (2,474,176) encrypting for retrotransposon protein, putative Ty3-gypsy subclass. Starch branching enzyme 3 (SBEIIb-Os02g32660) is essential for amylopectin synthesis in the endosperm. Several reports found that high amylose rice and maize reveals in amylose extender (ae) phenotype due to the inactivation of SBEIIb. These results clearly confirmed that a network of genes are involved in the biosynthesis of amylose. Our association analysis study provides a stronger picture of how genetic network of starch synthesizing genes involved in cooking and eating quality improvement in rice. We also identified a association on chromosome 11, S11_28809145 on Os11g0704000 with the function of selt/selW/selH seleno-protein domain-containing protein, expressed slightly closer to Os11g31330.

Kernel length, kernel breadth, and LBR are the important traits to assess the grain size and dimensions for selecting appropriate breeding lines for their consumer’s preference and market acceptability. Medium slender...
rice is more preferable than bold rice varieties. For kernel length, an MTA was identified on chromosome 11 (S11_28909306) with 9 haplotypes. For kernel breadth, one MTA was observed on chromosome 5, two on chromosome 3 and one on chromosome 2. LBR determines the grain size and is controlled by a complex network of genes and also affected by environmental factors. Our study reported similar findings of 1. The reported QTLs were associated with grain size and grain width. The QTLs were also associated with the minimum haplotype diversity range of 4 to 15 haplotypes, respectively. The same QTL on chromosome 6 at the position of 6,726,252 associated with ALK gene was reported by 1. A study conducted with 258 accessions of 3 k panel also reported the same genetic location (S6_6752888) for GT and a candidate of ALK gene encoding SSIIa affecting the chain-length distribution of amylopectin causing alkali disintegration of rice grains.9,55,57. For gel consistency, we identified two candidates on chromosome 7 (Os07g0597400) and 10 (Os10g0469000) with minimum haplotype variations of 8 and 2. Previously, a qGC6 was reported close to the breakpoint of WAXY gene.12. More importantly, a new novel association for GC is identified on chromosome 7 had near vicinity and it was not previously reported. Linear elongation ratio is an important cooking quality trait and we identified an associations for LER on chromosomes 9. A QTL for LER was reported on chromosome 4 with the marker interval of C933 to C946 using RIL population of indica x indica hybrids.86

On a whole, amylose and amyllopectin production is a complex phenomenon governed by several networks of genes and their coordinating expression during grain filling and seed maturation affecting the grain textural properties and even prevailing environmental conditions also. Noteworthy, it is not an easy breeding protocol to improve the quality aspects of rice varieties by selecting a notable gene that may be either superior or inferior for a specific trait and introgression by markers assisted pyramiding and selection. In this, we identified several superior haplotypes by studying haplotype diversity and linking haplotype variations with the trait value. Assembling superior haplotype combinations for most of the grain quality traits viz., AC, ASV, GC, LBR and RS coupled with GI in to one background to maximize the textual properties of grains, yield and most of morpho-agronomic traits to develop tailor-made new generation rice with enhanced genetic gain compared to available popular mega rice varieties.81,82. For example, AC (Os06g18710 – 25.50%) possessing high amylose class, LBR (Os05g0187500) and ASV (Os05g0187500) with worthy haplotype diversity in the range of 4 to 15 haplotypes, respectively. The same QTL on chromosome 6 at the position of 6,726,252 associated with ALK gene was reported by 1. A study conducted with 258 accessions of 3 k panel also reported the same genetic location (S6_6752888) for GT and a candidate of ALK gene encoding SSIIa affecting the chain-length distribution of amylopectin causing alkali disintegration of rice grains.9,55,57. For gel consistency, we identified two candidates on chromosome 7 (Os07g0597400) and 10 (Os10g0469000) with minimum haplotype variations of 8 and 2. Previously, a qGC6 was reported close to the breakpoint of WAXY gene.12. More importantly, a new novel association for GC is identified on chromosome 7 had near vicinity and it was not previously reported. Linear elongation ratio is an important cooking quality trait and we identified an associations for LER on chromosomes 9. A QTL for LER was reported on chromosome 4 with the marker interval of C933 to C946 using RIL population of indica x indica hybrids.86

In this investigation we also identified better donors for RS, PGI and other quality-related traits which can be utilized in the haplotype-based breeding program to develop elite lines with low glycemic index value with desirable quality traits by assembling superior haplotypes suited for different situations (Table S5). Our results were compared with previous reports of glycemic index, we found UQUIHUA::IRGC117,037–1 had registered intermediate GI of 66.88 with the RS content of 3.23% whereas reported as low GI (< 55). Apart from this we also selected different classes of glycemic index lines from the 3 K sub-panel grown in 2019WS and different quality-related traits were phenotyped and result showed that four lines namely BAIANG 6::IRGC 6129–1 from Indonesia, MAKRO::IRGC 74,763–1 from India, AUS 329::IRGC 29,116–1 from Bangladesh and KOTTEYARAN::IRGC 47,383–1 from Srilanka possesses medium slender grains with good elongation ratio and medium to soft gel consistency after cooking (Table S6). The line MAKRO::IRGC 74,763–1 from India having intermediate RS content and low PGI of 52.91.

But when considering ASV for GT, the identified entries for the low glycemic index showed the score of 1 to 2 which represents high GT requires more water, cooking time and poor in texture not suitable for cooking and eating (Table S6). Amylose content alone does not describe the cooking and eating quality of rice grains, as varieties with the same range of AC possess variable variations in the cooking and eating quality.85. In our study, intermediate ASV of 4 to 5 score possesses high PGI values. Superior haplotypes associated with the low glycemic index, intermediate ASV and soft GC in the preferred combinations can be utilized for the development of low glycemic rice varieties with desirable grain quality. Hence, the approach of haplotype-based breeding is anticipated to assist in the development of -premium quality rice varieties with low GI to meet the increasing demands of the rice consuming population.

### Conclusion

Significant variations were observed for the grain quality traits, especially for AC, ASV, GC and RS. We captured several novel and significant associations of SNPs for the target traits and studied haplotype differences of identified candidate genes. Newly identified candidate genes might be useful for further functional characterization and pathway elucidation for grain quality traits in rice. Here we proposed an exclusive strategy on use of superior haplotype-possessing elite donors and incorporating superior/appropriate haplotypes for the majority of quality traits in a genetic background as a prominent way to develop high yielding and quality rich rice varieties suitable for the consumption of diabetics, obese population and as preferred by consumer’s needs and demands.
Material and methods

Genetic material and field trial. The genetic material used in this study comprised of 150 accessions having different duration groups of 3 K RGP evaluated under transplanted situation (TPR) during dry season 2017 (DS2017) and Wet season 2018 (WS2018) (Table S7). These accessions were evaluated for RS, PGI and 9 quality parameters including physico-chemical properties, cooking and eating properties. The analysis was carried out at the rice quality lab at IRRI SA Hub, ICRI SAT campus. The agronomic management of the transplanted experiments including was carried as per the procedure detailed.

Phenotyping of physicochemical parameters and cooking and eating quality of rice. Traits namely, KL, KB and LBR were measured with 10 kernels using Standard Evaluation System of Rice (SES) and the mean was calculated. Amylose content of the grains was estimated by the rapid protocol of cut grain dip method. For cooking-related parameters, 20 kernels from each entry were soaked in 5 ml of distilled water in 15 ml test tubes for 20 min and cooked in a boiling water bath at 100 °C intended for the time period of 8 min to determine the KLAC, KBAC and LER. Gel consistency was determined by the method formulated by, where known amount of rice flour was placed in the culture tubes and wetted with 0.2 ml 95% ethanol containing 0.03% thymol blue and 2 ml 0.2 N KOH added and mixed with a Vortex Genie mixer set at specified speed. Tubes were covered with glass marbles and heated in a vigorously boiling water bath for 8 min. Then, the tubes were removed from the water bath and kept at room temperature for 5 min, cooled in an ice-water bath for 20 min, and laid flat on a laboratory table and length of the blue-colored gel was measured in millimeters. The method separated different classes of GC into soft (61–100 mm), medium (41–60 mm) and hard gel consistency (25–40 mm) based on SES, 2013. GT was estimated based on alkali spreading value (ASV) of milled rice by. Six whole-grain, milled rice samples were placed in duplicate Petri plates containing 10 ml of 1.7% KOH. The Petri plates were covered and incubated for 23 h at 30 °C. The appearance and disintegration of grains were visually observed after incubation based on the scale given by, SES, 2013.

Phenotyping for RS and predicted GI. Measurement of RS. RS content of the rice samples was determined by using a resistant starch assay kit (K-RSTAR, Megazyme, Irishtown, Ireland) by ten-fold downsampling of sample and reagents with slight modifications. Fine powder of polished rice flour (10 mg) was taken in 2 ml eppendorf tube and it was digested with 400 μl of enzymatic mixture-containing 10 mg/ml of pancreatic α-amylase and 3.0U/ml of amylglucosidase (AMG) and the mixture was incubated at 37°C for 16 h with continuous shaking (200 strokes/minute). After incubation, 400 μl of 95% ethanol was added to the mixture and vortexed to stop the reaction. Then, the samples were centrifuged for 10 min at 12,000 rpm and the supernatant was collected and the pellets were washed repeatedly with 200 μl (vortexed) and 600 μl of 50% ethanol. Washing with 50% ethanol was made twice to remove all the non-resistant starch in the samples and supernatants were pooled for the measurement of non-resistant starch. Then, the pellets were air-dried for 30 min to remove moisture content in the residue. Around, 200 μl of 2 M KOH was added to the residue and mixed properly to avoid the formation of clumps and incubated at 5°C for 1 h in a shaker (200 strokes/minute). Then, 800 μl of 1.2 M sodium acetate buffer and 10 μl of AMG (3,300 U/ml), vortexed and tubes were placed in a water bath for 30 min at 50°C with intermittent mixing with the help of vortex mixer for every 10 min. After incubation, samples were cooled and centrifuged at 12,000 rpm for 10 min. Then, 100 μl of sample aliquot was taken in a fresh test tube (15 ml), 3 ml of GOPOD reagent (D-Glucose assay kit, Megazyme, Irishtown, Ireland) and incubated in a water bath at 50°C for 30 min. The absorbance was determined with the help of UV-1800 (Shimadzu Corporation, Japan) at 510 nm against a blank containing buffer and GOPOD reagent.

Estimation of predicted GI (PGI). The glycemic index was determined by the protocol given by with little modifications. Fine powder of polished rice flour (50 mg) was cooked in 5 ml of water for 30 min and 10 ml of HCl-KCl buffer (pH = 1.5) was added to the sample. Then, 0.2 ml of a solution containing 1 g of pepsin in 10 ml of HCl-KCl buffer was added to each sample and incubated at 40 °C for 1 h in a shaking water bath. After incubation, the volume was adjusted to 25 ml with Tris maleate buffer (pH = 6.9). Then, 5 ml of α–amylase (2.610^3 U/ml) was added to hydrolyze the digested starch to glucose. The samples were incubated at 60 °C for 45 mins in a shaking water bath and glucose content in each aliquot was estimated using the GOPOD kit (D-Glucose assay kit, Megazyme, Irishtown, Ireland). Glucose was converted into starch by multiplying with 0.9. Kinetics of starch digestion was estimated by using the resistant starch assay kit (K-RSTAR, Megazyme, Ireland). Glucose was converted into starch by multiplying with 0.9. Kinetics of starch digestion was estimated by non-linear first-order equation; C = Cα (1-e^{-kt}). Where, C = Concentration of starch hydrolyzed at t, Cα = equilibrium concentration (i.e., % of starch hydrolyzed after 180 min which is the glucose concentration at equilibrium (t180), tf is the final time (180 min), t0 is the initial time (0 min) and k is the kinetic constant. The area under the hydrolysis curve was calculated by comparison with the AUC of a reference food (white bread). A hydrolysis index (HI) was calculated by comparison with the AUC of a reference food (white bread). HI = AUCsample / AUCreference × 100.

GWAS, Haplo‑Pheno analysis and Haplo‑Pheno analysis. Mean data was used for GWAS analyses using approximately more than 500 K SNP data. The analysis was carried out by using GAPIT (MLM model) (https://CRAN.R-project.org/), R based approach considering both kinship values (k-values) and population structure (Q-matrix). Of the several methods suggested to correct false positive in association analysis even
keeping stringent p-value benchmark, the most stringent correction method called “Bonferroni Correction” was used in the present analysis. The bonferroni threshold calculated by the formula \(1/m\) where ‘\(m\)’ is the number of test performed which resulted in \(-\log_{10}(1/m) = 5.69\). As it is too conservative and suggestive value of \(p = 0.00001\) \((\log_{10}(\text{p}) = 5.30)\) was used for the identification of peak associations (MTAs) with the targeted grain quality traits. Multi-locus GWAS (mrMLM) was also conducted to validate the MTAs identified by single locus model of MLM by adopting the significant LOD score of \(3.9\). The phenotypic allele effect (\(a\)) was determined by the formula given by\(^{29}\) and the favorable alleles of each trait were subsequently identified according to the breeding objective. Strongly associated SNPs with the targeted traits were utilized to find the causative QTLs genes by the RAP database. In-built tool of SNP seek database\(^{26}\) was utilized to conduct haplotype analysis for entire candidate locus, by employing default parameters with Calinski criteria for k-group determination. Nipponbare was used as the reference genome. All of the 150 lines belonging to 8 sub-populations namely, aro, aus, admix, ind1A, ind1B, ind2, ind3, and indx was considered for the haplotype analysis. We utilized the ‘3kfiltered’ SNP set present in the SNP seek database for the entire analysis. The filtered was obtained from the Base SNP set by applying the following filtering criteria: (1) alternative allele frequency at least 0.01, (2) proportion of missing calls per SNP at most 0.2\(^{22}\) (http://snp-seek.irri.org/download.zul) and this SNP set was already available in the SNP seek database which was directly utilized in this study. Haplotype analysis for candidate locus have been carried out considering only the nonsynonymous SNPs and indels in the exon region that results in amino-acid change. The information regarding haplotype and their diversity was obtained from the SNP seek database (http://snp-seek.irri.org/download.zul) to detect the superior haplotype by categorizing haplotypes by using phenotyping data of concerned genotype trait means for the associated genes.

**Statistical analysis.** Mean, range and standard deviation were calculated with the help of a standard excel program. Duncan’s multiple range tests were carried out using XL Stat 2019. Correlation analysis among all the component traits related to quality was conducted by SPSS ver. 20. Boxplot and correlogram were constructed with the help of R-package (https://CRAN.R-project.org/). Significant difference among the haplotypes (only the haplotypes validated in at least two lines were considered) was studied with the help of F-statistic and Duncan’s test with the significance level of \(p < 0.05\).

Received: 7 October 2020; Accepted: 19 March 2021

**References**

1. Hao, W., Zhu, M., Gao, J., Sun, S. & Lin, H. Identification of quantitative trait loci for rice quality in a population of chromosome segment substitution lines. *J. Integr. Plant Biol.* **51**, 500–512 (2009).

2. Guo, X., Lei, C., Wang, J., Jiang, L. & Wan, J. Identification of a stable quantitative trait locus for percentage grains with white chalkiness in rice (*Oryza sativa*). *J. Integr. Plant Biol.* **53**, 598–607 (2011).

3. Khush, G.S., Paule, C.M. and Cruz, N.M. Rice grain quality evaluation and improvement at IRRI. In Proceedings of the Workshop on Chemical Aspects of Rice Grain Quality, IRRI, Los Banos, Philippines, pp. 21–31 (1979).

4. Barclay, A. W. *et al.* Glycemic index, glycemic load, and chronic disease risk—a meta-analysis of observational studies. *Am. J. Clin. Nutr.* **87**, 627–637 (2008).

5. Gnagnarella, P., Gandini, S., La Vecchia, C. & Maisonneuve, P. Glycemic index, glycemic load, and cancer risk: a meta-analysis. *Am. J. Clin. Nutr.* **87**, 1793–1801 (2008).

6. Fitzgerald, M. A. *et al.* Identification of a major genetic determinant of glycemic index in rice. *Rice* **4**, 66–74 (2011).

7. Parween, S. *et al.* Balancing the double-edged sword effect of increased resistant starch content and its impact on rice texture: its genetics and molecular physiological mechanisms. *Plant Biotechnol. J.* https://doi.org/10.1111/pbi.13339 (2020).

8. Abbai, R., Singh, V. K., Snowden, R. J., Kumar, A. & Schrubbus, T. Seeking crops with balanced parts for the ideal whole. *Plant Sci.* **25(12)**, 1189–1193 (2020).

9. Anacleto, R. *et al.* Integrating a genome-wide association study with a large-scale transcriptome analysis to predict genetic regions influencing the glycemic index and texture in rice. *Plant Biotechnol. J.* **17**, 1261–1275 (2019).

10. Goni, I., Garcia-Alonso, A. & Saura-Calixto, F. A starch hydrolysis procedure to estimate glycemic index. *Nutrition Res.* **17**, 427–437 (1997).

11. Lai, M. H., Liu, K. L., Chen, P. & Y. Ke N.J., Chen J., Sung J., Wu Y. L., Lin S.D.,. Predicted glycemic index and glycemic index of formula given by29 and the favorable alleles of each trait were subsequently identified according to the breeding objective. Strongly associated SNPs with the targeted traits were utilized to find the causative QTLs genes by the RAP database. In-built tool of SNP seek database was utilized to conduct haplotype analysis for entire candidate locus, by employing default parameters with Calinski criteria for k-group determination. Nipponbare was used as the reference genome. All of the 150 lines belonging to 8 sub-populations namely, aro, aus, admix, ind1A, ind1B, ind2, ind3, and indx was considered for the haplotype analysis. We utilized the ‘3kfiltered’ SNP set present in the SNP seek database for the entire analysis. The filtered was obtained from the Base SNP set by applying the following filtering criteria: (1) alternative allele frequency at least 0.01, (2) proportion of missing calls per SNP at most 0.2\(^{22}\) (http://snp-seek.irri.org/download.zul) and this SNP set was already available in the SNP seek database which was directly utilized in this study. Haplotype analysis for candidate locus have been carried out considering only the nonsynonymous SNPs and indels in the exon region that results in amino-acid change. The information regarding haplotype and their diversity was obtained from the SNP seek database (http://snp-seek.irri.org/download.zul) to detect the superior haplotype by categorizing haplotypes by using phenotyping data of concerned genotype trait means for the associated genes.

**Statistical analysis.** Mean, range and standard deviation were calculated with the help of a standard excel program. Duncan’s multiple range tests were carried out using XL Stat 2019. Correlation analysis among all the component traits related to quality was conducted by SPSS ver. 20. Boxplot and correlogram were constructed with the help of R-package (https://CRAN.R-project.org/). Significant difference among the haplotypes (only the haplotypes validated in at least two lines were considered) was studied with the help of F-statistic and Duncan’s test with the significance level of \(p < 0.05\).
23. Wang, S. et al. The OsSPL16GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nature Genet.* **47**, 949–954 (2015).
24. Liu, S. et al. OsMAPK6s, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J.* **84**, 672–681 (2015).
25. Si, L. et al. OsSPL13 controls grain size in cultivated rice. *Nature Genet.* **48**, 447–456 (2016).
26. Jin, J. et al. *Plant Cell* **28**, 2453–2463 (2016).
27. Zhao, K. et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**, 467 (2011).
28. Singh, R., Singh, A., Sharma, T., Singh, A. & Singh, N. Fine mapping of grain length QTLs on chromosomes 1 and 7 in Basmati rice (*Oryza sativa*) L. *J. Plant Biochem. Biotechnol.* **21**, 157–166 (2012).
29. Zhang, Y. et al. Variations and transmission of QTL alleles for yield and fiber qualities in upland cotton cultivars developed in China. *PLoS ONE* **8**, e57220 (2013).
30. Qiu, X. et al. Genome-wide association study of grain appearance and milling quality in a worldwide collection of Indica rice germplasm. *PLoS ONE* **10**, e0145577 (2015).
31. Feng, Y. et al. Genome wide association mapping for grain shape traits in *indica* rice. *Planta* **244**(4), 819–830 (2016).
32. Edzesi, W. et al. Genetic diversity and elite allele mining for grain traits in *rice* (*Oryza sativa* L.) by association mapping. *Front. Plant Sci.* **7**, 787 (2016).
33. Wang, J., Hu, P. & Lin, L. Gradually decreasing starch branching enzyme expression is responsible for the formation of heterogeneous starch granules. *Plant Physiol.* **176**, 582–595 (2018).
34. Zhang, G. et al. Double repression of soluble starch synthase genes SSIIa and SSIIId in rice (*Oryza sativa*) L. uncovers interactive effects on the physicochemical properties of starch. *Genome* **54**(6), 448–459 (2011).
35. Pandey, M. K. et al. Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.). *Biotechnol. Adv.* **30**, 1697–1706 (2012).
36. Zeng, Y. W. et al. Identification of QTLs for resistant starch and total alkaloid content in brown and polished rice. *Genet. Mol. Res.* **15**(3), 15037268 (2015).
37. Wang, S. B. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**, 19444. https://doi.org/10.1038/srep19444 (2016).
38. Wen, Y. J. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* **19**, 700–712 (2018).
39. Zhang, Y.W., Tambia, C.L., Wen, Y.J., Li, P., Ren, W.L., Ni, Y.L., Gao, J., Zhang, Y.M. mRMLM v4.0: An R platform for multi-locus genome-wide association studies. *Genomics, Proteomics & Bioinformatics*, online, DOI: https://doi.org/10.1016/j.gpb.2020.06.006 (2020).
40. Khush, G. S. What it will take to feed 5.0 billion rice consumers in 2030. *Scientific Reports*.
65. Gao, Z. et al. ALK, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. *J. Integr. Plant Biol.* **53**, 756–765 (2011).

66. Bao, J. S., Corke, H. & Sun, M. Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physico-chemical properties in non-waxy rice (*Oryza sativa L.*). *Theor. Appl. Genet.* **113**, 1185–1196 (2006).

67. Mogga, M., Sibiya, J., Shimelis, H., Lamo, J. & Yao, N. Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa L.*) using DArT markers. *PloS ONE* **13**, e0198012 (2018).

68. Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H. & Nakamura, Y. Starch biosynthesis in cereal endosperm. *Plant Physiol. Biochem.* **48**, 383–392 (2010).

69. Lu, F. H. et al. An SNP downstream of the OsBEIIb gene is significantly associated with amylose content and viscosity properties in rice (*Oryza sativa L.*). *J. Cereal Sci.* **56**, 706–712 (2012).

70. Wambugi, P., Ndijiondjop, M. N., Furtado, A. & Henry, R. Sequencing of bulks of segregants allows dissection of genetic control of amylose content in rice. *Plant Biotechnol. J.* **16**, 100–110 (2018).

71. Fan, 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.* **112**, 1164–1171 (2006).

72. Weng, J. et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18**, 1199–1209 (2008).

73. Li, Y. et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nature Genet.* **43**(12), 1266–1269 (2011).

74. Qi, P. et al. The novel quantitative trait locus GL31 controls rice grain size and yield by regulating Cyclin-T13. *Cell Res.* **22**, 1666–1680 (2012).

75. Duan, P. et al. Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. *Mol. Plant* **10**, 685–694 (2017).

76. Liu, Q. et al. G-protein β subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Naure Commun.* **9**(1), 852 (2018).

77. Misra, G. et al. Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. *Sci. Rep.* **7**, 12478 (2017).

78. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genet.* **42**, 961–967 (2010).

79. Umemoto, T., Yano, M., Satoh, H., Shomura, A. & Nakamura, Y. Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. *Theor. Appl. Genet.* **104**, 1–8 (2002).

80. Liu, L. L. et al. Identification of stably expressed quantitative trait loci for cooked rice elongation in non-Basmati varieties. *Genome* **51**(2), 104–112 (2008).

81. Revan, M. W. et al. Genomic innovation for crop improvement. *Nature* **543**, 346–354 (2017).

82. Abbai, R. et al. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* **17**, 1612–1612 (2019).

83. ICAR-Indian Institute of Rice Research (IIRR), 52nd All India Co-Ordinated Programme: Agronomy Technical Programme, Rajendranagar, Hyderabad, (2016–2017).

84. Mansueto, L. et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Rice* **10**, 203–208 (2006).

85. Bevan, M. W. *Genomic innovation for crop improvement*. Nature **543**, 346–354 (2017).

86. Cagampang, G. B., Perez, C. M. & Juliano, B. O. A gel consistency test for eating quality of rice. *J. Sci. Food Agric.* **24**, 1589–1594 (1973).

87. Little, R. R., Hilder, G. B. & Dawson, E. H. Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal chem.* **35**, 111–126 (1958).

88. Butardo, V. M. et al. Impact of down-regulation of starch branching enzyme Iib in rice by artificial microRNA- and hairpin RNA-mediated RNA silencing. *J. Exp. Bot.* **62**, 4927–4941 (2011).

89. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).

90. Misra, G. et al. Dissecting the genome-wide genetic variants of milling and appearance quality traits in rice. *J. Exp. Bot.* **70**(19), 5115–5130 (2019).

91. Zhang, Y. M., Jia, Z. & Dunwel, J. M. The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* **10**, 100. https://doi.org/10.3389/fpls.2019.00100 (2019).

92. Mansueto, L. et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res* **45**, 1075–1081 (2016).

Acknowledgements

Authors are grateful to the Department of Science and Technology (DST), Science and Engineering Research Board (SERB), Government of India and Department of Biotechnology, Government of India for providing financial support.

Author contributions

A.K. and V.K.S. conceived and designed the experiments. R.S., A.K.S., S.V.H. and U.M.S. performed the phenotyping of quality and related traits. R.S., A.K.S. and R.A. analysed and interpreted the result. R.S. and A.K.S. prepared the manuscript. A.K. and V.K.S. revised and finalized the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-87964-8.

Correspondence and requests for materials should be addressed to A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
