THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Genome wide association study on cooked rice texture in rice (*Oryza sativa* L.)

벼에서 쌀밥 조직감 전유전체연관분석 연구

February, 2023

Graduate School of Agriculture and Life Sciences
Seoul National University
Crop Science and Biotechnology Major

Dasol Kim

Genome wide association study on cooked rice texture in rice (*Oryza sativa* L.)

Under the direction of Dr. Hee-Jong Koh
Submitting a master's thesis of Science

November, 2022

Graduate School of Agriculture and Life Sciences
Seoul National University
Crop Science and Biotechnology Major

Dasol Kim

Confirming the master's thesis written by
Dasol Kim

February, 2023

Chair _____
Tae-Jin Yang, Ph.D
Vice Chair _____
Hee-Jong Koh, Ph.D
Examiner _____
Suk-Ha Lee, Ph.D

Genome wide association study on cooked rice texture in rice (*Oryza sativa* L.)

DASOL KIM

ABSTRACT

Texture is considered a multidimensional sensory property experienced by humans. The types of texture felt by food are various. Food texture can be an indicator of high-quality rice and is an essential consideration in its consumer acceptance and palatability. Phenotypic variations of texture in cooked rice are described in terms of hardness, tenderness, and adhesiveness by texture assessments. We investigated diverse cooked rice texture parameters that include hardness (HRD), adhesiveness (ADH), cohesiveness (COH), springiness (SPR), resilience (RES), and chewiness (CWN) in 248 rice accessions, including improved varieties and landraces using a texture profile analysis (TPA).

Short-reads data were produced from next-generation sequencing (NGS). The data was mapped against the rice reference

genome (Nipponbare, IRGSP v1.0) to identify genomic variants. A total of 1,254,682 high-quality single-nucleotide polymorphisms (SNPs) datasets derived from the panels and texture parameters obtained from the TPA method were used for genome-wide association study (GWAS). SNPs associated significantly with HRD were detected on chromosome 5 (Chr05:13917049), 6 (Chr06:1772858; Chr06:1772859), which included *Waxy* (*Wx*), *OsBDG* in linkage disequilibrium region, and 6 (Chr06:4518410). A total of 19 significant SNPs associated with ADH were identified on chromosome 6. Detected SNP that is involved in RES located on Chromosome 5 (Chr05:3455897). The linkage disequilibrium analysis determined the putative candidate region that included significant SNPs analyzed with p-value ≤ 1.572e-6 and inherited together.

We identified the phenotypic differences through linkage disequilibrium (LD)-based analysis using candidate gene (*GS5*) that affects RES of cooked rice. The results of this study could contribute to our understanding of texture of cooked rice. Detected SNPs associated with the texture parameters and candidate genes affecting texture parameters will be an effective material to improve the physical characteristics of cooked rice in breeding system.

Keyword : Cooked rice texture, Genome wide association study (GWAS), Texture profile analysis (TPA), Hardness, Adhesiveness, Resilience

Student Number : 2021-25835

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

HRD: Hardness

ADH: Adhesiveness

COH: Cohesiveness

SPR: Springiness

RES: Resilience

GUM: Gumminess

CWN: Chewiness

Wx: *Waxy*

GS5: *GRAIN SIZE 5*

GWAS: Genome wide association study

Fast-LMM: Factored spectrally transformed linear mixed models

LD: Linkage disequilibrium

MAF: Minor allele frequency

PCA: Principal component analysis

LIST OF TERMS IN KOREAN

Hardness: 경도

Adhesiveness: 부착성

Cohesiveness: 응집성

Springiness: 탄력성

Resilience: 순간복원력

Gumminess: 검성

Chewiness: 씹힘성

INTRODUCTION

Rice (*Oryza sativa* L.) is a common staple food grain for more than half of the world's population. Among most significant cereals, rice is the only one consumed in whole grains after cooking (Li, 2017).

Rice grain quality is separated into four parts that include milling quality (MQ), appearance quality (AQ), nutritional quality (NQ), and eating & cooking quality (ECQ). MQ refers to the integrity of rice during processing, including roughness rate, milled rice rate, and head rice rate. AQ usually includes grain shape, chalkiness, transparency, and other indicators. NQ is influenced by the quantity and quality of starch, protein, vitamins, minerals, and other phytochemicals beneficial to human health. ECQ, also known as sensory quality, mainly includes cooked rice characteristics and acceptability (Li et al., 2022b).

Recently, People are showing a preference for a better quality of cooked rice with good sensory properties(Calingacion et al., 2014). High-quality rice must match the taste of consumers, even the texture. Because people in countries that consume rice as a staple food are sensitive to the quality of rice, and there is various

palatability in each country (Champagne et al., 2010). For instance, soft or sticky texture of cooked rice is preferred in some Asian countries like China, Korea, and Japan. India and Pakistan consumers prefer soft and fluffy rice varieties. The hard texture of cooked rice is favored in Bangladesh, Sri Lanka, Indonesia, and Myanmar (Suwannaporn & Linnemann, 2008). Therefore, breeding high-quality rice is increasing and it is crucial to capture various consumers' attention in the market.

Texture of food was defined by the International Standards Organization (ISO, 1994) as 'All the rheological and structure (geometrical and surface) attributes of a food product perceptible by means of mechanical, tactile, and where appropriate, visual, and auditory receptors'. Various textural characters exist in cooked rice that is defined as sticky, hard, firm, and dry (Misra et al., 2018). Many factors, such as the amylose content (AC) (Bhattacharya & Juliano, 1985), gel consistency (GC) (Cagampang et al. 1973), gelatinization temperature (GT) (Little et al. 1958), seed characteristics (Champagne et al., 1998), and cooking method (Crowhurst & Creed, 2001) affect indirectly texture of cooked rice.

In general, starch composition and structure of rice grains are major factors in cooked rice texture (Ong & Blanshard, 1995). Rice grains consist of 90% or more starch. Grain starch is formed of two

different polysaccharides: amylopectin (Ap; $\alpha-1,6-$branched polyglucans) and amylose (Am; linear $\alpha-1,4-$polyglucans) (Lee et al., 2017). Differences in amylose-to-amylopectin ratio in rice varieties determine cooked rice texture. A high amylose content (AC) is often related with a high hardness and a low stickiness in cooked rice (Rani & Bhattachrya, 1995). Amylose is synthesized by granule-bound starch synthesis I (GBSSI), which is encoded by the *Waxy* (*Wx*) gene (Larkin & Park, 2003). *Wx* allele has multiple variations (e.g., Wx^{a}, Wx^{b}, Wx^{in}, Wx^{op}, Wx^{mp}, and *wx*) and contributes to texture in rice (Anacleto et al., 2019; Chen et al., 2008; Larkin and Park, 2003).

Cooked rice texture is commonly assessed by texture profile analysis (TPA) using a texture analyzer with sensory evaluation by trained sensory panels (Wee et al., 2018). TPA is an instrumental technique that has been applied to specify characteristics of food. TPA method, which runs double compression cycles, mimics the first and second bites during mastication in the mouth and provides information on instrumental responses (Stokes et al., 2013). Cooked rice texture can be measured as four primary characteristics (hardness (HRD), adhesiveness (ADH), cohesiveness (COH) and springiness (SPR)) and two secondary characteristics (resilience (RES) and chewiness (CWN)) by TPA (Rosenthal, 2010). TPA

results of cooked rice texture provide quantitative data inexpensively and quickly, and have stable reproducibility (Misra et al., 2018; Ramesh et al., 2000; Mestres et al., 2011).

Texture diversity of cooked rice has already been reported in rice (Bao et al., 2006). Cooked rice texture, like other grain quality characteristics, is a complex quantitative trait (Hori et al., 2016). Rice texture has been examined using quantitative trait locus (QTL) analysis. In a recombinant inbred population, ADH is highly related to quantitative trait loci on chromosome 1 and chromosome 4, and 5 for HRD (Cho et al., 2010). These texture-related QTLs have not been fine-mapped, nor have candidate genes been identified. AC, which is encoded by the *Wx* gene on chromosome 6, is commonly known to correlate positively with HRD and negatively with ADH (Rani & Bhattachrya, 1995). However, Cho et al., 2010, found no significant associations, indicating that other genes contribute to these textural characteristics in cooked rice. To identify genetic elements that influence texture, it is necessary to use a statistical method that is more powerful than standard techniques (Misra et al., 2018).

Genome wide association studies (GWAS) can be used to examine the genetic basis of complex agronomic characteristics by using the high-quality single nucleotide polymorphisms (SNPs) (Huang et al., 2012). At present, there is little GWAS research about

cooked rice texture parameters, rather than the indirect factors affecting texture. The novel haplotypes and putative candidate genes influencing textural properties beyond AC were identified in diverse set of 236 *Indica* accessions from 37 countries using single-locus (SL), and multi-locus (ML) GWAS (Misra et al., 2018). Moreover, the chromosome 6 area, found by merging a GWAS and a TWAS (transcriptome-wide association study) on a panel of 305 *indica* varieties, represents a strong hot spot correlated to the rice glycaemic index (GI) related to cooked rice texture, involving 26 genes including *Wx* (Anacleto et al., 2019).

In this study, we investigated natural differences in cooked rice texture in 248 rice accessions including improved varieties and landraces. Then, we performed linkage disequilibrium (LD) block-based analysis using the significant SNPs and candidate genes. We identified specific alleles displaying phenotypic variations of texture. The findings of this study will support the comprehension of texture in cooked rice. SNPs and candidate genes associated with texture will be useful for improving the texture parameters of rice in breeding programs.

MATERIALS AND METHODS

Plant Materials

We collected 248 diverse rice accessions including landraces and improved varieties. Paddy rice was harvested in 2020 at the Experimental farm of Seoul National University, Suwon, South Korea. Harvested rice grains were dehulled using a dehull machine, and brown rice grains were milled through a milling machine. White rice was milled to 92.2%. Milled samples were sealed in plastic bags and stored at 0℃ until the experiment.

Cooking methods

For each rice variety, five grams of milled rice were divided into a heat resistance polypropylene (pp) beaker (55mm deep, 45mm in diameters). Samples were washed five times with cold water, followed by straining to remove excess water. After washing, the samples were soaked for 20 minutes in double distilled water (20ml) with an aluminum foil cover to minimize water loss during soaking. After soaking, we completely removed remained water. Distilled water was added to the samples to give a rice-to-water ratio of 1:1.25. After that, the samples were covered with aluminum foil, and

placed on an inner pot of an electric rice cooker. Cooking was started and stopped by the automatic method programmed in the cooker. Cooked samples were allowed to stand for 15 minutes up to texture measurements. Samples not yet ready for the analysis were kept at warm conditions supported by the cooker to reduce starch retrogradation during texture measurements.

Texture profile Analysis (TPA)

Texture profile analysis were performed according to the method described by NAITO et al. (1998) with slight modifications. Two cycles of force-time compression and tension were used to measure and calculate texture parameters using a texture analyzer (TX-700, Lamy rheology, France) with a 20 mm cylindrical probe with a 50 N load cell. The probe was placed 20 mm above the base and allowed to move at a speed of 1 mm/s. A compression rate of 80% strain was established. Five replicates were undertaken for the analysis.

All calculations of texture parameters have complied with the manufacturer's guideline (Fig. 1). Investigated parameters were hardness (HRD; Max force (gf) of the first curve that is required to reach a preset deformation on the sample), adhesiveness (ADH; negative area (mJ) of the first curve, A3), cohesiveness (COH; the

strength of the bonding within the food, A2 / A1) was obtained automatically using RHEOTEX software for TX-700. Springiness (SPR; the property of a sample to return to its original height after the applied force is removed, D2 / D1), resilience (RES; the amount that a food struggles to recover its previous shape and size, A5 / A4), gumminess (GUM; the energy necessary to chew a semi-solid food until it can be swallowed, cohesiveness * springiness), and chewiness (CWN; the energy necessary to transform a solid food into a swallowable state, gumminess * springiness) were separately calculated in python v 3.8 environment based on the original file extracted from the device.

Genotyping dataset and Genome wide association study (GWAS)

The single nucleotide polymorphism (SNP) genotype dataset used in the study was obtained from the 248 re-sequenced genomes mapped against the rice reference genome (Nipponbare, IRGSP v1.0) (Sakai et al., 2013) using the BWA v0.7.17 MEM algorithm (Li & Durbin, 2009). The variant call files (VCF) were processed by using HaplotypeCaller function of GATK v4.1.2 (McKenna et al., 2010) retaining only the high-quality SNPs with a minimum quality score of 30, and sequencing depth over 5.

A total of 1,254,682 high-quality SNPs from the 248 GWAS

accessions with minor allele frequency (MAF) cut-off of 5% were associated with the texture parameters in a mixed model GWAS using the Factored spectrally transformed linear mixed model association (FaST-LMM v2.07) (Lippert et al., 2011).

The association results were complete, and we generated manhattan plots using R package rMVP (Yin et al., 2021). The significant p-value thresholds for GWAS were calculated by dividing the significance level 0.05 by the total number of SNP markers.

Principal component analysis (PCA)

Linkage disequilibrium (LD)-based SNP pruning was conducted using PLINK v1.9 (Purcell et al., 2007) with the --indep-pairwise 50 5 0.2 command. Principal component analysis (PCA) was carried out on a total of 50,755 LD-pruned SNPs. PCA was revealed by --cluster--pca command of PLINK v1.9. Scatter plot was created in Python 3.8v using the seaborn.sns module.

Linkage disequilibrium (LD) block-based analysis

LD heatmap was created using LDBlockShow (Dong et al., 2021) for r^2 calculation with block type 5. Strong associations between parameters and SNPs were sorted from the lowest significant p-value, and candidate genes were selected, including in

LD blocks. Annotations of candidate genes are based on IRGSP v1.0.

Using texture phenotypes and all variants, including SNPs and InDels, we analyzed LD block-based studies. If an individual contained at least one missing or/and heterozygous genotype, it was removed. Lead SNPs and haplotypes of candidate genes based on LD were related to phenotypes and plotted in Python v 3.8. For statistical analysis, one-way ANOVA was performed using the statsmodels.stats.anova module, and a Bonferroni post hoc test were performed using the scipy.stats module in Python 3.8v.

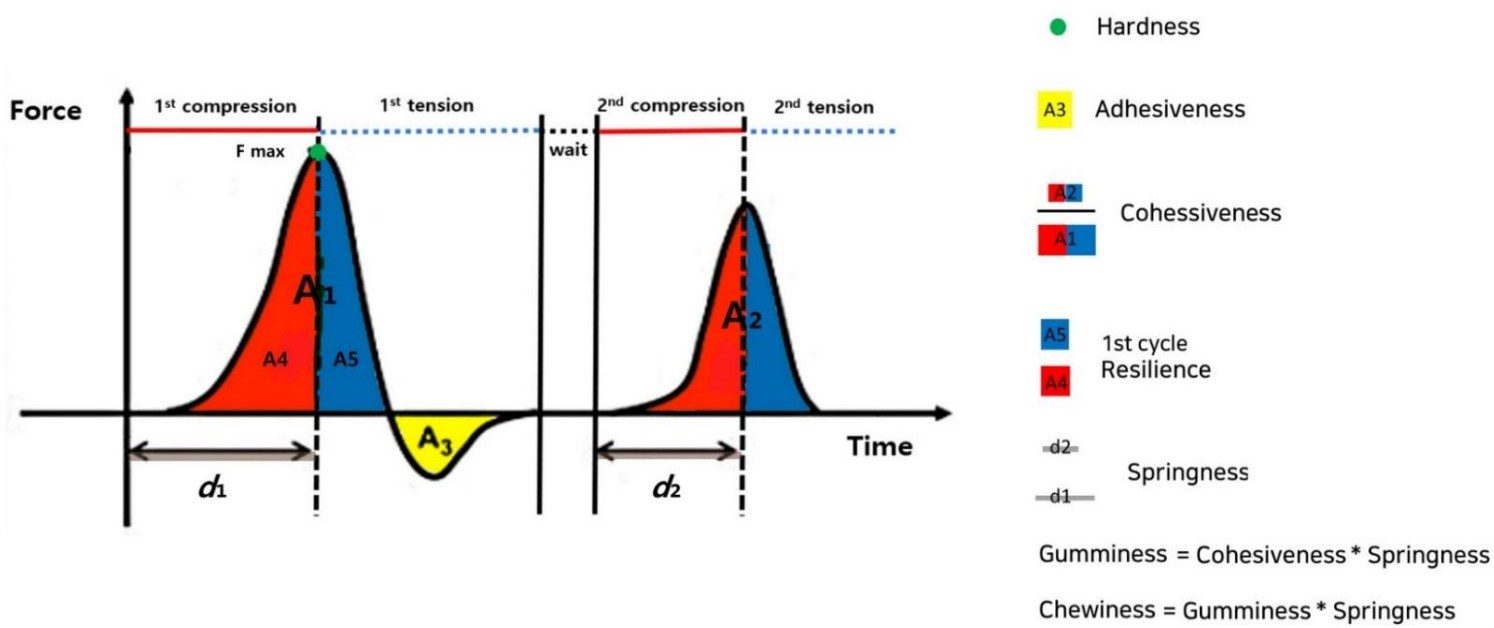


Figure. 1 Texture profile analysis (TPA) graph. The calculation method of texture parameters : Hardness (HRD) = Max force (N), Adhesiveness (ADH) = A3 (mJ), Cohesiveness (COH) = A2 / A1, Resilience (RES) = A5 / A4, Springiness (SPR) = d2 / d1, Gumminess (GUM) = COH * SPR, Chewiness (CWN) = GUM * SPR.

RESULTS

Variation in texture parameters of cooked rice

Texture of 248 rice GWAS accessions was evaluated using the TPA technique. As a result of PCA using 50,755 pruned variants to identify the structure of the GWAS accessions, it could be seen that most of them belong to the *japonica* group (Fig. 2a). Of the 248 GWAS accessions, 10 accessions were classed as *tropical japonica* and 230 as *temperate japonica*. In addition, its classification as five varieties of *indica*, two of *aus*, and one of *aromatic* was assessed. It is possible to examine the accession list, which includes improved varieties and landraces (Appendix. 1).

Associations between investigated phenotypes (HRD, ADH, COH, RES, SPR, CWN) were examined by correlation analysis, including AC, PC (Fig. 2b). The Pearson correlation coefficient was used to determine the correlation. According to the current research, a negative association was observed between AC and PC (r = −0.63, p = 2.26e−27). The strong positive correlation was found between CWN and ADH (r = 0.62, p = 4.27e−26).

HRD (Fig. 2c), ADH (Fig. 2d), COH (Fig. 2e), RES (Fig. 2f), SPR (Fig. 2g), and CWN (Fig. 2h) had patterns similar to the normal distribution. According to measurements of textural parameters

(Table 1), the mean values for HRD, ADH and COH were 728.03 (gf), 228.18 (mJ), and 0.21, respectively. The mean of RES was 0.04, SPR was 1.35, and CWN was 2222.38. These results showed that there are natural variations in cooked rice texture among GWAS accessions, and quantitative trait analysis was required to understand the basis of texture of cooked rice.

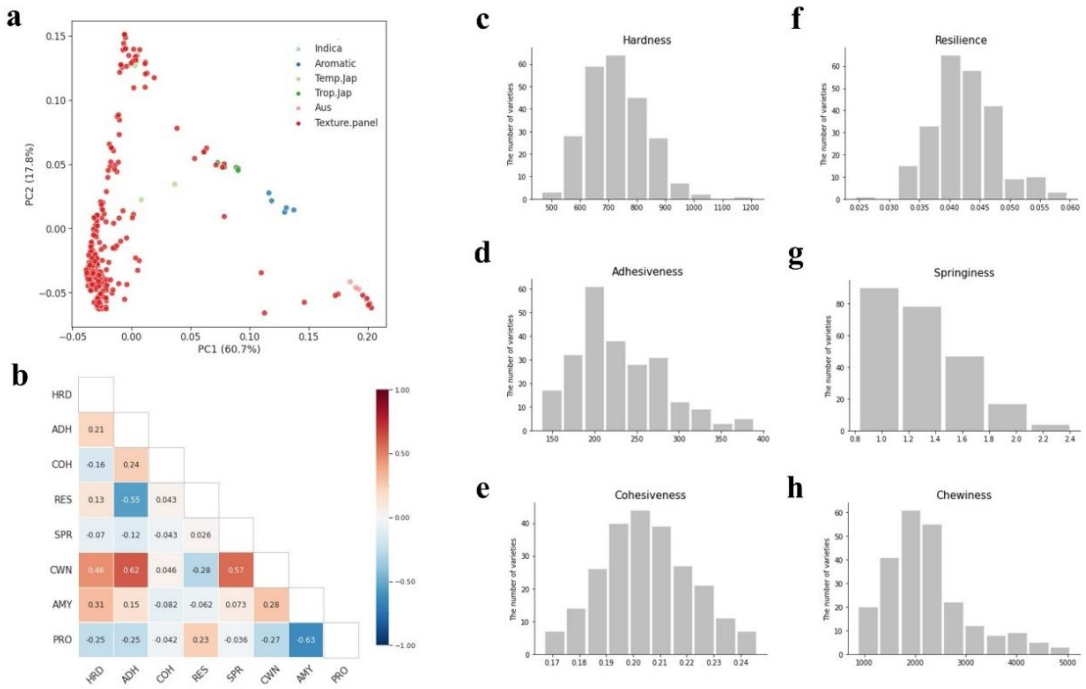


Figure. 2 Summary of variations in cooked rice texture parameters. (a) Principal component analysis (PCA) of GWAS accessions. (b) Correlation for texture parameters, amylose content (AC), and protein content (PC). Pearson's correlation coefficients were used to determine the correlation between traits. Positive and negative correlations were represented by red and blue, respectively. (c–h) Distributions for texture parameters (c) HRD, (d) ADH, (e) COH, (f) RES, (e) SPR, and (h) CWN.

Table 1. Texture variation among 248 GWAS accessions

Phenotype	Min	Max	Mean	SD	CV
HRD (gf)	463.3	1210.02	728.03	102.95	0.14
ADH (mJ)	136.32	389.54	228.18	51.9	0.23
COH	0.17	0.25	0.21	0.02	0.08
RES	0.02	0.06	0.04	0.005	0.12
SPR	0.82	2.41	1.35	0.34	0.25
CWN	859.52	5058.99	2222.38	827.74	0.37

Note: Min, Max, SD, CV represent minimum value, maximum value, standard deviation, and coefficient of variation, respectively.

GWAS of texture parameters

GWAS was carried out using Fast-LMM for marker trait associations in 248 germplasms using high-quality 1,254,682 SNPs. Association signals (p-values ≤ 1.572e-6) were identified in three in HRD (Table 1), nineteen in ADH (Table 2), and one in RES (Table 3). Three significant associations were detected for HRD and found for loci on chromosome 5 and 6 (Fig. 3a). The most significant association (chr06:14403140; p-values = 6.07e-08) in all texture phenotypes was detected for ADH on chromosome 6 (Fig. 3b). Significant SNPs in COH were detected in chromosome 5 (Appendix. 2a). There were no candidate genes that related to COH in the LD region containing the lead SNP. Additionally, no significant results were found in other parameters like SPR and CWN (Appendix. 2b, c).

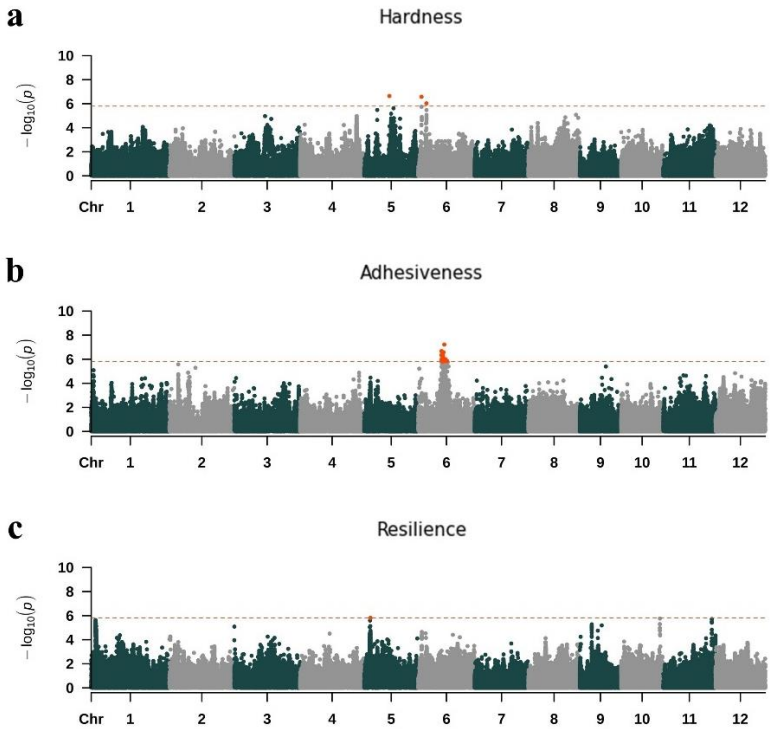


Figure. 3 Manhattan plots of genome wide association study (GWAS) for the texture parameters using FaST−LMM v2.07. For the Manhattan plot, the x−axis and the y−axis represents chromosome number and *p*−value, respectively. The genome wide significant threshold line is shown in orange. Red dots that have crossed the threshold line show SNPs strongly associated with texture. (a) HRD, (b) ADH, (c) RES.

Hardness (HRD)

In HRD, three SNPs revealed significant peaks. There was an association signal on chromosome 5 near 13.9 Mb, which was led by chr05:13917049 (Fig. 4a). *Os05g304900* and *MORE1E* were found within a 400kb region that included the lead SNP (chr05:13917049) (Table 2). A candidate gene, *Os05g304900*, in LD (r^2 > 0.69) was located from about 16 kb from the lead SNP (Fig. 4b). *Os05g304900* is like hydrolase, and hydrolyzing O−glycosyl compounds. *MORE1E* in LD (r^2 > 0.79) was identified about 15 kb from the lead SNP. *MORE1E* in is like hydrolase, and hydrolyzing O−glycosyl compounds. There was no significant difference in haplotypes of these two candidate genes associated to texture phenotypes.

We confirmed the significant differences in phenotypes using lead SNP. HRD was relatively low when the T allele at the Chr05:13917049 position on chromosome 5. HRD, on the other hand, was high when the C allele was present at the position (Fig. 4c). Among the used GWAS accessions, 226 had T allele, and 9 had C allele (Fig. 4d).

Association signals around 1.7 Mb on chromosome 6 were strongly associated with HRD, led by chr06:1772858 (Fig. 5a). *GBP1*, *Os06g0131500*, *OsBDG* and *Wx* were discovered in a 400kb region including the lead SNP (chr06:1772858) (Table 2). The LD block

containing the lead SNP was revealed to be approximately 4 kb in size (Fig. 5a), and *OsBDG* and *Wx* were found in that region. *OsBDG* was located within a region in low LD ($r^2 > 0.09$) with lead SNPs, and 3.6 kb apart from association signal (Table 2). *OsBDG* controls wax biosynthesis. *Wx* was in LD ($r^2 = 1$) and was around 0.7 kb from the lead SNP (Fig. 5b). *Wx* regulates granule-bound starch synthesis in endosperm (Wang et al., 1990; Sano, 1984). No significant difference was found in these two genes related to phenotype. We confirmed the significant differences in phenotypes using lead SNP. When Allele A was at the 1772858 positions on chromosome 6, HRD was relatively high; however, HRD was low when the G allele was represented (Fig. 5c). Between all resources we used, allele A was 14, and allele G was 229(Fig. 5d).

Strong association signals near 4 Mb on chromosome 6 exhibited phenotypic variations led by chr06:4218410 (Fig. 6a, b, c). *UGlcAE5*, *Os06g0192400*, and *OsUF3GT* were located within a 400 kb region including lead SNP (chr06:4218410). *UGlcAE5* is identical to nucleotide sugar epimerase-like protein (UDP-D-glucuronate 4- epimerase). *UGlcAE5* was found in a region in perfect LD ($r^2 = 1$) with the lead SNP, and about 11 kb away from the association peak (Fig. 6b). *Os06g0192400* is glycosyl transferase, and *OsUF3GT* is similar to UDP-glucose flavonoid 3-O-glucosyltransferase. The

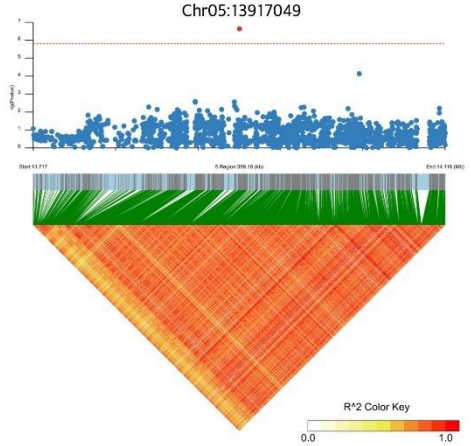
value of r^2 could not be determined due to the absence of variants in the region of *Os06g0192400* and *OsUF3GT*. There was no significance variance of HRD between these three candidate genes. Using the lead SNP, we confirmed the significant differences in phenotypes. When Allele G was present at the 4518410 positions on chromosome 6, HRD was comparatively lower than when it had C allele (Fig. 6c). 218 accessions had the G allele, while 11 contained the C allele (Fig. 6d).

Table. 2 SNPs and candidate genes associated with HRD by GWAS.

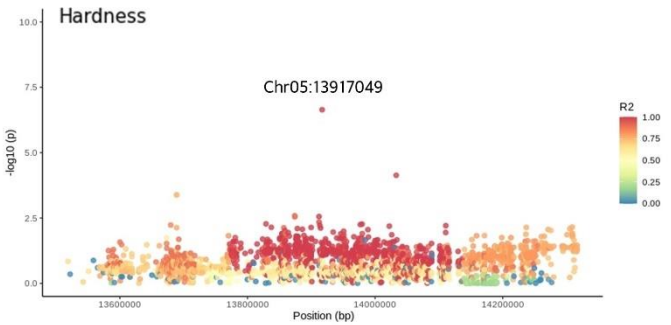
Chr	Lead SNP	*p*-value	Candidate gene	Gene position	Description
5	13917049	2.28E-07	*Os05g0304900*	5:13755590..13758581	Similar to hydrolase, hydrolyzing O-glycosyl compounds
			MORE1E(Os05g0305000)	5:13762103..13763364	Similar to hydrolase, hydrolyzing O-glycosyl compounds
6	1772858	2.65E-07	*GBP1(Os06g0130600)*	6:1636572..1640852	GAGA-binding transcription factor 1, Negative regulation of seedling growth and grain length development, Suppressor of flowering
			Os06g0131500	6:1691885..1694051	Glycoside hydrolase, family 17 protein
					Similar to glucan endo-1,3-beta-glucosidase 7
			OsBDG(Os06g0132500)	6:1736723..1742581	Control of wax biosynthesis
			Wx(Os06g0133000)	6:1765622..1770574	Granule-bound starch synthase, Synthesis of amylose in endosperm
6	4518410	9.37E-07	*UGlcAE5(Os06g0187200)*	6:4404127..4405602	Similar to Nucleotide sugar epimerase-like protein (UDP-D-glucuronate 4- epimerase)
			Os06g0192400	6:4653460..4655652	Glycosyl transferase, family 31 protein
			OsUF3GT(Os06g0192100)	6:4648131..4649391	Similar to UDP-glucose flavonoid-3-O-glucosyltransferase

Note: Information of candidate genes are based on IRGSP v1.0

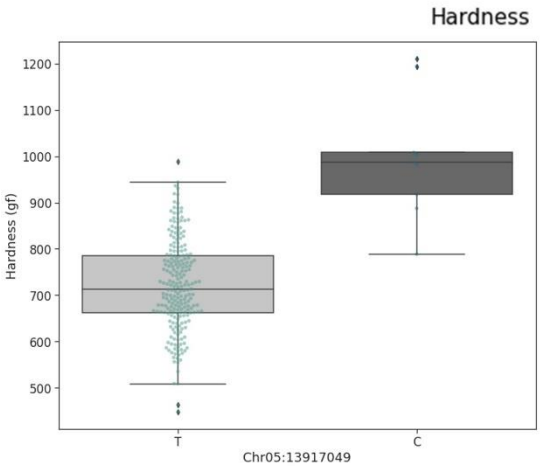
a



b



c

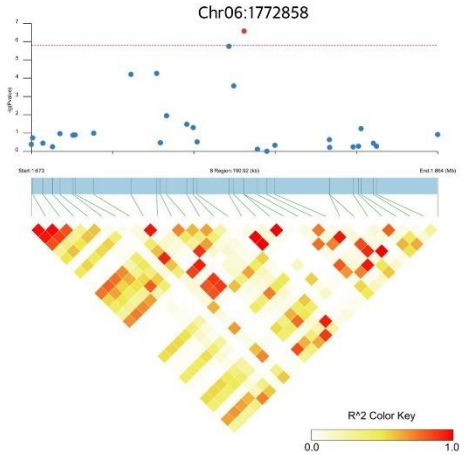


d

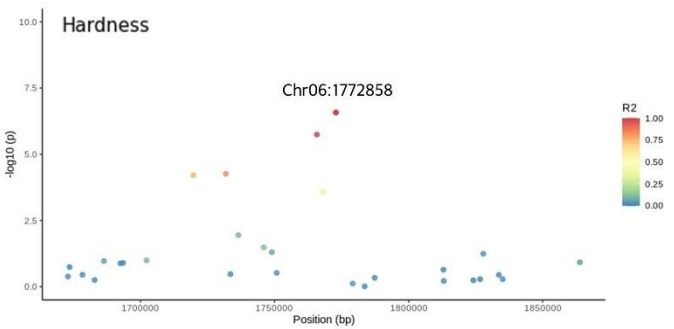
Position	Chr05:13917049	Number of Accessions
Allele T	T	226
Allele C	C	9

Figure. 4 Association results related to lead SNP (Chr05:13917049) for HRD. (a–b) Association region on chromosome 5 near the lead SNP for HRD (a) Local Manhattan plot (top) and LD heat map (bottom). Linkage block analysis of loci 13917049 on chromosome 5 using the LDBlockshow. White and red represents absent and perfect LD, respectively (r^2 = 0 to 1) (b) The zoomed locus plot of the GWAS peak region containing the lead SNP for HRD. Each dot represents a variant. SNPs of strong LD ($r^2 > 0.8$) with the lead SNP denote red color. (c) Box and swarm plots showing differences in HRD caused by Chr06:13917049 alleles ($p \leq 0.01$), and (d) Two types of alleles derived from the lead SNP.

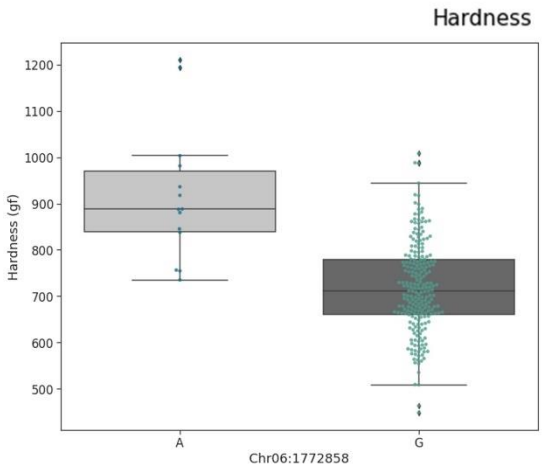
a



b



c

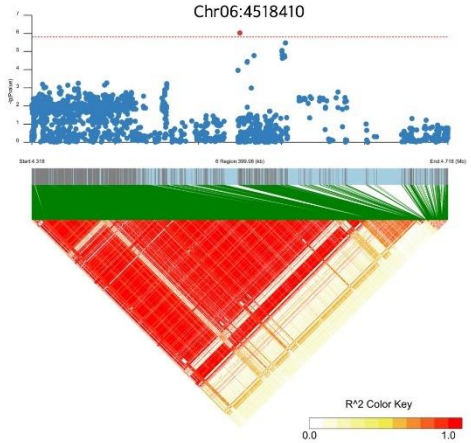


d

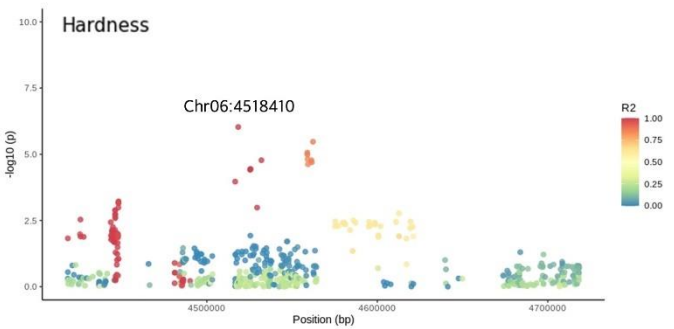
Position	Chr06:1772858	Number of Accessions
Allele A	A	14
Allele G	G	229

Figure. 5 Association results related to lead SNP(Chr06:1772858) for HRD. (a−b) Association region on chromosome 6 near the lead SNP for HRD (a) Local Manhattan plot (top) and LD heat map (bottom). Linkage block analysis of loci 1772858 on chromosome 6 using the LDBlockshow. White and red represents absent and perfect LD, respectively (r^2 = 0 to 1) (b) The zoomed locus plot of the GWAS peak region containing the lead SNP for HRD. Each dot represents a variant. SNPs of strong LD ($r^2 > 0.8$) with the lead SNP denote red color. (c) Box and swarm plots showing differences in HRD caused by Chr06:1772858 alleles ($p \leq 0.01$), and (d) Two types of alleles derived from the lead SNP.

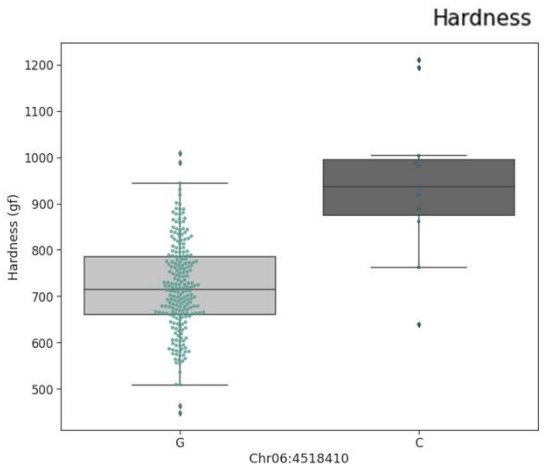
a



b



c



d

Position	Chr06:4518410	Number of Accessions
Allele G	G	218
Allele C	C	11

Figure. 6 Association results related to lead SNP(Chr06:4518410) for HRD. (a–b) Association region on chromosome 6 near the lead SNP for HRD (a) Local Manhattan plot (top) and LD heat map (bottom). Linkage block analysis of loci 4518410 on chromosome 6 using the LDBlockshow. White and red represents absent and perfect LD, respectively (r^2 = 0 to 1) (b) The zoomed locus plot of the GWAS peak region containing the lead SNP for HRD. Each dot represents a variant. SNPs of strong LD ($r^2 > 0.8$) with the lead SNP denote red color. (c) Box and swarm plots showing differences in HRD caused by Chr06:4518410 alleles ($p \leq 0.01$), and (d) Two types of alleles derived from the lead SNP.

Adhesiveness (ADH)

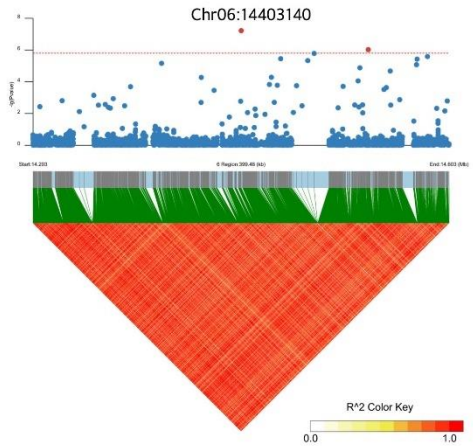
The association signal (Chr06:14403140; p−values = 6.07e−08) for ADH showed near 14.4 Mb on chromosome 6 (Fig. 7a, b). *Os06g0356800* were located within a 700 kb region including lead SNP (Chr06:14403140) (Table 3). *Os06g0356800* is like xylanase inhibitor protein I precursor. It was discovered in a region in low LD (r^2 < 0.01) with the lead SNP, and around 24 kb away from the association signals (Fig. 7b). In 700 kb region, there were eleven SNP variants related to the lead SNP, and the value of r^2 was over than 0.85. No significant phenotypic difference was observed in candidate gene. To use the lead SNP, we identified the significant differences in phenotype. G allele at the 14403140 on chromosome 6, ADH was relatively lower than C allele (Fig. 7c). There were 218 accessions of the G allele and 11 of the C allele (Fig. 7d).

Table. 3 SNPs and candidate genes with ADH by GWAS.

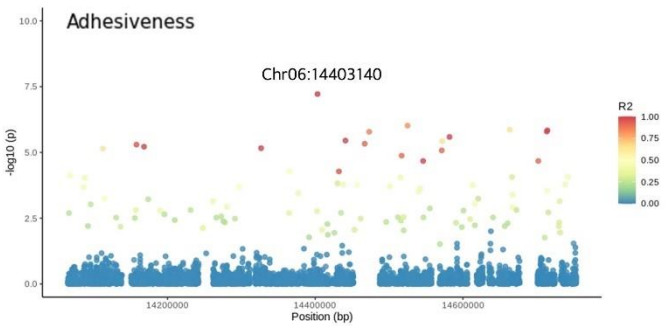
Chr	Lead SNP	*p*-value	Candidate gene	Gene position	Description
6	14403140	6.07E-08	*Os06g0356800*	6:14646987..14648089	Similar to Xylanase inhibitor protein I precursor.

Note: Information of candidate genes are based on IRGSP v1.0

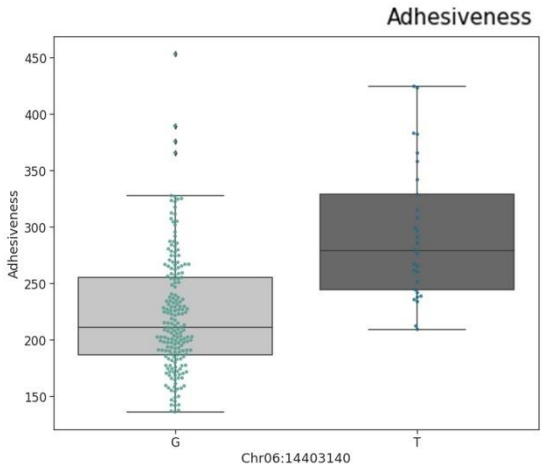
a



b



c



d

Position	Chr06:14403140	Number of Accessions
Allele G	G	218
Allele C	C	11

Figure. 7 Association results related to lead SNP(Chr06:14403140) for ADH. (a−b) Association region on chromosome 6 near the lead SNP for ADH (a) Local Manhattan plot (top) and LD heat map (bottom). Linkage block analysis of loci 14403140 on chromosome 6 using the LDBlockshow. White and red represents absent and perfect LD, respectively (r^2 = 0 to 1) (b) The zoomed locus plot of the GWAS peak region containing the lead SNP for ADH. Each dot represents a variant. SNPs of strong LD ($r^2 > 0.8$) with the lead SNP denote red color. (c) Box and swarm plots showing differences in ADH caused by Chr06:14403140 alleles ($p \leq 0.01$), and (d) Two types of alleles derived from the lead SNP.

Resilience (RES)

The signal (chr05:3455897) strongly related with RES found on near 3 Mb on chromosome 5 (Figure 8a, b). *Chalk5*, *GS5*, *DARX1*, and *OsLTPd8* were located within a 400 kb region with lead SNP (chr05:3455897) (Table 4). *Chalk5* regulates grain chalkiness and placed about 12 kb from the lead SNP. The value of r^2 could not be determined because of the absence of variants in the region of *Chalk5*. *GS5* is positive regulator of grain size and located approximately 1.6 kb from the lead SNP. The r^2 value was 0.18 in LD pairwise with the lead SNP (Fig. 8b). *DARX1* is modulation of secondary wall formation. *DARX1* had perfect LD (r^2 = 1) and located apart 4.3 kb from the lead SNP (Fig. 8b). *OsLTPd8* regulates seed storage domain containing protein. *OsLTPd8* was in weak LD (r^2 < 0.01) including the lead SNP and placed on 3.53 Mb to 3.54 Mb (Fig. 8b). No significant phenotypic differences were found in *Chalk5*, *DARX1*, *OsLTPd8*. We confirmed the significant differences in RES using the lead SNP. When Allele C was at the Chr05:3455897 positions on chromosome 5, RES was high, but RES was low when it had the A allele (Fig. 8c). Allele C was 178, and Allele A was 53 between all resources we used (Fig. 8d).

In *GS5*, a splice region variant (AA, CA) was found at position Chr05:3442417 (Fig. 9b, c). The number of accessions with ‘Allele

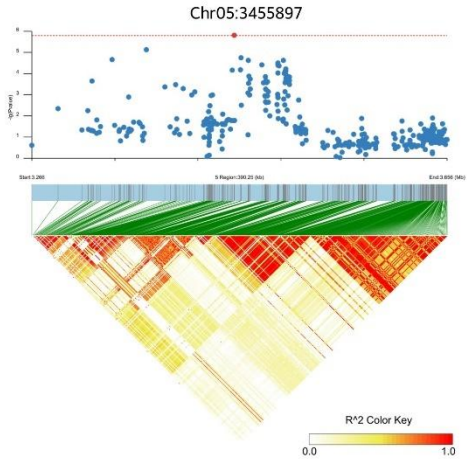
1' was 78, and RES was found to be significantly lower than that of 'Allele 2' (Fig 9a, c).

Table. 4 SNPs and candidate genes with RES by GWAS.

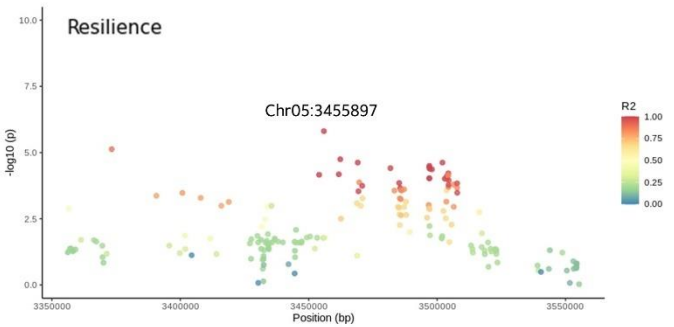
Chr	Lead SNP	*p*-value	Candidate gene	Gene position	Description
5	3455897	1.55E-06	*Chalk5(Os05g0156900)*	5:3335380..3339817	Vacuolar H+-translocating pyrophosphatase, Regulation of grain chalkiness
			GS5(Os05g0158500)	5:3439304..3443769	Serine carboxypeptidase, Positive regulator of grain size
			DARX1(Os05g0159300)	5:3499107..3507407	Arabinosyl deacetylase, GDSL esterase, Modulation of secondary wall formation
			OsLTPd8(Os05g0160300)	5:3539620..3540966	Bifunctional inhibitor/plant lipid transfer protein/seed storage domain containing protein

Note: Information of candidate genes are based on IRGSP v1.0

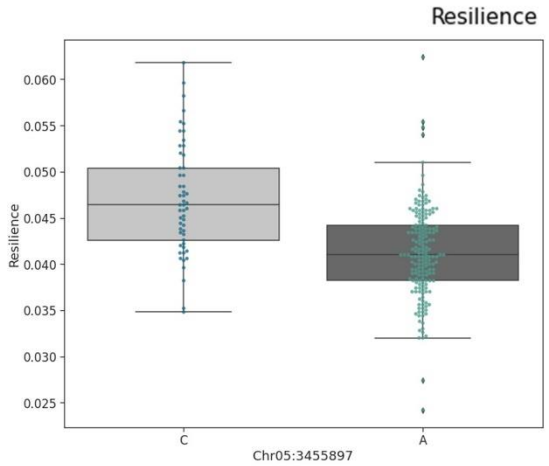
a



b



c

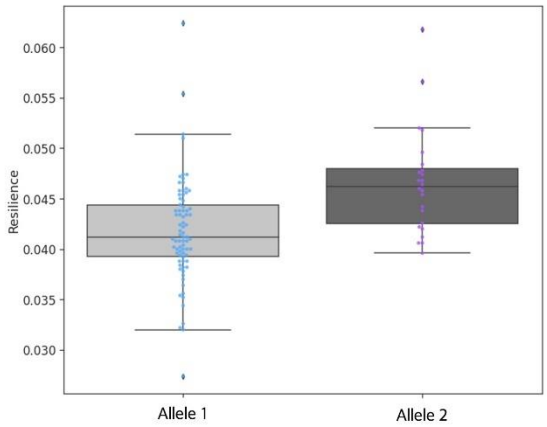


d

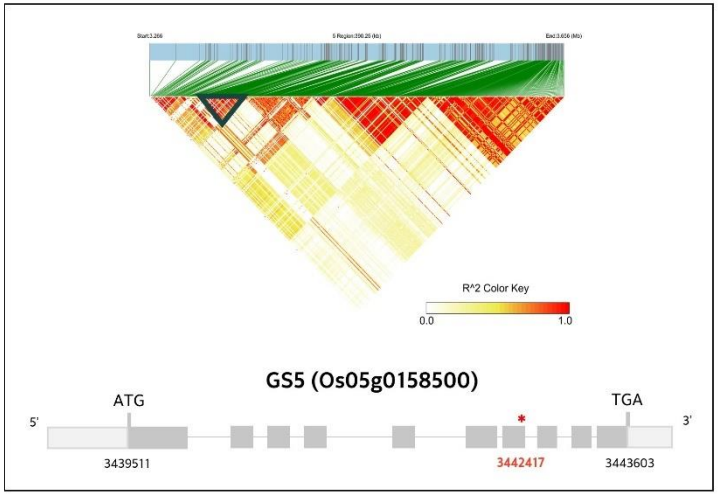
Position	Chr05:3455897	Number of Accessions
Allele C	C	178
Allele A	A	53

Figure. 8 Association results related to lead SNP(Chr05:3455897) for RES. (a−b) Association region on chromosome 5 near the lead SNP for RES (a) Local Manhattan plot (top) and LD heat map (bottom). Linkage block analysis of loci 3455897 on chromosome 5 using the LDBlockshow. White and red represented absent and perfect LD, respectively (r^2 = 0 to 1) (b) The zoomed locus plot of the GWAS peak region containing the lead SNP for RES. Each dot represented a variant. SNPs of strong LD (r^2 > 0.8) with the lead SNP denoted red color. (c) Box and swarm plots showing differences in RES caused by Chr05:3455897 alleles ($p \leq 0.01$), and (d) Two types of alleles derived from the lead SNP.

a



b



c

Splice region variant

Position	3442093	3442414	3442415	3442416	3442417	3442782	3442785	3442909	3443283	3443491	3443568	3443576	3443653	3443824	Number of Accessions
Allele 1	GC	C	AAC	AC	AA	TGG	T	GA	GA	TCGCCGC	TCGCCGCCTCGC	T	TGCA	G	78
Allele 2	GC	C	AAC	AC	CA	TGG	T	GA	GA	TCGCCGC	TCGCCGCCTCGC	T	TGCA	G	24

Figure. 9 About the candidate gene (*GS5*) for RES located near the lead SNP(Chr05:3455897). (a) The phenotypic variations were compared between allele one and allele two detected in accessions ($p \leq 0.01$), (b) LD heatmap of the lead SNP for RES (top) and *GS5* gene structure (bottom). The green triangle represented the LD region, including *GS5*. An asterisk (*) denoted the *GS5* splice region (Chr05:3442417). (c) Two difference types of *GS5* allele in Chr05:3442417. The position marked in red indicated a splice region variant.

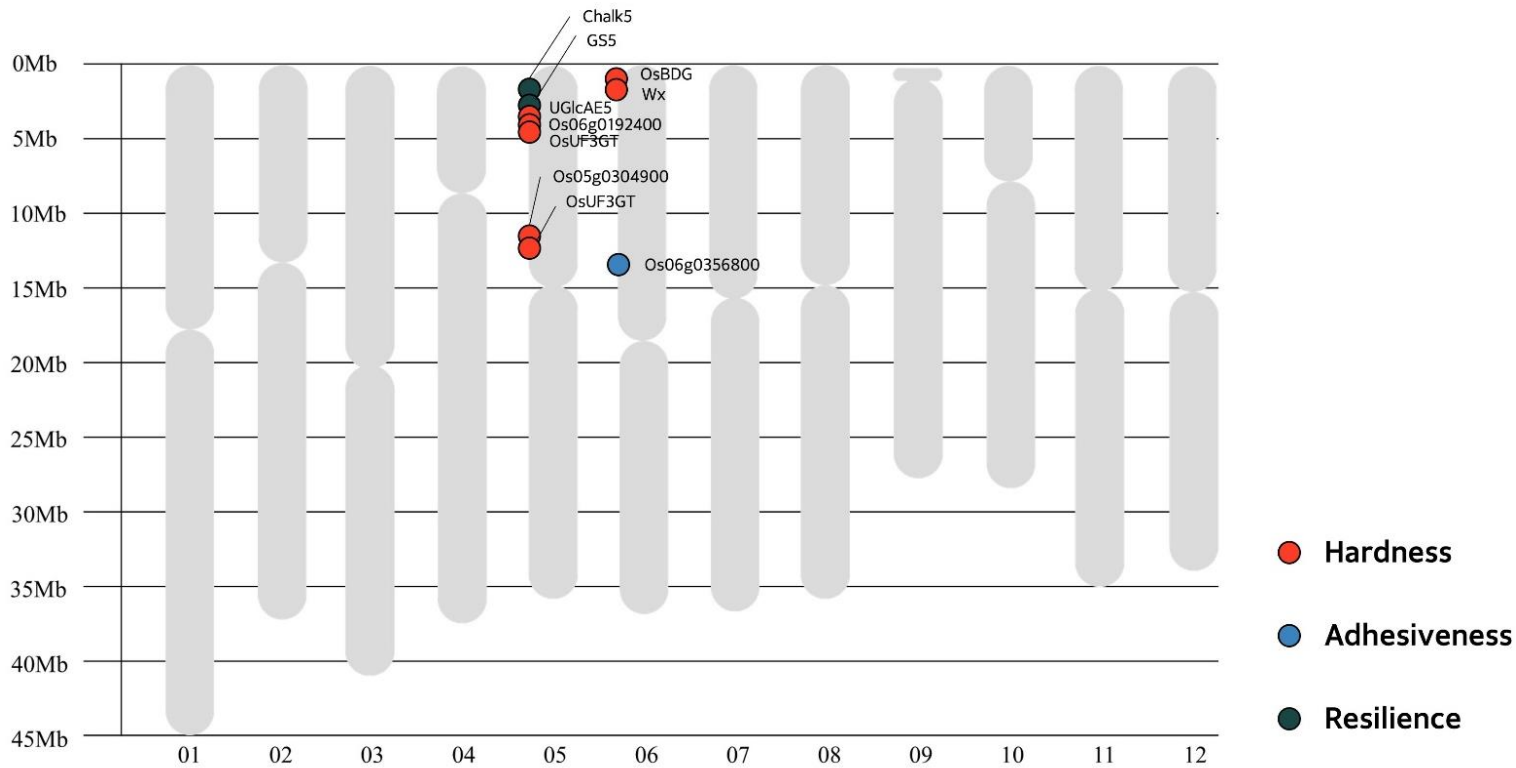


Figure. 10 Physical map of candidate genes with strong associations in 400kb region, including the lead SNPs for texture parameters. The scale on the left represents the mega base pair (Mb) distance. Orange, blue, green circles indicated the candidate genes in LD region for HRD, for ADH, for RES, respectively.

DISCUSSION

The texture of cooked rice has a significant impact on consumers' preferences. High-quality rice can be distinguished by including textural characteristics, which is a crucial step in fine-tuning products in order to capture the rice market (Anacleto et al., 2015).

Evaluation of cooked rice texture was a commonly indirect method. Texture was not evaluated directly, but it was predicted using factors that could affect texture. In the case of amylose composed over 90% in milled rice grain, it has been a significant factor considered for determining the grain quality of rice in breeding for a long time and is known to affect texture in various ways (Li & Gilbert, 2018; Juliano, 1984; Juliano et al., 1981; Kumar, Upadhyay, & Bhattacharya, 1976; Lorenz, Fong, Mossman, & Saunders, 1978).

Researchers were constantly developing strategies to measure and predict reliably the cooked rice texture to improve the quality of rice. Cooked rice texture could be assessed by TPA method and descriptive sensory method. TPA is a semi-throughput method for measuring the physical reaction during double compression and tension cycles. In comparison to the sensory test, TPA could be applied constantly, did not require sensory panelists training, and

only needed a small number of samples to investigate (Misra et al., 2018).

248 GWAS accessions was used to obtain phenotypic data using TPA for diverse texture parameters of cooked rice (Table 1, Appendix 1). Five replicates were conducted for each variety during the TPA process. The mean values for HRD, ADH and COH were 728.03 (gf), 228.18 (mJ), and 0.21, respectively. The mean of RES was 0.04, SPR was 1.35, and CWN was 2222.38. AC and PC displayed a significant negative correlation (Fig 2b). In the findings of previous investigations (Bian et al., 2020), when the AC and PC were high in grain, texture was hardened and palatability was decreased. Our result was different because the panels we used for texture analysis were different, and the factors involved in cooked rice texture were diverse.

We conducted GWAS for identifying the genetic backgrounds related to cooked rice texture. In this study, three lead SNPs, chr05:13917049 (Fig. 4), Chr06:1772858 (Fig. 5), and Chr06:4518410 (Fig. 6) were associated with HRD. Chr06:14403140 was the lead SNP strongly associated with ADH (Fig. 7). The lead SNP associated with RES was chr05:3455897 (Fig. 8). The lead SNPs that detected in study, were not positioned on genes exactly. We confirmed the significant differences in phenotypes using the lead

SNPs. HRD was relatively low when the T allele was at the Chr05:13917049 position on chromosome 5 (Fig. 4c). When Allele A was at the 1772858 positions on chromosome 6, HRD was relatively high; however, HRD was low when the G allele was represented (Fig. 5c). When Allele G was identified at the 4518410 positions on chromosome 6, HRD was significantly lower than when C allele was present (Fig. 6c). For ADH, G allele at the 14403140 on chromosome 6, ADH was relatively lower than C allele (Fig. 7c). Allele C was present at Chr05:3455897 on chromosome 5, RES was significantly high, whereas RES was low when allele A was present (Fig. 8c).

According to the manufacturer of the texture analyser, HRD was the maximum force (gf) of the first TPA cycle curve, ADH referred to a negative area (mJ) of the first curve, and RES was the extent to which a food struggles to recover its original shape and size.

We identified candidate genes related to texture parameters (Table 2, 3, 4). Association signals around 1.7 Mb on chromosome 6 were strongly associated with HRD, led by chr06:1772858 (Fig. 5a). *Wx* was in LD ($r^2 = 1$) and was around 0.7 kb from the lead SNP (Fig. 5b). *Wx* controls granule-bound starch synthesis in endosperm (Wang et al., 1990; Sano, 1984). No significant phenotypic differences were observed in *Wx* gene.

Os06g0356800 around 24 kb away from the association

signals (Chr06:14403140) for ADH. *Os06g0356800* is like xylanase inhibitor protein I precursor. Xylanases can also be discovered in plants, insects, snails, crabs, marine algae, and protozoa (Dekker & Richards, 1976). Cereals generate xylanases for re-modeling and growth of cereal cell walls during normal cell growth and for a more severe destruction of cell walls during seed germination (Dornez, 2009). There were no phenotypic changes in *Os06g0356800*.

GS5 is positive regulator of grain size (Li et al. 2011) and located approximately 1.6 kb from the lead SNP (chr05:3455897) for RES. *GS5* gene, encoding a putative serine carboxypeptidase, was fine mapped from a QTL responsible for grain size (Lee, 2015; Li et al. 2011). A splice variant was discovered in *GS5* at Chr05:3442417 (Fig. 9b, c) in our study. In a previous study, natural variations in the promoter region distinguished three alleles of the *GS5* gene. RES distribution of accessions with 'Allele 1' was significantly lower compared to that of accessions with 'Allele 2' (Fig 9a, c). Low-RES foods include ‘mushy’ foods such as mashed carrots, papaya slices, but harder food such as boiled potatoes, and thin ham slices were more resilient (Wee, 2018). It was the same as the meaning that the number of chews taken per bite and RES of food were positively correlated. In current study, it is difficult to ascertain the functional relationship between RES and *GS5*. In addition, we suggested the

need to study further the relationship between rice grain shape and cooked rice texture.

In conclusion, the results of this study included various phenotypic information on rice texture in 248 rice accessions. We suggested the primary data of textural parameters. Lead SNPs associated with cooked rice texture and information of candidate genes located within the range of LD, could be used to the markers and indicators for improving the texture in the breeding process.

REFERENCE

Anacleto, R., Badoni, S., Parween, S., Butardo Jr, V. M., Misra, G., Cuevas, R. P., Kuhlmann, M., Trinidad, T. P., Mallillin, A. C., & Acuin, C. (2019). Integrating a genome-wide association study with a large-scale transcriptome analysis to predict genetic regions influencing the glycaemic index and texture in rice. *Plant biotechnology journal*, *17*(7), 1261-1275.

Anacleto, R., Cuevas, R. P., Jimenez, R., Llorente, C., Nissila, E., Henry, R., & Sreenivasulu, N. (2015). Prospects of breeding high-quality rice using post-genomic tools. *Theoretical and Applied Genetics*, *128*, 1449-1466.

Bao, J., Shen, S., Sun, M., & Corke, H. (2006). Analysis of genotypic diversity in the starch physicochemical properties of nonwaxy rice: apparent amylose content, pasting viscosity and gel texture. *Starch-Stärke*, *58*(6), 259-267.

Bhattacharya, K., & Juliano, B. (1985). Rice: Chemistry and technology. *AACC, St Paul, MN*, *289*.

Bian, J.-l., Ren, G.-l., Chao, H., Xu, F.-f., Shi, Q., Tang, J.-h., Zhang, H.-c., Wei, H.-y., & Hui, G. (2020). Comparative analysis on grain quality and yield of different panicle weight indica-japonica hybrid rice (Oryza sativa L.) cultivars. *Journal of Integrative Agriculture*, *19*(4), 999-1009.

Cagampang, G. B., Perez, C. M., & Juliano, B. O. (1973). A gel consistency test for eating quality of rice. *Journal of the Science of Food and Agriculture*, *24*(12), 1589-1594.

Calingacion, M., Laborte, A., Nelson, A., Resurreccion, A., Concepcion, J. C., Daygon, V. D., Mumm, R., Reinke, R., Dipti, S., & Bassinello, P. Z. (2014). Diversity of global rice markets and the science required for consumer-targeted rice breeding. *PloS one*, *9*(1), e85106.

Champagne, E. T., Bett-Garber, K. L., Fitzgerald, M. A., Grimm, C. C., Lea, J., Ohtsubo, K. i., Jongdee, S., Xie, L., Bassinello, P. Z., & Resurreccion, A. (2010). Important sensory properties

differentiating premium rice varieties. *Rice*, *3*(4), 270-281.

Champagne, E. T., Lyon, B. G., Min, B. K., Vinyard, B. T., Bett, K. L., Barton, F. E., Webb, B. D., McClung, A. M., Moldenhauer, K. A., & Linscombe, S. (1998). Effects of postharvest processing on texture profile analysis of cooked rice. *Cereal Chemistry*, *75*(2), 181-186.

Chen, M. H., Bergman, C., Pinson, S., & Fjellstrom, R. (2008). Waxy gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. Journal of cereal science, 47(3), 536-545.

Cho, Y. G., Kang, H. J., Lee, Y. T., Jong, S. K., Eun, M. Y., & McCouch, S. R. (2010). Identification of quantitative trait loci for physical and chemical properties of rice grain. Plant Biotechnology Reports, 4, 61-73.

Crowhurst, D. G., & Creed, P. G. (2001). Effect of cooking method and variety on the sensory quality of rice. *Food Service Technology*, *1*(3), 133-140.

Dekker, R. F., & Richards, G. N. (1976). Hemicellulases: their occurrence, purification, properties, and mode of action. Advances in carbohydrate chemistry and biochemistry, 32, 277-352.

Dornez, E., Gebruers, K., Delcour, J. A., & Courtin, C. M. (2009). Grain-associated xylanases: occurrence, variability, and implications for cereal processing. Trends in food science & technology, 20(11-12), 495-510.

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., & Yang, T.-L. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Briefings in Bioinformatics, 22(4).

Hori, K., Suzuki, K., Iijima, K., & Ebana, K. (2016). Variation in cooking and eating quality traits in Japanese rice germplasm accessions. *Breeding Science*, *66*(2), 309-318.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo,

Y., Deng, L., & Zhu, C. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature genetics*, *44*(1), 32-39.

Juliano, B. O., Perez, C. M., Blakeney, A. B., Castillo, T., Kongseree, N., Laignelet, B., ... & Webb, B. D. (1981). International cooperative testing on the amylose content of milled rice. Starch? St rke, 33(5), 157-162.

Juliano, B. O., Perez, C. M., Alyo-Shin, E. P., Romanov, V. B., BLAKENEY, A. B., WELSH, L. A., ... & KIMURA, H. (1984). International cooperative test on texture of cooked rice. Journal of Texture Studies, 15(4), 357-376.

Kumar, B. M., UPADHYAY, J. K., & Bhattacharya, K. R. (1976). Objective tests for the stickiness of cooked rice. Journal of Texture Studies, 7(2), 271-278.

Larkin, P. D., & Park, W. D. (2003). Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (Oryza sativa L.). *Molecular Breeding*, *12*, 335-339.

Lee, Y., Choi, M.-S., Lee, G., Jang, S., Yoon, M.-R., Kim, B., Piao, R., Woo, M.-O., Chin, J. H., & Koh, H.-J. (2017). Sugary endosperm is modulated by starch branching enzyme IIa in rice (Oryza sativa L.). *Rice*, *10*(1), 1-13.

Li, H. (2017). Understanding the texture of cooked rice from the molecular, instrumental and sensory levels.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.

Li, H., & Gilbert, R. G. (2018). Starch molecular structure: The basis for an improved understanding of cooked rice texture. *Carbohydrate Polymers*, *195*, 9-17.

Li, P., Chen, Y.-H., Lu, J., Zhang, C.-Q., Liu, Q.-Q., & Li, Q.-F. (2022b). Genes and their molecular functions determining seed structure, components, and quality of rice. *Rice*, *15*(1), 1-27.

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., ... & Zhang, Q. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nature genetics, 43(12), 1266-1269.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature methods*, *8*(10), 833-835.

Little, R. R. (1958). Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem.*, *35*, 111-126.

Lorenz, K., RY, F., AP, M., & RM, S. (1978). Long, medium, and short grain rices-enzyme activities and chemical and physical properties.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.

Mestres, C., Ribeyre, F., Pons, B., Fallet, V., & Matencio, F. (2011). Sensory texture of cooked rice is rather linked to chemical than to physical characteristics of raw grain. Journal of Cereal Science, 53(1), 81-89.

Misra, G., Badoni, S., Domingo, C. J., Cuevas, R. P. O., Llorente, C., Mbanjo, E. G. N., & Sreenivasulu, N. (2018). Deciphering the genetic architecture of cooked rice texture. *Frontiers in Plant Science*, *9*, 1405.

NAITO, S., & OGAWA, T. (1998). Tensipresser precision in measuring cooked rice adhesiveness. Journal of texture studies, 29(3), 325-335.

Ong, M., & Blanshard, J. (1995). Texture determinants in cooked, parboiled rice. I: Rice starch amylose and the fine stucture of amylopectin. *Journal of Cereal Science*, *21*(3), 251-260.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M.

J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575.

Ramesh, M., Bhattacharya, K. R., & Mitchell, J. R. (2000). Developments in understanding the basis of cooked-rice texture. Critical Reviews in Food Science and Nutrition, 40(6), 449-460.

Rani, M. S., & Bhattachrya, K. (1995). Microscop of rice starch granules during cooking. *Starch-Stärke*, *47*(9), 334-337.

Rosenthal, A. J. (2010). Texture profile analysis-how important are the parameters? *Journal of texture studies*, *41*(5), 672-684.

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C. C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., & Itoh, T. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol*, *54*(2), e6.

Sano, Y. (1984). Differential regulation of waxy gene expression in rice endosperm. Theoretical and applied genetics, 68, 467-473.

Stokes, J. R., Boehm, M. W., & Baier, S. K. (2013). Oral processing, texture and mouthfeel: From rheology to tribology and beyond. *Current Opinion in Colloid & Interface Science*, *18*(4), 349-359.

Suwannaporn, P., & Linnemann, A. (2008). Rice-eating quality among consumers in different rice grain preference countries. *Journal of Sensory Studies*, *23*(1), 1-13.

Wang, Z. Y., Wu, Z. L., Xing, Y. Y., Zheng, F. G., Guo, X. L., Zhang, W. G., & Hong, M. M. (1990). Nucleotide sequence of rice waxy gene. Nucleic Acids Research, 18(19), 5898.

Wee, M. S. M., Goh, A. T., Stieger, M., & Forde, C. G. (2018). Correlation of instrumental texture properties from textural profile analysis (TPA) with eating behaviours and macronutrient composition for a wide range of solid foods.

Food & function, *9*(10), 5301-5312.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., & Li, X. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, proteomics & bioinformatics*, *19*(4), 619-628.

APPENDIX

Appendix 1. List of GWAS accessions and their subpopulation used in this study.

ID	Variety	Sub-population	ID	Variety	Sub-population
DB	도봉	Temp.Jap	J112	강찬	Temp.Jap
Giho	기호	Temp.Jap	J113	친농	Temp.Jap
Hwacheong	화청	Temp.Jap	J116	고아미	Temp.Jap
Hwaseong	화성	Temp.Jap	J118	고운	Temp.Jap
Ilpoom	일품	Temp.Jap	J12	산벼	Trop.Jap
J1	Mojodo	Indica	J120	해찬물결	Temp.Jap
J10	재래륙도	Trop.Jap	J121	해오르미	Temp.Jap
J101	칠보	Temp.Jap	J123	하남	Temp.Jap
J103	대찬	Temp.Jap	J125	한들	Aus
J104	대립벼 1	Temp.Jap	J126	한마음	Temp.Jap
J105	다미	Temp.Jap	J127	미품	Temp.Jap
J107	드래찬	Temp.Jap	J13	산두도	Trop.Jap
J108	동보	Temp.Jap	J139	황금노들	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J14	산도	Temp.Jap	J18	육월조	Temp.Jap
J141	황금누리	Temp.Jap	J180	녹양	Trop.Jap
J15	밭나락	Trop.Jap	J183	온누리	Temp.Jap
J152	진보	Temp.Jap	J184	풍미	Temp.Jap
J153	진부	Temp.Jap	J186	새계화	Temp.Jap
J155	진미	Temp.Jap	J188	삼광	Temp.Jap
J156	진수미	Temp.Jap	J19	올뭇게	Temp.Jap
J158	조안	Temp.Jap	J192	서안 1	Temp.Jap
J16	재래조도	Temp.Jap	J193	서간	Temp.Jap
J164	조운	Temp.Jap	J199	신동진	Temp.Jap
J166	주남	Temp.Jap	J20	보리벼	Temp.Jap
J167	주남조생	Temp.Jap	J202	신운봉 1	Temp.Jap
J171	말그미	Temp.Jap	J203	소비	Temp.Jap
J174	미광	Temp.Jap	J204	수안	Temp.Jap
J177	낙동	Temp.Jap	J208	태성	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J21	백지청벼	Temp.Jap	J231	무다래기	Temp.Jap
J211	운미	Temp.Jap	J232	무산도	Temp.Jap
J214	영호진미	Temp.Jap	J233	무안도	Temp.Jap
J218	구중도 99	Temp.Jap	J234	무주도	Temp.Jap
J219	나도	Temp.Jap	J235	미도	Temp.Jap
J22	마향조도	Temp.Jap	J236	미조	Temp.Jap
J221	남조	Temp.Jap	J237	백장군 22	Temp.Jap
J222	느스벼	Temp.Jap	J238	사두초	Indica
J224	당도	Temp.Jap	J239	산다다기도	Temp.Jap
J226	마마콩	Temp.Jap	J24	두도	Temp.Jap
J227	맥도	Temp.Jap	J243	쌍봉	Temp.Jap
J228	맥조	Temp.Jap	J244	쌍두조	Temp.Jap
J229	모도 52	Temp.Jap	J245	서간도도	Temp.Jap
J23	땅벼	Temp.Jap	J246	석산조	Temp.Jap
J230	모조	Temp.Jap	J247	소두조	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J248	쇠머리벼	Temp.Jap	J263	와방	Temp.Jap
J249	쇠머리지장	Temp.Jap	J264	외국벼	Temp.Jap
J25	도립	Indica	J265	왜조	Temp.Jap
J251	쇠벤치기	Temp.Jap	J270	정기조생	Temp.Jap
J252	수중조	Temp.Jap	J271	정조	Temp.Jap
J254	여벼	Temp.Jap	J272	정종화	Temp.Jap
J255	여수벼	Temp.Jap	J273	조나조	Temp.Jap
J256	열술벼	Temp.Jap	J274	조두조	Temp.Jap
J257	예조	Temp.Jap	J275	조선도	Indica
J258	오리도	Temp.Jap	J276	조타조	Temp.Jap
J259	오백조	Temp.Jap	J277	종조백조	Temp.Jap
J26	대골벼	Temp.Jap	J278	쫄장벼	Temp.Jap
J260	오정근	Temp.Jap	J279	중앉은뱅이	Temp.Jap
J261	올벼	Temp.Jap	J280	쥐잎파리벼	Temp.Jap
J262	올왜두	Temp.Jap	J282	진안도	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J284	차나락 76	Temp.Jap	J300	조도	Temp.Jap
J285	최부지	Temp.Jap	J301	청군벼	Temp.Jap
J286	팔다도	Temp.Jap	J302	몰라벼	Temp.Jap
J287	팟벼	Temp.Jap	J303	옥경	Temp.Jap
J288	평북 4	Temp.Jap	J306	무모조적조	Temp.Jap
J289	평북 7	Temp.Jap	J307	달골못	Temp.Jap
J29	냉조(B)70	Temp.Jap	J31	냉도	Temp.Jap
J290	평북 8	Temp.Jap	J311	율조조	Temp.Jap
J291	평양	Temp.Jap	J315	장안	Temp.Jap
J292	포천 장망 메벼	Temp.Jap	J316	농광	Temp.Jap
J294	풍우조	Temp.Jap	J317	조동지	Temp.Jap
J295	피벼	Temp.Jap	J318	화중	Temp.Jap
J296	한기부지	Temp.Jap	J319	화영	Temp.Jap
J297	황토조	Temp.Jap	J32	구황도	Temp.Jap
J298	선달	Temp.Jap	J320	팔공	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J321	청호	Temp.Jap	J337	호품	Temp.Jap
J322	선서	Temp.Jap	J338	친들(익산 529 호)	Temp.Jap
J324	안중	Temp.Jap	J339	일미	Temp.Jap
J325	금남	Temp.Jap	J340	남평	Temp.Jap
J326	광안	Temp.Jap	J341	화랑	Temp.Jap
J327	하이아미	Temp.Jap	J342	영남	Temp.Jap
J328	영안	Temp.Jap	J343	동진	Temp.Jap
J329	팔굉	Temp.Jap	J344	추청	Temp.Jap
J330	청운(수원 537 호)	Temp.Jap	J345	새누리	Temp.Jap
J331	농백	Aus	J346	청품(수원 567 호)	Temp.Jap
J332	청명	Temp.Jap	J349	상주 48 호	Temp.Jap
J333	화신	Temp.Jap	J35	객주조도	Temp.Jap
J334	계화	Temp.Jap	J350	Mutsunishiki	Temp.Jap
J335	섬진	Temp.Jap	J352	Mutsukaori	Temp.Jap
J336	고품	Temp.Jap	J353	Kantou 51	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population	ID	Variety	Sub-population
J354	Fukunishiki	Temp.Jap	J373	Mineasahi	Temp.Jap
J355	Homarenishiki	Trop.Jap	J374	Yamahikari	Temp.Jap
J356	Hitomebore	Temp.Jap	J375	Yamasenishiki	Temp.Jap
J357	Kitakogane	Aromatic	J376	Norin22	Temp.Jap
J358	Oochikara	Temp.Jap	J377	Yumepirika	Temp.Jap
J359	Hokuriku 130	Temp.Jap	J378	Sinboi3	Temp.Jap
J36	깨벼	Temp.Jap	J380	Hua 99115	Temp.Jap
J360	Yumetsukushi	Temp.Jap	J381	공주령 06	Temp.Jap
J361	Chiyominori	Trop.Jap	J382	공주령 07	Temp.Jap
J362	Akinishiki	Temp.Jap	J386	Yanfeng 47	Temp.Jap
J364	Akitakomachi	Temp.Jap	J388	Yunjingyou 15	Temp.Jap
J368	Hinohikari	Temp.Jap	J390	Zhongzuo 321	Temp.Jap
J37	설악벼	Temp.Jap	J392	공주령 03	Temp.Jap
J370	Itadaki	Temp.Jap	J393	Huayu 13	Temp.Jap
J371	Kinuhikari	Temp.Jap	J394	Lunhui 422	Trop.Jap

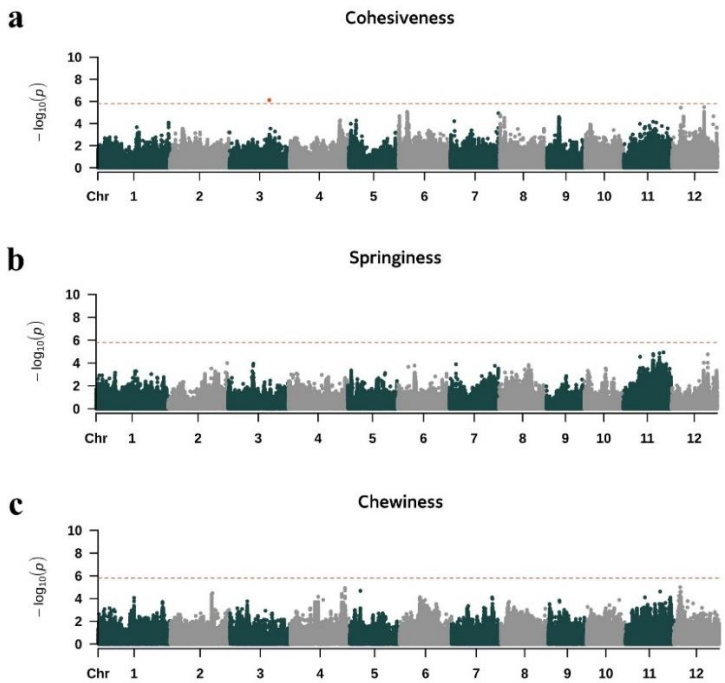
Appendix 1. *Cont.*

ID	Variety	Sub population	ID	Variety	Sub-population
J395	Chujing 27	Temp.Jap	J51	남원벼	Temp.Jap
J397	Qingjiao 301	Temp.Jap	J53	신운봉벼	Temp.Jap
J398	Ningjing 14	Temp.Jap	J54	대야벼	Temp.Jap
J400	Yanjing 9967	Temp.Jap	J57	화남벼	Temp.Jap
J401	Taichung 178	Temp.Jap	J59	삼백벼	Temp.Jap
J402	Taipei 309	Temp.Jap	J66	안산벼	Temp.Jap
J403	Taichung 65	Temp.Jap	J67	삼천벼	Temp.Jap
J404	Khaohsiung 64	Temp.Jap	J69	서진벼	Temp.Jap
J405	Kaoshiung 142	Temp.Jap	J70	화삼벼	Temp.Jap
J406	Taichung 16	Indica	J77	상미	Temp.Jap
J408	Calrose 76	Trop.Jap	J78	수라	Temp.Jap
J42	영산벼	Temp.Jap	J79	원황	Temp.Jap
J45	화진벼	Temp.Jap	J80	인월	Temp.Jap
J46	탐진벼	Temp.Jap	J82	안성벼	Temp.Jap
J49	오봉벼	Temp.Jap	J84	해평	Temp.Jap

Appendix 1. *Cont.*

ID	Variety	Sub-population
J88	백진주	Temp.Jap
J9	화성밭찰	Trop.Jap
J97	청담	Temp.Jap
J98	청해진미	Temp.Jap
Jinheung	진흥	Temp.Jap
Joryeong	조령	Temp.Jap
Koshihikari	Koshihikari	Temp.Jap
Nipponbare	Nipponbare	Temp.Jap
Samnam	삼남	Temp.Jap
Sasanishiki	Sasanishiki	Temp.Jap
Singeumo	신금오	Temp.Jap
Suwonjo	수원조	Temp.Jap

Note: Temp.Jap, Trop.Jap represent temperate japonica, and tropical japonica, respectively



Appendix 2. Manhattan plots of GWAS for the texture parameters using FaST−LMM v2.07 and linkage. For the Manhattan plot, the x−axis and the y−axis represents chromosome number and *p*−value, respectively. The genome−wide significant threshold line is shown in orange. Red dots that have crossed the threshold line show SNPs strongly associated with texture. (a) COH, (b) SPR, (c) CWN.

초록

벼에서 쌀밥 조직감 전유전체연관분석 연구

조직감은 인간에 의해 지각되는 다면적인 관능 특성이다. 식품에서 느낄 수 있는 조직감 유형은 다양하다. 조직감은 고품질 쌀을 직접적으로 평가하는 지표로 활용될 수 있으며, 사람들의 기호에 영향을 미치는 요소이므로 육종 과정에서 중요하게 고려해야 한다. 쌀밥에서 조직감 표현형은 조직감 분석 및 평가에 의하여 단단함, 부드러움, 찰기가 있음 등과 같은 단어로 표현할 수 있다.

육성종과 재래종을 포함한 248개의 벼 패널에서 텍스쳐 프로파일 분석(Texture Profile Analysis, TPA)을 이용하여 쌀밥 조직감을 조사하였다. 패널에서 대규모 병렬형 염기서열 분석법(Next Generation Sequencing, NGS)으로 숏 리드(Short reads)를 얻었고, 이를 벼의 참조 유전체(Nipponbare, IRGSP v1.0)과 대조하였다. Re-sequencing 을 통해 조사된 고품질 단일 염기 다형성(Single nucleotide polymorphism, SNP) 1,254,682개와, TPA에서 평가된 경도, 부착성, 응집성, 탄력성, 순간 복원력, 씹힘성을 포함하는 조직감 항목을 이용하여 전유전체연관분석(Genome wide association study, GWAS)을 실시하였다. 경도와 유의하게 연관된 단일 염기 다형성 마커는 4개로, 벼 염색체 5번(Chr05:13917049), 6번(Chr06:1772858, 1772859, 4518410)에서 탐지되었다. 부착성과 연관된 단일 염기 다형성은 염색체 6번에서 총 19개를 확인하였다. 염색체 5번에는 순간 복원력과 연관된 단일 염기 다형성 마커

(Chr05:3455897)를 찾을 수 있었다. 연관 불균형 기반 분석(Linkage Disequilibrium (LD)-based analysis)을 통해, 각 쌀밥의 조직감 항목에서 p-value ≤ 1.572e-6로 유의하게 분석된 마커 위치를 포함하며, 함께 유전되는 지역의 범위와 후보 유전자를 확인하였다. 특히, 순간 복원력 형질 관련 유전자인 *GS5*의 일배체형 분석을 통해 표현형 차이를 식별하였고, 이를 통해 가장 강력하게 쌀밥 조직감 항목에 영향을 주는 후보 유전자를 제시하였다.

본 연구 결과에서는 벼 재래종과 육성종에서 쌀밥 조직감의 다양한 표현형 정보를 포함한다. 탐색된 쌀밥 조직감 연관 단일 염기 다형성과, 연관 불평형 범위 내 위치하고 있는 후보 유전자군 정보와 같은 조직감 항목의 유전적 기초 자료는 향후 벼에서 조직감 개선 육종 과정에 이용할 수 있을 것으로 사료된다.

주요 단어 : 쌀밥 조직감, 전유전체연관분석(GWAS), 텍스쳐 프로파일 분석(TPA), 경도, 부착성, 순간 복원력

학번 : 2021-25835

ACKNOWLEDGEMENT

구석구석 도움이 닿지 않은 곳이 없습니다.

2년의 석사 과정 동안, 저에게 도움 주신 모든 분들께 감사드립니다.

늘 온화하고 인자한 모습으로 석사 과정 내내 아낌없는 조언해 주셨던,

존경하는 고희종 지도교수님,

매일 지치지 않고 열심히 실험하시고, 일하시는 모습으로 저의 열심을

당겨 주셨던 백기 박사님,

연구와 관련된 반짝이는 아이디어를 제공해주시고, 고민되는 부분에 여러

경우의 수를 만들어 해결해 주셨던 수 오빠,

연구를 시작할 수 있도록 고생하며 준비하셨던 백미재료를 내어 주시고,

실험실 생활 내내 세심하게 품어 주셨던 윤경 언니,

타지에서 공부하시느라 여러 힘든 일이 있을 텐데도, 먼저 도와줄 일이

없냐고 물으시며 밝게 웃으시던 퓨 언니,

학부도 달랐고, 인턴도 하지 않은 채 실험실에 들어와 마음 붙일 곳 없던

저에게 좋은 동기로, 졸업까지 함께 해준 지영 언니.

수원 농장에서 실험 재료들을 보살펴 주시는 김홍열 박사님과

강미경 여사님께도 진심으로 감사드립니다. 지금은 실험실을 떠나셨지만,

바보 같은 질문을 해도 늘 친절하게 알려주셨던 선배님들, 모두 감사합니다.

육종실에서 공부할 수 있었던 건 큰 행운이었습니다.

욕심 많은 딸이 마음껏 꿈 꾸고, 실현할 수 있도록 가장 가까운 곳에서 지지해주시는 가족들에게 뜨거운 사랑과, 감사를 보냅니다. 제가 하고 싶어하는 것들, 그저 알아서 잘 하겠지 굳게 믿어 주신 마음이 저를 더욱 단단하게 만듭니다. 오롯이 건네받은 사랑이 큰 덕분입니다.

석사과정 중에도, 연기를 놓지 않을 수 있게 배려해 주시고, 단단했던 틀에서 벗어나는 방법을 일러주신, 나의 사랑하는 어른, 윤성원 배우님, 늘 평일 저녁시간을 쓸 수 없었고, 여러모로 많은 부분에서 부족한 나를 기다려주고 이해해준 '배우집단 몽돌' 여러분, 고맙고 사랑합니다.

바쁘다는 핑계로, 공부하는 내내 만나기 어려운 양 굴었던
제 곁을 떠나지 않고 지켜준 친구들에게도 짙은 애정을 보냅니다.

덕분에 2년을 후회 없이 보냅니다.
저를 만들어 주신 모든 분들께, 감사와 사랑을 전합니다.
받은 마음 잊지 않고, 따뜻하게 그리고 다정하게 살아가겠습니다.

여전히 겨울이 한창입니다.
봄이 다가오기 전까지 바람이 차가우니, 부디 건강 유의하세요.
고맙습니다.