Ph.D. Dissertation of Agriculture

Physicochemical and Genomic Studies on Eating Quality of *Japonica* Rice (*Oryza sativa* L.)

자포니카 벼의 식미에 관한
이화학적 및 유전체 특성 연구

February 2023

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

Yoon Kyung Lee

Physicochemical and Genomic Studies on Eating Quality of *Japonica* Rice (*Oryza sativa* L.)

Under the direction of Dr. Hee-Jong Koh

Submitting a Ph.D. Dissertation of Agriculture
February 2023

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

Yoon Kyung Lee

Confirming the Ph.D. Dissertation written by
Yoon Kyung Lee
December 2022

Chair	Nam-Chon Paek, Ph.D.
Vice Chair	Hee-Jong Koh, Ph.D.
Examiner	Hak-Soo Seo, Ph.D.
Examiner	Do-Soon Kim, Ph.D.
Examiner	Joong Hyoun Chin, Ph.D.

Physicochemical and Genomic Studies on Eating Quality of *Japonica* Rice (*Oryza sativa* L.)

Yoon Kyung Lee

General Abstract

Eating quality is a major trait in rice (*Oryza sativa* L.) that is significantly and directly related to the consumers' satisfaction, hence achieving high eating quality has been a goal of many breeding programs. However, widely known major genetic factor, Wx, cannot explain all the variations in eating and cooking quality. Moreover, for complicated traits like eating quality, identifying genetic architecture of a trait and minor effect genetic factors are crucial in unravelling the genetic factors controlling the trait to utilize it in breeding programs.

Firstly, 10 eating and cooking qualities were assessed in 284 *japonica* rice accessions, and the genome-wide association of traits and genetic variants were studied with year replicates to dissect the genetic architecture of eating and cooking qualities and to discover novel genes that control the traits. With transcriptome analysis, the candidate genes for cooked rice glossiness and starch pasting properties were identified in chromosome 6 and 2, respectively. A candidate gene for glossiness, Os06g0256500, is related to the interconversion of glucose 6-phosphate to fructose 6-phosphate which is an important substrate beginning the starch biosynthesis. A candidate gene for maximum, minimum, and final viscosity, Os02g0224300, is related to one of the storage protein subunits. The transcripts expression levels of the genes and different phenotypic variations based on the haplotypes were determined. Moreover, the functions of the genes to the eating quality properties were verified using homozygote T-DNA insertion lines. These newly identified genetic factors are related to the eating quality properties other than amylose and protein content, and it is expected to be usefully applied in breeding programs for acquiring high eating quality rice cultivars.

Secondly, high eating quality cultivar, Koshihikari, was studied for its cultivar-specific genome and genetic regions significantly related to eating quality

were identified. Koshihikari de novo genome was assembled using Oxford nanopre long-read sequences and Illumina short-read sequences. Highly contiguous genome was obtained, and structural variations were examined. Among the gaps between Koshihikari and Nipponbare (reference genome, IRGSP-1.0), previously identified eating quality related QTLs were found in chromosome 3 and 11. From the sequence variations in chromosome 11, eating quality related marker P5 was located. From the study, high eating quality cultivars generated from Koshihikari possessed P5 sequences, on the other hand, low eating quality cultivars generated from Koshihikari did not. The eating quality of near-isogenic lines of Samnam genetic background with P5 segment was improved when measured by Toyo taste meter. The identified Koshihikari-specific genomic region was associated with eating quality, so introgression of the region could further increase the eating quality of breeding lines and facilitate the breeding program by means of MAS.

Lastly, aroma and flavor are studied in japonica rice cultivars which is an important sensory attribute affecting eating quality. In this chapter, volatile organic compounds potentially affecting the eating quality of non-aromatic *japonica* rice were identified using gas chromatography mass spectrometry, sensory panel test, and Toyo taster meter analyses of 14 rice cultivars. Partial least squared discriminant analysis demonstrated an outstanding classification of the volatile compounds on eating quality discrimination. Among the compounds, 1-octen-3-ol, 1-ethyl-3,5-dimethylbenzene, 2,6,11-trimethyldodecane, 3-ethyloctane, 2,7,10-trimethyldodecane, methyl salicylate, 2-octanone, and heptanal were selected as important compounds. Characterization of rice eating quality based on volatile compounds suggests the application of metabolite profiling data for rice breeding of high eating quality.

Keywords : japonica, rice, *Oryza sativa*, eating quality, genomics, GWAS, de novo assembly, volatile organic compounds

Student Number : 2018-31862

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BUSCO	Benchmarking universal single-copy orthologs
DEGs	Differentially expressed genes
EQ	Eating quality
GC/MS	Gas chromatography/mass spectrometry
GWAS	Genome wide association study
HS-SPME	Headspace solid phase microextraction
InDels	Insertion/deletion mutations
LD	Linkage disequilibrium
logCPM	Logarithm of counts per million reads
MAS	Marker assister selection
NIL	Near isogenic line
PLS-DA	Partial least square discrimination analysis
PCR	Polymerization chain reaction
PCA	Principal component analysis
QTLs	Quantitative trait loci
RVA	Rapid viscosity analysis
SNP	Single nucleotide polymorphism
VIP	Variable importance in projection
VOCs	Volatile organic compounds

CHAPTER I

Genome wide association study on the
eating quality of *japonica* rice (*Oryza sativa* L.)

Abstract

Market demand for high eating quality rice has been increasing due to improvement in living standards. However, its complicated nature hinders clear understanding of the genetic architecture and mechanism of controlling the eating quality trait. With the significance of identifying genetic architecture of eating quality and minor effect genetic factors, candidate genes for eating quality traits in *japonica* rice were identified by genome wide association study of 284 rice accessions in total. Commonly detected lead SNPs for Toyo taste value through out of 3 years replication were found from chromosome 2, 6, and 8. In the case of rapid viscosity analysis, common lead SNPs were detected from chromosome 2, 8, and 11 from 2 years of replication. Transcriptome analysis revealed clusters of genes that showed different expression patterns based on the phenotypes. The genes from clusters which were located within LD blocks of lead SNPs from GWAS were identified, Os06g0256500 and Os02g0224300 for glossiness and viscosity, respectively. The T-DNA lines of candidate genes associated with Toyo taste value and RVA were verified by comparing the phenotypes to the parental cultivars. The haplotype analyses and phenotypes of T-DNA insertion lines confirmed the functions of the genes.

Introduction

Rice grain is mainly composed of starch, approximately 80-85%, in endosperm, some storage proteins, 4-10%, in its aleurone layer, and storage proteins and lipids (1%) in embryo (Champagne et al., 2004). Amylose and amylopectin comprise a highly ordered structure in the starch granule, molecular and granular differences generated by different amylose and amylopectin ratio affect the functional properties in cereals (Li et al. 2017). It is not surprising that eating quality (EQ) of rice is affected by various starch related physico-chemical properties. EQ is an abstract concept that is subjectively perceived by the individuals with empirical and cultural backgrounds. Evaluation of EQ encompass all sensory attributes, including appearance, texture, taste, flavor, and aroma. Hence, all the major and minor contributing factors related to the EQ should be inclusively considered to understand the trait.

Starch metabolism has been extensively studied; ADP-glucose produced by ADP-glucose pyrophosphorylase (AGPase) is transferred to amylopectin and amylose (Smith et al., 1997, Tetlow et al., 2004) by starch synthases like soluble starch synthases (SSI, SSII, SSIII) and granule-bound starch synthases (GBSSI, GBSSII). Starch-branching enzymes (SBEI, SBEII) catalyze the hydrolysis of α-1,4 glucosidic linkages and form α-1,6 glucosidic linkages and starch-debranching enzymes (ISA1, ISA2, PUL) determine the structure of amylopectin (Morell et al., 2003, Li et al., 2003, Slattery et al. 2000, Tetlow and Emes 2017). It is known that amylose content is directly related to EQ, and other starch properties like gel consistency, viscosity gelatinization temperature are also considered in determining EQ. However, amylose content and allele type of GBSSI does not explain all the variations of EQ. Hence, various attributes like physical, textural, and aromatic properties and other major or minor genetic factors should be considered. Furthermore, the complexity of this trait makes it difficult to elucidate the molecular mechanisms. It may explain why only few iconic cultivars have made although many rice breeding programs worldwide aimed to attain high EQ cultivars (Custodio et al. 2019).

To rapidly unravel quantitative traits like EQ, genome-wide association study has been widely adopted. GWAS of amylose and protein content using various sources of rice accessions have been conducted yet specifying and cloning other related genes was not successful than detecting GBSSI gene region as one of lead SNPs (Xu et al 2016, Wang et al. 2017a, Bao et al. 2017). Many studies challenged to dissect the genetic bases of other EQ traits like grain appearance and milling quality (Wang et al. 2017b), cooked rice texture (Misra et al. 2018), starch viscosity (Buenafe et al. 2021), then complemented the candidate genes from GWAS with haplotype analyses, transcriptome analyses, and functional studies for verification.

To efficiently attain high EQ cultivars by applying molecular breeding techniques, understanding the genetic architecture of EQ traits and molecular functions of related genes are essential. In this study, GWAS of 10 EQ-related traits for japonica rice accessions were conducted with year replicates to identify associated QTLs of the traits. RNA sequencing and transcriptome analysis was performed to narrow down the candidate genes. Functions of the identified candidate genes were studied using haplotype analysis and T-DNA insertion lines.

Materials and methods

1.1 Plant materials

Japonica rice including improved cultivars and landraces which its heading date is within August was selected for comprising the panels. The panels composed of 135, 197, and 284 japonica rice accessions (Table 1-1) were cultivated in an experimental field at Seoul National University, Suwon, South Korea in 2018, 2019, and 2020, respectively. The general lowland cultivation method was applied.

1.2 Phenotypic analyses

The eating quality was evaluated in 2018, 2019, and 2020. The heading date of each accession was recorded, and plants were harvested when reaching 45-50 days after heading. The harvested plants were air dried until reaching the grain water content of 13–14%, and subsequently threshed using a thresher. The grains were dehulled (using a dehull machine), milled to 92.2% (using a milling machine), and stored at 12°C storage-room. Amylose and protein content adjusted to 15% moisture content were measured using Near Infrared Grain Tester (AN-820, Kett, Japan) from polished rice.

Starch pasting properties were measured using Rapid Visco Analyzer (Newport Scientific, Warriewood, Australia). Finely ground polished rice flour (4 g) was added to 15 ml of distilled water, spindle speed and temperature setting followed the method described in the AACC Method 61-02 (American Association of Cereal Chemists, 2000); heating cycle (50-95 °C) - hold (95 °C) - cooling cycle (95-50 °C). Maximum viscosity, minimum viscosity, final viscosity, setback viscosity, and breakdown viscosity were measured in Rapid Viscosity Unit (RVU). Moreover, peak time (in mins) and pasting temperature (in °C) were measured as well.

Head rice (33 g) was cooked at 80°C for 10 min, and allowed to sit at room temperature for 5 min. Then, the surface glossiness of cooked rice, which is highly correlated with its palatability, was quantified into Toyo value using the Toyo taste-meter (MA-30A; Toyo, Japan). All measurements of each accession were obtained in triplicate and the values were averaged for further analysis.

1.3 Genotypic data

Total DNA was extracted from young leaves of 6-week-old plants of each accession according to the CTAB method (Murray and Thompson, 1980). Sheared DNA into 450-500 bp fragments were used for preparing DNA library using TruSeq Nano DNA Library Prep kits (Illumina, San Diego, CA, USA) following the instructions of the manufacturer's protocol. The constructed DNA library was sequenced using an Illumina HiSeq X system and obtained 2 × 150 bp paired end reads with sequencing depth of > 10× for each sample. The adaptors were removed from the raw reads, and low-quality bases were eliminated using Trimmomatic v0.38 (Bolger et al 2014). Quality-trimmed reads were mapped against rice reference genome, Nipponbare IRGSP v1.0, using bwa-mem with default parameters of BWA software v0.7.17 (Li and Durbin 2009). The mapped reads were sorted using Samtools v1.9 (Li et al. 2009) and removed the duplicates using Picard v2.20.2 (http://broadinstitute.github.io/picard/). Nucleotide variants were called using HaplotypeCaller function of GATK v4.1.2 (McKenna et al. 2010). The heterozygous genotypes were filtered out. After removing the variants with missing rates > 0.20 and minor allele frequency < 0.05, total of 1,254,682 SNPs was identified.

1.4 Association analysis and candidate genes identification

GWAS of EQ of the accessions were performed using Factored Spectrally Transformed Linear Mixed Model using FastLmm v2.07 (Lippert et al. 2014). The outputs of association analysis subsequently identified lead SNP loci in the regions that exhibit significant association with trait variation at highest R2 and lowest false discovery rate adjusted P-values (threshold 0.01 and 0.05). Manhattan plots and Q-Q plots for each trait were generated using "rMVP" in the R package (Yin L. et al. 2021). Single-SNP associations were considered true positive when a peak of multiple SNPs was detected at lower -log10 (P-values) in the Manhattan plot. For identification of the candidate genes for each EQ traits, the LD heatmaps in the vicinity of the peaks in Manhattan plots were constructed using "LDBlockShow" (Dong et al. 2021), the LD was measured in R2, and the blocks were detected using PLINK method.

1.5 Transcriptome analysis

Accessions that showed extreme values in each phenotype criterion were selected. For Toyo value, Jaeraeryukdo, Koshiji-wase, Jokwang, Ilpum, Samkwang, Gopum were selected and for RVA, Cheonggunbyeo, Mojo, Jaeraeryukdo, Younghojinmi, Chiyominori, Hwanggeumnodeul were selected. Total RNA was extracted from developing endosperms at 7-9 days after heading while rice hulls and embryos were removed (The cell deaths of endosperm start to occur by 10-13 days after heading). Each of two samples per accession was composed of about 6-8 rice endosperms. In total, 12-15 endosperms were used in RNA extraction of the accession. TaKaRa MiniBEST Plant RNA Extraction Kit was used following the protocol of manufacturer. RNA sequencing was performed using size-selected and quality-checked samples generated from TruSeq Stranded mRNA Sample Prep Kit by Illumina NovaSeq 6000 platform for 150 bp paired-end reads. About 6 Gb of outputs were generated for each sample. Generated raw sequences of samples were processed for adaptor trimming and quality trimming using Trimmomatic v0.38. The trimmed fastq files were subjected for quality check using FastQC v0.11.9 (Andrews, S. 2010). The sequences were aligned to Nipponbare reference sequence (IRGSP-1.0) using HISAT2 v2.2.1 (Kim et al. 2019) and generated the output into SAM files. Using Samtools v1.9 (Li et al. 2009), the SAM files were converted into BAM files and indexed for further analyses. The mapped sequencing reads were aligned to IRGSP-1.0 transcripts file using featureCounts (Liao et al. 2013) function of R, and the genomic features were counted. The read counts were generated, and differential gene expressions were normalized using edgeR (Robinson et al. 2010) function with CPM method. The TMM normalized log CPM values of genes for the samples were subjected for clustering analysis using Clust (Abu-Jamous and Kelly 2018), and clusters of genes that are consistently co-expressed were identified. The gene lists of identified clusters were used in gene ontology enrichment analysis using PANTHER database v17.0 (Thomas et al. 2003, Thomas et al. 2006) for biological process and molecular function. The ontology was visualized using REVIGO (Supek et al. 2011) with removing the redundant GO terms by clustering and reduction algorithms.

Table 1-1. List of accessions used in this study

	Accession name	Year replicates	Classification
1	도봉	2018, 2019, 2020	improved variety
2	기호	2018	improved variety
3	화청	2018	improved variety
4	화성	2018, 2019	improved variety
5	일품	2018	improved variety
6	모조도	2018, 2019, 2020	landrace
7	재래륙도	2018	landrace
8	칠보	2018	improved variety
9	대찬	2018	improved variety
10	대립벼 1	2018	improved variety
11	다미	2018	improved variety
12	드래찬	2018	improved variety
13	동보	2018	improved variety
14	강찬	2018	improved variety
15	친농	2018	improved variety
16	고아미	2018	improved variety
17	고운	2018	improved variety
18	산벼	2018	landrace
19	해찬물결	2018	improved variety
20	해오르미	2018	improved variety
21	하남	2018	improved variety
22	한들	2018, 2019, 2020	improved variety
23	한마음	2018, 2019, 2020	improved variety
24	미품	2018	improved variety
25	산두도	2018, 2019, 2020	landrace
26	황금노들	2018	improved variety
27	산도	2018, 2019, 2020	landrace
28	황금누리	2018	improved variety
29	밭나락	2018	landrace
30	진보	2018	improved variety
31	진부	2018	improved variety
32	진미	2018, 2019, 2020	improved variety
33	진수미	2018	improved variety
34	조안	2018	improved variety
35	조광	2018	improved variety
36	재래조도	2018	landrace
37	조운	2018	improved variety
38	주남	2018	improved variety
39	주남조생	2018	improved variety
40	인부지도	2018	landrace
41	말그미	2018	improved variety

42	미광	2018	improved variety
43	목양	2018	improved variety
44	낙동	2018	improved variety
45	남일	2018	improved variety
46	육월조	2018, 2019, 2020	landrace
47	녹양	2018	improved variety
48	오대벼	2018	improved variety
49	온누리	2018	improved variety
50	풍미	2018	improved variety
51	새계화	2018	improved variety
52	삼광	2018	improved variety
53	올뭇게	2018, 2019, 2020	landrace
54	서안 1	2018	improved variety
55	서간	2018	improved variety
56	신동진	2018	improved variety
57	보리벼	2018, 2019, 2020	landrace
58	신운봉 1	2018	improved variety
59	소비	2018	improved variety
60	수안	2018	improved variety
61	태봉	2018	improved variety
62	태성	2018	improved variety
63	백지청벼	2018, 2019, 2020	landrace
64	운광	2018	improved variety
65	운미	2018	improved variety
66	영호진미	2018	improved variety
67	구중도 99	2018	landrace
68	나도	2018, 2019, 2020	landrace
69	마향조도	2018	landrace
70	남조	2018	landrace
71	느스벼	2018, 2019, 2020	landrace
72	당도	2018	landrace
73	마마콩	2018, 2019, 2020	landrace
74	맥도	2018, 2019, 2020	landrace
75	맥조	2018	landrace
76	모도 52	2018, 2019, 2020	landrace
77	땅벼	2018, 2019, 2020	landrace
78	모조	2018	landrace
79	무다래기	2018, 2019, 2020	landrace
80	무산도	2018	landrace
81	무안도	2018	landrace
82	무주도	2018	landrace
83	미도	2018	landrace
84	미조	2018, 2019, 2020	landrace
85	백장군 22	2018	landrace

86	사두초	2018, 2019, 2020	landrace
87	산다다기도	2018	landrace
88	두도	2018	landrace
89	상도	2018, 2019, 2020	landrace
90	쌍봉	2018, 2019, 2020	landrace
91	쌍두조	2018	landrace
92	서간도도	2018, 2019, 2020	landrace
93	석산조	2018	landrace
94	소두조	2018, 2019, 2020	landrace
95	쇠머리벼	2018	landrace
96	쇠머리지장	2018, 2019, 2020	landrace
97	도립	2018	landrace
98	쇠벤치기	2018, 2019, 2020	landrace
99	수중조	2018	landrace
100	숙나	2018	landrace
101	여벼	2018, 2019, 2020	landrace
102	여수벼	2018, 2019, 2020	landrace
103	열술벼	2018, 2019, 2020	landrace
104	예조	2018	landrace
105	오리도	2018, 2019, 2020	landrace
106	오백조	2018	landrace
107	대골벼	2018, 2019, 2020	landrace
108	오정근	2018	landrace
109	올벼	2018, 2019, 2020	landrace
110	올왜두	2018, 2019, 2020	landrace
111	와방	2018, 2019, 2020	landrace
112	외국벼	2018, 2019, 2020	landrace
113	왜조	2018, 2019, 2020	landrace
114	적박나	2018, 2019, 2020	landrace
115	녹두도	2018	landrace
116	정기조생	2018	landrace
117	정조	2018, 2019, 2020	landrace
118	정종화	2018	landrace
119	조나조	2018	landrace
120	조두조	2018	landrace
121	조선도	2018	landrace
122	조타조	2018	landrace
123	종조백조	2018	landrace
124	쫄장벼	2018	landrace
125	중앉은뱅이	2018, 2019, 2020	landrace
126	쥐잎파리벼	2018, 2019	landrace
127	진도	2018, 2019, 2020	landrace
128	진안도	2018, 2019, 2020	landrace
129	진화	2018, 2019, 2020	landrace

130	차나락 76	2018	landrace
131	최부지	2018	landrace
132	팔다도	2018	landrace
133	팟벼	2018, 2019, 2020	landrace
134	평북 4	2018, 2019, 2020	landrace
135	평북 7	2018	landrace
136	냉조(B)70	2018	landrace
137	평북 8	2018, 2019, 2020	landrace
138	평양	2018	landrace
139	포천 장망 메벼	2018	landrace
140	표도	2018	landrace
141	풍우조	2018	landrace
142	피벼	2018, 2019, 2020	landrace
143	한기부지	2018, 2019	landrace
144	황토조	2018, 2019, 2020	landrace
145	선달	2018	landrace
146	조도	2018	landrace
147	청군벼	2018	landrace
148	몰라벼	2018	landrace
149	옥경	2018, 2019, 2020	landrace
150	백곡나	2018	landrace
151	돈나	2018	landrace
152	무모조적조	2018, 2019, 2020	landrace
153	달골못	2018, 2019, 2020	landrace
154	재래종나	2018, 2019	landrace
155	냉도	2018	landrace
156	원조나	2018, 2019, 2020	landrace
157	율조조	2018	landrace
158	치차벼	2018, 2019, 2020	landrace
159	호미나	2018	landrace
160	영덕	2018, 2019	improved variety
161	장안	2018, 2019	improved variety
162	농광	2018, 2019	improved variety
163	조동지	2018, 2019	improved variety
164	화중	2018, 2019	improved variety
165	화영	2018, 2019	improved variety
166	구황도	2018	landrace
167	팔공	2018, 2019	improved variety
168	청호	2018, 2019	improved variety
169	선서	2018, 2019	improved variety
170	안중	2018, 2019	improved variety
171	금남	2018, 2019	improved variety
172	광안	2018, 2019	improved variety
173	하이아미	2018, 2019	improved variety

174	영안	2018, 2019	improved variety
175	팔굉	2018, 2019	improved variety
176	강원도	2018	landrace
177	청운(수원 537 호)	2018, 2019	improved variety
178	농백	2018, 2019	improved variety
179	청명	2018, 2019	improved variety
180	화신	2018, 2019	improved variety
181	계화	2018, 2019	improved variety
182	섬진	2018, 2019	improved variety
183	고품	2018, 2019	improved variety
184	호품	2018, 2019	improved variety
185	친들(익산 529 호)	2018, 2019	improved variety
186	일미	2018, 2019	improved variety
187	강릉도	2018, 2019, 2020	landrace
188	남평	2018, 2019	improved variety
189	화랑	2018, 2019	improved variety
190	영남	2018, 2019	improved variety
191	동진	2018, 2019	improved variety
192	추청	2018, 2019	improved variety
193	새누리	2018, 2019	improved variety
194	청품(수원 567 호)	2018, 2019	improved variety
195	해담쌀	2018, 2019	improved variety
196	상주 48 호	2018, 2019	improved variety
197	객주조도	2018	landrace
198	Mutsunishiki	2018, 2019, 2020	improved variety
199	Mutsukaori	2018, 2019, 2020	improved variety
200	Kantou 51	2018, 2019	improved variety
201	Fukunishiki	2018, 2019	improved variety
202	Homarenishiki	2018, 2019	improved variety
203	Hitomebore	2018, 2019	improved variety
204	Kitakogane	2018, 2019	improved variety
205	Oochikara	2018, 2019	improved variety
206	Hokuriku 130	2018, 2019	improved variety
207	깨벼	2018	landrace
208	Yumetsukushi	2018, 2019	improved variety
209	Chiyominori	2018, 2019	improved variety
210	Akinishiki	2018, 2019	improved variety
211	Akitakomachi	2018, 2019, 2020	improved variety
212	Hatsuboshi	2018, 2019	improved variety
213	Hinohikari	2018, 2019	improved variety
214	Hounenwase	2018, 2019	improved variety
215	설악벼	2018	improved variety
216	Itadaki	2018, 2019	improved variety
217	Kinuhikari	2018, 2019	improved variety

218	KoshijiWase	2018, 2019	improved variety
219	Mineasahi	2018, 2019	improved variety
220	Yamahikari	2018, 2019	improved variety
221	Yamasenishiki	2018, 2019, 2020	improved variety
222	Norin22	2018, 2019	improved variety
223	Yumepirika	2018, 2019	improved variety
224	Sinboi3	2018, 2019	improved variety
225	Zhongzuo 9936	2018, 2019	improved variety
226	Hua 99115	2018, 2019, 2020	improved variety
227	공주령 06	2018, 2019	improved variety
228	공주령 07	2018, 2019	improved variety
229	Jingxiang 1	2018, 2019	improved variety
230	Yanfeng 47	2018, 2019	improved variety
231	Shennong 265	2018, 2019	improved variety
232	Yunjingyou 15	2018, 2019	improved variety
233	Yunjing 20	2018, 2019	improved variety
234	Zhongzuo 321	2018, 2019	improved variety
235	공주령 03	2018, 2019	improved variety
236	Huayu 13	2018, 2019	improved variety
237	Lunhui 422	2018, 2019	improved variety
238	Chujing 27	2018, 2019	improved variety
239	Qingjiao 301	2018, 2019	improved variety
240	Ningjing 14	2018, 2019	improved variety
241	Yanjing 9967	2018, 2019	improved variety
242	Taichung 178	2018, 2019	improved variety
243	Taipei 309	2018, 2019	improved variety
244	Taichung 65	2018, 2019	improved variety
245	Khaohsiung 64	2018, 2019	improved variety
246	Kaoshiung 142	2018, 2019	improved variety
247	Taichung 16	2018, 2019	improved variety
248	S-201	2018, 2019	improved variety
249	Calrose 76	2018, 2019	improved variety
250	영산벼	2018	improved variety
251	화진벼	2018	improved variety
252	탐진벼	2018	improved variety
253	상남밭벼	2018, 2019, 2020	improved variety
254	오봉벼	2018	improved variety
255	남원벼	2018	improved variety
256	신운봉벼	2018	improved variety
257	대야벼	2018	improved variety
258	화남벼	2018	improved variety
259	삼백벼	2018	improved variety
260	중화벼	2018	improved variety
261	안산벼	2018	improved variety

262	삼천벼	2018	improved variety
263	서진벼	2018	improved variety
264	산두메벼	2018, 2019, 2020	landrace
265	화삼벼	2018	improved variety
266	상미	2018	improved variety
267	수라	2018	improved variety
268	원황	2018	improved variety
269	인월	2018	improved variety
270	안성벼	2018	improved variety
271	해평	2018	improved variety
272	백진주	2018	improved variety
273	화성밭찰	2018	landrace
274	청청진미	2018	improved variety
275	청담	2018	improved variety
276	청해진미	2018	improved variety
277	진흥	2018, 2019	improved variety
278	조령	2018, 2019	improved variety
279	Koshihikari	2018	improved variety
280	Nipponbare	2018, 2019	improved variety
281	삼남	2018	improved variety
282	Sasanishiki	2018, 2019	improved variety
283	신금오	2018, 2019, 2020	improved variety
284	수원조	2018, 2019	landrace

Results

1.1 Population structure and PCA

In the current study, total of 284 rice accessions were used. The rice panel was composed of 171 improved varieties and 113 landraces. Before conducting genetic analyses, the accessions were studied for its population analysis. Principal component analysis (PCA) of the accessions with 25 check varieties (5 accessions each of aromatic, aus, indica, temperate japonica, and tropical japonica rice) identified most of the accessions as temperate japonica (Figure 1-1) while PC1 explained 60.5 % of the variations, and PC2 explained 18.1 % of the variations. The population structure analysis classified the accessions to mainly 2 populations, temperate japonica type and admixture/tropical japonica type (Figure 1-2). The temperate japonica population was further classified into 3 clusters; landraces, improved cultivars, and improved cultivars that were recently generated comparatively.

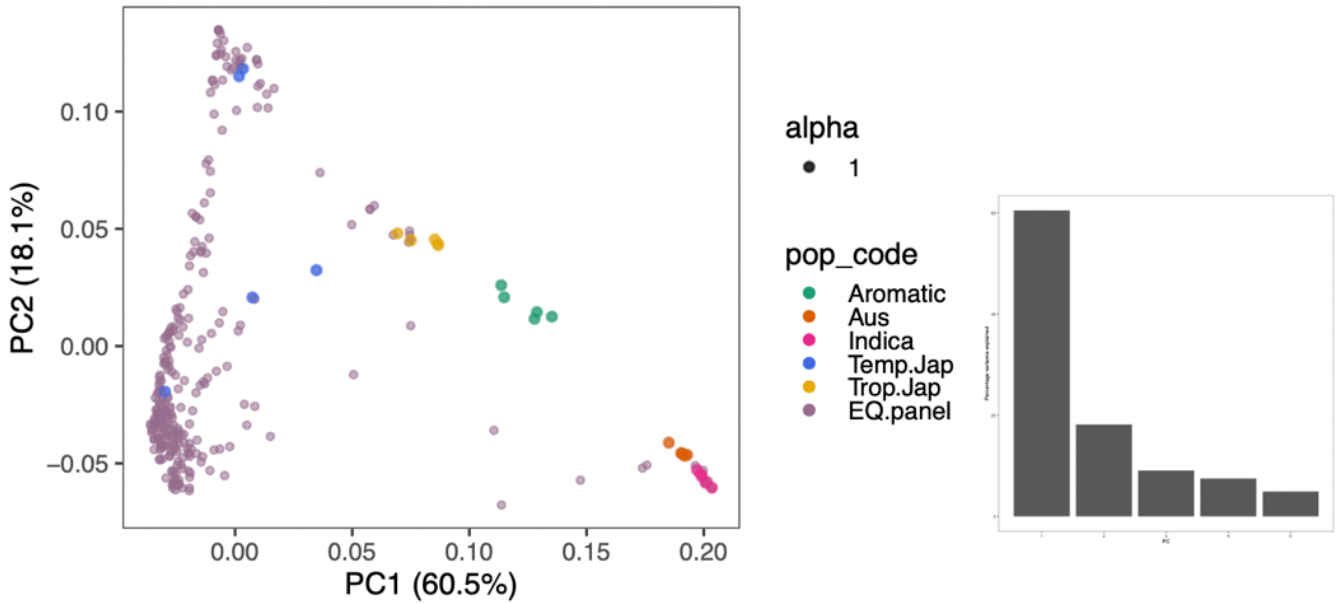


Figure 1-1. Principal Component Analysis (PCA) of the accessions used in the study.
The color-coded dots indicate each of accession. The bar graph present explained variances of the PC.

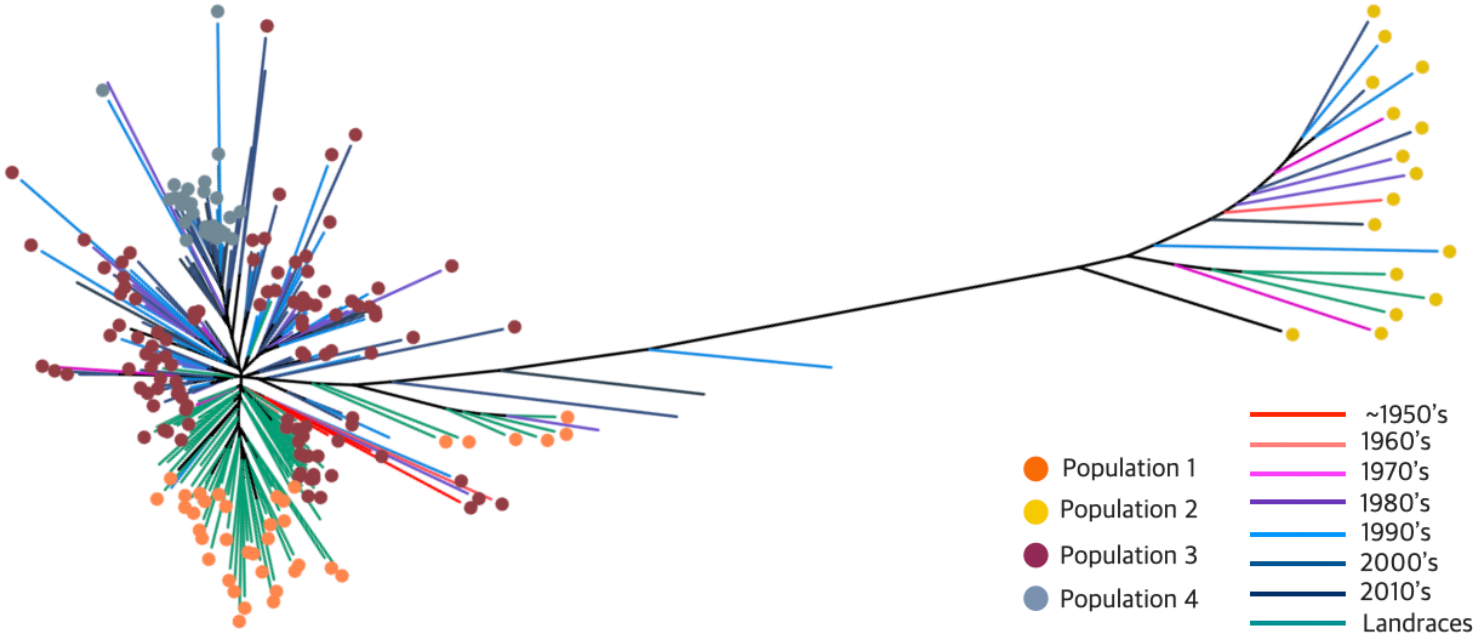


Figure 1-2. Unrooted neighbor-joining tree.
Branch color represents landraces and breeding year. Dot labels indicate the subpopulations when K=4.

1.2 Phenotyping

To assess the eating quality of rice accessions, amylose content, protein content, and Toyo value were measured in 3 years replicates from 2018 to 2020. In case of 7 starch pasting properties, rapid viscosity analysis (RVA) was conducted for 2 years, 2019 and 2020. The histograms of phenotypic traits of each year (Figure 1-3) mostly showed normal distributions, suggesting that the traits are related to and appropriate for the qualitative trait analyses.

Pearson correlations among the measured traits were studied (Figure 1-4). Notably, protein content showed negative correlation with amylose content as is generally known, and were negatively correlated to Toyo taste value. Among the pasting properties, maximum viscosity, minimum viscosity, final viscosity, and breakdown viscosity showed strong positive correlations. On the other hand, setback viscosity and pasting temperature showed negative correlations to the maximum viscosity and breakdown viscosity.

The environmental effects and year replicate variances of each trait were compared by performing ANOVA and calculating broad sense heritability (Table 1-2). Among 10 EQ traits, amylose content showed the lowest heritability to 0.0616 which thought to be affected by environmental and other factors. The heritability of starch pasting properties was fairly high, from 0.4226 to 0.9551.

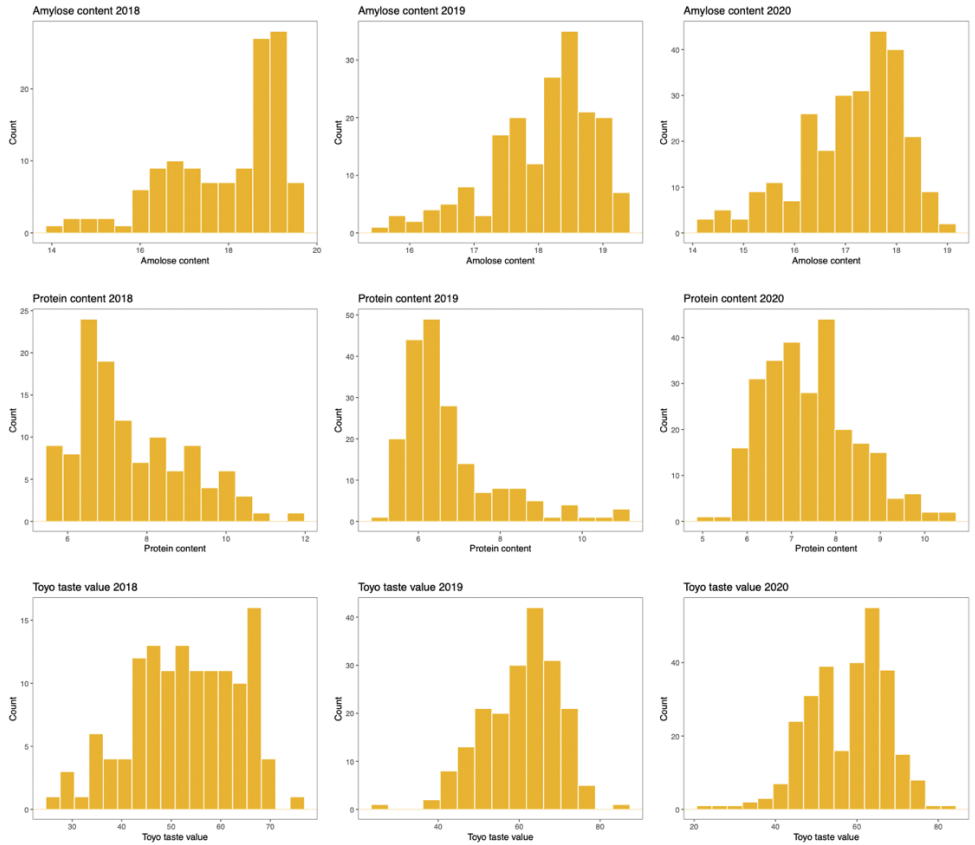


Figure 1-3. The histogram of observed phenotypes related to the eating quality.

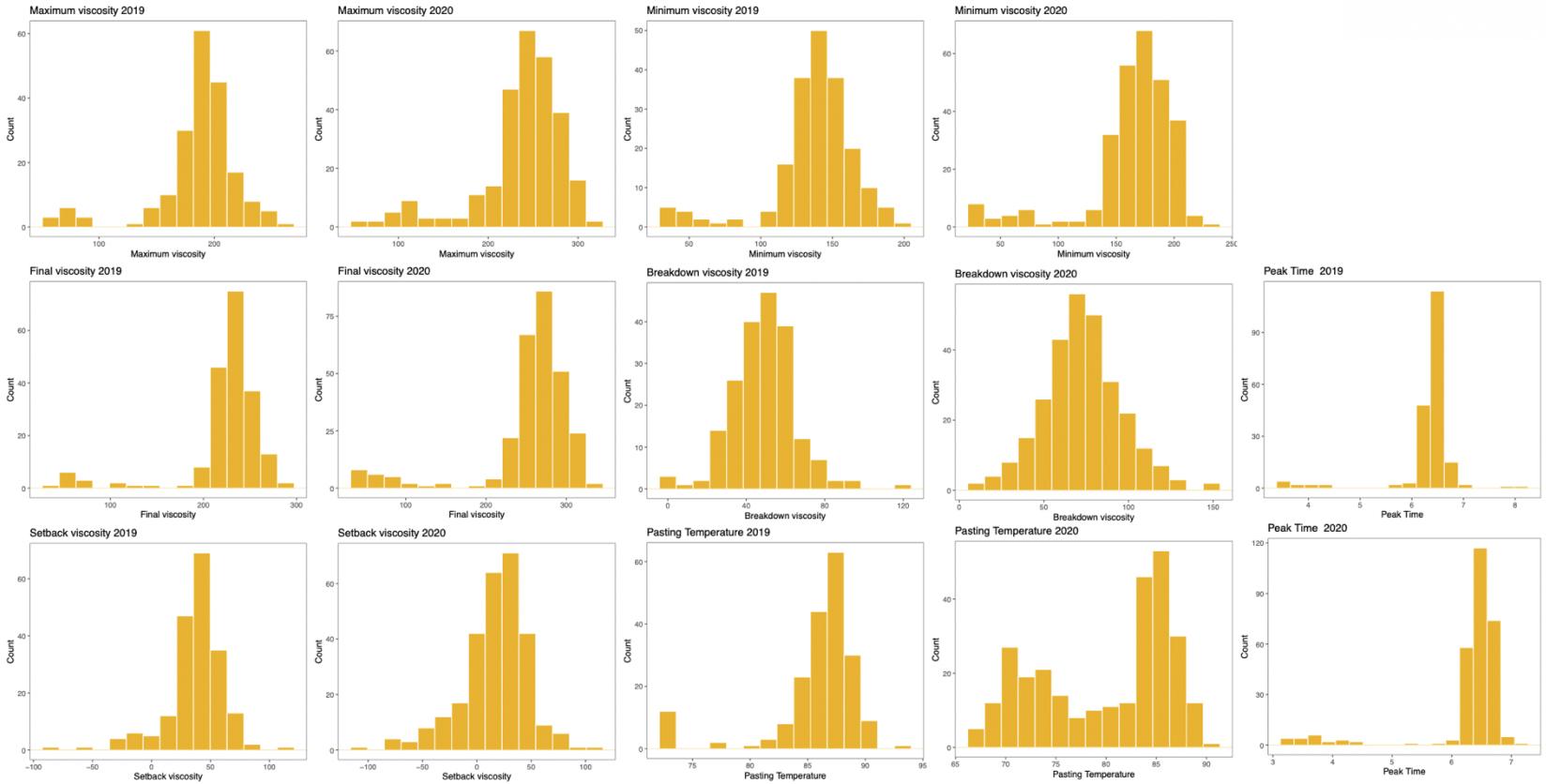


Figure 1-3. Continued.

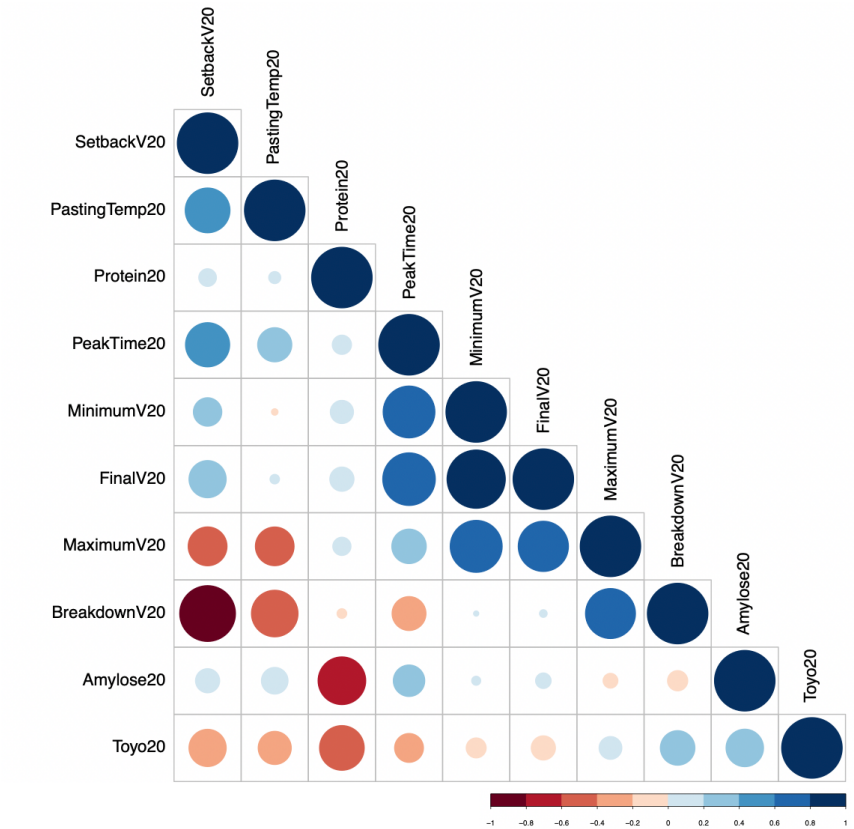


Figure 1-4. Correlation analysis of the measured eating quality traits in 2020.

Table 1-2. The heritability of eating quality related traits

Trait	Heritability (H^2)
Amylose content	0.0616
Protein content	0.5156
Toyo taste value	0.8801
Maximum viscosity	0.7853
Minimum viscosity	0.8627
Final viscosity	0.9043
Breakdown viscosity	0.7116
Setback viscosity	0.8720
Peak time	0.9551
Pasting temperature	0.4226

1.3 GWAS of cooked rice glossiness

The panel composed of 135 accessions was used for GWAS of Toyo taste value in 2018. A total of 5 lead SNPs were detected (Figure 1-5) when a threshold of the lowest false discovery rate adjusted P-values were 0.05, one from chromosome 4, 3 lead SNPs from chromosome 6, and one from chromosome 8. In 2019, total of 197 accessions were used, and generated lead SNPs from chromosomes 2, 3, 6, 8, 9, and 11. In 2020, lead SNPs were detected from chromosomes 2, 4, 6, 8, and 9 from 284 rice accessions. The commonly detected lead SNPs from all the year replicates were in chromosomes 6 and 8. Considering the LD blocks, there still were thousands of candidate genes within the region that could be associated with glossiness and EQ of rice although only commonly detected lead SNPs were considered.

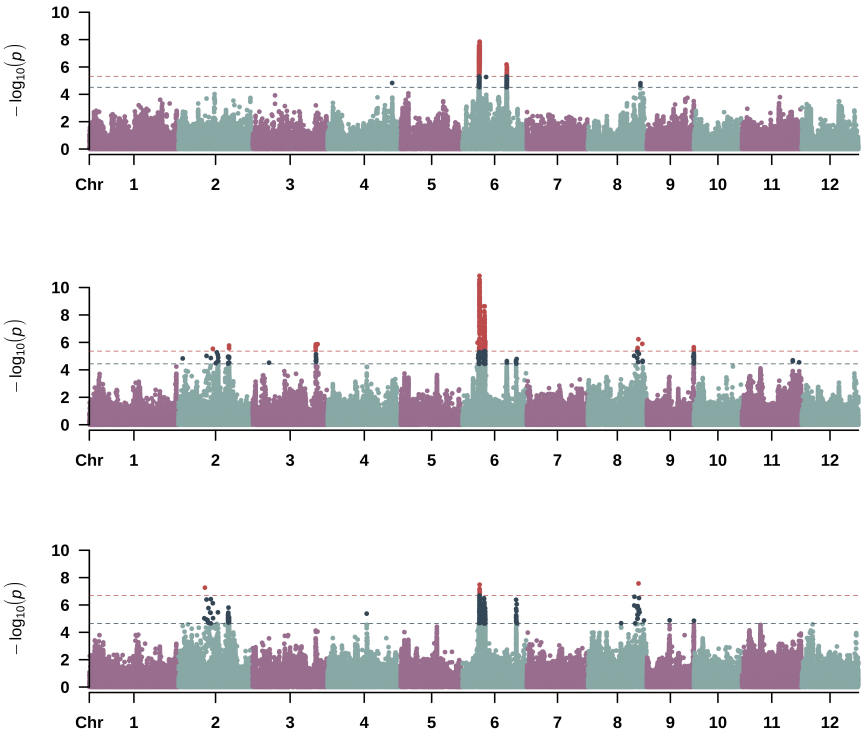


Figure 1-5. Manhattan plots of GWAS for glossiness of cooked rice using Toyo taste-meter in 3-year replicates

1.4 Transcriptome analysis for glossiness

Among the rice accessions, 3 accessions that showed lowest Toyo taste value, Jaeraeryukdo, Koshiji-wase, Jokwang, and 3 accessions that showed highest Toyo taste value, Ilpum, Samkwang, Gopum were selected to compare gene expression and transcriptome analysis in groups. The average Toyo taste value of Jaeraeryukdo, Koshiji-wase, Jokwang was 44.49, and that of Ilpum, Samkwang, Gopum was 72.36. The normalized read counts per transcript were computed and used in clustering analysis. As a result, two clusters were observed (Figure 1-6) with clear co-segregation of log of normalized read counts (logCPM) and Toyo taste values. Cluster 1 was composed of 1,466 genes that showed accessions with low Toyo taste values having high logCPM and accessions with high Toyo taste values having low logCPM values. Gene ontology enrichment analysis revealed that the molecular functions of the genes belong to Cluster 1 were mainly related to the structural molecule activity, organic cyclic compound binding, nucleic acid binding, and structural constituent of ribosome, and the related biological processes were mainly metabolic processes like small molecule metabolic process, cellular metabolic process, and cellular amide metabolic process. Cluster 2 was composed of 1,378 genes which accessions with low Toyo taste values had low logCPM values and accessions with high Toyo taste values had high logCPM values. Gene ontology enrichment analysis classified the genes' molecular function into binding activities like lipid binding, unfolded protein binding, GTP binding, and catalytic activities like glycosyltransferase activity, GTPase activator activity, metallochaperone activity, and structural constituent of cytoskeleton. The related biological processes of genes in cluster 2 were cellular processes like vesicle-meditated transport, localization, and cellular component organization and metabolic processes like primary metabolic process. The heatmap that show logCPM values of individual accessions for two clusters clearly depict the different expression patterns of genes belong to the clusters that could be related to the glossiness. The list of genes belong to these clusters was compared with the genes within LD blocks of lead SNPs from GWAS results to specify candidate gene associated with the glossiness.

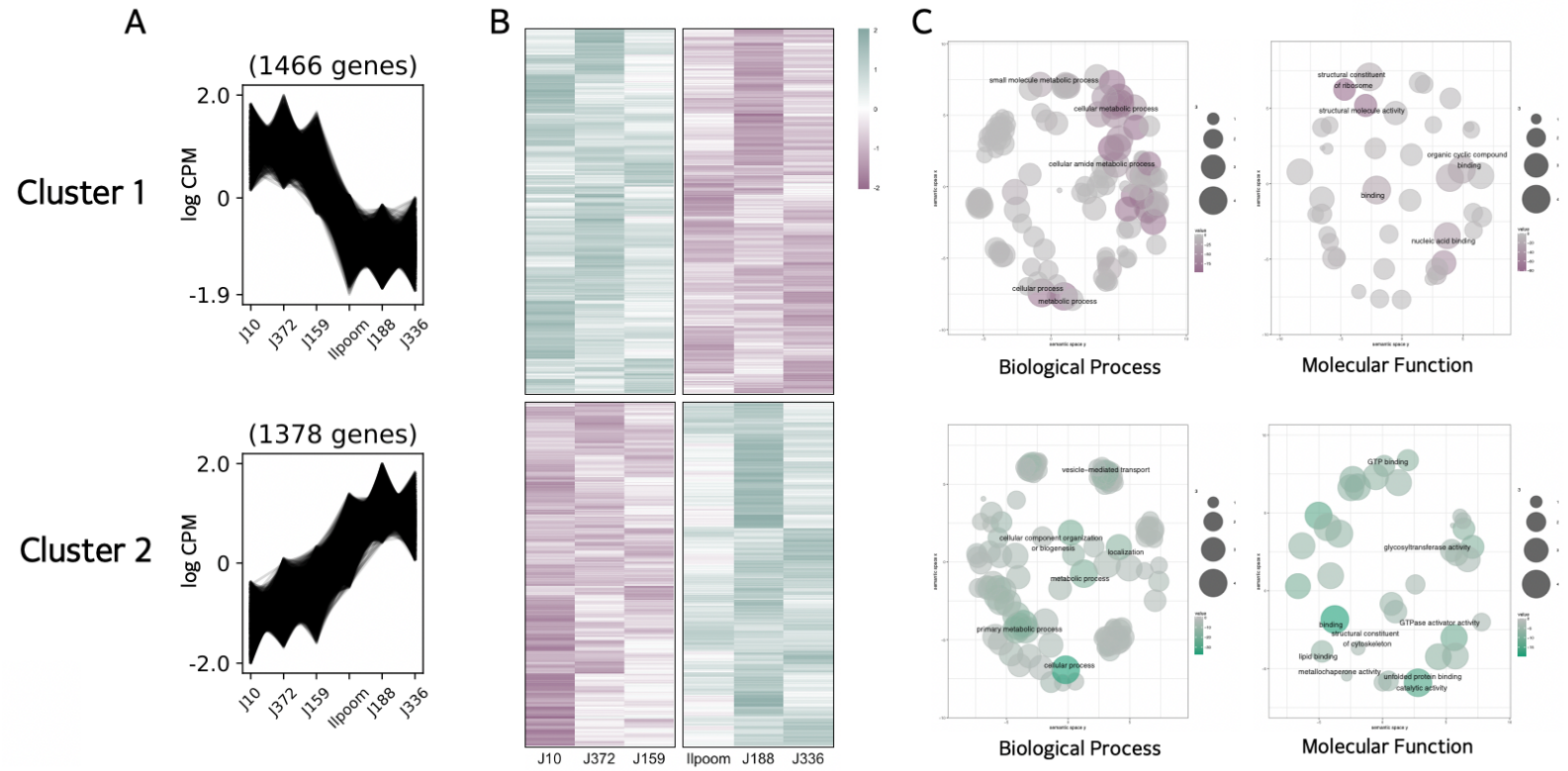


Figure 1-6. Transcriptome analysis of cultivars showing extreme Toyo taste values.
A. Clustering analysis. B. Heatmap analysis of the gene expression levels of the cultivars. C. Gene Ontology enrichment analysis.

1.5 T-DNA lines verification of candidate gene

The candidate gene for the glossiness of cooked rice were identified to Os06g0256500 (*pgi-b*). The gene is from cluster 1 of transcriptome analysis and belongs to LD block of lead SNP chr06:8280642. Two T-DNA lines within the gene location were identified and donated by Kyunghee University. The homozygotes were selected by HPT, LP and RP, and LB and RP PCR combinations (Figure 1-7 A). The T-DNA insertion of *pgi-b-1* was located in the promoter region, and that of *pgi-b-2* was from the 3rd to the 4th exon. The phenotypes of T-DNA lines were observed and compared with wild type cultivar, Dongjin, to verify the function of *pgi-b*. The average Toyo taste value of Dongjin was 66, while that of *pgi-b-1* was 69.6 and that of *pgi-b-2* was 74 (Figure 1-7). It confirmed that as the gene lose its function, eating quality was higher. The result was coherent with transcriptome analysis that the high Toyo taste value cultivars showed lower gene expression, while low Toyo taste value cultivars showed higher gene expression.

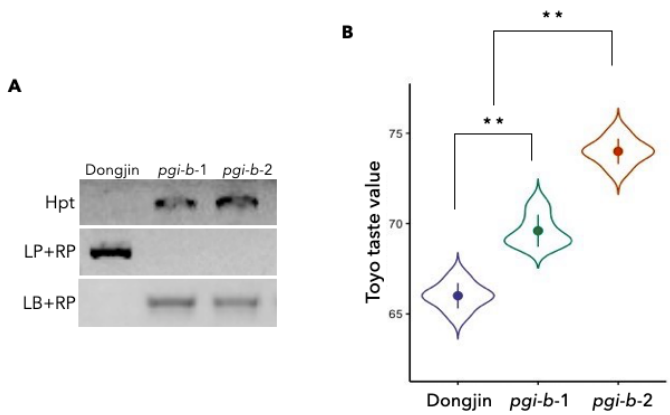
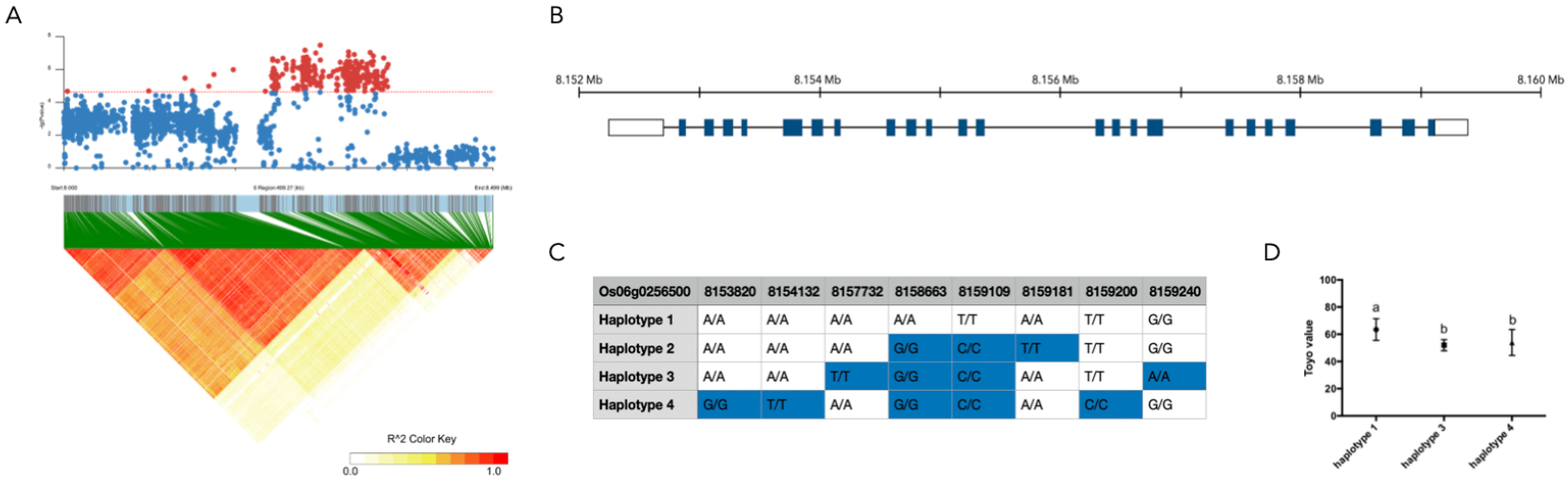


Figure 1-7. The phenotypic evaluation of T-DNA lines of *pgi-b*.

A. marker selection of homozygote individuals. B. Toyo taste values of the T-DNA insertion mutation lines and wild type. Significance indicates results of ANOVA test at 0.01.

1.6 Haplotype analysis for glossiness

As annotated from a database (RAP-DB), the description of gene function was similar to Glucose-6-phosphate isomerase (Figure 1-8). The gene is 6,869 bp in length and composed of 24 exons, located in short arm of chromosome 6, from 8,152,444 bp to 8,159,313 bp in positive strand. The haplotypes that could cause non-synonymous variants were identified, and the Toyo taste values of each haplotype were compared in natural population. As a result, haplotype 1 which is a reference type showed the highest Toyo taste value, and haplotype 3 were significantly lower than that.



Os06g0256500	8153820	8154132	8157732	8158663	8159109	8159181	8159200	8159240
Haplotype 1	A/A	A/A	A/A	A/A	T/T	A/A	T/T	G/G
Haplotype 2	A/A	A/A	A/A	G/G	C/C	T/T	T/T	G/G
Haplotype 3	A/A	A/A	T/T	G/G	C/C	A/A	T/T	A/A
Haplotype 4	G/G	T/T	A/A	G/G	C/C	A/A	C/C	G/G

Figure 1-8. The candidate gene associated with the glossiness of cooked rice.
A. Regional association and LD heat map of candidate region. B. The gene structure of Os06g0256500. C. Classification of the haplotypes of Os06g0256500. D. Toyo taste values of each haplotype.

1.7 GWAS of starch pasting properties using RVA

A total of 7 traits related to rice starch pasting properties including maximum viscosity, minimum viscosity, final viscosity, breakdown viscosity, setback viscosity, peak time, and pasting temperature were measured in 2 year-replicate. In 2019, 197 accessions were used for phenotyping and GWAS analysis. Except for the breakdown viscosity, waxy gene was commonly detected from the manhattan plots of all the other 6 traits (Figure 1-8). Moreover, lead SNPs on chromosomes 2, 7, 8, and 11 were commonly detected from maximum viscosity, minimum viscosity, and final viscosity. In 2020, 284 accessions were used. Similarly, waxy region was detected from all the 7 starch pasting properties. Common lead SNPs were also detected from chromosomes 2 and 8 from maximum viscosity, minimum viscosity, and final viscosity. In conclusion, the novel lead SNPs in chromosomes 2 and 8 were commonly detected from 2-year replicates of maximum viscosity, minimum viscosity, and final viscosity.

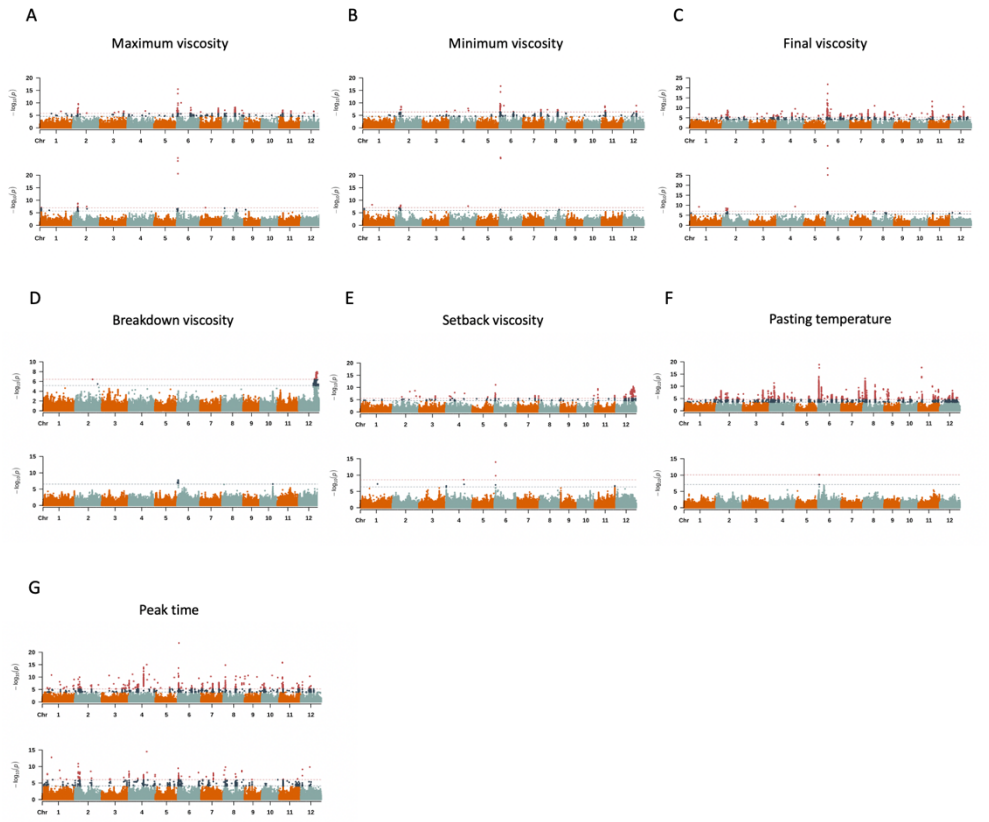


Figure 1-9. Manhattan plots of starch pasting properties for 2-year replicates.
A. Maximum viscosity, B. Minimum viscosity, C. Final viscosity, D. Breakdown viscosity, E. Setback viscosity, F. Pasting temperature, G. Peak time. The blue dots above blue dashed lines indicate the SNPs above the lowest false discovery rate adjusted P-values were 0.05. The red dots above red dashed lines indicate < the threshold of 0.01.

1.8 Transcriptome analysis for RVA

To further narrow down the associated genes for starch pasting properties, RNA sequencing and transcriptome analysis of accessions with extreme phenotypes were conducted. Cheonggunbyeo, Mojo, and Jaeraeryukdo were selected for all having lowest maximum, minimum, and final viscosities, and Younghojinmi, Chiyominori, and Hwanggeumnodeul were selected for having highest values in maximum, minimum, and final viscosities. The average maximum, minimum, and final RVU of Cheonggunbyeo, Mojo, and Jaeraeryukdo were 200.06, 145.41, and 240.93, respectively. Those of Younghojinmi, Chiyominori, and Hwanggeumnodeul were 289.18, 194.37, and 298.55, respectively.

The log values of normalized read counts were obtained, and clustering analysis identified groups of genes that co-segregate with the phenotypes (Figure 1-9). Cluster 1 was composed of 488 genes which showed low logCPM values from low RVA cultivars and high logCPM values from high RVA cultivars. Cluster 2 was composed of 534 genes with high logCPM values from low RVA cultivars and low logCPM values from high RVA cultivars. The genes constitute these clusters showed clearly different expression patterns in the heatmap of the cultivars. The gene ontology enrichment analysis was performed for each cluster. The major biological process of genes from cluster 1 were cellular process and metabolic process like macromolecule metabolic process, cellular metabolic process, nitrogen compound metabolic process, and organonitrogen compound metabolic process. The identified molecular functions of the genes were catalytic activity, transferase activity, ion binding, and small molecule binding. In case of cluster 2, the biological processes of genes were mainly metabolic process and cellular process, like heterocycle metabolic process, organic cyclic compound metabolic process, and nucleobase-containing compound biosynthetic process. The molecular functions of the genes were identified as heterocyclic compound binding, catalytic activity acting on nucleic acid etc. The list of genes composing cluster 1 and 2 were compared with the genes within LD blocks of commonly detected lead SNPs.

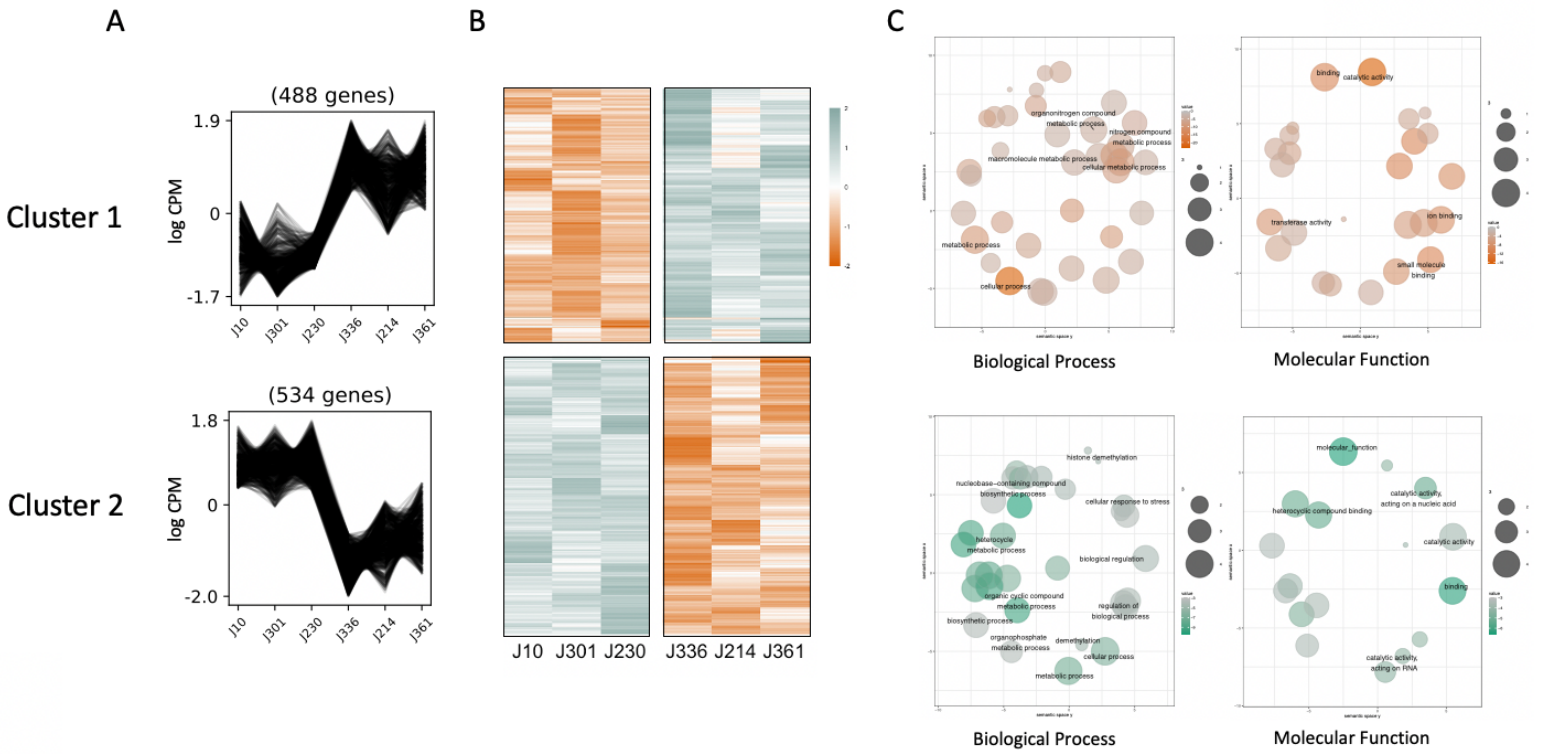


Figure 1-10. Transcriptome analysis of cultivars showing extreme RVA.
A. Clustering analysis. B. Heatmap analysis of the gene expression levels of the cultivars. C. Gene Ontology enrichment analysis.

1.9 Haplotype analysis of RVA

The candidate gene for pasting properties narrowed from clustering analysis and LD block analysis was Os02g0224300. The gene is 1,392 bp in length composed of 1 coding exon, located in short arm of chromosome 2 from 6,983,500 bp to 6,984,891 bp in positive strand, and translation length is 421 amino acids (Figure 1-10). The annotated function of the gene is high molecular weight glutenin family. From the sequences of GWAS panel, non-functional alleles were found. Among the 4 haplotypes, haplotype 3 showed significantly lower maximum and minimum viscosity.

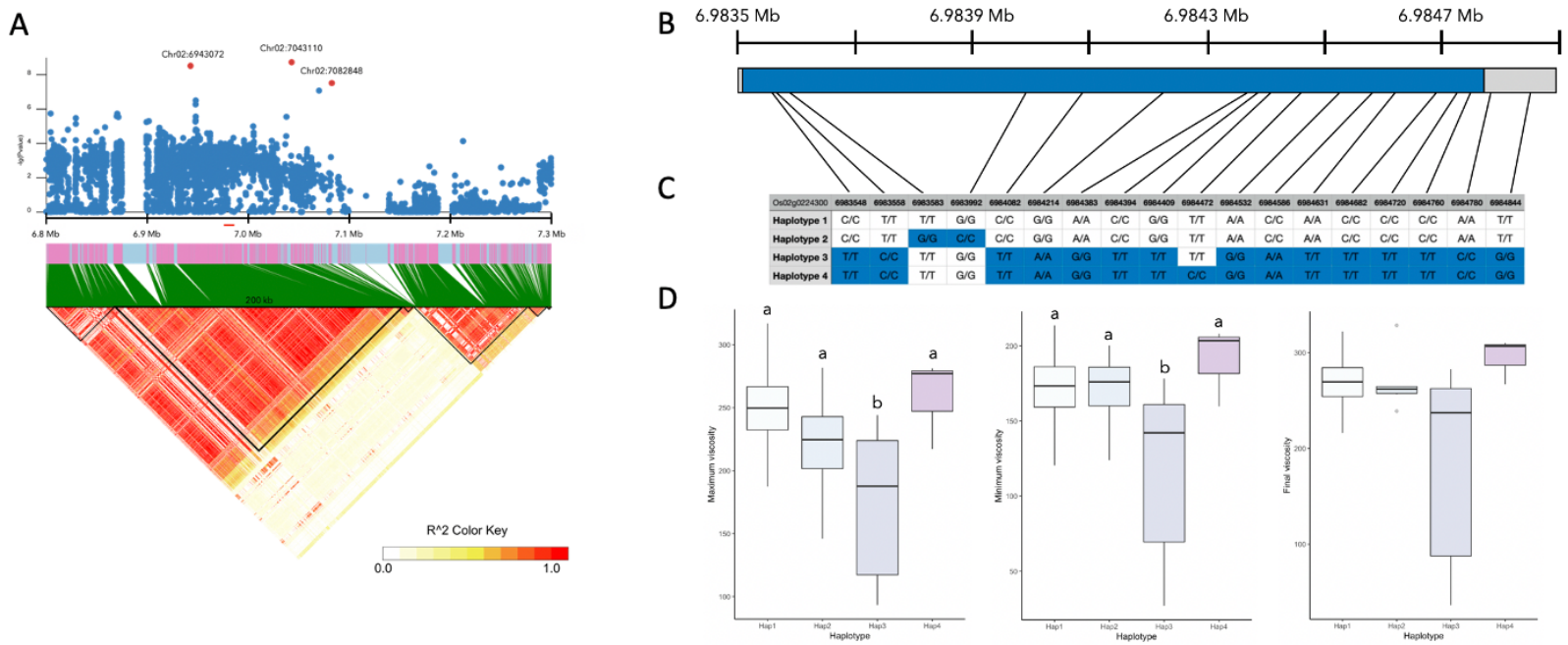


Figure 1-11. The candidate gene associated with the starch pasting properties.

A. Regional association and LD heat map of candidate region. B. The gene structure of Os02g0224300. C. Classification of the haplotypes of Os02g0224300. D. Maximum, minimum and final viscosities of each haplotype.

1.10 T-DNA lines verification

The T-DNA insertion lines were taken from Kyunghee University. The insertion was confirmed using PCR analysis of HPT marker and vector sequence. The homozygote plants were genotyped using markers designed from T-DNA sequence information. RVA was conducted for wild type plants and T-DNA lines. As a result, the maximum viscosity, minimum viscosity, final viscosity, and setback viscosity were significantly lower in T-DNA insertion line. While breakdown viscosity was slightly increased, the observed pasting properties indicated that T-DNA insertion negatively affected to the EQ of rice.

Table 1-2. The comparison of pasting properties of T-DNA line and wild-type using RVA.

	Maximum viscosity	Minimum viscosity	Final viscosity	Breakdown viscosity	Setback viscosity	Peak time	Pasting temp.
Wild type	174.67	108.02	193.0	66.65	18.33	6.22	86.39
T-DNA line	189.50	129.10	218.56	60.40	29.06	6.32	86.78

Viscosities are in rapid viscosity unit (RVU), peak time in minutes, and pasting temperature in °C.

Discussion

Several major genes and QTLs are known to be related to the eating quality of rice. Most prominently, genes for starch biosynthesis such as GBSS I, branching enzymes, debranching enzyme, starch synthases etc. play pivotal role by controlling amylose content. Amylose content of rice grain determine stickiness when cooked. On the other hand, protein content is known to be related to the hardness or toughness of cooked rice. In this study, a lead SNP in GBSS I, also known as waxy gene, region was also identified as associated important genetic factor for controlling amylose and protein content in japonica rice panel. However, the manhattan plots for GWAS of amylose and protein content seemed spurious with many false positive associations over all the chromosomes. Furthermore, considering the fact that amylose and protein content only itself cannot explain all the EQ variations, seeking other major or minor genetic factors related to the EQ properties are essential in unraveling the architecture of EQ and controlling the trait.

Glossiness of cooked rice is an attractive physical property that affects the consumers' satisfaction. It is even included in the sensory panel test as one of the criteria evaluating the EQ of cooked rice together with texture properties and tastes. The glossiness of cooked rice could be related to various factors such as lipid content, water containing and retaining power of rice grain while and after cooking procedure, cytomorphological features that could affect the physical properties of rice grain, etc. However, the rational explanations and specific evidence for the genetic factors associated with glossiness is hard to find. The glossiness of cooked rice can be measured using machines like Toyo taste meter and Satake taste analyzer which irradiate specific ray and convert the reflection into numerical values. The measured values are studied to be positively correlated to the EQ evaluated by sensory panel test. The GWAS of glossiness using Toyo taste values generated consistent results throughout of 3-year replication. The lead SNPs in chromosomes 2, 6, and 8 were commonly detected for 3 consecutive years even when the composition of panels was different each year. It could be inferred that the detected QTLs could be considered as major genetic factors for glossiness that get less effects from environmental factors.

To further narrow down the glossiness associated genome regions into genetic level, the expression levels of cultivars with extreme Toyo taste values were compared. Differential expression genes should be analyzed by comparing multiple samples to multiple samples, so the conventional DEG analysis that compare mutant type from wild type or time series samples were not suitable. Instead, clustering analysis was performed, and gene expression patterns of low Toyo taste value cultivars and high Toyo taste value cultivars were identified. Among the genes of differential expression clusters, Os06g0256500 was within LD block of commonly detected lead SNP. Toyo taste values were significantly different not only from the cultivars of different expression level of the gene, but also from identified haplotypes and T-DNA lines. The annotated function of the protein is glucose-6-phosphate isomerase, which is known to be related to the catalysis of the conversion of glucose 6-phosphate to fructose 6-phosphate, at the second step in glycolysis, and the reverse reaction during gluconeogenesis. The amino acids sequence is 567 in length, and the annotated gene ontology molecular function terms are carbohydrate derivative binding, isomerase activity, and monosaccharide binding, while annotated biological process terms are gluconeogenesis, glucose 6-phosphate metabolic process, and glycolytic process. According to the Gramene database, the function of the protein is reported to be related to ascorbate biosynthesis (R-OSA-1119410), cytosolic glycolysis (R-OSA-1119570) and especially starch biosynthesis (R-OSA-1119477). Glucose 6-phosphate is the first substrate in starch synthesis and glycolysis. According to Zhang et al. (2020) who revealed the differentially expressed proteins related to gluconeogenesis pathway and result in synthesis and accumulation of amylose content in rice grain, glucose 6-phosphate isomerase (EC 5.3.1.9) were downregulated as well. Os06g0256500 is said to be controlling the glossiness of rice grain while getting involved in the primary part of starch metabolism and glycolysis by controlling the level of glucose 6-phosphate and fructose 6-phosphate.

Starch pasting property is closely related to the cooking properties and affect eating quality of rice. Previous studies on GWAS of starch paste viscosity properties for non-glutinous rice revealed QTLs nearby isoamylase 3 gene in chromosome 9 and starch synthase IV-1 gene in chromosome 1 (Xu et al. 2016). Although several studies revealed how the starch pasting and viscosity properties affect the textures, grain rigidity, and cooking quality (Balet et al., 2019, Pang et al.,

2016, Yan et al., 2005, Bao et al., 2000), causal genes for variations of the property are nearly identified except for waxy gene in chromosome 6. Waxy gene was detected in several viscosity properties in this study as well, however there were also consistently detected QTLs in other regions like chromosome 2 and 8.

Commonly detected lead SNPs were identified from several pasting properties, maximum viscosity, minimum viscosity and final viscosity, and the genes within the LD blocks were studied. Thousands of genes were narrowed down to hundreds owing to differential expression clustering analysis. The identified candidate gene, Os02g0224300, exhibited low gene expression levels in the cultivars with low maximum, minimum, and final viscosity units, and high gene expression levels in the cultivars with high maximum, minimum, and final viscosity units. Moreover, haplotype analysis revealed different allele type led to significant differences in maximum and minimum viscosity. The gene is annotated to encode high molecular weight glutenin subunit x-like protein, one of the storage proteins. A recent study by Li et al. (2022) revealed OsPHYB mutation led to down-regulation of storage protein genes including Os02g0224300, Os02g0456100, OsEnS-51, Os06g0507150, and explained it for the increase in the chalkiness of the mutant grains. Multiple genes are reported to be involved in protein metabolism up to date, and it is closely related to the chalkiness of grains (Xie et al. 2021). Also, regulation of the contents of amino acids and storage proteins could affect the quality of rice (Guo et al. 2020, Lin et al. 2017a, Lu et al. 2018). Although the composition of rice grain had been extensively studied, only one major QTL has been cloned for grain protein content (Peng et al., 2014). High molecular weight glutenin subunits are storage gluten proteins present in the starchy endosperm cells of wheat grain (Li et al, 2021). In hexaploidy wheat, Glu-1 loci are known to encode high molecular weight glutenin subunits (Payne et al., 1984, Lawrence and Shepherd, 1981) and highly related to end use quality. Moreover, the mechanism regulating the expression of this high molecular weight glutenin subunit is unclear up to date. In rice, an attempt had been made to exert wheat glutenin subunits accumulate in the endosperm of transgenic rice seeds (Jo et al., 2017). Thus, finding of this study on association of viscosity and candidate gene, Os02g0224300, could disclose more about how seed storage protein content affect EQ.

CHAPTER II

Cultivar specific genomic regions associated with eating quality of Koshihikari

Abstract

The *japonica* rice (*Oryza sativa* L.) cultivar Koshihikari is considered an important breeding material due to its good eating quality (EQ). To effectively utilize Koshihikari in molecular breeding programs, determining its whole genome sequence including cultivar specific segment is necessary. Therefore, the Koshihikari genome was sequenced using Nanopore and Illumina platforms, and *de novo* assembly was performed. A highly contiguous Koshihikari genome sequence was compared with Nipponbare, the reference genome of japonica rice. Genome-wide synteny was observed, as expected, without large structural variations. However, several gaps in alignment were detected on chromosomes 3, 4, 9, and 11. It was notable that previously identified EQ related QTLs were found in these gaps. Moreover, sequence variations were identified in chromosome 11 at a region flanking the P5 marker, one of the significant markers of good EQ. The Koshihikari-specific P5 region was found to be transmitted through the lineage. High EQ cultivars derived from Koshihikari possessed P5 sequences, while Koshihikari-derived low EQ cultivars did not contain the P5 region, implying that the P5 genomic region affect the EQ of Koshihikari progenies. The EQ of near-isogenic lines (NILs) of Samnam (a low EQ cultivar) genetic background harboring the P5 segment was improved compared to that of Samnam in Toyo taste-value. The structure of the Koshihikari-specific P5 genomic region associated with good EQ was analyzed, so it is expected to facilitate the molecular breeding of rice cultivars with superior EQ.

Introduction

Rice (*Oryza sativa* L.) is the staple food of 3.5 billion people worldwide, and has been extensively studied as a model crop in genomics research. *O. sativa* is divided into two subspecies, japonica and indica (Izawa and Shimamoto 1996). According to a recent within-species diversity analysis, *O. sativa* could be further divided into nine subpopulations, based on genetic variation and population structure (Wang et al. 2018). Advances in sequencing technologies have enabled the identification of different kinds of sequence variations, thus unraveled the genomic diversity within a given species. The 3,000 Rice Genomes Project was established to comprehend the genomic diversity within *O. sativa*. The genomes of 3,024 rice accessions were sequenced using the Illumina platform by preparing 500-bp-insert DNA libraries, generating 17 × 109 bp of data (Li et al. 2014). Numerous single nucleotide polymorphisms (SNPs) and other forms of variations such as structural variations and gene presence/absence variations were identified when aligned with the Nipponbare reference genome sequence (Hu et al. 2018).

Koshihikari is a *japonica* rice cultivar developed in 1956 in Japan. With its superior agronomic characteristics such as adaptability to diverse environments, tolerance to pre-harvest sprouting, cold tolerance during booting stage, and most importantly good EQ and stickiness of cooked rice, Koshihikari is one of the most widely cultivated rice cultivars in Japan (Kobayashi et al. 2018). Because of its popularity in the market, the unique agronomic features of Koshihikari have been studied extensively. Ohtsubo et al. (2002) developed molecular markers including P5, B43, and M11 to differentiate Koshihikari and Koshihikari-derived cultivars from other rice genotypes and to utilize the former cultivars for molecular breeding purposes. Several major-effect quantitative trait loci (QTLs) underlying important traits such as heading date (Matsubara et al. 2012), grain quality (Takeuchi et al. 2008), and other physiological characteristics (Hori et al. 2010) have been identified using segregating populations, the generation of which is time consuming and labor intensive. In an attempt to better understand the unique superior characteristics of Koshihikari, its whole genome sequencing was conducted using next-generation sequencing (NGS) techniques and their genome composition was studied using SNP

information. Although approximately 67,000 SNPs were discovered between Koshihikari and Nipponbare, the draft genome sequence of Koshihikari was highly fragmented with thousands of scaffolds (Yamamoto et al. 2010). Therefore, there still is limitation in explaining the cultivar specific characteristics only with the haplotypes by means of genome-wide SNPs from fragmented scaffolds.

Of the several criteria used to determine the quality of rice, EQ is the most important trait. Physicochemical properties determine the cooking and EQ of rice, and granule-bound starch synthase (Wang et al. 1995) and starch synthase II (Gao et al. 2011) are mainly responsible for these properties. Several studies have attempted to identify the genetic regions related to the good EQ of Koshihikari. For example, two QTLs associated with the stickiness of cooked rice were identified using the double-haploid lines of Akihikari and Koshihikari (Takeuchi et al. 2007); 21 QTLs associated with the EQ of rice were discovered using the Koshihikari/Kasalath//Koshihikari backcross inbred lines and chromosome segment substitution lines (CSSLs) (Ebitani et al. 2005); and 43 QTLs responsible for various physicochemical properties were detected in the recombinant inbred lines (RILs) derived from Moritawase and Koshihikari (Wada et al. 2008). Additionally, molecular markers strongly associated with the EQ of rice were identified (Lestari et al. 2009), and one marker, P5, was specifically detected from Koshihikari and its related cultivar Hitomebore.

Advances in scientific knowledge and the related technologies lead to limitless possibilities in understanding the genetic basis of agronomic traits. The genome sequence of *O. sativa* was the first to be assembled using the Sanger sequencing technique (International Rice Genome Sequencing Project and Sasaki 2005; Michael and VanBuren 2015). Completion of the Nipponbare reference genome sequence enabled the resequencing of important rice cultivars, which in turn led to the identification of SNPs and short insertion/deletion mutations (InDels) associated with important agricultural traits. However, resequencing is not applicable to highly diversified regions. Recently, long-read sequencing has been used to obtain chromosome-level genome sequences of important cultivars. The whole genome sequence of IR64 was determined using linked-read and nanopore sequencing approaches. The de novo genome assembly technique produced a highly contiguous genome of IR64, with an estimated size of 367 Mb (Tanaka et al. 2020).

Additionally, high-quality reference genomes of Basmati 334 and Dom Sufid were successfully generated using nanopore sequencing (Choi et al. 2020). The assembled genomes were highly contiguous, and structural variations and presence/absence variations were well characterized.

In this study, we constructed a high-quality de novo assembly of the Koshihikari genome using both Nanopore long-read and Illumina short-read sequencing. Furthermore, cultivar specific genomic regions associated with the good EQ of Koshihikari were identified from the assembled genome sequences of Koshihikari, and were verified using NILs.

Materials and methods

2.1 Plant materials and DNA extraction

Oryza sativa ssp. japonica cv. Koshihikari was used in this study. Koshihikari seeds (accession number IT002752; NAC, RDA, South Korea) were germinated in the dark for 3 days. The seedlings were transplanted into pots, and grown in the growth chamber at 24°C under 16-h light/8-h dark photoperiod and 60% relative humidity. Young leaves of 3-week-old plants were harvested, flash frozen in liquid nitrogen, and stored at -80°C. DNA was extracted from the frozen leaf samples according to the high-molecular-weight gDNA protocol of Oxford Nanopore Technologies.

To identify Koshihikari-specific genomic regions associated with its eating quality, Koshihikari was crossed with Samnam, a Korean japonica cultivar with poor eating quality, and BC3F4 and BC4F4 lines of Samnam//Samnam/Koshihikari were developed. Genotyping was performed using the Fluidigm SNP type genotyping assay following the methods of Seo et al. (2020) using the SNP marker set. All populations were cultivated in the field under the conventional cultivation conditions at the Experimental Farm of Seoul National University.

2.2 Nanopore sequencing

A DNA library for GridION sequencing was prepared using the ligation sequencing kit (SQK-LSK109; Oxford Nanopore Technologies). R9.4.1 flowcells (Oxford Nanopore Technologies) were used for sequencing. Base calling was performed using Guppy (version 4.2.3) with a high accuracy method.

2.3 Illumina sequencing

Libraries with insert size of 500 bp were prepared from the extracted DNA, according to the instructions provided in the Illumina TruSeq DNA Library Preparation Kits v2 Guide. Short-read sequencing was performed on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 (2 × 250 bp paired-end reads). To improve analysis accuracy, raw data were preprocessed using Trimmomatic (version

0.33) (Bolger et al. 2014), with the following parameters: 3 minimum quality base, 4 sliding window, 20 average quality, and 50 minimum read size.

2.4 *De novo* assembly

The Koshihikari genome was de novo assembled, as described by Choi et al. (2020) with slight modifications. Adaptor sequences were trimmed using Porechop. The raw sequence reads were corrected using Canu (version 2.1.1) (Koren et al. 2017), and the initial assembly was performed using Flye (version 2.8.2) (Kolmogorov 2019).

The initial assembly was created using the hybrid method to minimize the error rate. The raw nanopore reads were subjected to four rounds of polishing using Minimap2 (Li 2018) and Racon (version 1.4.20) (Vaser et al. 2017), followed by one round of polishing using Medaka (version 1.2.1) (https://github.com/nanoporetech/medaka). Then Illumina reads were subjected to four rounds of polishing using Pilon (version 1.22) (Walker et al. 2014).

The contigs were scaffolded using RagTag (version 1.0.2) (https://github.com/malonge/RagTag), with a reference-guided scaffolding approach. The Nipponbare IRGSP-1.0 genome was used as a reference.

2.5 Gene prediction and functional annotation

Gene models were annotated using MAKER (version 2.31.11) (Cantarel et al. 2008). Publicly available IRGSP-1.0 transcript and protein data (https://rapdb.dna.affrc.go.jp/) were used as evidence for the pipeline. RepeatModeler (version 1.0.8) (https://github.com/Dfam-consortium/RepeatModeler) identified the repetitive elements, and RepeatMasker (version 4.1.1) (https://www.repeatmasker.org/) detected the repetitive regions. The initial MAKER analysis was followed by ab initio gene prediction, SNAP (Korf 2004) and Augustus (Stanke et al. 2008), generating datasets, which were used to train gene models. Then, a second iteration of MAKER was run. The gene models were visualized using JBrowse 2 (version 1.0.4) (Buels et al. 2016), a customizable genome browser. The functional annotation of a region of interest was performed using BLASTP against UniProt SwissProt. Psi-BLAST was conducted against

EggNOG database 4.5 to predict protein sequences with EggNOG annotation descriptions. tRNA and rRNA were predicted using tRNAscan-SE 2.0 and Barrnap 0.9, respectively. The annotations were merged using Annie (version 4bb3980) and GAG (version d80f3fa).

2.6 Validation and structure analysis of the Koshihikari genome

The statistics of the assembled genome were generated using the bbmap stats.sh script of BBTools suite. The gene completion of assembly and annotation was evaluated using BUSCO (version 5.0.0). The Koshihikari genome was aligned to the Nipponbare genome using D-GENIES, and synteny was visualized. Genome alignment and structure variation were visualized using Mauve (version 2.4.0) (Darling et al. 2004). The identified structural variations were confirmed using PCR analysis (Table 2-1).

2.7 Variant calling and read depth comparison

The cultivars which have Koshihikari in their pedigree were selected and used in whole genome sequencing; Hatsuboshi, Fukunishiki, Homarenishiki, Kitakogane, Sasanishiki, Mineasahi, Kantou 51, Hokuriku 130, Hitomebore, Kinuhikari, Itadaki, Akinishiki, Yumetsukushi, Yumepirika, Sinboi 3, Hinohikari. DNA library was constructed with insert size of 450-500 bp using TruSeq Nano DNA Library Prep kits (Illumina, San Diego, CA, USA) following the manufacturer's guide. Prepared libraries were quantified by qPCR according to the Illumina qPCR quantification protocol. The sequencing data of 2 × 150 bp paired-end reads were generated using Illumina HiSeq X system with a sequencing depth of >10× per sample. The adaptors and low-quality bases were removed using Trimmomatic v0.38 (Bolger et al. 2014) using parameters of ILLUMINACLIP:2:30:10 SLIDINGWINDOW:4:15 MINLEN:50. Trimmed reads were aligned to the generated Koshihikari de novo genome as a reference genome using BWA v0.7.17 MEM with default parameters (Li and Durbin 2009). Samtools v1.9 (Li et al. 2009) was used in sorting the aligned reads, and Picard v2.20.2 (http://broadinstitute.github.io/picard/) was used in removing the duplicates. The nucleotide variants calling was performed using HaplotypeCaller function of GATK v.4.1.2 (McKenna et al. 2010), and the

heterozygous genotypes were filtered. The aligned reads of each cultivar on the interested region were enumerated using Integrative Genomics Viewer (IGV) v2.11.9 (Helga et al. 2013) and compared the read depths.

2.8 Assessment of the eating quality of rice

To evaluate the genic region associated with eating quality, the eating quality of Samnam/Koshihikari NILs was assessed over two consecutive years. The harvested rice samples were dehulled and polished to 92.2%, and moisture content was measured to 14%. Head rice (33 g) was cooked at 80°C for 10 min, and allowed to sit at room temperature for 5 min. Then, the surface glossiness of cooked rice, which is highly correlated with its palatability, was quantified using the Toyo taste-meter (MA-30A; Toyo, Japan). Values of 5 measurements were averaged to obtain one value per sample. Pasting properties were measured using Rapid Visco Analyzer (Newport Scientific, Warriewood, Australia) following the method described in the AACC Method 61-02 (American Association of Cereal Chemists Inc, 2000); heating cycle (50-95 °C) - hold (95 °C) - cooling cycle (95-50 °C). The measurements were taken in triplicate per sample. Moisture, protein, and amylose content were measured using Near Infrared Grain Tester (AN-820, Kett, Japan) from polished rice.

Table 2-1. The PCR markers designed for confirmation of important structural variation

Region	Forward primer (5' → 3')	Physical position	Reverse primer (5' → 3')	Physical position	Annealing temperature (℃)	Expected size (bp)
Chr3_1	CTAGGTGAGTGTCCCATGAA	1248640	CTCTCTTCTCGCTTCTTCTTT	1249065	56	426
Chr3_2	ATCATTCTGAAGATGAAGCAAC	2324612	ACTAAACTTTGTGGAGGGAAAT	2324915	53	304
Chr3_3	CTGGTATCCATTTTGAGTTTTT	2445210	ATGGTCAAATCCTACAAAGAAA	2445621	53	412
Chr11_1	GGGATTGAACTCGACAATAG	20203607	TCCCTTAGGTGCTAAATCTG	20203799	57	193
Chr11_2	ATGTATGAGCGAAGAGACTCA	20252485	GAACTTTCAGGGCACATATC	20252574	57	90
Chr11_3	TAAACGGTTTCATAGGTTGAAG	20381638	TCTAAAATGGATTGCTAGGGTA	20381911	52	274
Chr11_4	ATTAGGCTCAACTCTTCAAACA	20423338	TCTCCTCTCTTGTAGCCAAATA	20423733	52	396
Chr11_5	GCAAAGGGACTATATCAAAACA	20437951	TCTAAACACGCCTTAGTTCAGT	20438331	58	381
Chr11_6	CAAACGAAAGCAATGTTCTA	20459661	CCAGCCTGCTTAACTAATTT	20459997	53	337

Results

2.1 *De novo* assembly of Koshihikari genome using Nanopore and Illumina sequencing reads

The Koshihikari genome was sequenced using the Oxford Nanopore Technologies GridION platform and Illumina MiSeq platform. Long-read sequencing generated 3,510,702 reads (~16 Gb), while short-read paired-end resequencing generated 34,940,048 reads (18.4 Gb) (Table 2-2).

Table 2-2. Statistics of sequencing data.

	ONT GridION	Illumina MiSeq
Number of reads	3,510,702	34,940,048
Total bases (bp)	15,985,579,646	18,418,445,998
Sequencing depth	42X	49X

To obtain a high-quality genome sequence, long and short reads were corrected prior to the initial assembly and used in polishing the draft genome. Since Koshihikari is a *japonica* cultivar and share similarities with Nipponbare in their whole genomes, a reference-guided scaffolding was adopted. Consequently, a 348.7-Mb genome assembly of Koshihikari, with 1,530 contigs and 161 scaffolds, was obtained (Table 2-3). Numerous contigs, ranging from 228 bp to 123 kb in size, remained unscaffolded. The Benchmarking Universal Single-Copy Orthologs (BUSCO) gene completion of assembly was 98.5% of Embryophyta gene groups, which is similar to that of Nipponbare (98.4%) (International Rice Genome Sequencing Project and Sasaki 2005).

Table 2-3. Summary of the Koshihikari genome assembly.

	Koshihikari
Number of contigs	1,530
Number of scaffolds	161
Total number of bases in contigs	348,716,585
Total number of scaffolded bases	348,853,485
Contig N50 length	536.46 kb
Contig L50	178
Scaffold N50 length	27.74 Mb
Scaffold L50	6
Maximum contig length	2.95 Mb
Maximum scaffold length	41.47 Mb
GC content	43.24%
BUSCO gene completion	98.5%

Based on the high-quality Koshihikari draft genome sequence, protein-coding genes were predicted using the MAKER program (version 2.31.11) (Cantarel et al. 2008) and a dataset publicly available at the Rice Annotation Project Database (RAP-DB) (Sakai et al. 2013). A total of 46,275 genes were annotated, and the BUSCO gene completion of gene annotation was 89.4% of 1,614 total gene groups from the Embryophyta dataset.

2.2 Genome comparison and structure analysis

Alignment of the Koshihikari draft genome with the Nipponbare reference genome revealed a highly conserved genome structure (Fig. 2-1A). No large structural variations were detected between the two genomes, although gaps in alignment were observed on chromosomes 3, 4, 9, and 11. Among these gaps, QTLs in EQ and stickiness, qOE3, was previously reported in chromosome 3 in Fig. 2-1B (Takeuchi et al. 2008), and Koshihikari-specific molecular marker associated with EQ, P5, was reported in chromosome 11 (Lestari et al. 2009). Additionally, several regions of low similarity between Nipponbare and Koshihikari genomes were identified in chromosome 11 (Fig. 2-1C).

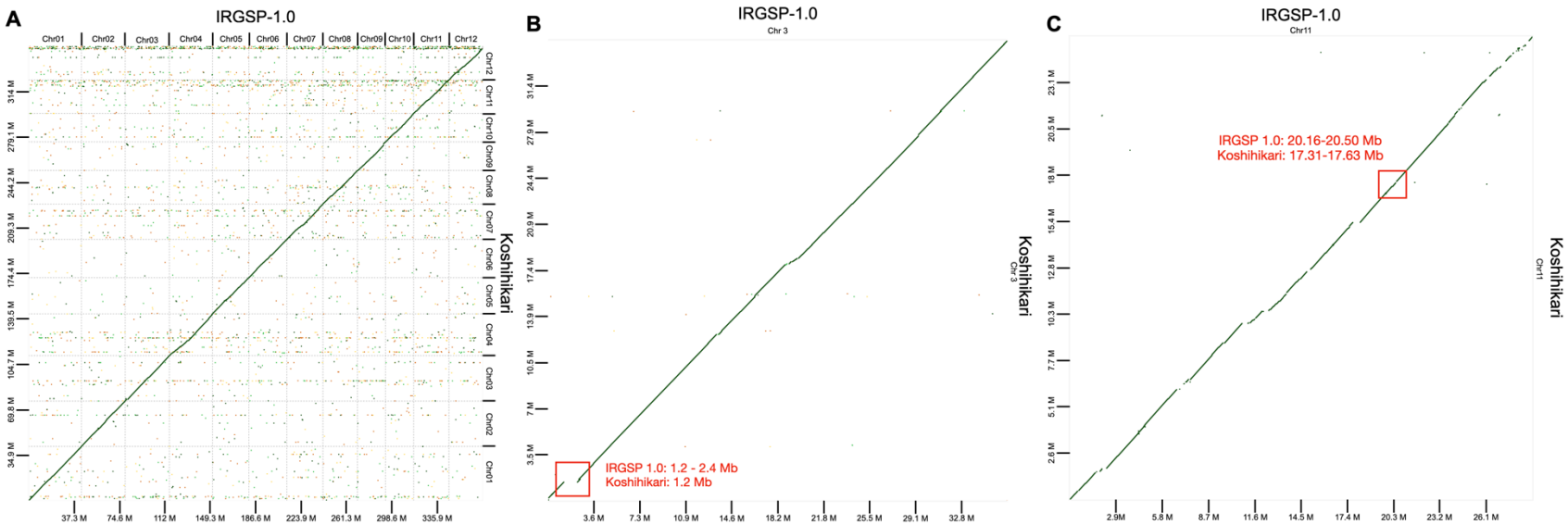


Figure 2-1. Alignment of the Koshihikari draft genome and Nipponbare reference genome sequence.
A. Alignment of all chromosomes, B. chromosome 3, and C. chromosome 11. Alignment blocks with greater than 90 % sequence identity are shown. Red box indicates the candidate region.

A structure of gapped region in qOE3 was depicted in detail by comparison of Koshihikari and Nipponbare sequences using Mauve (Figure 2-2A), and the sequence differences were validated using PCR analysis (Figure 2-3). By comparing the structures, the physical position of 1.25 Mb to 2.33 Mb of Nipponbare sequence was deleted in Koshihikari draft genome. In this considerable size of deleted sequence, 172 annotated genes based on RAP-DB were included. Notably, genes like Os03g0128100, 1,3-beta-glucan synthase component family protein, Os03g0129300, beta subunit of glyceraldehyde-3-phosphate dehydrogenase, and Os03g0141200, beta-amylase PCT-BMYI were part of the deleted region. Moreover, there was about 100 kb of Koshihikari specific insertion sequence based on Nipponbare sequence at 2.46 Mb. The gene prediction analysis of Koshihikari draft genome revealed that 18 genes were predicted in the 100 kb of Koshihikari insertion sequence which functions include glucosaminyl(N-acetyl) transferase (Table 2-4).

The sequence of P5 marker was searched using BLAST and aligned to one of the gaps on chromosome 11 of Koshihikari. The detailed structure of P5 region was explored to determine the sequence variation within the region. The Koshihikari draft genome sequence contained several deletions, insertions, and inversions compared with the Nipponbare reference (Fig. 2-2B) and confirmed with PCR analysis (Figure 2-3). A 17.55–17.64-Mb region on chromosome 11 of Koshihikari, which contained the P5 marker, showed low sequence similarity with the corresponding region in the Nipponbare genome. PCR was previously performed using P5 markers to Koshihikari related cultivars and Norin 1, one of the parental cultivars of Koshihikari, also had P5 segment indicating that P5 marker region of Koshihikari was derived from Norin 1. This region was also detected from other Koshihikari-derived improved cultivars such as Akitakomachi, Kinuhikari, Yumehikari, and Itadaki (data not shown).

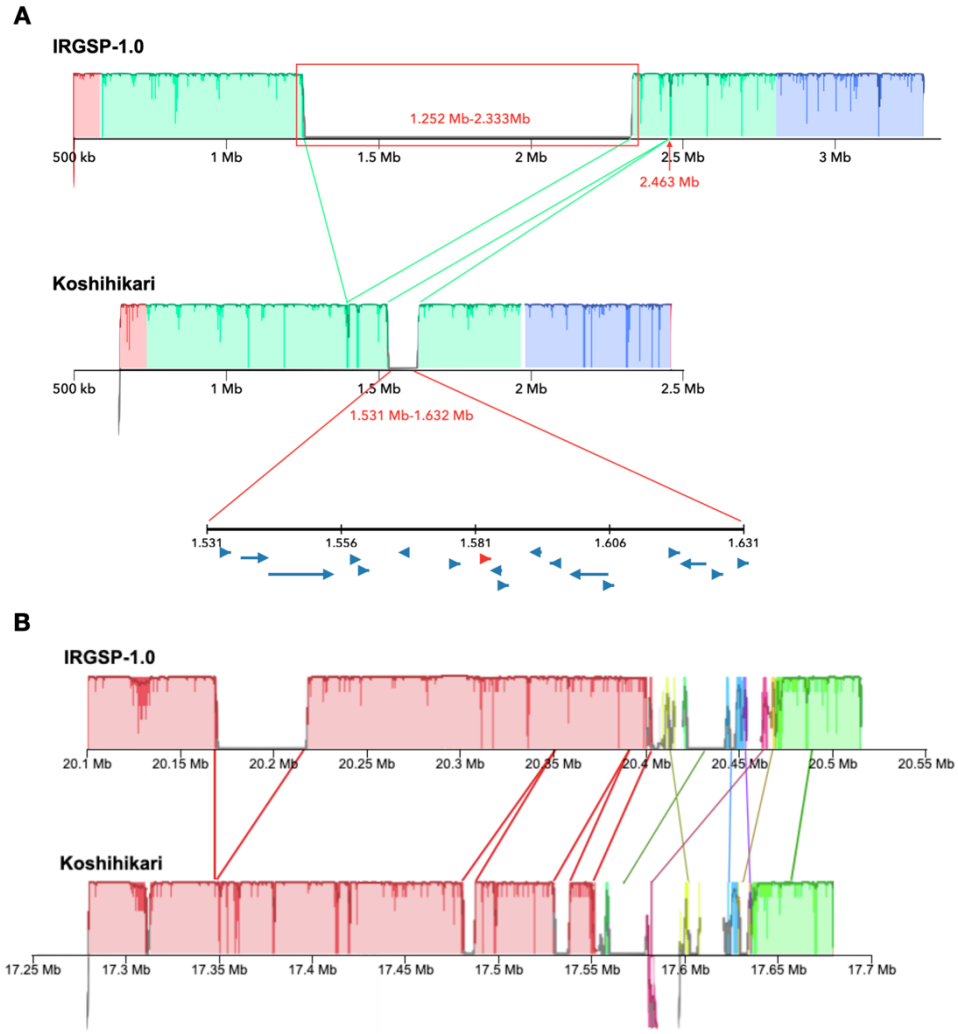


Figure 2-2. Structure analysis of gapped regions.
A. Region nearby qOE3 in chromosome 3. B. Region flanking P5 marker in chromosome 11. Genomic regions showing significant differences in structure are compared. Lines with different colors indicate regions containing deletions, insertions, and inversions. Blue arrows indicate the predicted genes in the region. Gray line labeled with P5 indicate the position of P5 marker in the genome.

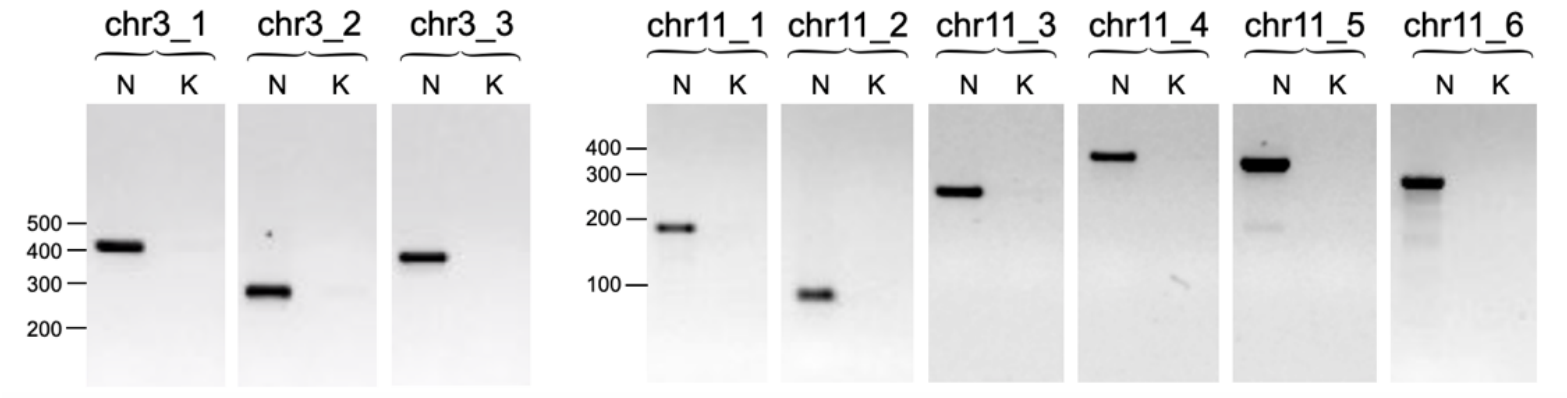


Figure 2-3. Validation of structural variations of Nipponbare and Koshihikari using PCR analysis.
Gel electrophoresis of PCR products in 2% agarose gel. N represents Nipponbare and K represents Koshihikari. The size marker is indicated with the numbers and lines on the left of the gel images.

Table 2-4. Functional annotation of genes in Koshihikari specific region in chromosome 3.

Annotated ID	Start position	End position	Predicted function
K05667	1534174	1534701	Signal peptidase i
K05668	1537838	1542682	Exostosin
K05669	1542915	1552858	Hypothetical protein
K05670	1557361	1558955	Exostosin
K05671	1558991	1559179	Exostosin
K05672	1567964	1568977	leucine Rich Repeat
K05673	1576200	1577013	Transcription factor
K05674	1583326	1584277	Glucosaminyl (N-acetyl) transferase
K05675	1583905	1585964	expressed protein
K05676	1585825	1586043	hypothetical protein
K05677	1592282	1593362	Domain-Containing protein
K05678	1595487	1595900	Domain-Containing protein
K05679	1599198	1600478	F-box and WD repeat domain containing 7, E3 ubiquitin protein ligase
K05680	1601299	1607765	hypothetical protein
K05681	1617230	1618057	disease resistance-responsive family protein
K05682	1618412	1623079	hypothetical protein
K05683	1624488	1627373	hypothetical protein
K05684	1631885	1632412	Signal peptidase i

The genetic diversity of the novel Koshihikari-specific sequences could contribute to phenotypic differences associated with the EQ of cooked rice. Gene prediction analysis revealed that this 17.55–17.64-Mb region harbored numerous protein-coding genes (Figure 2-4), some of which were present in the Nipponbare genome, while others were novel. The function of these genes was annotated as mostly hypothetical or unknown in rice. However, a couple of the predicted genes were annotated as related to the cell wall proteins and polysaccharides, such as glycine-rich cell wall structural protein, peptidase A1 domain-containing protein, and glycosyl transferase family protein (Table 2-5).

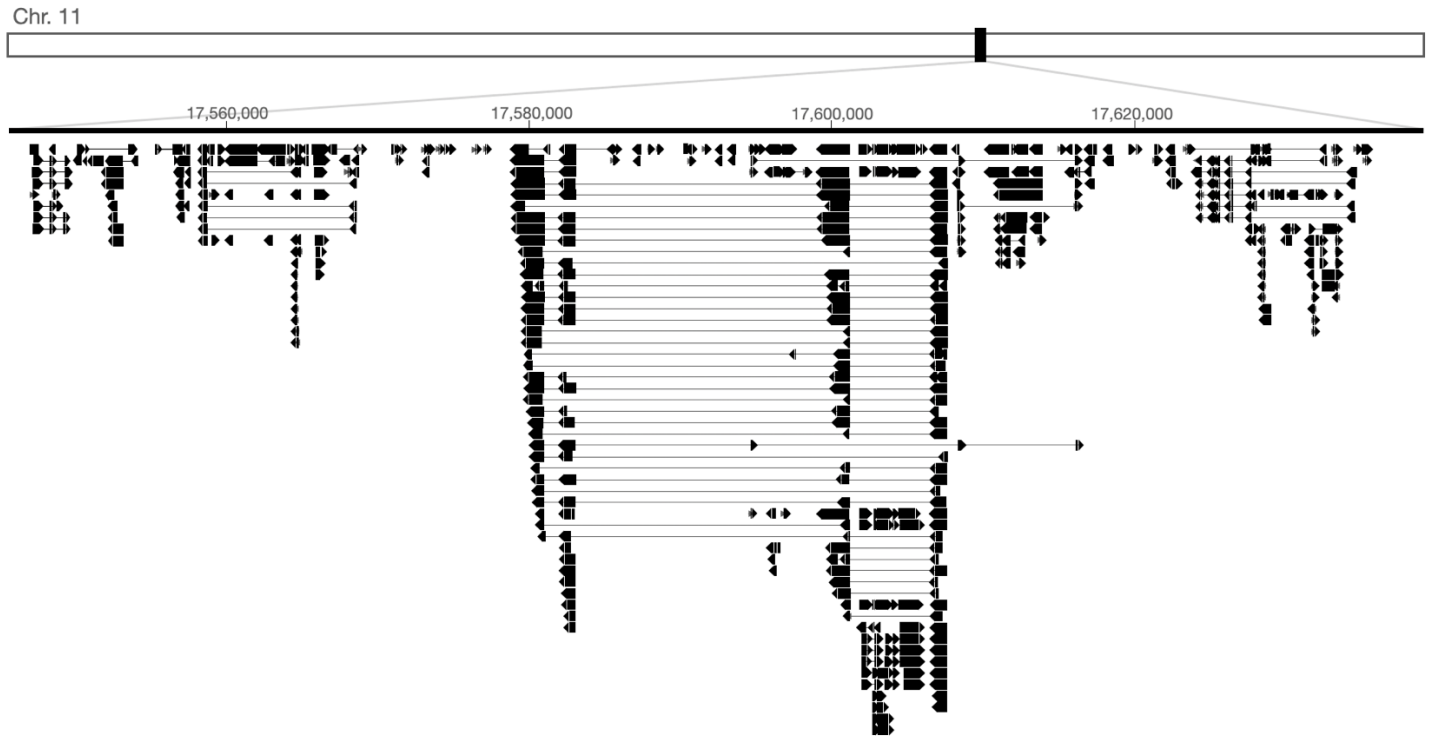


Figure 2-4. Prediction of genes in the novel Koshihikari-specific sequences in chromosome 11.
Visualization of the predicted and gene forms within the candidate region in the Koshihikari draft genome is shown.

Table 2-5. Functional annotation of genes in the candidate P5 locus of the Koshihikari genome.

Annotated ID	Length	Koshihikari position	Nipponbare position	Functional annotation (Oryza)	Other species
Os11t0549900-01	2,373	17,547,213..17,549,585 (+)	chr11:20433314..20437020 (+)	Similar to tubulin alpha-6 chain.	
snap_ gene-58.104-mRNA-1	3,251	17,550,645..17,553,895 (+)			dTDP-4-dehydrorhamnose reductase (Kineosphaera limosa NBRC 100…), Phosphodiesterase (Nematostella vectensis), Secretory lipase (Amycolatopsis cihanbeyliensis)
augustus_ gene-58.34-mRNA-1	3,981	17,550,191..17,554,171 (-)		Protein FAR1-RELATED SEQUENCE (Oryza sativa subsp. japonica)	
Os01t0852650-00	1,224	17,552,022..17,553,245 (-)	chr01:36727945..36729912 (+)	FAR1 DNA binding domain domain containing protein.	
Os07t0627100-01	276	17,552,316..17,552,591 (-)	chr07:25986046..25987423 (+)	MULE transposase, conserved domain domain containing protein.	
Os04t0127900-00	336	17,552,460..17,552,795 (-)	chr04:1760660..1762304 (+)	Conserved hypothetical protein.	
Os10t0571900-00	786	17,552,487..17,553,272 (-)	chr10:22677765..22679745 (-)	Similar to Malic enzyme (Fragment).	
snap_ gene-58.105-mRNA-1	1,457	17,555,396..17,556,852 (+)			Glycosyltransferase family 4 protein (Desertihabitans brevis)
Os07t0191900-01	370	17,556,868..17,557,237 (-)	chr07:4967544..4969280 (+)	Conserved hypothetical protein.	
Os10t0399901-01	271	17,557,319..17,557,589 (-)	chr10:13478227..13482415 (-)	Conserved hypothetical protein.	
Os05t0504800-01	330	17,556,908..17,557,237 (-)	chr05:24835787..24839449 (-)	Similar to Protein kinase domain containing protein.	
Os12t0185900-01	126	17,557,552..17,557,677 (-)	chr12:4334013..4338337 (-)	Similar to Zinc knuckle family protein	
Os09t0321000-01	243	17,556,926..17,557,168 (-)	chr09:9245915..9246868 (+)	Conserved hypothetical protein.	
Koshihikari35799	523	17,556,979..17,557,501 (-)		Uncharacterized protein	UBC core domain-containing protein (Perca flavescens) 51.6%
snap_ gene-58.140-mRNA-1	327	17,558,407..17,558,733 (-)		PGG domain-containing protein (Oryza glumipatula), Os01g0254850 protein (Oryza sativa subsp. japonica), Os01g0254400 protein (Oryza sativa subsp. japonica)	
snap_ gene-58.141-mRNA-1	246	17,558,971..17,559,216 (-)			Lipid A biosynthesis lauroyl acyltransferase (Methylobrevis pamukkalensis), Uncharacterized protein (Sorghum bicolor)
augustus_ gene-58.35-mRNA-1	716	17,558,407..17,559,122 (-)		PGG domain-containing protein (Oryza glumipatula), Uncharacterized protein (Oryza barthii)	PGG domain-containing protein (Zea mays)
Koshihikari35800	6,178	17,558,407..17,564,584 (-)		PGG domain-containing protein (Oryza glumipatula)	
Os01t0254850-00	10,234	17,558,425..17,568,658 (-)	chr01:8471334..8472108 (-)	Hypothetical conserved gene.	
Os01t0254400-00	10,198	17,558,455..17,568,652 (-)	chr01:8455342..8456022 (+)	Hypothetical conserved gene.	
Os03t0255900-01	228	17,558,473..17,558,700 (-)	chr03:8241532..8242113 (-)	Conserved hypothetical protein.	

snap_ gene-58.142-mRNA-1	1,797	17,560,246..17,562,042 (-)		Uncharacterized protein (Oryza meridionalis)	Reverse transcriptase Ty1/copia-type domain-containing protein (Aegilops tauschii subsp. str…), Retrovirus-related Pol polyprotein from transposon TNT 1-94 (Glycine soja), BHLH domain-containing protein (Capsicum chinense)
augustus_ gene-58.36-mRNA-1	3,666	17,560,246..17,563,911 (-)		CCHC-type domain-containing protein (Oryza sativa subsp. indica)	Reverse transcriptase Ty1/copia-type domain-containing protein (Aegilops tauschii subsp. str…), Retrovirus-related Pol polyprotein from transposon TNT 1-94 (Glycine soja), Peptidase A1 domain-containing protein (Capsicum baccatum), BHLH domain-containing protein (Capsicum chinense)
Os05t0442550-01	228	17,560,252..17,560,479 (-)	chr05:21657847..21661906 (-)	Similar to H0801D08.16 protein.	
snap_ gene-58.143-mRNA-1	1,581	17,562,331..17,563,911 (-)		CCHC-type domain-containing protein (Oryza sativa subsp. indica)	Anthocyanidin 3-O-glucosyltransferase 2-like (Hibiscus syriacus)
Os03t0327700-01	258	17,562,841..17,563,098 (-)	chr03:11962246..11966697 (+)	Similar to H0716A07.9 protein.	
Os04t0138432-01	142	17,564,101..17,564,242 (+)	chr04:2539347..2540950 (+)	Conserved hypothetical protein.	
Os05t0247400-01	210	17,564,784..17,564,993 (-)	chr05:8935703..8937728 (-)	Conserved hypothetical protein.	
Os07t0152301-01	854	17,565,685..17,566,538 (+)	chr07:2765292..2770654 (+)	Similar to Calcineurin B-like protein.	
snap_ gene-58.106-mRNA-1	591	17,565,807..17,566,397 (+)		Os07g0152301 protein (Oryza sativa subsp. japonica), Os05g0293600 protein (Oryza sativa subsp. japonica),	
Os05t0293600-01	703	17,565,807..17,566,509 (+)	chr05:12949524..12955005 (+)	Hypothetical protein	
Os05t0354075-00	267	17,565,977..17,566,243 (+)	chr05:16735951..16739911 (+)	Conserved hypothetical protein.	
augustus_ gene-58.37-mRNA-1	1,896	17,566,924..17,568,819 (-)		CCHC-type domain-containing protein (Oryza sativa subsp. indica), PGG domain-containing protein (Oryza glumipatula),	
snap_ gene-58.144-mRNA-1	327	17,567,874..17,568,200 (-)			Alanine and glycine-rich protein (Mytilus californianus), Putative glycine-rich cell wall protein (Arabidopsis thaliana)
snap_ gene-58.145-mRNA-1	198	17,568,622..17,568,819 (-)		PGG domain-containing protein (Oryza glumipatula)	O-fucosyltransferase family protein (Ananas comosus), Putative endo-1,3(4)-beta-glucanase (Beauveria bassiana D1-5)
Koshihikari35801	198	17,568,622..17,568,819 (-)		PGG domain-containing protein (Oryza glumipatula)	low-density lipoprotein receptor-related protein 11 isoform X1 (Pipra filicauda) 59.3%,
snap_ gene-58.107-mRNA-1	213	17,570,978..17,571,190 (+)		Uncharacterized protein (Oryza rufipogon)	
Os01t0612350-01	147	17,573,326..17,573,472 (-)	chr01:24261982..24267594 (-)	Hypothetical gene.	
snap_ gene-58.146-mRNA-1	2,187	17,579,160..17,581,346 (-)		Uncharacterized protein (Oryza nivara), Os11g0550500 protein (Oryza sativa subsp. japonica	Uncharacterized protein (Triticum aestivum)

augustus_ gene-58.38-mRNA-1	3,977	17,579,160..17,583,136 (-)		Uncharacterized protein (Oryza sativa subsp. indica), AAA domain-containing protein (Oryza nivara)	
Koshihikari35802	3,977	17,579,160..17,583,136 (-)		Uncharacterized protein	NB-ARC domain-containing protein (Triticum aestivum)
Os11t0550500-01	28,548	17,579,163..17,607,710 (-)	chr11:20461110..20464673 (+)	Similar to LZ-NBS-LRR class RGA.	
Os02t0301800-02	28,008	17,579,703..17,607,710 (-)	chr02:11701995..11708370 (-)	NB-ARC domain containing protein	
Os02t0272900-01	28,002	17,579,718..17,607,719 (-)	chr02:9923292..9928340 (-)	NB-ARC domain containing protein.	
Os11t0550100-01	27,348	17,579,856..17,607,203 (-)	chr11:20445822..20450284 (-)	Similar to NB-ARC domain containing protein, expressed.	
Os11t0238000-00	27,855	17,579,865..17,607,719 (-)	chr11:7365098..7367861 (+)	Similar to NB-ARC domain containing protein.	
Os12t0553200-01	2,801	17,580,336..17,583,136 (-)	chr12:22488914..22492254 (-)	NB-ARC domain containing protein.	
Os11t0588600-01	825	17,582,312..17,583,136 (-)	chr11:22328967..22333982 (+)	NB-ARC domain containing protein.	
Os11t0589800-00	672	17,582,450..17,583,121 (-)	chr11:22382356..22384989 (+)	Similar to NB-ARC domain containing protein.	
snap_ gene-58.147-mRNA-1	3,197	17,582,291..17,585,487 (-)		Uncharacterized protein (Oryza nivara)	Rx_N domain-containing protein (Triticum turgidum subsp. durum)
Os12t0484100-00	231	17,587,215..17,587,445 (-)	chr12:17865895..17866566 (+)	Hypothetical conserved gene	
Os03t0573001-00	123	17,592,692..17,592,814 (-)	chr03:20886344..20888055 (+)	Conserved hypothetical protein.	
Os12t0502600-01	189	17,593,492..17,593,680 (-)	chr12:19151501..19152088 (+)	Conserved hypothetical protein	
Koshihikari35803	452	17,596,209..17,596,660 (+)		Uncharacterized protein	Starch synthase, chloroplastic/amyloplastic (Tetradesmus obliquus), Polysaccharide deacetylase (Pyrenophora seminiperda CCB06)
snap_ gene-58.108-mRNA-1	663	17,596,781..17,597,443 (+)		Uncharacterized protein (Oryza sativa subsp. indica), Os07g0582700 protein (Oryza sativa subsp. japonica)	Glycoside hydrolase family 16 protein (Streptomyces spectabilis), Glutamate synthase large subunit (Micromonospora sp. HM5-17), glycine-rich cell wall structural protein 2 (Prunus avium)
augustus_ gene-58.13-mRNA-1	1,338	17,597,027..17,598,364 (+)		Uncharacterized protein (Oryza nivara), Os07g0582700 protein (Oryza sativa subsp. japonica)	Glycoside hydrolase family 16 protein (Streptomyces spectabilis), Sugar ABC transporter permease (Streptomyces sp. HC44), glycine-rich cell wall structural protein 2 (Prunus avium)
Os09t0477550-00	610	17,596,077..17,596,686 (-)	chr09:18278585..18280223 (-)	Similar to Mov34/MPN/PAD-1 family protein.	
snap_ gene-58.148-mRNA-1	1,844	17,599,337..17,601,180 (-)		Uncharacterized protein (Oryza nivara), Os11g0550500 protein (Oryza sativa subsp. japonica)	NB-ARC domain-containing protein (Triticum turgidum subsp. durum)
augustus_ gene-58.39-mRNA-1	1,875	17,599,337..17,601,211 (-)		Uncharacterized protein (Oryza nivara), Uncharacterized protein (Oryza sativa subsp. indica)	Putative disease resistance protein (Panicum miliaceum), NB-ARC domain-containing protein (Triticum aestivum)
Koshihikari35804	1,875	17,599,337..17,601,211 (-)		Uncharacterized protein	NB-ARC domain-containing protein (Triticum aestivum)

Os11t0550100-01	7,117	17,600,051..17,607,167 (-)	chr11:20445822..20450284 (-)	Similar to NB-ARC domain containing protein, expressed.	
Os11t0238000-00	7,648	17,600,063..17,607,710 (-)	chr11:7365098..7367861 (+)	Similar to NB-ARC domain containing protein.	
Os02t0272900-01	7,468	17,599,940..17,607,407 (-)	chr02:9923292..9928340 (-)	NB-ARC domain containing protein.	
Os09t0357400-00	6,868	17,600,258..17,607,125 (-)	chr09:11540699..11556180 (-)	Disease resistance protein domain containing protein.	
Os02t0286850-00	6,712	17,600,402..17,607,113 (-)	chr02:10778463..10780758 (-)	Similar to Stripe rust resistance protein Yr10.	
Os06t0716600-00	2,320	17,601,892..17,604,211 (+)	chr06:30439245..30441564 (+)	Hypothetical conserved gene.	
Koshihikari35805	2,342	17,601,892..17,604,233 (+)		Os06g0716600 protein (Oryza sativa subsp. japonica), Uncharacterized protein (Oryza sativa subsp. indica)	NAM-associated domain-containing protein (Panicum hallii var. hallii)
snap_ gene-58.149-mRNA-1	1,191	17,602,032..17,603,222 (-)		Uncharacterized protein (Oryza sativa subsp. indica)	Protein phosphatase (Streptomyces rubidus), Oxalate decarboxylase (Bradyrhizobium erythrophlei)
Os11t0506300-00	2,095	17,602,051..17,604,145 (+)	chr11:18036285..18046195 (-)	Hypothetical conserved gene.	
augustus_ gene-58.14-mRNA-1	2,161	17,602,051..17,604,211 (+)		Uncharacterized protein (Oryza sativa subsp. indica), Os11g0506300 protein (Oryza sativa subsp. japonica)	NAM-associated domain-containing protein (Panicum hallii var. hallii), Myb-like domain-containing protein (Panicum miliaceum)
Os06t0521000-01	659	17,602,735..17,603,393 (+)	chr06:19186999..19192365 (-)	Homeodomain-like containing protein.	
Os01t0373400-01	767	17,602,753..17,603,519 (+)	chr01:15389760..15392040 (-)	Homeodomain-like containing protein.	
Os01t0549300-01	1,130	17,602,753..17,603,882 (+)	chr01:20518681..20521502 (-)	Homeodomain-like domain containing protein.	
Os08t0178900-00	1,130	17,602,753..17,603,882 (+)	chr08:4617579..4620400 (-)	Homeodomain-like domain containing protein.	
Koshihikari35806	1,155	17,604,443..17,605,597 (+)		Uncharacterized protein, GRAS domain-containing protein (Oryza brachyantha)	
snap_ gene-58.109-mRNA-1	1,155	17,604,443..17,605,597 (+)		GRAS domain-containing protein (Oryza brachyantha), Os06g0597250 protein (Oryza sativa subsp. japonica), Os11g0433800 protein (Oryza sativa subsp. japonica)	Ribosomal protein (Trifolium pratense)
augustus_ gene-58.15-mRNA-1	1,155	17,604,443..17,605,597 (+)		Uncharacterized protein (Oryza sativa subsp. indica), GRAS domain-containing protein (Oryza brachyantha)	Ribosomal protein (Trifolium pratense)
Os07t0546750-00	1,338	17,604,536..17,605,873 (+)	chr07:21674780..21678262 (-)	Hypothetical conserved gene.	
Os11t0433800-01	1,314	17,604,560..17,605,873 (+)	chr11:14038904..14042499 (+)	Protein of unknown function DUF635 family protein.	
Os06t0597250-00	1,290	17,604,575..17,605,864 (+)	chr06:23524605..23526405 (+)	Similar to B protein	
Os01t0872900-01	1,065	17,604,800..17,605,864 (+)	chr01:37860133..37861943 (-)	Protein of unknown function DUF635 family protein.	
augustus_ gene-58.40-mRNA-1	861	17,606,829..17,607,689 (-)		Uncharacterized protein (Oryza nivara)	Disease resistance RPP13-like protein 4 isoform X1 (Panicum miliaceum), Rx_N domain-containing protein (Triticum aestivum), Auxin-responsive protein (Hordeum vulgare subsp.

					vulgare), NB-ARC domain-containing protein (Triticum aestivum)
snap_ gene-58.150-mRNA-1	882	17,606,829..17,607,710 (-)		Uncharacterized protein (Oryza nivara)	Disease resistance RPP13-like protein 4 isoform X1 (Panicum miliaceum), Rx_N domain-containing protein (Triticum turgidum subsp. durum)
Koshihikari35807	882	17,606,829..17,607,710 (-)		Uncharacterized protein (Oryza nivara)	Rx_N domain-containing protein (Triticum aestivum)
Os08t0170100-01	837	17,606,874..17,607,710 (-)	chr08:4105460..4107105 (+)	NB-ARC domain containing protein.	
Os02t0301800-01	834	17,606,877..17,607,710 (-)	chr02:11701995..11708370 (-)	NB-ARC domain containing protein.	
Os11t0588600-01	834	17,606,877..17,607,710 (-)	chr11:22328967..22333982 (+)	NB-ARC domain containing protein.	
Os10t0124400-00	834	17,606,877..17,607,710 (-)	chr10:1545297..1548037 (-)	Hypothetical conserved gene	
Os02t0301800-01	834	17,606,877..17,607,710 (-)	chr02:11701995..11708370 (-)	NB-ARC domain containing protein.	
Os11t0589800-00	681	17,607,015..17,607,695 (-)	chr11:22382356..22384989 (+)	Similar to NB-ARC domain containing protein.	
Os04t0385650-01	195	17,608,334..17,608,528 (-)	chr04:18924957..18928259 (+)	Conserved hypothetical protein.	
Os12t0542601-01	211	17,608,349..17,608,559 (-)	chr12:21783684..21786014 (-)	Conserved hypothetical protein.	
Os11t0697600-00	186	17,608,365..17,608,550 (+)	chr11:28560432..28561333 (+)	Hypothetical gene.	
Os09t0414600-01	156	17,608,431..17,608,586 (+)	chr09:14812210..14816389 (-)	Conserved hypothetical protein.	
Koshihikari35808	1,667	17,611,194..17,612,860 (-)		Os12g0550800 protein (Oryza sativa subsp. japonica), Uncharacterized protein (Oryza nivara)	
Os12t0550800-06	1,323	17,610,477..17,611,799 (-)	chr12:22291145..22294711 (-)	Non-protein coding transcript.	
Os12t0550800-01	2,978	17,611,053..17,614,030 (-)	chr12:22291145..22294711 (-)	Hypothetical conserved gene.	
Os12t0550800-05	1,450	17,611,534..17,612,983 (-)	chr12:22291145..22294711 (-)	Conserved hypothetical protein.	
snap_ gene-58.151-mRNA-1	462	17,612,399..17,612,860 (-)		Uncharacterized protein (Oryza rufipogon)	Rhomboid domain-containing protein (Acanthochromis polyacanthus)
snap_ gene-58.110-mRNA-1	219	17,611,933..17,612,151 (+)		Uncharacterized protein (Oryza barthii), Os02g0702351 protein (Oryza sativa subsp. japonica)	
Os12t0550800-02	1,741	17,612,286..17,614,026 (-)	chr12:22291145..22294711 (-)	Conserved hypothetical protein.	
snap_ gene-58.152-mRNA-1	228	17,615,800..17,616,027 (-)			Uncharacterized protein (Eucalyptus grandis), Glutamine amidotransferase type-2 domain-containing protein (Kalmanozyma brasiliensis (st…)
augustus_ gene-58.41-mRNA-1	1,492	17,615,800..17,617,291 (-)		Uncharacterized protein (Oryza nivara), PGG domain-containing protein (Oryza rufipogon)	PGG domain-containing protein (Zea mays)
Os11t0697600-00	187	17,616,119..17,616,305 (+)	chr11:28560432..28561333 (+)	Hypothetical gene.	
Os04t0385650-01	163	17,616,158..17,616,320 (+)	chr04:18924957..18928259 (+)	Conserved hypothetical protein.	
Os09t0414600-01	139	17,616,185..17,616,323 (+)	chr09:14812210..14816389 (-)	Conserved hypothetical protein.	
snap_ gene-58.153-mRNA-1	291	17,617,043..17,617,333 (-)		Uncharacterized protein (Oryza nivara), PGG domain-containing protein (Oryza glumipatula)	PGG domain-containing protein (Zea mays)

snap_ gene-58.154-mRNA-1	231	17,617,361..17,617,591 (-)		PGG domain-containing protein (Oryza punctata)	PGG domain-containing protein (Triticum aestivum), Mannan endo-1,6-alpha-mannosidase (Rachicladosporium antarcticum)
Os01t0254850-00	387	17,617,076..17,617,462 (-)	chr01:8471334..8472108 (-)	Hypothetical conserved gene.	
Os01t0254400-00	360	17,617,103..17,617,462 (-)	chr01:8455342..8456022 (+)	Hypothetical conserved gene.	
augustus_ gene-58.42-mRNA-1	354	17,618,308..17,618,661 (-)		Nicotianamine synthase 2 (Oryza sativa subsp. japonica), Nicotianamine synthase 1 (Oryza sativa subsp. japonica)	Uncharacterized protein (Triticum aestivum)
snap_ gene-58.155-mRNA-1	432	17,618,308..17,618,739 (-)		Nicotianamine synthase (Oryza meyeriana var. granulata), Nicotianamine synthase (Oryza sativa subsp. indica)	Uncharacterized protein (Triticum aestivum)
snap_ gene-58.112-mRNA-1	294	17,621,401..17,621,694 (+)		Nicotianamine synthase 1 (Oryza sativa subsp. japonica)	
augustus_ gene-58.43-mRNA-1	336	17,622,457..17,622,792 (-)		Uncharacterized protein (Oryza barthii), Expressed protein (Oryza sativa subsp. japonica) Os11g0550800 protein	50S ribosomal protein 6, chloroplastic (Panicum miliaceum), Plastid-specific ribosomal protein 6 (Zea mays)
snap_ gene-58.156-mRNA-1	387	17,622,457..17,622,843 (-)		Uncharacterized protein (Oryza rufipogon), Os05g0370300 protein (Oryza sativa subsp. japonica)	
Os11t0550800-01	384	17,622,460..17,622,843 (-)	chr11:20478291..20478867 (-)	Similar to plastid-specific ribosomal protein 6.	
snap_ gene-58.113-mRNA-1	4,298	17,623,684..17,627,981 (+)			Uncharacterized protein (Actinomyces sp. S6-Spd3)
Koshihikari35809	3,967	17,624,150..17,628,116 (-)		Uncharacterized protein (Oryza nivara), Os11g0551350 protein (Oryza sativa subsp. japonica)	Magnesium transporter MRS2-I-like (Hibiscus syriacus)
Os11t0550900-01	10,654	17,624,150..17,634,803 (-)	chr11:20482685..20488052 (-)	Conserved hypothetical protein.	
augustus_ gene-58.44-mRNA-1	3,673	17,624,444..17,628,116 (-)		Uncharacterized protein (Oryza rufipogon), Uncharacterized protein (Oryza sativa subsp. indica)	Uncharacterized protein (Triticum aestivum), Uncharacterized protein (Zea mays)
Os11t0551350-01	388	17,627,729..17,628,116 (-)	chr11:20501771..20503074 (-)	Hypothetical conserved gene.	
Os06t0324000-01	129	17,628,580..17,628,708 (-)	chr06:12651333..12652636 (-)	Conserved hypothetical protein.	
Koshihikari35810	531	17,628,622..17,629,152 (-)		ABC transporter domain-containing protein (Oryza barthii), Os06g0206401 protein (Oryza sativa subsp. japonica), Uncharacterized protein (Oryza nivara), Glycylpeptide N-tetradecanoyltransferase (Oryza glumipatula)	1-phosphatidylinositol 4-kinase (Oryza rufipogon), ULP_PROTEASE domain-containing protein (Oryza rufipogon)
snap_ gene-58.157-mRNA-1	531	17,628,622..17,629,152 (-)		Uncharacterized protein (Oryza barthii), Os06g0206401 protein (Oryza sativa subsp. japonica), ABC transporter domain-containing protein (Oryza barthii), Glycylpeptide N-tetradecanoyltransferase (Oryza glumipatula)	

snap_ gene-58.114-mRNA-1	294	17,628,061..17,628,354 (+)		Uncharacterized protein (Oryza punctata)	
snap_ gene-58.158-mRNA-1	279	17,629,769..17,630,047 (-)		Os09g0447866 protein (Oryza sativa subsp. japonica), Uncharacterized protein (Oryza rufipogon)	Uncharacterized protein (Arabidopsis thaliana)
Os02t0159000-00	3,958	17,628,910..17,632,867 (-)	chr02:3207416..3208323 (+)	Conserved hypothetical protein.	
snap_ gene-58.159-mRNA-1	384	17,630,055..17,630,438 (-)		DUF834 domain-containing protein (Oryza meridionalis)	Glycine-rich cell wall structural protein 1.8 (Phaseolus vulgaris)
augustus_ gene-58.45-mRNA-1	468	17,630,055..17,630,522 (-)		DUF834 domain-containing protein (Oryza barthii),	Glycine-rich cell wall structural protein 1.8 (Phaseolus vulgaris), Histone-glutamine methyltransferase (Peniophora sp. CONT)
snap_ gene-58.160-mRNA-1	1,109	17,630,491..17,631,599 (-)		Os10g0208750 protein (Oryza sativa subsp. japonica), Os05g0182650 protein (Oryza sativa subsp. japonica)	Thiamine pyrophosphate-dependent dehydrogenase E1 component subunit alpha (Baekduia soli), Peptidoglycan-associated protein (Lampropedia puyangensis), Voltage-gated potassium channel subunit beta-2 (Xenopus laevis)
augustus_ gene-58.16-mRNA-1	1,731	17,631,598..17,633,328 (+)		Uncharacterized protein (Oryza sativa subsp. indica)	CRC domain-containing protein (Porphyra umbilicalis), Synapsin I (Lipotes vexillifer), Peptidyl-prolyl cis-trans isomerase (Raphidocelis subcapitata)
Os04t0336801-00	330	17,631,651..17,631,980 (-)	chr04:15822077..15823819 (+)	Conserved hypothetical protein.	
Os04t0335200-01	330	17,631,651..17,631,980 (-)	chr04:15688728..15690470 (+)	Conserved hypothetical protein.	
Os04t0225400-01	192	17,631,818..17,632,009 (-)	chr04:8370325..8373574 (+)	Hypothetical protein.	
Os02t0286891-01	126	17,631,864..17,631,989 (+)	chr02:10787293..10789508 (+)	Similar to OSIGBa0161P06.1 protein.	
Os12t0432301-01	138	17,631,875..17,632,012 (+)	chr12:13964313..13968145 (-)	Conserved hypothetical protein.	
Os01t0392600-01	181	17,632,335..17,632,515 (+)	chr01:16693961..16700566 (+)	Conserved hypothetical protein.	
snap_ gene-58.115-mRNA-1	852	17,632,477..17,633,328 (+)		Os04g0446401 protein (Oryza sativa subsp. japonica), Uncharacterized protein (Oryza sativa subsp. indica),	DNA helicase (Oreochromis niloticus)
Koshihikari35811	852	17,632,477..17,633,328 (+)		Uncharacterized protein (Oryza sativa subsp. indica)	CRC domain-containing protein (Porphyra umbilicalis), DNA helicase (Oreochromis niloticus)
Os06t0324000-01	144	17,633,383..17,633,526 (+)	chr06:12651333..12652636 (-)	Conserved hypothetical protein.	
Os01t0349000-01	150	17,633,393..17,633,542 (+)	chr01:13911234..13912822 (-)	Conserved hypothetical protein.	
Os02t0216150-00	174	17,633,395..17,633,568 (+)	chr02:6496937..6498264 (+)	Conserved hypothetical protein.	
Os07t0112600-01	141	17,633,402..17,633,542 (+)	chr07:697413..698927 (-)	Conserved hypothetical protein.	
snap_ gene-58.116-mRNA-1	366	17,635,141..17,635,506 (+)		Uncharacterized protein (Oryza nivara)	Glutaminase (Acanthamoeba castellanii str…)

2.3 Read depth analysis

Koshihikari has been widely used as breeding material, however not all Koshihikari-derived improved cultivars show good eating quality as Koshihikari. To identify Koshihikari specific sequences that are significantly associated to the eating quality, the whole genome sequencing read depths of Koshihikari-derived cultivars in the sequence variation regions were demonstrated and compared with the Toyo taste-values of the cultivars (Figure 2-5 and Figure 2-6). The improved cultivars which Koshihikari was used as breeding material in generation of the variety were selected, and divided into two groups based on the EQ. The Toyo taste-value of low EQ group was ranged from 47 to 58, 54.8 in average, and that of high EQ group was from 64 to 76, 68.9 in average (Figure 2-5B). As a result, only P5 region showed coherent segregation of presence/absence and eating quality evaluation (Figure 2-5A). Out of 8 low EQ cultivars, only Hatsuboshi was aligned to Koshihikari draft genome in P5 marker region. On the other hand, 7 out of 8 high EQ cultivars were aligned to Koshihikari draft genome, except for Sinboi 3 (Figure 2-5A, C). The results clearly indicated that the presence of P5 sequence in the genome could lead to the improvement in EQ. Other sequence variation regions like qOE3 region in chromosome 3 and regions nearby P5 marker region in chromosome 11 didn't show clear cosegregation as P5 region (Figure 2-6).

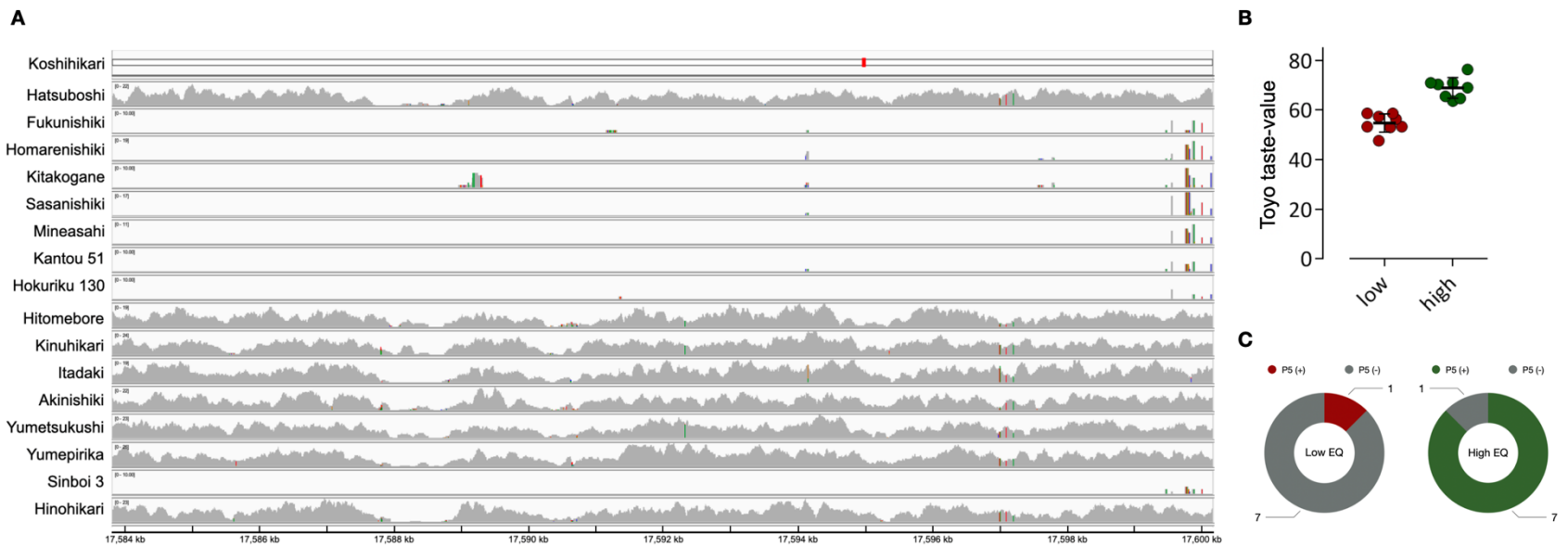


Figure 2-5. Comparison of EQ and presence/absence of P5 sequence in Koshihikari-derived cultivars.
A. Sequencing reads alignment depth of Koshihikari-derived cultivars in P5 marker region B. Toyo taste-values of EQ groups. C. Comparison of presence/absence based on EQ groups.

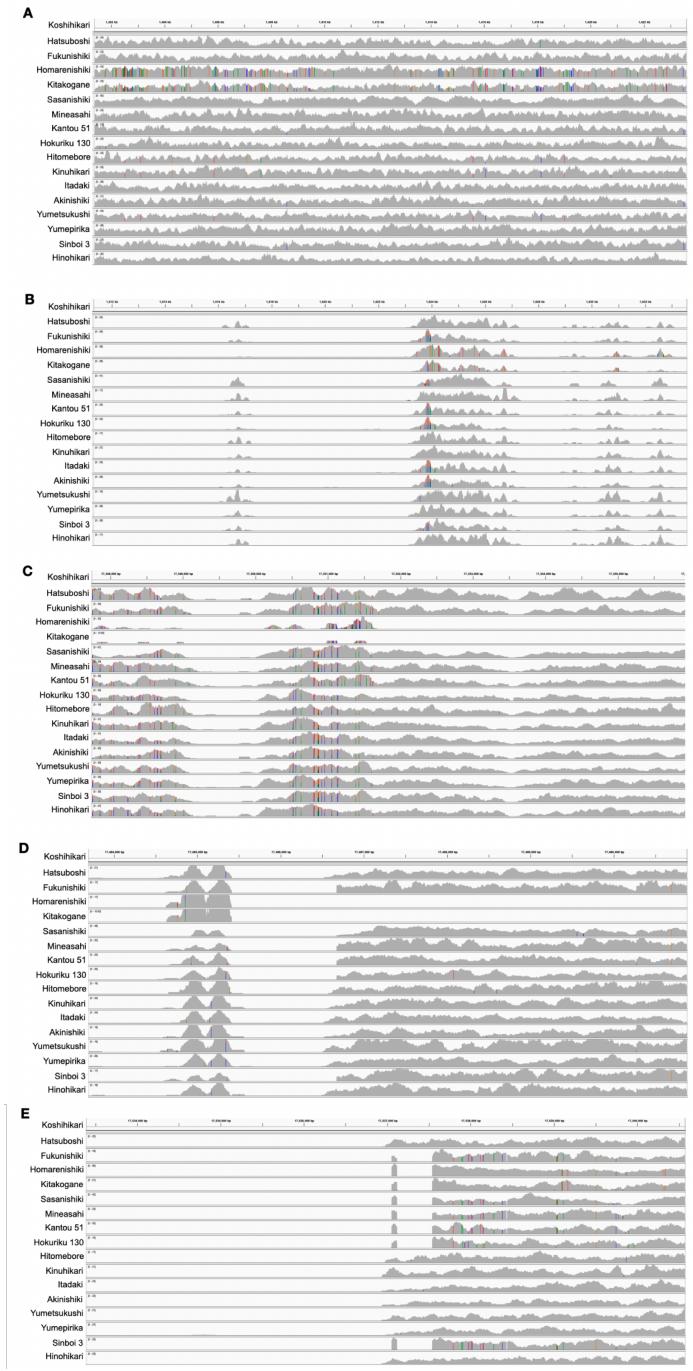


Figure 2-6. Comparison of presence/absence in sequence variation regions.
A. Deletion region in Chr. 3 B. Insertion region in Chr. 3 C. Deletion region in Chr.11 D. First insertion region in Chr. 11 E. Second insertion region in Chr. 11.

2.4 Association of the P5 region with eating quality

To validate the association of the Koshihikari-specific sequence of the P5 marker region with its good EQ, NILs of Samnam (a low EQ cultivar) harboring the Koshihikari-specific P5 region were developed. NILs that show the highest homogeneity to Samnam, 95.8 %, 96.8 %, and 96.9 % of recurrent parent genome recovery (Additional file 1: Table S5), respectively, were selected, and several traits related to the EQ of NILs were compared with that of the parental cultivars (Table 2-6). The differences between Samnam and Koshihkari were noticeable in most of the traits except for the amylose content and breakdown viscosity. NILs didn't show significant differences compared to that of Samnam in most of the traits. However, the Toyo-taste values of NILs were notably higher than Samnam, ranged from 51.7 to 57.7, which indicate that the EQ of NILs was generally improved by Koshihikari-specific P5 segments.

Table 2-6. The phenotypic comparison of NILs and parent cultivars.

	Samnam	Koshihikari	NIL-1	NIL-2	NIL-3
Amylose content (%)	16.6[a]	17.05[a]	16.55[a]	17.45[a]	17.0[a]
Protein content (%)	7.8[a]	6.55[b]	7.5[a]	7.6[a]	7.5[a]
Maximum viscosity (RVU)	250.6[b]	290.1[a]	248.9[b]	242.3[b]	241.8[b]
Minimum viscosity (RVU)	149.4[c]	176.8[a]	143.9[c]	149.2[c]	154.1[b]
Final viscosity (RVU)	251.2[b]	274.7[a]	247.0[c]	257.9[b]	255.6[b]
Breakdown viscosity (RVU)	101.1[a]	113.4[a]	104.9[a]	93.1[b]	87.7[c]
Gelatinization temperature (℃)	84.4[b]	82.1[a]	84.1[b]	85.2[b]	85.2[b]
Toyo taste-value	50.7[c]	65.3[a]	51.7[c]	55.3[b]	57.7[b]

The different subscripts indicate significant differences of NILs and parental cultivars at $p < 0.05$ on each phenotype by Tukey's HSD test.

Discussion

With the increasing popularity of long-read sequencing in genomics studies, the related pipelines and methods have been developed (Pennisi 2017; Huddleston et al. 2014; Mahmoud et al. 2019; Amarasinghe et al. 2020). The unresolved genomic regions of Koshihikari were analyzed using a long-read and short-read hybrid approach. With a moderate sequencing depth, sequence correction using Illumina short reads, and reference guided hybrid assembly, a significantly improved, high-quality Koshihikari draft genome sequence was assembled. The draft genome of Koshihikari assembled in this study showed high contiguity. However, generating a satisfactory genome assembly requires the use of additional technologies such as Bionano optical mapping and Hi-C sequencing, especially for more complex genomes (Choi et al. 2020; Etherington et al. 2020). Although the Koshihikari draft genome assembled in this study contains numerous contigs, the Nipponbare reference-guided approach compensated for these contigs and led to chromosome-level scaffolding.

Cultivars can be distinguished based on their unique features. The most genomics and genetics studies are conducted to understand the biology behind such variations. The current study attempted to elucidate the causal cultivar-specific genomic regions that contribute to the good EQ of Koshihikari by taking advantage of long-read sequencing. Several genomic regions of Koshihikari showed structural variations compared with Nipponbare. The gapped region in short arm of chromosome 3 was previously identified QTL from Koshihikari and Nipponbare backcross inbred lines (BILs), qOE3, related to the overall eating quality and stickiness of cooked rice. The specific genome structure of the region was observed in detail in this study. A large deletion from 1.2 Mb to 2.33 Mb and insertion of sequences in size of 100 kb was observed in Koshihikari. The consequences of deletion of some genes like 1,3-beta-glucan synthase component family protein, beta subunit of glyceraldehyde-3-phosphate dehydrogenase, beta-amylase PCT-BMYI and etc. to EQ trait should further be assessed. Takeuchi et al. in 2008 confirmed that the Koshihikari alleles of these major QTLs in short arm of chromosome 3 increased

EQ using chromosome segement substitution line containing the forementioned Koshihikari segment in the Nipponbare background.

One of the gapped regions containing the P5 molecular marker on chromosome 11 was studied closely because the P5 marker was previously developed and used to distinguish the EQ of cooked rice among Koshihikari-derived *japonica* varieties (Ohtsubo et al. 2002; Lestari et al. 2009). The known P5 sequences were aligned to a region which is considerably different from the Nipponbare sequences. Several protein-coding genes were predicted and annotated within a 90-kb candidate region, ranging from 17.55 to 17.64 Mb on the long arm of chromosome 11 (Table 2-5). The functions of some of these genes were annotated as hypothetical. Some of these genes contained domains with unknown functions, such as proline-glycine-glycine conserved motif containing (PGG), no apical meristem (NAM), and nucleotide-binding adaptor shared by apoptotic protease-activating factor-1, R proteins and *Caenorhabditis elegans* death-4 protein (NB-ARC) domains. Little is known about the effects of these domains on the physicochemical properties EQ of rice. However, some genes within the candidate region were also annotated as related to the cell wall proteins and polysaccharides, based on their sequence similarity with proteins from other species including eukaryotes and prokaryotes. Among the eukaryotes, peptidase A1 domain-containing protein from *Capsicum baccatum*, a putative glycine-rich cell wall protein from *Arabidopsis thaliana*, and the glycine-rich cell wall structural protein 1.8 from *Phaseolus vulgaris* showed high sequence similarity with the predicted genes in the region. Similarly, among the prokaryotes, glycosyltransferase family 4 protein which is involved in catalyzing the lipid cycle reactions in cell wall peptidoglycan biosynthesis (Higashi et al. 1967) from the actinobacterium *Desertihabitans brevis*, and Lipid A biosynthesis lauroyl acyltransferase from *Methylobrevis pamukkalensis*, a type of bacterial lipopolysaccharides, which are glycolipids that constitute the outer monolayer of the membranes of most Gram-negative bacteria (Preston et al. 1996; Vorachek-Warren et al. 2002) were annotated. This suggests that cell wall related protein-encoding genes within the P5 molecular marker region might affect the grain quality of rice. In comparison with its parental cultivar, there was no significant difference in various EQ traits of NILs, except for Toyo taste-value. Since Toyo taste-meter measures the glossiness of cooked grain, this suggests that P5 segment is more

associated with other EQ-related factors like structural elements including cell wall components than starch biosynthesis. However, the relationship between the cell wall composition of the endosperm and the EQ of cooked rice needs to be verified.

Koshihikari has been extensively used in breeding program for its superior EQ trait. As a result, numbers of cultivars were generated from Koshihikari or its parents, however, not all the Koshihikari-derived cultivars showed good EQ performance as Koshihikari. In the current study, sequencing reads of most of the high EQ cultivars generated from Koshihikari were aligned to P5 marker region, on the other hand, low EQ cultivars were nearly aligned. In spite of the fact that the cultivars were generated from Koshihikari, presence/absence of P5 segment was associated with the EQ. We also developed NILs to validate the minor effects of Koshihikari-specific P5 segment to the EQ of Koshihikari rice. The Toyo taste-meter measures the glossiness of cooked rice, and the Toyo taste-value is known to be highly correlated with the overall eating quality of cooked rice (Saika 1992). In this study, the Toyo taste-values of NILs ranged from 51.7 to 57.7, which explained approximately 13.8 % of phenotypic variance by the effects of the P5 Koshihikari-specific segments to EQ at most. Although each NIL showed variation in the power of explanation, as expected, it can be explained by environmental factors. Moreover, application of the results to rice breeding will be feasible by developing additional molecular markers in the candidate region. The results of this study are expected to facilitate the molecular breeding of rice cultivars with good EQ, and to help understanding the molecular basis of the quality of rice.

CHAPTER III

Volatile organic compounds related to the eating quality of *japonica* rice (*Oryza sativa* L.)

Abstract

Eating quality (EQ) of rice has a complex nature composed of physicochemical properties. Nevertheless, breeding programs evaluating EQ through sensory test or taste-evaluation instruments have been laborious, time-consuming, and inefficient. EQ is affected by both taste and aroma. However, in actual breeding programs, aroma of cooked rice has been considered the least due to lack of information. Here we identified a total of 41 volatile compounds potentially affecting the EQ of non-aromatic, cooked *japonica* rice, identified by GC-MS, sensory panel test, and Toyo taste-meter analyses. Partial least squares discriminant analysis demonstrated significantly distinctive classification of eating-quality based on the contents of identified volatile compounds. Several volatile compounds related to lipid oxidation and fatty acid degradation were identified to affect the EQ in *japonica* rice. Of them, 1-octen-3-ol, 1-ethyl-3,5-dimethylbenzene, 2,6,11-trimethyldodecane, 3-ethyloctane, 2,7,10-trimethyldodecane, methyl salicylate, 2-octanone, and heptanal were selected as important compounds. The discriminant model for the classification of the quality of cultivars was robust and accurate, an r-squared value was 0.91, a q squared value was 0.85, and an accuracy was 1.0. From the analysis of head rice, total of 55 volatile compounds were profiled. In head rice analysis, 1-Hexanol, Methyl myristate, Palmitic acid, Benzene ethanol, 2,4,7,9-Tetramethyl-5-decyne-4,7-diol, Triisobutyl phosphate, and 2,2,4-Trimethyl-1,3-pentanediol diisobutyrate were identified as important volatile compounds identifying EQ in uncooked rice samples. Overall, characterization of EQ of rice cultivars based on volatile compounds, suggests the application of metabolite profiling data for rice breeding of high eating quality.

Introduction

Rice (*Oryza sativa* L.) is one of the most important agricultural crops and serves as a staple food worldwide. Because of improvements in living standards, the market demand for high -quality rice has been increasing (Custudio et al. 2019). Sensory panel test is a direct method for evaluating the EQ of cooked rice. Preferably gender balanced trained panelists individually rate the cooked rice samples for intensities and preferences of attributes, like appearance, hardness, stickiness, taste or flavor, texture, and overall eating quality (Srisawas and Jindal 2006, Lyon et al. 2000). Despite of its direct and intuitive evaluation in EQ, sensory panel test is time-consuming, labor-intensive, requiring large volume of samples, so it is not applicable to early generation testing in breeding programs. Therefore, it has been replaced by the analysis of physicochemical properties of rice. Starch is a main component of rice endosperm, which consists of amylose and amylopectin (Jeon et al. 2010). Hence starch related traits such as amylose content, gel consistency, and gelatinization temperature have been widely studied (Lu et al. 2019, Tan et al. 1999, Okpala et al. 2022, Ball et al. 1998) and their genetic backgrounds have been determined such as granule-bound starch synthase and starch branching enzymes, representatively (Septiningsih et al. 2003, Xu et al. 2015, Wang et al. 1995, Satoh et al. 2003). However, EQ determined by taste-evaluation instruments and physicochemical analysis have not been so satisfying so far. Considering that EQ is a complicated trait, all attributes related to the human sense are significant in discriminating EQ. While the olfactory sense primarily perceives the information, aroma and flavor are considered two of the main EQ-related factors in rice sensory properties (Del Mundo and Juliano 1981, Meilgaard et al. 2007, Champagne et al. 2010). Empirically, rice breeders recognize that accurate sensory evaluation of cooked rice is hard to achieve when having a stuffy nose.

Attempts have been made to screen and profile the volatile organic compounds (VOCs) of rice; however, little is known about the relationship between these compounds and the flavor or EQ of cooked rice (Champagne 2008). Consequently, aroma and flavor have been restrictively considered in application of identified chemical compounds in evaluation of EQ. Moreover, most of the previous

studies focused on aromatic rice cultivars, such as basmati and tropical japonica rice. For instance, 2-acetyl-1-pyrroline was identified as the VOC responsible for the specific popcorn-like aroma and the characteristic flavor of fragrant rice. In addition, a molecular marker, based on an 8 bp deletion in the fragrance gene, was developed to distinguish between fragrant and non-fragrant rice cultivars Lorieux et al. 1996, Chen et al. 2006, Bradbury et al. 2008).

Over 300 volatile compounds have been identified in rice via analytical chemistry. Among these compounds, a few have been identified as oxidation products and are considered as possible negative contributors to rice flavor (Yajima et al. 1978, Buttery et al. 1988, Zheng et al. 2008). A recent study aimed to detect volatile compounds in cooked japonica rice using solid phase microextraction (SPME) with gas chromatography–resonance-enhanced multiphoton ionization time-of-flight mass spectrometry (GC/REMPI-TOFMS) focused on decreasing the extraction time and comparing the method in detection of particular compounds, 4-vinylphenol and indole (Shinoda et al. 2020). Zhang et al. (2020) performed metabolite profiling via headspace (HS)-SPME GC/MS and HS GC/ion mobility spectrometry (IMS) to discriminate between white and yellow rice using partial least squares discriminant analysis (PLS-DA). Consequently, hexanal, nonanal, octanal, 1-pentanol, and 2-pentyl-furan, involved in fatty acid oxygenation, were identified as compounds with high variable importance in projection (VIP) scores. Thus, further research on VOCs of cooked japonica rice cultivar is needed in order to characterize important compounds that affect the general sensory properties, and finally to utilize it for breeding desirable high EQ rice cultivars.

This study aims to identify volatile compounds in cooked rice which significantly affect the EQ of non-aromatic japonica rice cultivars to provide a fundamental information to set the evaluation standard of EQ in association with aromatic compounds for genetic improvement.

Materials and methods

3.1 Plant materials and growth conditions

A total of 24 non-aromatic temperate japonica rice cultivars were selected, based on prior knowledge of their putative EQ: Koshihikari, Chucheong, Gopum, Samkwang, Sindongjin, Ilpum, Saenuri, Cheongpum, Gyehwa, Dongjin, Seomjin, Hwaseong, Nampyeong, Hopum, Yeongdeok, Giho, Nakdong, Nongbek, Hwacheong, Samnam, Palgong, Junam, Saegyehwa, and Namil. All the accessions were retained at the Agricultural Genetic Resource Center, Seoul National University, Suwon, South Korea, and followed the relevant institutional guidelines.

All rice cultivars were cultivated in 2020 in an experimental farm of Seoul National University located in Suwon. The general lowland cultivation method was applied. The heading date of each cultivar was recorded, and plants were harvested in 45-50 days after heading. The harvested plants were air dried until reaching the grain water content of 13–14%, and subsequently threshed using a thresher. The grains were dehulled (using a dehull machine), milled to 92.2% (using a milling machine), and immediately stored in the low-temperature storage at 12°C until the experiment.

3.2 Sensory panel test

Samples were prepared and cooked according to the protocol of the National Institute of Crop Science (NICS), Rural Development Administration (RDA), Korea (Kwon et al. 2011). The milled rice samples were weighed to 200 g, washed five times with tap water, and soaked for 20 min. The water was drained off for 10 min, and the macerated rice was cooked in 1.2 volumes of water (rice: water = 1:1.2 w/w) using the automatic cooking cycle of an electric rice cooker. Subsequently, the cooked rice was mixed thoroughly in the rice cooker and allowed to sit for 15 min. Sensory evaluation was performed by a panel of 14 trained members. The EQ of each cultivar was scored from +2 (very good) to -2 (very bad) in comparison with that of Chucheong (reference sample; score = 0), and the average value was computed for each cultivar.

3.3 Toyo taste-meter analysis

The Toyo taste-meter (MA-30A; Toyo Rice Corporation, Wakayama, Japan) was used to measure the EQ of cooked rice. The Toyo taste-value is known to be significantly correlated with the palatability of cooked rice (Kwak et al. 2015). Head rice (33 g) was cooked at 80 ℃ for 10 min, and then allowed to sit at room temperature for 5 min. The glossiness of the surface of the cooked rice grains was measured in triplicate using certain electromagnetic waves, and then converted into the Toyo taste-value.

3.4 Untargeted profiling of volatile compounds

For the profiling of volatile compounds in rice, head rice and cooked rice were used and compared. The milled rice samples were cooked according to the NICS protocol, with some modification. Briefly, milled rice samples (3 g) were quantified, washed, and soaked in tap water for 20 min. After draining the water, each rice sample was transferred to a 20 ml glass vial (CTC, Perkin Elmer & Agilent), to which 3.6 ml distilled water and 2 μl of 2,000 ppm 1,2,3-trichloropropane (Sigma-Aldrich, St. Louis, MO; internal standard) were added. The vials were closed tightly with a magnetic crimp cap using PTFE/silicone septa, and vortexed vigorously. Then, the sealed glass vials were placed in an electronic rice cooker, and samples were cooked for 25 min.

VOCs were analyzed using the HS-SPME injector-equipped Thermo Scientific Trace 1310 Gas Chromatograph, TSQ 8000 Triple quadrupole Mass Spectrometer and TriPlus RSH autosampler (Waltham, MA, USA) with a DB-Wax capillary column (60 m × 0.25 mm, 0.50 μm film thickness; Agilent Technologies). The samples were incubated for 10 min at 70 ℃. Headspace volatiles of the cooked rice sample were adsorbed by inserting DVB/CAR/PDMS StableFlex SPME fiber (2 cm, 50/30 μm; Supelco, Bellefonte, PA, USA) into the vial for 50 min at 70 ℃ with agitation, and then desorbed for 2 min. Blanks were run after every 10 samples as a control. The injector temperature was 250 ℃. The oven temperature of GC was as follows: initial temperature of 40 ℃ held for 2 min; increased to 150 ℃ at 3.0℃/min and held for 10 min; increased to 200 ℃ at 3.0℃/min and held for 5 min; and increased to 230 ℃ at 6.0℃/min and held for 5 min. Research-grade helium was

used as the carrier gas at a constant flow rate of 2.0 ml/min in the splitless with purge mode. The mass spectrometer was set to the electron impact mode at 230 ℃ with 70 eV and scanned at 35–550 m/z. Analysis was performed in triplicate. The compounds were identified by comparing the chromatogram and retention indices with the reference in the database, NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library, version 2.0g (National Institute of Standards and Technology, Gaithersburg, MD, USA), with a match score cutoff of at least 80%. The data obtained were processed using the Xcalibur software (Thermo Fisher Scientific, Waltham, MA, USA). The ion peak area of the identified compound was divided by that of 1,2,3-trichloropropane (internal standard), and the calculated area ratio was used for further statistical analysis.

3.5 Statistical analysis

All data were presented as mean values. Pearson's correlation and Spearman correlation analyses were performed using a package from the RStudio 1.1.453 software (R Foundation for Statistical Computing, Vienna, Austria). To perform multivariate statistical analysis, the relative peak area of the identified volatile compounds was normalized to quantile, scaled to mean centered, and divided by the standard deviation of each variable (see Supplementary Figure S3 and S4 online). The resultant data were then subjected to PLS-DA using MetaboAnalyst 5.0 (Pang et al. 2021).

Results and Discussion

3.1 EQ evaluation

The EQ of 24 non-aromatic temperate japonica rice cultivars was evaluated in this study, based on the sensory panel test and Toyo taste-meter readings (Figure 3-1). According to the results of the sensory panel test conducted by 14 trained panelists, Samkwang was the best EQ cultivar, followed by Ilpum, Gopum, Koshihikari, and Cheongpum. The least favored cultivars were Namil, followed by Samnam and Palgong. Meanwhile, EQ measured by Toyo taste-meter was the highest in Saenuri followed by Cheongpum, Gopum, Samkwang, and Ilpum. On the other hand, Yeongdeok showed the lowest Toyo taste-value, followed by Namil, Samnam, Saegyehwa, and Palgong. A study conducted by Lestari et al. (2009) evaluating the eating quality of 22 japonica rice used Toyo taste-values in marker-trait regression analysis. The used cultivars were not completely identical, but Gopum, Samkwang, Ilpum, and Koshihikari exhibited high Toyo taste-value above 70. Similarly, low Toyo taste-value was observed around 60 in Palgong and Samnam in the study. Another kind of taste-evaluation instrument, Satake taste analyzer (Satake, STA1B-RHS1A-RFDM1A, Japan) generates taste score of cooked rice for EQ evaluation (Champagne et al. 1996, Ma et al. 2020, Shi et al. 2022), and were successively used in evaluating 533 rice accessions for the genome-wide association study (Zhou et al. 2021). Although there was a strong correlation between EQs by the sensory panel test and Toyo taste-meter (Figure 3-2), Toyo taste-meter which is widely used to evaluate the EQ of cooked rice for its relative convenience in comparison with sensory test (Saika 1992, Sun et al. 2011) may not accurately evaluate the EQ of cooked japonica rice. The differences in ranking by two methods could be explained by the fact that Toyo taste-meter only consider the glossiness of the surface of cooked rice grains. This suggests that flavor and aroma should be considered together to overcome the limit of the conventional EQ evaluation methods like Toyo taste-meter which measures and generates the numerical value by the appearance of cooked rice.

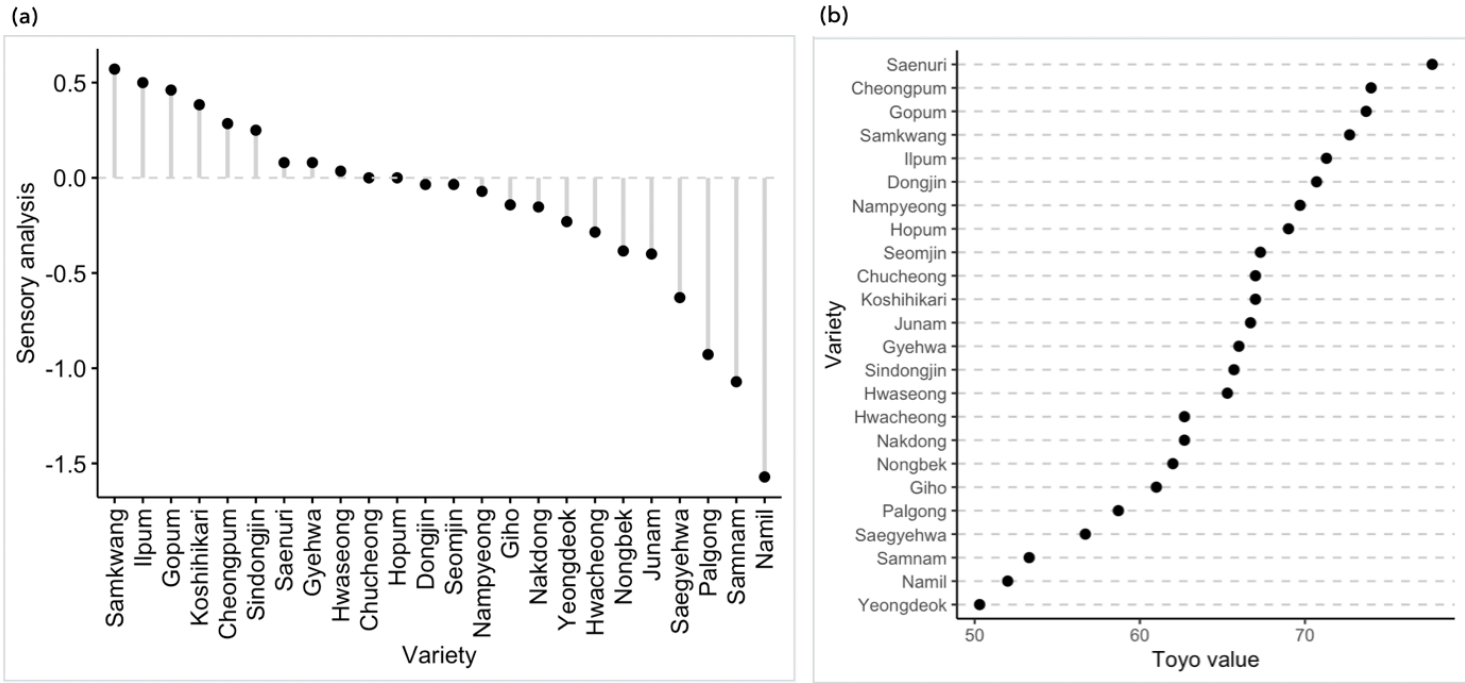


Figure 3-1. Evaluation of the EQ of 24 rice cultivars using the sensory panel test and Toyo taste-meter.
EQ of rice cultivars measured by the sensory panel test (A) and Toyo taste-meter (B).

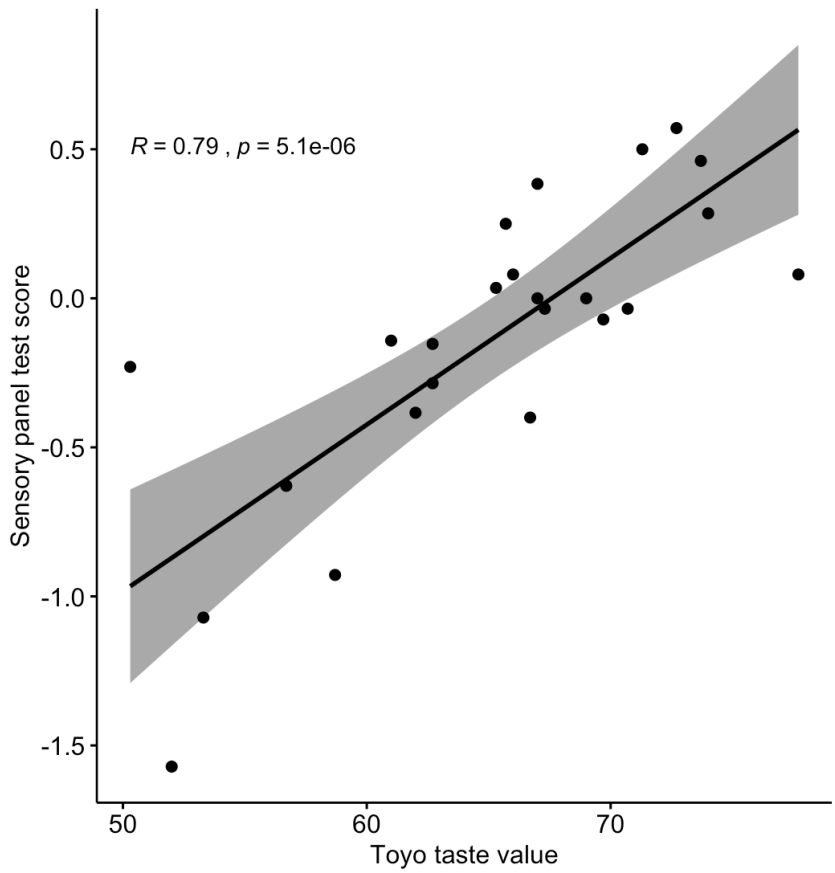


Figure 3-2. Rank correlation analysis of sensory panel test results and Toyo taste value readings using 24 cultivars.

3.2 Identification of volatile compounds in cooked rice

Based on the results in Figure 3-1A, seven cultivars with higher EQ (Samkwang, Ilpum, Gopum, Koshihikari, Cheongpum, Sindongjin and Saenuri), and seven cultivars with lower EQ (Namil, Samnam, Palgong, Nongbek, Hwacheong, Yungdeok, and Giho) compared to the check variety, were selected for analysis of the volatile compounds that affect the EQ of rice. The cooked rice of 14 japonica cultivars was analyzed using HS-SPME GC-MS/MS, and 41 volatile compounds were identified as Table 3-1 and quantified (Table 3-2). These compounds were classified into nine different classes: alcohols, aldehydes, carboxylic acid, ester, hydrocarbons, imine, ketones, phenols, terpenoid, and unknown compounds. The most frequently observed classes were hydrocarbons and aldehydes. Among the hydrocarbons, alkanes and alkenes are reportedly derived from lipid breakdown. Nonadecane was previously detected in aromatic and non-aromatic rice cultivars (Bryant and McClung, 2011). Although a number of alkanes and alkenes were detected in the current study, limited information is available on their effects on the flavor of cooked rice. Aldehydes usually exhibit a relatively low odor threshold and are considered one of the main factors affecting the overall flavor profile of cooked rice. Among the identified aldehydes, hexanal exhibits fruity, grass, and green attributes; however, lipid oxidation generates a large amount of hexanal, resulting in off-odors in rice (Bergman et al. 2000, Monsoor and Proctor 2004, Hu et al. 2020). Octanal, heptanal, and nonanal are derived from oleate hydroperoxide decomposition. Moreover, low odor threshold alcohols, such as lipid-derived 1-octen-3-ol and polyunsaturated fatty acid metabolism-related 1-hexanol, were also identified. All identified volatile compounds were subjected to further statistical analysis to identify important volatile features related to the EQ of cooked temperate japonica rice.

Table 3-1. Volatile compounds identified by HS-SPME GC/MS from cooked rice

ID	Compound name	Retention time (min)	Retention index
1	Decane	13.88	below 1000
2	2,7,10-Trimethyl-dodecane	15.23	1027
3	3-Ethyloctane	16.38	1052
4	2,6,11-Trimethyl-dodecane	17.63	1079
5	Hexanal	18.51	1098
6	Dodecane	23.42	1200
7	Heptanal	23.56	1203
8	Unknown_71	25.03	1234
9	Unknown_57	27.77	1291
10	Nonadecane	28.21	1300
11	2-Octanone	28.37	1304
12	Octanal	28.60	1309
13	Unknown_69	29.83	1335
14	(E)-2-Heptenal	30.48	1349
15	6-Methyl-5-hepten-2-one	30.89	1358
16	1-Hexanol	31.14	1363
17	2-Methyl-1-pentadecene	31.14	1364
18	1,3,5-Trimethyl-benzene	31.29	1367
19	1-Ethyl-3,5-dimethyl-benzene	31.89	1380
20	1-Methyl-4-(1-methylethyl)-benzene	32.23	1387
21	2-Ethyl-1,4-dimethyl-benzene	32.60	1395
22	Nonanal	33.47	1414
23	Unknown_57(2)	34.09	1428
24	Unknown_43	34.73	1443
25	1,3-Bis(1,1-dimethylethyl)-benzene	34.87	1446
26	1,2,3,5-Tetramethyl-benzene	35.32	1456
27	1-Octen-3-ol	35.51	1461
28	3-Methyl-tetradecane	35.60	1463
29	1,2,3,4-Tetramethyl-benzene	35.82	1467
30	2-(3-Methylbutyl)-thiophene	36.65	1486
31	1-Octanol	40.24	1566
32	2,6,10,15-Tetramethyl-heptadecane	41.77	1600
33	4-Methyl-benzaldehyde	46.74	1685
34	Methoxy phenyl oxime	51.57	1760
35	Methyl salicylate	54.96	1814
36	trans-Calamenene	57.74	1865
37	6,10-Dimethyl-5,9-undecadien-2-one	58.37	1876
38	2-methyl-propanoic acid, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester	59.25	1893
39	Butylated hydroxytoluene	61.27	1935
40	2-Methoxy-5-vinylphenol	74.41	2232
41	2,4-Bis(1,1-dimethylethyl)-phenol	77.72	over 2300

Table 3-2. The quantified VOCs from cooked rice calculated into area ratio using internal standard.

ID	1	2	3	4	5	6	7	8	9	10
Compound name	Decane	2,7,10-Trimethyl-dodecane	3-Ethyloctane	2,6,11-Trimethyl-dodecane	Hexanal	Dodecane	Heptanal	Unknown_71	Unknown_57	Nonadecane
Class	Alkane	Alkane	Alkane	Alkane	Aldehyde	Alkane	Aldehyde	Unknown	Unknown	Alkane
Koshihikari	0.0226667	0.004	0.004	0.0046667	0.07	0.0256667	0	0.0096667	0.0076667	0.0036667
Saenuri	0.0203333	0.004	0.0036667	0.004	0.0933333	0.02	0.0066667	0.0083333	0.006	0.0026667
Gopum	0.0193333	0.0043333	0.0033333	0.004	0.068	0.0213333	0.0046667	0.009	0.007	0.003
Sindongjin	0.0505	0.009	0.00725	0.00725	0.13475	0.0255	0.00675	0.00875	0.00625	0.00325
Ilpoom	0.0503333	0.0093333	0.007	0.008	0.1636667	0.0213333	0.0086667	0.0083333	0.0056667	0.002
Samkwang	0.0236667	0.0046667	0.0043333	0.0043333	0.1176667	0.026	0.0063333	0.0076667	0.0063333	0.002
Chungpoom	0.046	0.0093333	0.006	0.0076667	0.085	0.0163333	0.0056667	0.0073333	0.0046667	0.002
Hwacheong	0.0086667	0.001	0.0013333	0.002	0.13	0.0166667	0.008	0.007	0.007	0.0036667
Yungdeok	0.0126667	0.0013333	0.002	0.0026667	0.184	0.0253333	0.0126667	0.007	0.006	0.004
Samnam	0.0073333	0	0.001	0.0013333	0.144	0.0113333	0.0066667	0.0046667	0.004	0.002
Namil	0.0056667	0	0.001	0.001	0.112	0.008	0.0056667	0.0023333	0.002	0.002
Giho	0.0093333	0.0016667	0.0016667	0.0016667	0.1076667	0.0126667	0.0073333	0.0033333	0.003	0.0023333
Nongbek	0.011	0.002	0.002	0.0016667	0.0863333	0.0093333	0.0046667	0.0033333	0.003	0.002
Palgong	0.011	0.002	0.002	0.002	0.1273333	0.012	0.0063333	0.005	0.0043333	0.0023333

Table 3-2. Continued.

11	12	13	14	15	16	17	18	19	20
2-Octanone	Octanal	Unknown_69	(E)-2-Heptenal	6-Methyl-5-hepten-2-one	1-Hexanol	2-Methyl-1-pentadecene	1,3,5-Trimethyl-benzene	1-Ethyl-3,5-dimethyl-benzene	1-Methyl-4-(1-methylethyl)-benzene
Ketone	Aldehyde	Unknown	Aldehyde	Ketone	Alcohol	Hydrocarbon	Aromatic hydrocarbon	Aromatic hydrocarbon	Aromatic hydrocarbon
0.003	0.0073333	0.007	0.007333	0.004667	0.001	0.0013333	0.0133333	0.0086667	0.0113333
0.003	0.008	0.0066667	0.003333	0.006	0.002333	0.0023333	0.0113333	0.0073333	0.0096667
0.002	0.009	0.0043333	0.006667	0.003667	0.002	0.002	0.0086667	0.005	0.007
0.00325	0.012	0.0045	0.011	0.008	0.0045	0.0045	0.0175	0.012	0.01625
0.003667	0.014333	0.0036667	0.006667	0.006667	0.007667	0.0076667	0.0156667	0.0113333	0.0153333
0.003	0.009	0.002	0.008	0.007333	0.020667	0.0206667	0.0076667	0.005	0.0066667
0.002	0.007667	0.003	0.006	0.006	0.009	0.009	0.0126667	0.009	0.012
0.003667	0.008333	0.001	0.007667	0.004667	0.002667	0.0026667	0.007	0.003	0.0046667
0.008	0.02	0.0013333	0.007333	0.011	0.009	0.009	0.0083333	0.0036667	0.0053333
0.004667	0.009333	0.001	0.007333	0.006	0.005333	0.0053333	0.0066667	0.003	0.004
0.003333	0.007	0.001	0.005333	0.005667	0.002667	0.0026667	0.0056667	0.002	0.0023333
0.002667	0.01	0.002	0.007333	0.0056667	0.004	0.004	0.0046667	0.0026667	0.0036667
0.002667	0.005	0.002	0.005667	0.0043333	0.002333	0.0023333	0.0043333	0.0023333	0.0033333
0.003	0.009	0.0023333	0.006667	0.006667	0.010667	0.0106667	0.005	0.0033333	0.0046667

Table 3-2. Continued.

21	22	23	24	25	26	27	28	29	30	31
2-Ethyl-1,4-dimethyl-benzene	Nonanal	Unknown_57(2)	Unknown_43	1,3-Bis(1,1-dimethylethyl)-benzene	1,2,3,5-Tetramethyl-benzene	1-Octen-3-ol	3-Methyl-tetradecane	1,2,3,4-Tetramethyl-benzene	2-(3-Methylbutyl)-thiophene	1-Octanol
Aromatic hydrocarbon	Aldehyde	Unknown	Unknown	Aromatic hydrocarbon	Aromatic hydrocarbon	Alcohol	Alkane	Aromatic hydrocarbon	Hydrocarbon	Alcohol
0.0283333	0.039	0.009	0.002333	0.0706667	0.038	0.0103333	0.0063333	0.053	0.0133333	0.003667
0.0253333	0.0436667	0.009	0.003	0.061	0.034	0.0136667	0.008	0.0483333	0.0126667	0.004
0.0173333	0.0476667	0.008667	0.003	0.055	0.023	0.01	0.0066667	0.0323333	0.007	0.004
0.041	0.06475	0.00575	0.004	0.10025	0.0495	0.01975	0.0025	0.06875	0.0185	0.00425
0.0396667	0.053	0.005667	0.005667	0.0743333	0.046	0.0173333	0.003	0.0646667	0.0113333	0.004667
0.016	0.041	0.005333	0.004667	0.05	0.0203333	0.0146667	0.002	0.0293333	0.016	0.003667
0.0306667	0.0436667	0.005	0.003	0.0696667	0.0363333	0.01	0.003	0.0516667	0.011	0.003
0.0106667	0.0323333	0.013667	0.005667	0.0353333	0.0143333	0.0176667	0.0106667	0.02	0.0143333	0.003333
0.013	0.0686667	0.014	0.009	0.0403333	0.016	0.02	0.0096667	0.0226667	0.0136667	0.005667
0.0106667	0.0326667	0.007667	0.004667	0.0343333	0.0136667	0.019	0.0046667	0.0193333	0.0133333	0.003333
0.006	0.0273333	0.005	0.003667	0.0176667	0.0073333	0.012	0.005	0.0106667	0.007	0.002333
0.0086667	0.036	0.007	0.005	0.03	0.0116667	0.0133333	0.0056667	0.0166667	0.0163333	0.004
0.0083333	0.0203333	0.005667	0.002	0.032	0.0106667	0.0163333	0.003	0.0153333	0.0106667	0.002667
0.0123333	0.034	0.008667	0.006333	0.0363333	0.0153333	0.0186667	0.006	0.0223333	0.0226667	0.004667

Table 3-2. Continued.

32	33	34	35	36	37	38	39	40	41
2,6,10,15-Tetramethyl-heptadecane	4-Methyl-benzaldehyde	Methoxy phenyl oxime	Methyl salicylate	trans-Calamenene	6,10-Dimethyl-5,9-undecadien-2-one	2-methyl-propanoic acid, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester	Butylated hydroxytoluene	2-Methoxy-5-vinylphenol	2,4-Bis(1,1-dimethylethyl)-phenol
Alkane	Aldehyde	Imine	Ester	Terpene	Ketone	Carboxylic acid	Alkene	Phenol	Phenol
0.0103333	0	0.0393333	0.1456667	0.0133333	0.00566667	0.082	0.9656667	0.0066667	0.1596667
0.008	0	0.0173333	0.1006667	0.0106667	0.006	0.056333333	0.6346667	0.003	0.1243333
0.0066667	0	0.01	0.0713333	0.008	0.004	0.059	0.3623333	0.0026667	0.0866667
0.0065	0.00425	0.0135	0.10325	0.008	0.0065	0.076	0.19825	0.0035	0.05025
0.004	0.0033333	0.0153333	0.1053333	0.0043333	0.00466667	0.042666667	0.375	0.0033333	0.0246667
0.003	0.003	0.0106667	0.0426667	0.0086667	0.004	0.028	0.227	0.0026667	0.0203333
0.003	0.0023333	0.0126667	0.069	0.0106667	0.00333333	0.021333333	0.2796667	0.0033333	0.019
0.009	0.0076667	0.0056667	0.0426667	0.0113333	0.00433333	0.026333333	1.1193333	0.0036667	0.0676667
0.0053333	0.008	0.008	0.0466667	0.0036667	0.00633333	0.017	1.198	0.01	0.0503333
0.0033333	0.0066667	0.007	0.0416667	0.0056667	0.00433333	0.023666667	0.6263333	0.005	0.058
0.0026667	0.0076667	0.0053333	0.0153333	0.001	0.00233333	0.014333333	0.513	0.0056667	0.029
0.0066667	0	0.006	0.026	0.0096667	0.00466667	0.04	0.7966667	0.0056667	0.0363333
0.0063333	0	0.0053333	0.0296667	0.0053333	0.00233333	0.026666667	0.6316667	0.0043333	0.0386667
0.007	0	0.0053333	0.0313333	0.007	0.006	0.036333333	0.8896667	0.004	0.0636667

3.3 Statistical analysis of the EQ-related compounds from cooked rice

Prior to data analysis, the relative peak area ratio and sensory panel test score of each compound were normalized and scaled. Pearson's correlation analysis was performed to identify volatile compounds highly correlated to the sensory panel test result (Figure 3-3).

Of the 41 compounds, seven compounds, 3-ethyl-octane, 1-ethyl-3,5-dimethylbenzene, 2,7,10-trimethyldodecane, 2,6,11-trimethyldodecane, 1-methyl-4-(1-methylethyl)-benzene, decane, and 1,3,5-trimethylbenzene, showed significant positive correlations. Numerous benzene-derived aromatic hydrocarbons were previously identified in unprocessed rice samples (Bullard and Holguin 1977). Decane was previously identified in cooked scented-rice samples (Yajima et al. 1979, Singh et al. 2000). The above-mentioned volatile compounds have been identified in cooked non-aromatic rice for the first time. Ten of the 41 compounds, hexanal, butyl hydroxytoluene, 1-octen-3-ol, 2-methoxy-5-vinylphenol, 2-(3-methylbutyl)-thiophene, heptanal, 6-methyl-5-hepten-2-one, 1-octanol, nonanal, and 2,4-bis(1,1-dimethylethyl)-phenol, showed significant negative correlations with the sensory panel test. Aliphatic aldehydes, such as hexanal, heptanal, and nonanal, are generated from the degradation of fatty acids, and 1-octen-3-ol is a well-known lipid-derived alcohol (Monsoor and Proctor 2004, Cho et al. 2014, Frankel 1998). Thus, these lipid oxidation products could have negatively affected the EQ of rice.

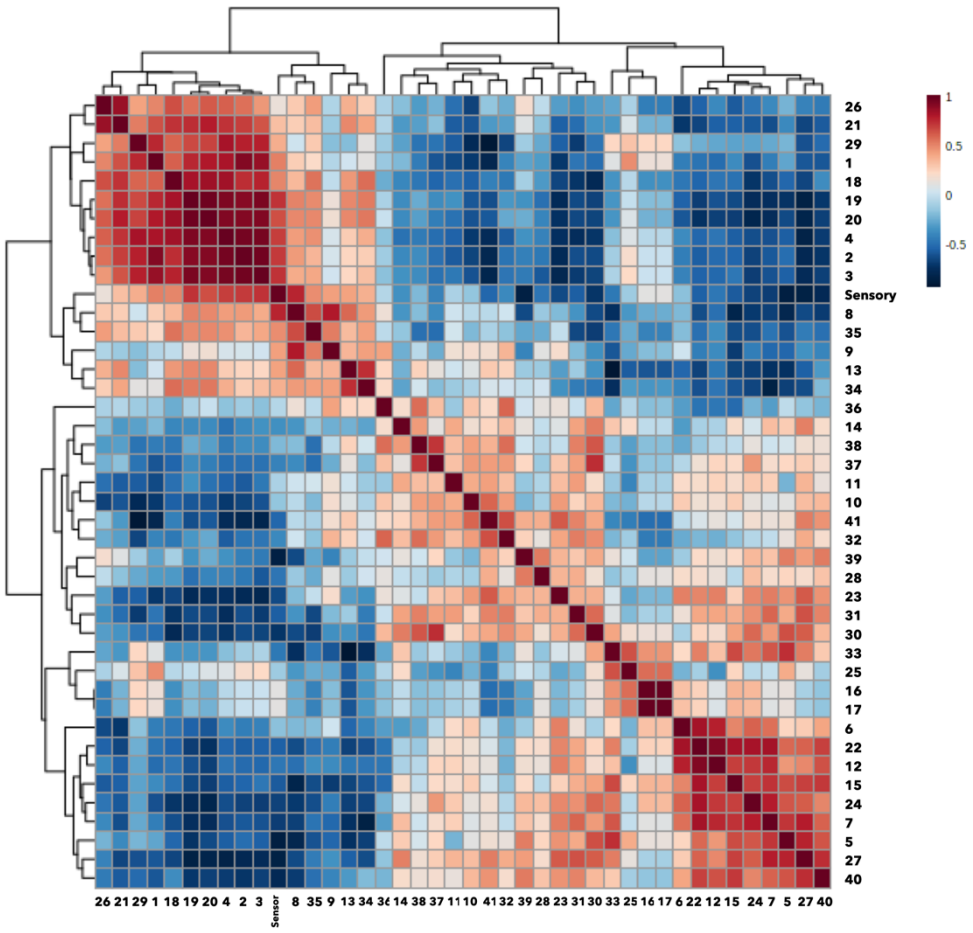


Figure 3-3. Pearson's correlation analysis of the sensory panel test scores and HS-SPME GC/MS-identified volatile compounds.
The ID numbers of VOCs are indicated. The color scale indicates the correlation coefficients ranging from 1 (red) to -1 (blue).

The PLS-DA models were used to discriminate between high and low EQ cultivars, based on the volatile compounds identified in cooked rice samples. The first latent variable explained 44.2% of the total variables (Figure 3-4A). The score plot indicated a clear segregation of rice cultivars based on their EQs. This implies that the detection method and the identified volatile compounds were appropriate for the identification of rice cultivars with superior EQ. The accuracy, goodness of fit, and goodness of prediction of this model were 1.0, 0.911, and 0.846, respectively, when the number of components was 1. These values signify that the model generated in this study is both accurate and robust.

The variable importance in projection (VIP) scores, which imply biomarkers that play important roles in the discrimination from the PLS-DA model, were also calculated (Figure 3-4B). Among the volatile compounds, 1-octen-3-ol displayed the highest VIP score (1.51), followed by 1-ethyl-3,5-dimethylbenzene, 1-methyl-4-(1-methylethyl)benzene, 2,6,11-trimethyldodecane, 3-ethyloctane, 2,7,10-trimethyldodecane, methyl salicylate, 2-octanone, and heptanal. Notably, the VIP scores of lipid oxidation products were higher than 1. Additionally, comparatively higher peak ratio of 1-octen-3-ol was detected through out of low EQ cultivars (Figure 3-5), which implies that 1-octen-3-ol potentially negatively affected on the EQ. On the other hand, higher peak ratios were shown from the benzene-derived aromatic hydrocarbons, 3-ethyloctane, and methyl salicylate in high EQ cultivars. These compounds can be considered as the VOCs that are positively related to the EQ. The higher the VIP scores, the clearer segregation patterns of the peak ratios between the low and high EQ cultivars were observed. The results imply that the content of certain VOCs could provide supplementary information regarding flavor and aroma in evaluation of EQ. A recent study on the evaluation of EQ of 6 japonica rice varieties presented that JR5, one of the varieties considered as high-EQ in the study, had higher 1-octen-3-ol and explained it as positively contributed VOC to the aroma of the variety (Li et al. 2022). However, results of the study lack statistical power and failed to present a scientific basis on the effects of the VOCs on determining EQ. On the other hand, the PLS-DA model of this study clearly presented and suggested VOCs that positively or negatively affect EQ.

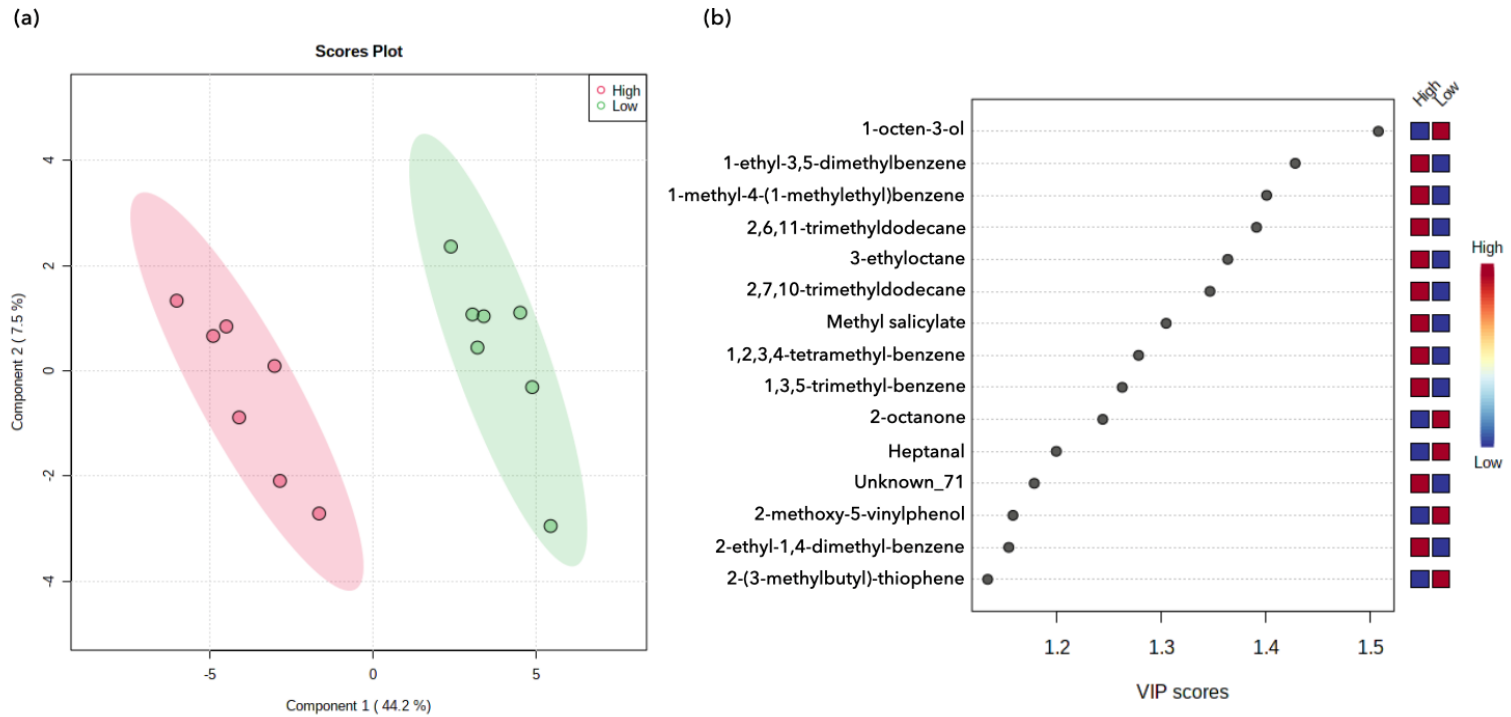


Figure 3-4. PLS-DA of the HS-SPME GC/MS data of non-aromatic cooked rice.
A. PLS-DA score plot. B. VIP scores of compounds. The color-coded boxes indicate peak area ratio as high (red) and low (blue), first column is high eating-quality cultivars, and second column is low eating-quality cultivars.

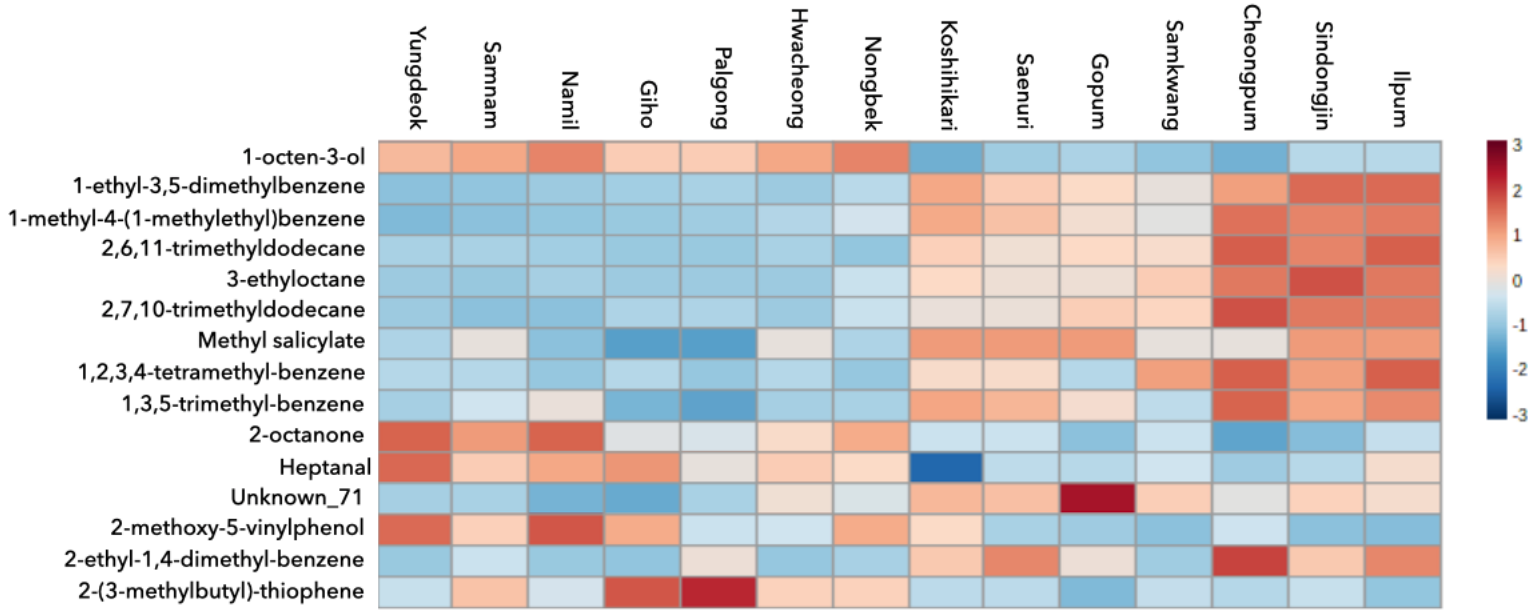


Figure 3-5. Heatmap analysis of the relative peak ratio of important VOCs according to the cultivars.
The color code indicates the differences in the relative peak ratio of significant VOCs, ranging from blue (negative) to the red (positive).

3.4 Identification of volatile compounds in head rice

The volatile organic compounds of uncooked head rice were also profiled. Total of 55 compounds were identified as Table 3-3, including 5 unidentified compounds. Similar to the analysis of cooked rice, the profiled volatile compounds were quantified by calculating relative peak area ratio compared to the internal standard, 1,2,3-propane (Table 3-4).

Comparing the results with the VOCs profiled from cooked rice, 6 compounds were also detected in the head rice: 1-hexanol, 1-octanol, decane, dodecane, hexanal, and nonanal. However, these overlapped compounds were not part of the key volatile compounds that significantly play role in differentiating EQ of japonica rice.

Table 3-3. Volatile compounds identified by HS-SPME GC/MS from head rice

ID	Component Name	Retention time (min)
1	1-Heptanol	35.76
2	1-Hexadecanol	27.04
3	1-Hexanol	31.14
4	1-Methoxy-2-propanol	20.55
5	1-Nonadecene	34.39
6	1-Nonanol	45.71
7	1-Octanol	40.24
8	1-Pentanol	26.28
9	1-Tricosene	34.12
10	1-Tridecene	28.87
11	2-Butyl-1-octanol	27.39
12	2-Hexanecarboxylic_acid	47.49
13	2-Methylundecane	23.99
14	2,2,4-Trimethyl-1,3-pentanediol_diisobutyrate	59.25
15	2,4,7,9-Tetramethyl-5-decyne-4,7-diol	68.94
16	2,4,7,9-Tetramethyl-5-decyne-4,7-diol(2)	68.67
17	3-Methylcyclopentyl_acetate	22.41
18	3-Methyltridecane	31.31
19	3,3-Diethyltridecane	45.51
20	5,5-Dibutylnonane	32.93
21	5,5-Diethyltridecane	41.77
22	6-Methyl-5-heptene-2-one	30.89
23	Benzeneethanol	61.53
24	Benzenemethanol	59.79
25	Butylated_Hydroxytoluene	61.29
26	Calamenene	57.74
27	Decane	13.88
28	dl-Menthol	45.02
29	Dodecane	23.42
30	Heptanoic_acid	63.46
31	Hexadecane	28.21
32	Hexanal	18.56
33	Hexanoic_acid	58.34
34	l-Limonene	24.21
35	m-Xylene	23.82
36	Methyl_hexanoate	23.53
37	Methyl_myristate	65.3
38	Neodecanoic_acid	67.5
39	Nonanal	33.47
40	o-Xylene	21.28
41	o-Xylene(2)	21.61
42	Palmitic_acid,_methyl_ester	74.46

43	Pentanoic_acid	52.21
44	Phenol	65.77
45	Phthalic_acid,_butyl_hept-4-yl_ester	73.62
46	Propanoic_acid,_2-methyl-,_3-hydroxy-2,2,4-trim	59
47	Tricyclopentadeca-3,7-dien[8.4.0.1(11,14)]	62.19
48	Triisobutyl_phosphate	60.48
49	UNIDENTIFIED	26.68
50	UNIDENTIFIED_2	28.37
51	UNIDENTIFIED_3	28.57
52	UNIDENTIFIED_4	37.3
53	UNIDENTIFIED_5	37.62
54	γ-Hexalactone	50.4
55	γ-n-Amylbutyrolactone	67.1

Table 3-4. The quantified VOCs from head rice calculated into area ratio using internal standard.

ID	1	2	3	4	5	6	7	8
Compound	1-Heptanol	1-Hexadecanol	1-Hexanol	1-Methoxy-2-propanol	1-Nonadecene	1-Nonanol	1-Octanol	1-Pentanol
Koshihikari	0.003	0	0.019	0.006	0.004	0.007	0.011	0.003
Saenuri	0.006	0	0.048	0.0045	0.006	0.005	0.012	0.0065
Gopum	0.008	0.001	0.0375	0.0035	0.005	0.0095	0.015	0.005
Sindongjin	0.007	0	0.05	0.003	0.003	0.008	0.017	0.006
Ilpum	0.0095	0.0005	0.0735	0.003	0.0055	0.01	0.0185	0.0075
Samkwang	0.0075	0	0.0595	0.0015	0.004	0.006	0.018	0.0065
Cheongpum	0.0085	0.001	0.0565	0.002	0.0075	0.007	0.0155	0.0065
Hwacheong	0.0065	0.001	0.0315	0.002	0.006	0.006	0.014	0.005
Yeongdeok	0.007	0.001	0.077	0.002	0.006	0.0075	0.0185	0.011
Samnam	0.005	0.001	0.06	0.002	0.006	0.0055	0.0125	0.0075
Namil	0.004	0.001	0.0375	0.0025	0.0055	0.006	0.0135	0.0065
Giho	0.0065	0.001	0.0445	0.002	0.0065	0.0105	0.022	0.006
Nongbek	0.0055	0	0.0315	0.002	0.0035	0.0075	0.0165	0.0035
Palgong	0.0045	0	0.0255	0.0015	0.0055	0.0045	0.0175	0.003

Table 3-4. Continued

9	10	11	12	13	14	15	16	17
1-Tricosene	1-Tridecene	2-Butyl-1-octanol	2-Hexanecarboxylic_acid	2-Methylundecane	2,2,4-Trimethyl-1,3-pentanediol_diisobutyrate	2,4,7,9-Tetramethyl-5-decyne-4,7-diol	2,4,7,9-Tetramethyl-5-decyne-4,7-diol(2)	3-Methylcyclopentyl_acetate
0.007	0.001	0.001	0.001	0.001	0.009	0.002	0.001	0.001
0.0115	0.001	0.001	0.001	0.001	0.011	0.002	0.0015	0.002
0.01	0.001	0.001	0.001	0.001	0.012	0.002	0.002	0.003
0.006	0	0.001	0.0015	0.0005	0.0145	0.002	0.002	0.003
0.0095	0.001	0.001	0.002	0.001	0.01	0.0025	0.002	0.003
0.006	0.001	0.001	0.002	0.001	0.0115	0.002	0.002	0.002
0.0105	0.0015	0.002	0.002	0.0015	0.0095	0.002	0.002	0.002
0.01	0.001	0.001	0.002	0.001	0.006	0.003	0.0025	0.0025
0.0075	0.0015	0.002	0.001	0.002	0.0075	0.0035	0.003	0.004
0.0075	0.001	0.002	0.001	0.001	0.0075	0.003	0.003	0.003
0.0075	0.001	0.0015	0.001	0.001	0.0075	0.0025	0.002	0.002
0.0115	0.001	0.001	0.002	0.001	0.011	0.003	0.0025	0.0025
0.01	0.0005	0.001	0.001	0.001	0.0095	0.0025	0.002	0.001
0.0095	0.001	0.001	0.003	0.001	0.0125	0.002	0.002	0.001

Table 3-4. Continued

18	19	20	21	22	23	24	25	26
3-Methyltridecane	3,3-Diethyltridecane	5,5-Dibutylnonane	5,5-Diethyltridecane	6-Methyl-5-heptene-2-one	Benzeneethanol	Benzenemethanol	Butylated_Hydroxytoluene	Calamenene
0.004	0.002	0.001	0.004	0.001	0.005	0.002	0.729	0.003
0.005	0.0025	0.002	0.0045	0.002	0.007	0.002	1.153	0.003
0.005	0.003	0.001	0.005	0.002	0.0095	0.002	0.8945	0.003
0.003	0.0025	0.001	0.0045	0.0025	0.01	0.0025	0.357	0.003
0.005	0.0025	0.002	0.005	0.0025	0.0095	0.0025	1.2175	0.002
0.0035	0.001	0.001	0.002	0.002	0.0075	0.001	0.552	0.0045
0.006	0.002	0.002	0.0045	0.002	0.012	0.0015	1.193	0.0065
0.005	0.002	0.002	0.0035	0.002	0.006	0.002	0.8295	0.0045
0.005	0.001	0.0015	0.0025	0.0065	0.0075	0.002	0.698	0.001
0.0045	0.0015	0.001	0.002	0.003	0.007	0.002	0.6945	0.003
0.004	0.001	0.001	0.002	0.003	0.005	0.001	0.694	0.0005
0.006	0.003	0.0015	0.005	0.002	0.008	0.002	0.974	0.0045
0.0035	0.002	0.001	0.0045	0.001	0.006	0.0025	0.607	0.0025
0.005	0.003	0.0015	0.005	0.002	0.0065	0.001	0.763	0.004

Table 3-4. Continued

27	28	29	30	31	32	33	34	35	36
Decane	dl-Menthol	Dodecane	Heptanoic_acid	Hexadecane	Hexanal	Hexanoic_acid	l-Limonene	m-Xylene	Methyl_hexanoate
0.004	0.001	0.002	0.001	0.003	0.002	0.012	0.002	0.003	0.009
0.0035	0.002	0.002	0.001	0.003	0.004	0.0185	0.001	0.003	0.013
0.004	0.002	0.002	0.001	0.004	0.002	0.0165	0.002	0.004	0.0105
0.004	0.002	0.002	0.0015	0.0035	0.0045	0.028	0.002	0.0035	0.012
0.004	0.002	0.002	0.001	0.0035	0.0045	0.0225	0.002	0.004	0.0155
0.0045	0.002	0.002	0.002	0.003	0.0065	0.0305	0.002	0.004	0.0175
0.004	0.001	0.003	0.002	0.0045	0.004	0.0335	0.003	0.004	0.0205
0.0045	0.002	0.002	0.001	0.004	0.003	0.023	0.001	0.0035	0.01
0.005	0.001	0.003	0.002	0.004	0.0095	0.0395	0.002	0.006	0.0385
0.004	0.001	0.002	0.001	0.003	0.005	0.0255	0.002	0.004	0.015
0.0055	0.0015	0.002	0.001	0.003	0.003	0.025	0.002	0.005	0.0145
0.0035	0.002	0.003	0.001	0.005	0.003	0.0235	0.002	0.004	0.014
0.004	0.002	0.002	0.001	0.0035	0.003	0.014	0.0015	0.003	0.0075
0.0035	0.001	0.002	0.0015	0.004	0.0065	0.026	0.001	0.003	0.014

Table 3-4. Continued

37	38	39	40	41	42	43	44	45
Methyl_myristate	Neodecanoic_acid	Nonanal	o-Xylene	o-Xylene(2)	Palmitic_acid,_methyl_ester	Pentanoic_acid	Phenol	Phthalic_acid,_butyl_hept-4-yl_ester
0.001	0.001	0.011	0.002	0.004	0.017	0.001	0.002	0.003
0.001	0.002	0.0195	0.002	0.005	0.0155	0.002	0.001	0.0395
0.0035	0.002	0.0175	0.0025	0.006	0.035	0.002	0.001	0.0315
0.006	0.0025	0.031	0.0025	0.006	0.056	0.0025	0.001	0.0465
0.002	0.0025	0.0285	0.002	0.006	0.025	0.0025	0.001	0.0415
0.001	0.003	0.0355	0.0025	0.0065	0.0115	0.0035	0.003	0.029
0.003	0.003	0.026	0.0025	0.0065	0.0355	0.004	0.002	0.033
0.006	0.003	0.025	0.002	0.006	0.0655	0.002	0.001	0.036
0.0155	0.0035	0.033	0.004	0.01	0.13	0.004	0.001	0.0575
0.0075	0.003	0.0195	0.003	0.006	0.0845	0.0025	0.001	0.052
0.004	0.0025	0.022	0.003	0.0085	0.0515	0.003	0.002	0.044
0.0135	0.003	0.0395	0.0025	0.0065	0.1675	0.0025	0.001	0.0535
0.006	0.0025	0.029	0.002	0.005	0.0505	0.001	0.001	0.0395
0.0075	0.002	0.047	0.002	0.004	0.048	0.003	0.001	0.033

Table 3-4. Continued

46	47	48	49	50	51	52	53	54	55
Propanoic_acid,_2-methyl-,_3-hydroxy-2,2,4-trim	Tricyclopentadeca-3,7-dien[8.4.0.1(11,14)]	Triisobutyl_phosphate	UNIDENTIFIED	UNIDENTIFIED_2	UNIDENTIFIED_3	UNIDENTIFIED_4	UNIDENTIFIED_5	γ-Hexalactone	γ-n-Amylbutyrolactone
0.002	0.001	0.002	0.001	0.001	0.003	0.007	0.003	0.001	0.001
0.002	0.002	0.003	0.001	0.002	0.004	0.008	0.004	0.003	0.002
0.002	0.0015	0.0035	0.001	0.001	0.005	0.008	0.004	0.003	0.0015
0.003	0.002	0.004	0.001	0.002	0.0045	0.009	0.003	0.004	0.0045
0.0025	0.001	0.0045	0.001	0.002	0.0045	0.0095	0.0045	0.004	0.003
0.0025	0.0025	0.0045	0.001	0.0025	0.005	0.01	0.003	0.0035	0.0075
0.0025	0.0025	0.004	0.001	0.003	0.0055	0.009	0.0045	0.004	0.0045
0.002	0.001	0.0055	0.001	0.002	0.006	0.0075	0.004	0.003	0.001
0.002	0.001	0.006	0.002	0.004	0.008	0.0085	0.0035	0.005	0.002
0.0025	0.001	0.0055	0.001	0.002	0.006	0.009	0.003	0.003	0.002
0.002	0.0005	0.0045	0.0015	0.001	0.007	0.0085	0.003	0.003	0.001
0.002	0.001	0.005	0.001	0.002	0.0055	0.009	0.0045	0.0035	0.0025
0.0025	0.001	0.005	0.001	0.001	0.0045	0.008	0.003	0.002	0.001
0.002	0.0055	0.004	0.001	0.002	0.004	0.0075	0.0045	0.003	0.009

3.5 Statistical analysis of the EQ-related compounds from head rice

The important volatile compounds identified by cooked rice are more directly related to the organic compounds associated with EQ than that from uncooked head rice. However, there is still great advantage in identifying important volatile compounds from head rice in identifying EQ, that it's easier to conduct the experiment with shortened labor and time. Hence, in order to identify key biomarker from uncooked rice for efficient use in breeding programs, PLS-DA using compounds identified from head rice was performed.

As a result, PLS-DA model discriminating cultivars based on the volatile compounds from head rice samples was generated (Figure 3-6). The first and second latent variable explained 21.3% and 15% of the total variables, respectively (Figure 3-6A). The score plot indicated a clear segregation of rice cultivars based on their EQs. The accuracy, goodness of fit, and goodness of prediction of this model were 0.857, 0.808, and 0.451, respectively, when the number of components was 1. That when the number of components was 5, the cross validation results complemented to 0.929, 0.999, and 0.816, respectively.

The variable importance in projection (VIP) scores, which imply biomarkers that play important roles in the discrimination from the PLS-DA model, were also calculated (Figure 3-6B). Among the volatile compounds, 1-Hexanol, Methyl myristate, Palmitic acid, Benzene ethanol, 2,4,7,9-Tetramethyl-5-decyne-4,7-diol, Triisobutyl phosphate, and 2,2,4-Trimethyl-1,3-pentanediol diisobutyrate were higher than 1.6 in VIP scores. 1-Hexanol and Benzene ethanol potentially positively affected on the EQ, and other compounds negatively affected on the EQ.

Meanwhile, there still is grey area to practically utilize the results, for instance the interactions among the VOCs should be further studied and considered. Furthermore, an unknown compound that potentially affects the EQ of rice should be investigated further and identified. Either positively or negatively affect, the compounds with VIP scores greater than 1 are considered as key VOCs affecting the general sensory properties of cooked japonica rice. These could further be used as quality evaluation criteria to consider aroma and flavor attributes in rice breeding programs.

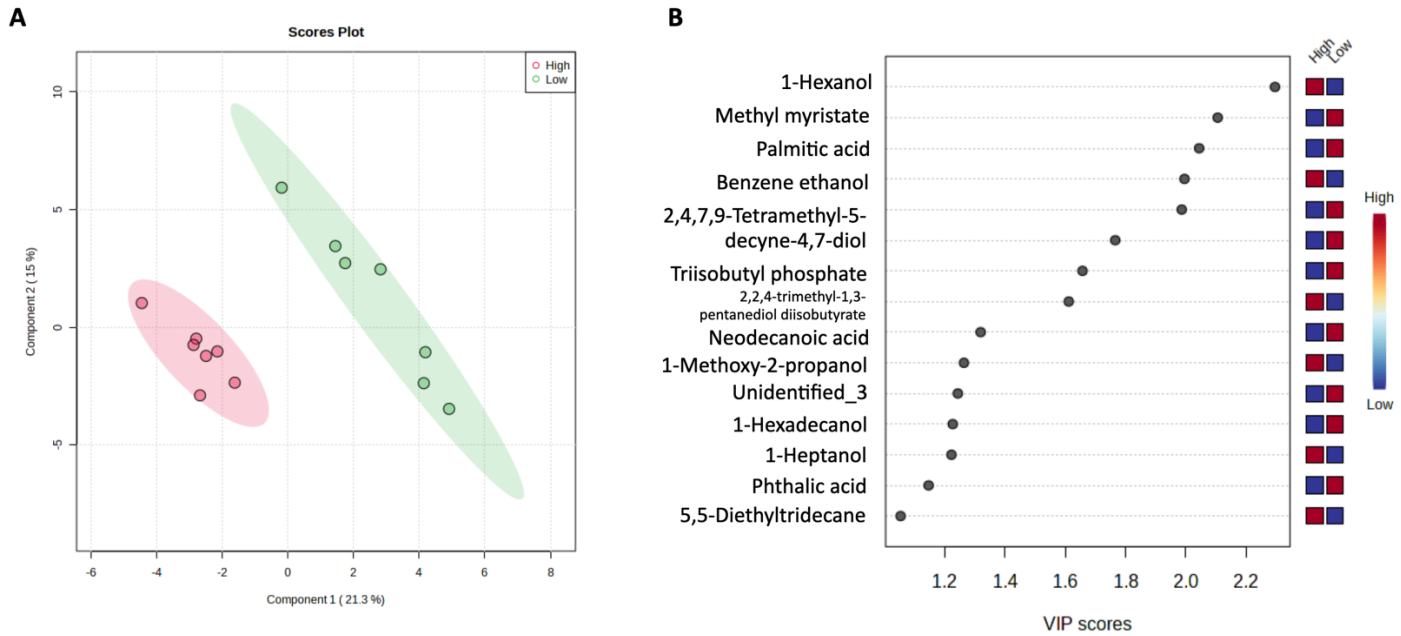


Figure 3-6. PLS-DA of the HS-SPME GC/MS data of non-aromatic head rice.
A. PLS-DA score plot. B. VIP scores of compounds. The color-coded boxes indicate peak area ratio as high (red) and low (blue), first column is high eating-quality cultivars, and second column is low eating-quality cultivars.

Conclusion

In this study, the EQ of 14 non-aromatic japonica rice cultivars was evaluated using the sensory panel test, Toyo taste-meter, and volatile compound profiling. A number of volatile compounds were identified for the first time in cooked and uncooked non-aromatic japonica rice, and these compounds showed a strong correlation with the EQ of rice. Notably, lipid-derived compounds (e.g., 1-octen-3-ol), fatty acid degradation-related compound (heptanal), and other compounds (e.g., 2-octanone, methyl salicylate, and other benzene-derived compounds) were identified as important variables that discriminate rice cultivars based on EQ in cooked rice analysis. In head rice analysis, 1-Hexanol, Methyl myristate, Palmitic acid, Benzene ethanol, 2,4,7,9-Tetramethyl-5-decyne-4,7-diol, Triisobutyl phosphate, and 2,2,4-Trimethyl-1,3-pentanediol diisobutyrate were identified as important volatile compounds identifying EQ in uncooked rice samples. To take into account the aroma and flavor characteristics of cooked rice in evaluating its EQ, a highly accurate discriminant model was generated. The contents of listed significant VOCs could be suggested for the new EQ standard. The results could serve as a foundation for future research on integrated rice EQ, and could facilitate the development of high-quality rice varieties.

Bibliography

Abu-Jamous, B., Kelly, S. *Clust*: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol* 19, 172 (2018). https://doi.org/10.1186/s13059-018-1536-8

Amarasinghe S, Su S, Dong X et al. (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21:30 https://doi.org/10.1186/s13059-020-1935-5

American Association of Cereal Chemists Inc (2000) Approved methods of the American Association of Cereal Chemists, 10th edn. American Association of Cereal Chemists, St. Paul

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Balet, S., Guelpa, A., Fox, G., & Manley, M. (2019). Rapid Visco Analyser (RVA) as a tool for measuring starch-related physiochemical properties in cereals: A review. Food Analytical Methods, 12, 2344– 2360. https://doi.org/10.1007/s12161-019-01581-w

Ball, S. G., van de Wal, M., Visser R. G. Progress in understanding the biosynthesis of amylose. *Trends Plant Sci.* 3, 462-467 (1998).

Bao, J. S., Zheng, X. W., Xia, Y. W., He, P., Shu, Q. Y., Lu, X., Chen, Y., & Zhu, L. H. (2000). QTL mapping for the paste viscosity characteristics in rice (Oryza sativa L.). Theoretical Applied Genetics, 100, 280–284. https://doi.org/10.1007/s001220050037

Bao, J. Zhou, X. Xu, F. He, Q. Park, Y. Genome-wide association study of the resistant starch content in rice grains. Starch 69(7-8):1600343 (2017) https://doi.org/10.1002/star.201600343

Bergman, C. J. *et al.* Rapid gas chromatographic technique for quantifying 2-acetyl-1-pyrroline and hexanal in rice (Oryza sativa L.) *Cereal chem*. 77(4), 454-458 (2000).

Bolger A, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30(15):2114–2120 https://doi.org/10.1093/bioinformatics/btu17030

Bradbury, L.M.T. *et al.* Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. *Plant Mol Biol.* 68, 439–449 (2008).

Bryant, R. J. and McClung, A. M. Volatile profiles of aromatic and non-aromatic rice cultivars using SPME/GC-MS. *Food Chem*. 124, 501-513 (2011).

Buels R, Yao E, Diesh CM et al. (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66 https://doi.org/10.1186/s13059-016-0924-1

Bullard, R. W. and Holguin, G. Volatile components of unprocessed rice (Oryza sativa L.). *J. Agric. Food Chem.* 25, 99-103 (1977).

Buttery, R. G., Turnbaugh, J.G., Ling, L.C. Contribution of volatiles to rice aroma. *J. Agric. Food Chem*. 36, 1006-1009 (1988).

Cantarel B, Korf I, Robb S, Parra G, Ross E, Moore B, Holt C, Sánchez A, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188-96 doi: 10.1101/gr.6743907

Champagne, E. T. *et al.* Quality evaluation of U.S. medium-grain rice using a Japanese taste analyzer. *Cereal Chem.* 73, 290-294 (1996).

Champagne, E.T., Wood, D.F., Juliano, B.O., Bechtel, D.B. 2004. The Rice Grain and its Gross Composition. In: Rice Chemistry and Technology. Third Edition, Chapter 4. American Association of Cereal Chemists Press. Minneapolis, MN. p77-107.

Champagne, E. T. Rice Aroma and Flavor: A Literature Review. *Cereal Chem.* 85(4), 445-454 (2008).

Champagne, E. T. *et al.* Important sensory properties differentiating premium rice varieties. *Rice*. 3, 270-281 (2010).

Chen, S. H., Wu, J., Yang, Y., Shi, W. W., Xu, M. L. The fgr gene responsible for rice fragrance was restricted within 69 kb. *Plant Sci*. 171, 505-514 (2006).

Chen P. et al. Genetic basis of variation in rice seed storage protein (albumin, globulin, prolamin, and glutelin) content reveald by genome-wide association analysis. Front. Plant Sci. (2018)

Chen Y, Lun AAT, Smyth GK (2016). "From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline." *F1000Research*, 5, 1438. doi: 10.12688/f1000research.8987.2.

Cho, S., Nuijten, E., Shewfelt, R. L., Kays, S. J. Aroma chemistry of African Oryza glaberrima and Oryza sativa rice and their interspecific hybrids. *J. Sci. Food Agric.* 94, 727-735 (2014).

Choi J, Lye Z, Groen S et al. (2020) Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. Genome Biol 21:21 https://doi.org/10.1186/s13059-020-1938-2

Custudio, M. C., Cuevas, R. P., Ynion, J., Laborte, A. G., Velasco, M. L., Demont, M. Rice quality: How is it defined by consumers, industry, food scientists, and geneticists? *Trends in Food Sci. & Tech.* 92, 122-137 (2019).

Darling A, Mau B, Blattner F, Perna N (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14(7):1394-1403 doi:10.1101/gr.2289704

Del Mundo, A. M. and Juliano, B. O. Consumer preference and properties of raw and cooked milled rice. *J. Texture Stud.* 12, 107-120 (1981).

Ebitani T, Takeuchi Y, Nonoue N, Yamamoto T, Takeuchi K, Yano M (2005) Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar 'Kasalath' in a genetic background of japonica elite cultivar 'Koshihikari'. Breed Sci 55:65-73

Etherington G, Heavens D, Baker D, Lister A, McNelly R, Garcia G, Clavijo B, Macaulay I, Haerty W, Palma F (2020) Sequencing smart: De novo sequencing and assembly approaches for a non-model mammal, GigaScience 9(5):giaa045 https://doi.org/10.1093/gigascience/giaa04528

Frankel, E. N. Lipid Oxidation. The Oily Press, Dundee, UK (1998).

Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y, Yan M, Jiang H, Dong G, Huang Y, Han B, Li J, Qian Q (2011) ALK, the key gene for gelatinization temperature, is a modifier gene for gel consistency in rice. J Integr Plant Biol 53(9):756-65 doi: 10.1111/j.1744-7909.2011.01065.x

Guo N. et al. Rice OsLHT1 functions in leaf-to-panicle nitrogen allocation for grain yield and quality. Frontiers in Plant Science 11:1150 (2020)

He, W., Wang, L., Lin, Q., Yu, F. Rice seed storage proteins: Biosynthetic pathways and the effects of environmental factors. J of Int. Plant Biol. 63(12):1999-2019 (2021)

Higashi Y, Strominger J, Sweeley C (1967) Structure of a lipid intermediate in cell wall peptidoglycan synthesis: a derivative of a C55 isoprenoid alcohol. Proc Natl Acad Sci USA 57:1878-1884

Hori K, Sugimoto K, Nonoue Y et al. (2010) Detection of quantitative trait loci controlling pre-harvest sprouting resistance by using backcrossed populations of japonica rice cultivars. Theor Appl Genet 120:1547–1557. https://doi.org/10.1007/s00122-010-1275-z

Hu, X, Lu, L., Guo, Z., Zhu, Z. Volatile compounds, affecting factors and evaluation methods for rice aroma: A review. *Trends in Food Science & Tech*. 97, 136-146 (2020).

Hu Z, Wang W, Wu Z et al. (2018) Novel sequences, structural variations and gene presence variations of Asian cultivated rice. Sci Data 5:180079 https://doi.org/10.1038/sdata.2018.79

Huddleston J, Ranade S, Malig M et al. (2014) Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res 24(4):688-96 doi: 10.1101/gr.168450.113

International Rice Genome Sequencing Project, Sasaki T (2005) The map-based sequence of the rice genome. Nature 436:793–800 https://doi.org/10.1038/nature03895

Izawa T, Shimamoto K (1996) Becoming a model plant: The importance of rice to plant science. Trends in Plant Science 1(3):95-99 https://doi.org/10.1016/S1360-1385(96)80041-0

Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H., Nakamura, Y. Starch biosynthesis in cereal endosperm. *Plant Physiol. Biochem.* 48(6), 383-392 (2010).

Jo, Y. Cho, K. Lee, H. Lim, S. Kim J. Kim, Y. Lee, J. Cellular localization of wheat high molecular weight glutenin subunits in transgenic rice grain. Int. J. Mol. Sci. 18(11):2458 (2017)

Keeling, P. and Myers, A. Biochemistry and genetics of starch synthesis. Annu Rev Food Sci Technol. 1(1) 271-303 (2010)

Kim, D., Paggi, J.M., Park, C. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907–915 (2019). https://doi.org/10.1038/s41587-019-0201-4

Kobayashi A, Hori K, Yamamoto T et al. (2018) Koshihikari: a premium short-grain rice cultivar – its expansion and breeding in Japan. Rice 11:15 https://doi.org/10.1186/s12284-018-0207-4

Kolmogorov M, Yuan J, Lin Y, Pevzner P (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37(5):540-546 doi: 10.1038/s41587-019-0072-8

Koren S, Walenz B, Berlin K, Miller J, Bergman N, Phillippy A (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27(5):722-736 doi: 10.1101/gr.215087.116

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59 https://doi.org/10.1186/1471-2105-5-59

Kwak, H.S., Kim, M., Lee, Y., Jeong, Y. Identification of key sensory attributes for consumer acceptance and instrumental quality of aseptic-packaged cooked rice. *Int. J. Food Sci. Technol.* 50, 691-699 (2015).

Kwon, S.W. *et al.* Identification of quantitative trait loci associated with rice eating quality traits using a population of recombinant inbred lines derived from a cross between two temperate *japonica* cultivars. *Mol Cells.* 31, 437–445 (2011).

Lawrence G. and Shepherd K. Chromosomal location of genes controlling seed proteins in species related to wheat. Theor. Appl. Genet. 1981;59:25–31. doi: 10.1007/BF00275771

Lestari P, Ham T, Lee H, Woo M, Jiang W, Chu S, Kwon S, Ma K, Lee J, Cho Y, Koh H (2009) PCR Marker-Based Evaluation of the Eating Quality of Japonica Rice (Oryza sativa L.) Journal of Agricultural and Food Chemistry 57(7):2754-2762 DOI: 10.1021/jf803804k

Li F. et al. Transcriptome and metabolome analyses reveals the pathway and metabolites of grain quality under phutochrome B in rice (Oryza sativa L.) Rice 15:52 (2022)

Li J, Wang J, Zeigler R (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. Gigascience 28:3-8 doi: 10.1186/2047-217X-3-8

Li M, Dhital S, Wei Y (2017) Multilevel structure of wheat starch and its relationship to noodle eating qualities. Comp Rev Food Sci Food Safety 16:1042–1055

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094-3100 doi: 10.1093/bioinformatics/bty191

Li, S. *et al.* New perspective to guide rice breeding: Evaluating the eating quality of japonica rice. *Cereal chem.* 99, 603-614 (2022).

Li, Y. Fu, J. Shen, Q. Yand, D. High-molecular-weight glutenin subunits: genetics, structures and relation to end use qualities. Int. J. Mol. Sci. 22(1):184 (2021)

Li Z, Sun F, Xu S, Chu X, Mukai Y, Yamamoto M, Ali S, Rampling L, Kosar-Hashemi B, Rahman S (2003) The structural organization of the gene encoding class ii starch synthase of wheat and barley and the evolution of the genes encoding starch synthases in plants. Funct Integr Genomics 3:76–85

Lilin Yin, Haohao Zhang, Zhenshuang Tang, Jingya Xu, Dong Yin, Zhiwu Zhang, Xiaohui Yuan, Mengjin Zhu, Shuhong Zhao, Xinyun Li, Xiaolei Liu, rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study, Genomics, Proteomics & Bioinformatics, Volume 19, Issue 4, 2021, Pages 619-628

Lorieux, M., Petrov, M., Huang, N., Guiderdoni, E., Ghesquiere, A. Aroma in rice: Genetic analysis of a quantitative trait. *Theor. Appl. Gen.* 93, 1145-1151 (1996).

Lu, L., Fang, C., Hu, Z., Hu, X., Zhu, Z. Grade classification model tandem BpNN method with multi-metal sensor for rice eating quality evaluation. *Sensors & Actuators B: Chemical*. 281, 22-27 (2019).

Lyon, B. G., Champagne, E. T., Vinyard, B. T., Windham, W. R. Sensory and instrumental relationships of texture of cooked rice from selected cultivars and postharvest handling practices. *Cereal Chem.* 77(1), 64-69 (2000)

Ma, Z. *et al.* Biochemical composition distribution in different grain layers is associated with the edible quality of rice cultivars. *Food Chem.* 311, 125896 (2020).

Mahmoud M, Zywicki M, Twardowski T, Karlowski W (2019) Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. Genomics 111(1):43-49 doi: 10.1016/j.ygeno.2017.12.011

Monsoor, M. A. and Proctor, A. Volatile component analysis of commercially milled head and broken rice. *Journal of Food Science.* 69(8), C632-C636 (2004).

Lin Z. et al. Complementary proteome and transcriptome profiling in developing grains of a notched-belly rice mutant reveals key pathways involved in chalkiness formation. Plant Cell Physiol. 58:560-573 (2017)

Matsubara K, Ogiso-Tanaka E, Hori K, Ebana K, Ando T, Yano M (2012) Natural Variation in Hd17, a Homolog of Arabidopsis ELF3 That is Involved in Rice Photoperiodic Flowering, Plant and Cell Physiology 53(4):709–716 https://doi.org/10.1093/pcp/pcs028

McCarthy DJ, Chen Y, Smyth GK (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." *Nucleic Acids Research*, 40(10), 4288-4297. doi: 10.1093/nar/gks042.

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

Meilgaard, M., Civille, G. V., Carr, B. T. Sensory Evaluation Techniques. CRC Press. Boca Raton, FL (2007).

Michael T, VanBuren R (2015) Progress, challenges and the future of crop genomes. Curr Opin Plant Biol 24:71-81 doi: 10.1016/j.pbi.2015.02.002

Misra,G. S. Badoni, C.J. Domingo, R.P.O. Cuevas, C. Llorente, E.G.N. Mbanjo, N. Sreenivasulu Deciphering the genetic architecture of cooked rice texture Front. Plant Sci., 9 (2018), p. 1405, 10.3389/fpls.2018.01405

Morell MK, Kosar-Hashemi B, Cmiel M, Samuel MS, Chandler P, Rahman S, Buleon A, Batey IL, Li Z (2003) Barley *sex6* mutants lack starch synthase iia activity and contain a starch with novel properties. The Plant J 34:173–185

Ohtsubo K, Nakamura S, Imamura T (2002) Development of the primer sets for identification of a rice variety, Koshihikari, by PCR. Nippon Nogeikagaku Kaishi 76:388-397

Okpala, N. E., Aloryi, K. D., An, T., He, L., Tang, X. The roles of starch branching enzymes and starch synthase in the biosynthesis of amylose in rice. *J. Cereal Sci.* 104, 103393 (2022).

Pang, A. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*. 49(W1), W388-W396 (2021).

Pang, Y., Ali, J., Wang, X., Franje, N. J., Revilleza, J. E., Xu, J., & Li, Z. (2016). Relationship of rice grain amylose, gelatinization temperature and pasting properties for breeding better eating and cooking quality of rice varieties. PLOS ONE, 11. https://doi.org/10.1371/journal. Pone.0168483

Paul D. Thomas, Michael J. Campbell, Anish Kejariwal, Huaiyu Mi, Brian Karlak, Robin Daverman, Karen Diemer, Anushya Muruganujan, Apurva Narechania. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, 13: 2129-2141 .

Paul D. Thomas, Anish Kejariwal, Nan Guo, Huaiyu Mi, Michael J. Campbell, Anushya Muruganujan, and Betty Lazareva-Ulitsky. 2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucl. Acids Res.*34: W645-W650.

Payne P. Holt L. Jackson E. Law C. Wheat storage proteins: Their genetics and their potential for manipulation by plant breeding. Philos. Trans. R. Soc. Lond. B Biol. Sci. 304:359-371 (1984)

Peng, B. et al., OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice. Nat. Commun. 5:4847 (2014)

Pennisi E (2017) New technologies boost genome quality. Science 357(6346):10-11 doi: 10.1126/science.357.6346.10

Preston A, Mandrell R, Gibson B, Apicella M (1996) The lipooligosaccharides of pathogenic gram-negative bacteria. Crit Rev Microbiol 22:139-180

Robinson MD, McCarthy DJ, Smyth GK (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics*, 26(1), 139-140. doi: 10.1093/bioinformatics/btp616.

Saika K (1992) The structure of a rice taste sensor and its application. "TOYO" Taste-Meter", a practical measurement system for rice. Food Industry 35:37-43

Sakai H, Lee S, Tanaka T, Numa H, Kim J, Kawahara Y, Wakimoto H, Yang 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 doi: 10.1093/pcp/pcs183

Satoh H. *et al.* Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. *Plant Physiol.* 133, 1111-1121(2003).

Seo J, Lee G, Jin Z, Kim B, Chin J, Koh H (2020) Development and application of indica-japonica SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding. Mol Breeding 40:39 https://doi.org/10.1007/s11032-020-01123-x

Septiningsih, E. M., Trijatmiko, K. R., Moeljopawiro, S., McCouch, S. R. Identification of quantitative trait loci for grain quality in an advanced backcross population dericed from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon. Theor. Appl. Genet.* 107, 1433-1441 (2003).

Shi, S. *et al.* Use of protein content, amylose content, and RVA parameters to evaluate the taste quality of rice. *Front. Nutr.* 8, 758547 (2022).

Shinoda, R. *et al.* Using SPME-GC/REMPI-TOFMS to measure the volatile odor-active compounds in freshly cooked rice. *ACS Omega*. 5, 20638-20642 (2020).

Singh, R. K., Singh, U. S., Khush, G. S. Aromatic Rices. Int. Rice Res. Inst., New Delhi, India (2000).

Slattery CJ, Kavakli IH, Okita TW (2000) Engineering starch for increased quality and quantity. Trends Plant Sci 5:291–298

Smith AM, Denyer K, Martin C (1997) The synthesis of the starch granule. Annu Rev Plant Biol 48:67–87

Srisawas, W. and Jindal, V. K. Sensory evaluation of cooked rice in relation to water-to-rice ratio and physicochemical properties. *J. Texture Studies.* 38, 21-41 (2006).

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5):637-44 doi: 10.1093/bioinformatics/btn013

Sun, M. M. *et al.* Molecular aspect of good eating quality formation in japonica rice. *PLOS ONE*. 6(4), e18385 (2011).

Supek F, Bošnjak M, Škunca N, Šmuc T. "*REVIGO summarizes and visualizes long lists of Gene Ontology terms*" PLoS ONE 2011. doi:10.1371/journal.pone.0021800

Takeuchi Y, Nonoue Y, Ebitani T, Suzuki K, Aoki N, Sato H, Ideta O, Hirabayasi H, Hirayama M, Ohta H (2007) QTL detection for eating quality including glossiness, stickiness, taste and hardness of cooked rice. Breed Sci 57:231-242

Takeuchi Y, Hori K, Suzuki K, Nonoue Y. Takemoto-Kuno Y, Maeda H, Sato H, Hirabayashi H, Ohta H, Ishii T et al. (2008) Major QTLs for eating quality of an elite Japanese rice cultivar, Koshihikari, on the short arm of chromosome 3. Breed Sci 58:437-445

Tan, Y. F., Li, J. X., Yu, S. B., Xing, Y. Z., Xy, C. G., Zhang, Q. The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theor. Appl. Genet.* 99, 642-648 (1999).

Tanaka T, Nishijima R, Teramoto S, Kitomi Y, Hayashi T, Uga Y, Kawakatsu T (2020) De novo genome assembly of the indica rice variety IR64 using linked-read sequencing and nanopore sequencing. G3: GENES, GENOMES, GENETICS 10(5)1495-1501 https://doi.org/10.1534/g3.119.400871

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. J Exp Bot 55:2131–2145

Tetlow IJ, Emes MJ (2017) Starch biosynthesis in the developing endosperms of grasses and cereals. Agronomy 7(4):81

Thorvaldsdóttir, H, Robinson J, Mesirov J. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Briefings in Bioinformatics 14(2):178-192

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27(5):737-746 doi: 10.1101/gr.214270.116

Vorachek-Warren M, Carty S, Lin S, Cotter R, Raetz C (2002) An Escherichia coli mutant lacking the cold shock-induced palmitoleoyltransferase of lipid A biosynthesis: absence of unsaturated acyl chains and antibiotic hypersensitivity at 12 degrees C. J Biol Chem 277:14186-14193

Wada T, Ogata T, Tsubone M, Uchimura Y, Matsue Y (2008) Mapping of QTLs for eating quality and physicochemical properties of the japonica rice 'Koshihikari'. Breed Sci 58:427-435

Walker B, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo C, Zeng Q, Wortman J, Young S, Earl A (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963 doi: 10.1371/journal.pone.0112963

Wang, X. Pang, Y. Zhang, J. et al. Genome wide and gene-based association mapping for rice eating and cooking characteristics and protein content. Sci Rep 7:17203(2017) doi: 10.1038/s41598-017-17347-5

Wang, X. Pang, Y. Wang, C. Chen, K. Zhu, Y. et al. New candidate genes affecting rice grain appearance and milling quality detected by genome-wide and gene-based association analyses. Front Plant Sci. (2017) https://doi.org/10.3389/fpls.2016.01998

Wang Z, Zheng F, Shen G, Gao J, Snustad D, Li M, Zhang J, Hong M (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J 7(4):613-22 doi: 10.1046/j.1365-313x.1995.7040613.x

Wang W, Mauleon R, Hu Z et al. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43-49 https://doi.org/10.1038/s41586-018-0063-9

Xie Q et al. Dynamic formation and transcriptional regulation mediated by phytohormones during chalkiness formation in rice. BMC Plant Biol. 21:308 (2021)

Xu, F., Sun, C., Huang, Y., Chen, Y., Tong, C., Bao, J. QTL mapping for rice grain quality: a strategy to detect more QTLs within sub-populations. *Mol. Breed.* 35, 105 (2015).

Xu, F., Bao, J., He, Q. *et al.* Genome-wide association study of eating and cooking qualities in different subpopulations of rice (*Oryza sativa* L.). *BMC Genomics* 17, 663 (2016). https://doi.org/10.1186/s12864-016-3000-z

Yajima, I., Yanai, T., Nakamura, M., Sakakibara, H., Habu, T. Volatile flavor components of cooked rice. *Agric. Biol. Chem.* 42, 1229-1233 (1978).

Yajima, I., Yanai, T., Nakamura, M., Sakakibara, H., Hayashi, K. Volatile Flavor Components of Cooked Kaorimai (Scented Rice, *O. sativa japonica*). *Agric. and Biol.Chem*. 43(12), 2425–2429 (1979).

Yamamoto T, Nagasaki H, Yonemaru J. et al. (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267 https://doi.org/10.1186/1471-2164-11-267

Yan, C., Li, X., Zhang, R., Sui, J. M., Liang, G. H., Shen, X. P., Gu, S. L., & Gu, M. H. (2005). Performance and inheritance of rice starch RVA profile characteristics. Rice Science, 12, 39–47

Yang Liao, Gordon K Smyth and Wei Shi. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. Bioinformatics, 30(7):923-30, 2014.

Zhang H. et al. Proteomic profiling reveals differentially expressed proteins associated with amylose accumulation during rice grain filling. BMC genomics 21,714 (2020)

Zhang, X. *et al.* A study on volatile metabolites screening by HS-SPME-GC-MS and HS-GC-IMS for discrimination and characterization of white and yellowed rice. *Cereals & Grains*. 97, 496-504 (2020).

Zheng, Z., Zhang, H., Chen, J. Y., Zhang, T., Matsunaga, R. Flavor volatiles of rice during cooking analyzed by modified headspace SPME/GC-MS. *Cereal Chem*. 85, 140-145 (2008).

Zhou, H. *et al.* The origin of Wxla provides new insights into the improvement of grain quality in rice. *J. Integr. Plant Biol.* 63, 878-888 (2021).

초록

자포니카 벼의 식미에 관한 이화학적 및 유전체 특성 연구

벼의 품질은 다양한 이화학적 특성을 아우르는 복잡한 형질이다. 그 중에서도 식미 형질은 시장에서 소비자에게 직접적 영향을 주는 주요 형질로 분류된다. 밥을 섭취하며 느껴지는 모든 감각 관련 속성은 식미 파악에 중요한 요소로서, 이와 관련된 이화학적 특성들이 식미의 직·간접적 평가에 활용되고 있다. 하지만 식미는 개인의 주관적 판단에 의해 결정되고, 경험적·문화적 배경이 그 판단에 영향을 미쳐 문화권마다 선호하는 벼 품질의 항목 또한 서로 상이하게 다르므로 소비층에 맞는 식미 형질을 파악하고 조절할 수 있어야 한다. 더불어서 식미 관련 주요 유전자인 Waxy 유전자의 대립유전형에 따른 변이 만으로는 다양한 식미가 설명되고 있지 않아 그 이외의 관여 유전자들을 파악하는 것이 중요하다. 본 연구의 목표는 자포니카 벼의 식미 결정 요인을 방향성분, 물리적 특성, 화학성분 등으로 세분화하여 이화학적·유전적 품종육성 지표를 확립하는데 있다. 이를 위해 밥의 세분화된 식미 특성 지표를 확립하고, 관여 요인의 전유전체 연관분석 및 전사체분석으로 관여 유전자를 발굴해 선발지표를 개발을 했다.

자포니카 재래종과 육성종으로 구성된 284 개 자원을 대상으로 10 개의 식미 관련 형질들을 연차반복하여 조사했다. 도요 식미치 분석 결과 3 년차 반복에서 공통적으로 2 번 염색체 상단과 8 번 염색체 하단부에서 lead SNP 이 발견되었고, 최고 및 최저 도요 식미치의 자원을 선별해 출수 후 10일차 배유에서 전사체분석한 결과 고식미 자원과 저식미 자원에서 공통의 발현량 차이를 보이는 유전자를 발굴했다. Os06g0256500 은 glucose-6-

phosphate 와 fructose-6-phosphate 를 호환시키는 효소로, 전분생합성 경로에서 중요한 기질이다. 이 유전자의 기능을 T-DNA 삽입 계통을 이용해 검증한 결과, knock-out 개체와 wild-type 간에 유의한 도요 식미치 차이를 보였다. 신속점도계를 활용해 전분호화도를 측정한 후 전유전체 연관분석 및 전사체분석 결과, 2 번 염색체 상단에서 최고점도, 최저점도, 최종점도에 유의하게 관여하는 유전자로 Os02g0224300 을 동정했다. 이 유전자는 high molecular weight glutenin family protein 을 encoding 하는데, 35S enhancer tagging line 을 이용해 기능을 확인했다. 그 결과 wild-type 에 비해 enhancer tagging line 에서 유의하게 낮은 전분호화특성을 보였다. 식미 형질에 기여하는 유전요인을 밝히고 기능을 확인하여 이를 통해 기존 육종과정에서 식미를 향상 시킬 때 MAS 를 활용하지 못하던 것을 가능하게 기여 할 수 있다.

고시히카리 특이적 염기서열에서 식미관련 품종 특이적 유전 구간을 파악하고 육종에 활용하기 위해 Nanopore 와 Illumina platform 을 이용해 고시히카리 유전체를 해독했다. 자포니카 품종인 니폰바레와 고시히가리의 유전체를 비교 했을 때 3 번 염색체와 11 번 염색체에서 특이적 구간을 확인 했다. 고시히카리 특이적 염기서열 구간에 고시히카리를 육종 소재로 이용해 육성된 품종 16 개의 NGS read sequence 를 고시히카리 유전체에 정렬했을 때, 11 번 염색체의 P5 마커 구간에서 유의하게 고식미 품종에서 정렬이 되었고 저식미 품종에서 정렬이 되지 않아 해당 구간의 식미와 연관성을 확인했다. 더불어서 저식미 품종인 삼남벼에 P5 마커 구간을 이입한 근동질계통에서 식미 특성 중 도요 식미치에서 유의하게 삼남벼에 비해 향상 되었다. 이 결과는 기존의 고식미 품종에 P5 구간의 이입을 통해 윤기를 향상시켜 식미를 증진하는데 활용 될 수 있다.

향미가 아닌 자포니카 쌀에서 식미에 영향을 주는 주요 방향성 물질을 발견하고 이것을 벼 육종의 새로운 기준으로 활용하고자 식미 평가와 방향성분 물질 분석을 수행했다. 식미 평가와 도요 식미 분석을 통해 선별한 고식미 품종 7 개와 저식미 품종 7 개의 밥과 백미에서 headspace SPME GC/MS 분석으로 휘발성 방향물질을 프로파일링 했고 PLS-DA 분석을 통해 식미 구분에 결정적 작용을 하는 주요 물질을 선별 했다. 밥의 식미 구분에 큰 영향을 미친 물질로는 1-octen-3-ol, 1-ethyl-3,5-dimethylbenzene, 3-ethyloctane 등이 있었고, 백미에서 분석된 주요 물질로는 1-hexanol, methyl myristate, palmitic acid, benzene ethanol 등이 있었다. 밥과 백미의 프로파일링 결과 탐지된 방향성 물질들 중 두 시료 상에서 공통적으로 발견된 물질은 6 개가 있었지만 모두 식미 구분에 영향을 주는 주요 물질로 분류 되지는 않았다. 해당 연구에서는 밥의 휘발성 방향물질을 분석해 식미와 유관한 밥의 향에 대해 직접적인 접근을 취했다. 백미에서 분석된 식미 결정 주요 물질은 향후 자포니카 쌀의 향취 형질에 대해 새로운 육종 기준으로 활용 될 수 있다. 본 연구에서 취한 다양한 식미 구성요소와 복합적 식미 특성에 대해 통합적인 접근을 통해 종합적으로 연구 결과를 식미 향상 분자육종 프로그램에 활용할 수 있다.

주요단어: 자포니카, 벼 식미, 유전체, 전유전체연관분석, 방향성물질

학번: 2018-31862