# GENOME-WIDE ASSOCIATION MAPPING FOR RICE GRAIN QUALITY AND NUTRITION

A Dissertation

by

A S M FARIDUL ISLAM

Submitted to the the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Michael Thomson |
| Committee Members, | William Park |
| | Endang Septiningsih |
| | Joseph Awika |
| Head of Department, | David D. Baltensperger |

August 2021

Major Subject: Plant Breeding

Copyright 2021 ASM Faridul Islam

ABSTRACT

Being a staple food for half of the world population, rice plays a crucial role in terms of food security, economic impact, and nutrition. Superior rice grain quality is considered as one of the key characteristics selected by rice breeders during variety development and demanded by consumers when buying the rice product. Rice grain quality has four facets, including milling, appearance, eating and cooking, and nutritional quality. Genetically, grain quality traits are complex and quantitatively inherited, controlling by pleiotropic genes/QTLs with environmental effect. The present study was conducted to identify the genetic basis (QTLs) of grain quality traits by using single-locus (three models) and multi-locus (six models) methods of genome-wide association studies (GWAS) using 174 diverse rice accessions and 6,565 SNP markers. A total of 147 QTLs were identified for mineral elements, 216 QTLs were identified for grain appearance, milling, eating and nutritional traits, and 17 QTLs for resistant starch (RS) content. While 43 (29%), 28 (13%) and 1 (5%) of the identified QTLs co-located with the positions of known genes, QTLs, and markers reported previously, the remaining 104, 188 and 16 QTLs were novel. While single-locus methods alone detected 110, 106 and 9 SNPs for minerals, appearance and nutritional and RS content, respectively, 22, 64 and 3 SNPs were found by multi-locus methods alone, and 15, 46 and 5 QTLs were found by both GWAS methods. By conducting in silico gene expression analysis to identify candidate genes for the respective traits, 792 of 3129, 1329 of 4609 and 122 of 381 genes that were mined within the linkage disequilibrium distance of significant SNPs (250kb) were found being expressed in the rice reproductive stage. This study demonstrated the usefulness of using multiple GWAS

methods as a complement to each other to identify as accurate QTLs for the corresponding traits. This study also suggests that there are still many QTLs yet to be discovered across the diversity of rice accessions. The identified common QTLs along with novel QTLs will be valuable resources for further gene functional characterization that can help accelerate the genetic improvement of rice grain quality traits.

## CONTRIBUTORS AND FUNDING SOURCES

This research study work was supervised by a dissertation committee consisting of Professor Dr. Michael Thomson, Associate Professor Dr. Endang Septiningsih ( Dept. Soil and Crop Sciences), Professor Dr. Joseph Awika (Dept. Nutrition and Food Science), and Professor Dr. William Park (Dept. Biochemistry and Biophysics).

This research work was collaborated with Dr. Ming Chen, Research Chemist at Dale Bumpers National Rice Research Center, Stuttgart, AR, who supported by giving laboratory protocol, advising on phenotype data analysis.

All the other work related to this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by Texas A&M University, Texas A&M Agrilife Research.

TABLE OF CONTENTS

Page

ABSTRACT .....ii

CONTRIBUTORS AND FUNDING SOURCES.....iv

TABLE OF CONTENTS .....v

LIST OF FIGURES..... viii

LIST OF TABLES .....x

1. INTRODUCTION.....1

1.1. References .....7

2. IDENTIFYING THE GENETIC BASIS OF MINERAL ELEMENTS (CU, FE, MN, MG, K, AND ZN) IN RICE GRAIN USING A GENOME-WIDE ASSOCIATION MAPPING STUDY (GWAS) .....9

2.1. Introduction .....9
2.2. Materials And Methods.....12
2.2.1. Plant Materials.....12
2.2.2. Sample Preparation For Phenotyping.....12
2.2.3. Phenotypic Measurements.....14
2.2.4. Analysis Of Phenotypic Data .....16
2.2.5. Genotype Data Preparation .....16
2.2.5.1. Genotype Marker.....16
2.2.5.2. Analysis Of Population Structure And Kinship Coefficient .....16
2.2.5.3. Linkage Disequilibrium (LD) Analysis.....18
2.2.6. Genome-Wide Association Analysis (GWAS) .....18
2.2.7. In-Silico Gene Expression Analysis.....19
2.3. Results .....19
2.3.1. Population Structure And Linkage Disequilibrium (LD) Pattern .....19
2.3.2. Phenotypic Variation Analysis.....21
2.3.3. GWAS analysis .....21
2.3.4. In-Silico Gene Expression Analysis.....29
2.4. Discussion .....31
2.4.1. Population Structure, LD, And Phenotypic Variation.....31
2.4.2. Performance Of Single And Multi-Locus GWAS Models .....32

2.4.3. Comparison And Reliability Of Our GWAS Studies....................................40
2.5. Conclusion..................................................................................................46
2.6. References ..................................................................................................46

3. DISCOVERING CANDIDATE GENOMIC REGIONS FOR MILLING AND HEAD RICE YIELD, GRAIN SIZE AND SHAPE, CHALKINESS, APPARENT AMYLOSE CONTENT, PROTEIN CONTENT, AND GELATINIZATION TEMPERATURE BY USING GWAS ..........................................................49

3.1. Introduction ..................................................................................................49
3.2. Materials And Methods...................................................................................52
3.2.1. Plant Materials And Sample Preparation ......................................................52
3.2.2. Phenotypic Measurements.............................................................................52
3.2.3. Analysis Of Phenotypic Data .........................................................................53
3.2.4. GWAS Analysis ............................................................................................53
3.2.5. In-Silico Gene Expression Analysis..............................................................54
3.3. Results ..........................................................................................................54
3.3.1. Phenotypic Variation And Correlation Analysis............................................54

..........................................................................................................................55

3.3.2. Population structure and GWAS analysis .....................................................56
3.3.3. In-Silico Gene Expression Analysis..............................................................81
3.4. Discussion ......................................................................................................83
3.4.1. Population Structure, LD, And Phenotypic Variation...................................83
3.4.2. Performance Of Single And Multi-Locus GWAS Models .............................88
3.4.3. Comparison And Reliability Of Our GWAS Studies.....................................88
3.5. Conclusion......................................................................................................93
3.6. References ......................................................................................................94

4. GENOME-WIDE ASSOCIATION MAPPING FOR RESISTANT STARCH OF COOKED RICE USING DIVERSE GERMPLASM ......................................................97

4.1. Introduction ....................................................................................................97
4.2. Materials And Methods....................................................................................99
4.2.1. Plant Materials And Sample Preparation .......................................................99
4.2.2. Phenotypic Measurements..............................................................................99
4.2.3. Analysis Of Phenotypic Data ........................................................................100
4.2.4. GWAS Analysis ...........................................................................................100
4.2.5. *In-Silico* Gene Expression Analysis .........................................................100
4.3. Results ..........................................................................................................101
4.3.1. Phenotypic Variation Analysis.....................................................................101
4.3.2. Population structure and GWAS analysis ....................................................104
4.3.3. In-Silico Gene Expression Analysis.............................................................106
4.4. Discussion .....................................................................................................106

4.4.1. Phenotypic Analysis ........................................................................ 106
........................................................................ 108
4.4.2. Performance Of Single And Multi-Locus GWAS Models And Comparison GWAS Result ........................................................................ 108
4.5. Conclusion ........................................................................ 111
4.6. References ........................................................................ 112

5. CONCLUSIONS ........................................................................ 114

APPENDIX A DESCRIPTION OF THE RICE GERMPLASM USED IN THE STUDY ........................................................................ 116

APPENDIX B PHYSICAL MAP FOR SIX MINERAL ELEMENTS ........................ 125

LIST OF FIGURES


Page

Figure 2.1 Origin of rice accessions used in the study. 13

Figure 2.2 Outline of the GWAS analysis of the study 14

Figure 2.3 Population structure analysis result. 15

Figure 2.4 Phenotypic variation of the six mineral elements of rice grain 17

Figure 2.5 Correlation matrix for the six mineral elements. 20

Figure 2.6 Manhattan plots of GWAS for six minerals 27

Figure 2.7 Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for the six mineral elements. 28

Figure 2.8 All the significant SNPs found in the study for the six mineral elements 30

Figure 2.9 Heatmap of In-silico gene expression analysis results for the six mineral elements of the study. 41

Figure 3.1 Phenotypic variation of rice grain appearance, milling, eating and cooking, and nutritional quality traits 55

Figure 3.2 Correlation matrix for ten measured traits belong to rice grain appearance, milling, eating and cooking, and nutritional quality. 57

Figure 3.3 Manhattan plots of GWAS for ten measured traits used in the GWAS study 69

Figure 3.4 Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for ten traits of rice grain appearance, milling, eating and cooking, and nutritional quality.. 70

Figure 3.5 All the significant SNPs found in the study for ten rice grain appearance, milling, eating and cooking, and nutritional quality 71

Figure 3.6 Heatmap of In-silico gene expression analysis results for the ten measured traits of the study.. 82

Figure 4.1 Phenotypic variation of resistant starch (RS) of rice grain 101

Figure 4.2 Manhattan plots of GWAS for resistant starch (RS) 103

Figure 4.3 Physical map of the significant SNPs associated with resistant starch (RS).. 105

Figure 4.4 Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for resistant starch (RS) 107

Figure 4.5 Heatmap of In-silico gene expression analysis result of resistant starch (RS)... 108

LIST OF TABLES

Page

Table 2.1 Parameters used during microwave digestion.. 12

Table 2.2 Phenotypic variation in whole, *Indica* and *Japonica* rice accessions. 13

Table 2.3 List of loci detected in the study... 39

Table 2.4 Comparison of the GWAS result with the previous studies... 45

Table 3.1 Phenotypic variation in whole, *Indica* and *Japonica* rice accessions.. 58

Table 3.2 List of loci detected in the study.. 80

Table 3.3 Comparison of the GWAS result with the previous studies.. 87

Table 4.1 Phenotypic variation in whole, *Indica* and *Japonica* rice accessions.. 102

Table 4.2 Regression models for resistant starch (RS) content.. 102

Table 4.3 List identified QTLs for resistant starch (RS) content.. 110

Table 4.4 Comparison of the GWAS result with the previous studies.. 111

# 1. INTRODUCTION

Rice is the most important crop plant in the world in terms of food security, economic impact, and nutrition. It belongs to the genus *Oryza*, which is a member of the grass-type plant family (Gramineae family). Although the origin of rice is still a debated issue, it is assumed that the first domestication of the two most important *Oryza* species, *Oryza sativa* (Asian rice) and *Oryza glaberrima* (African rice), began between 8,200 and 13,500 years ago in China and 2,000-3,000 years ago in Mali, respectively (Global Rice Science Partnership, 2013). Since then, rice cultivation has spread to each continent except Antarctica; Asian rice species have been growing worldwide, whereas African rice is cultivated in parts of West Africa.

Rice is a diploid plant with 12 pairs of chromosomes, and its genome size is about 389 Mb (Song, Tian, Zhang, Hu, & Yu, 2018). Its genome is considered as a model genome for plant molecular biology and agricultural research due to its small genome size, fully sequenced genome with high accuracy, high-efficiency transformation technology, and large germplasm collections (Li et al., 2018). The availability of a high-quality sequenced rice genome, along with next generation sequencing (NGS) technologies, brought a revolution in rice breeding and genetic research aimed at to fulfill the demand of rice production (Jackson, 2016). To date, the following molecular technologies have been used to study rice genetics: Genome-wide association mapping, Whole Genome SNP Array, Genomic-based genotyping platforms and re-sequencing, Genome-guided RNA-seq, Transcriptome profiling, Map-based cloning approach, gene editing, Sequencing-By-Synthesis (SBS), and Next generation sequencing (NGS) technologies (Mahender, Anandan, Pradhan, & Pandit, 2016). These genomic approaches are especially effective for

complex traits controlled by multiple genes which are affected by the environment (Mahender et al., 2016). Most rice grain quality traits are complex quantitative traits. To dissect the genetic causes of complex traits, the genome-wide association study (GWAS), with high-resolution genome-wide markers, has become a common tool. It is a strategy of making associations between phenotype and genotype using diverse germplasm (Li et al., 2018).

Rice grain quality traits are complex and quantitatively inherited, controlled by multiple genes/QTL with environmental effect. To identify QTLs for rice quality traits, biparental QTL mapping, with two parents of contrasting performance, has been widely utilized. Great progress from major gene cloning and functional characterization to fine mapping of QTLs to identifying QTLs has occurred in the last two decades (Bao, 2014). However, QTL mapping using a single biparental population has some drawbacks, including the ability to detect only two alleles per locus and a low-resolution power because of having limited recombination events.

Association mapping, as an alternative approach to linkage mapping, is a method to link phenotypic and genotypic variation in unrelated germplasm accessions (Mezmouk et al., 2011). This approach mainly depends on linkage disequilibrium (LD), or the non-independence of alleles in a population, and requires high density markers to identify significant marker-trait associations (Qiu et al., 2015). The high-density marker issue has been solved by high-throughput NGS sequencing and SNP array techniques which generate large numbers of SNP markers throughout the genome. This approach has several advantages, such as: (1) it permits the use of current, available populations instead of the cross-controlled population that takes time and money; (2) it can detect more than two alleles per locus; and (3) it allows high resolution of mapping. Though

it is a promising technique, there are still some drawbacks, for example, rare alleles reduce the statistical power to identify associations, the necessity for large numbers of markers, and the population structure and/or relatedness between two samples need to be controlled (Mezmouk et al., 2011).

Initially, the genome-wide associate study (GWAS) technique was applied in human genetics and subsequently introduced in various plant species successfully (Mezmouk et al., 2011). This technique has also been widely used in rice. At first, X. Huang et al. (2010) used GWAS in rice for 14 agronomic traits, including six grain quality traits (grain width, grain length, grain weight, gelatinization temperature, protein content, and amylose content) using 517 rice landraces and ~3.6 million SNPs. Since then, GWAS has been used to dissect the genetic basis of rice traits ranging from morphological traits to physiological to biochemical to secondary metabolites traits (Chen et al., 2014; Magwa, Zhao, & Xing, 2016; Matsuda et al., 2015; Q. Wang et al., 2015; Q. Wang et al., 2017; Yang et al., 2018). Rice grain quality traits were included during the initial application of this approach and subsequent GWAS studies have been conducted for grain quality traits by other groups. Two and five QTLs were detected for GW and GL, respectively in the first GWAS studies in rice done by X. Huang et al. (2010). Using the same approach, but a different germplasm collection consisting of 950 worldwide rice accessions, Xuehui Huang et al. (2012) found two, four and two QTL for GW, GL, and 1000-grain weight (TGW), respectively. Zhao et al. (2011) identified eight, 10 and 15 QTLs for GW, GL and GLWR, respectively, using 44,100 SNP and 413 diverse accessions from 82 countries. Begum et al. (2015) used advanced lines from IRRI's irrigated breeding program as well as 71,710 SNPs generated by Genotyping-by-sequencing (GBS), helping to identify seven, 10 and 11 QTLs for GW, GL and GLWR,

respectively. With a GWAS study consisting of 533 *O. sativa* landrace and elite accessions and 4,358,600 SNP derived from GBS. Yang et al. (2018) found 10, 11, 19 and 21 QTLs for GLWR, TGW, GW and GL, respectively. Qiu et al. (2015) conducted GWAS mapping for 10 appearance and milling qualities together, including grain length, grain width, grain length to width ratio (GLWR), grain thickness, 1000-grain weight (TGW), degree of endosperm chalkiness, percentage of grains with chalkiness, brown rice rate, milled rice rate and head milled rice rate using 18,824 SNPs on 272 worldwide Indica accessions and identified 38 QTLs for ten characters. Similarly, Wang et al. (2016) detected 72 QTL affecting the nine traits using 258 accessions selected from 3K Rice Genome Project and 22,488 SNPs. By using the same accessions and SNP markers, X. Wang et al. (2017) conducted another GWAS study on eating and cooking quality, identifying eight, five, three and three QTL for amylose content (AC), gel consistency (GC), gelatinization temperature (GT) and protein content, respectively. Mogga et al. (2018) used 59 rice genotypes including upland and lowland rice cultivar and DArTseq platform derived 525 SNPs to perform GWAS study on AC and GT and 22 markers were significantly associated with two traits in which 2 markers positions were reported previously and 20 markers loci were novel. Regarding nutritional quality, few studies have been reported using GWAS approach. Bao, Zhou, Xu, He, and Park (2017) performed GWAS study for resistant starch using 137 rice accessions and 295 whole genome resequencing data and they identified four QTLs. Few studies using the GWAS method have been found for mineral's QTL identification. A GWAS mapping on As, Cu, Mo and Zn in whole grain rice was studied in five environments by Norton et al. (2014) using 300 accessions and 36.9 k single nucleotide polymorphisms (SNPs) and 17 QTLs were identified for four minerals. Another study was also reported by Nawaz et al. (2015), who used USDA mini-core subset and 155 SSR markers to conduct GWAS for eight minerals (Zn, Fe, Mn, Cu, P, Ca, K,

and Mg) in two environments, identifying 60 QTLs for eight elements. All of these reports proved that GWAS is an effective method for dissecting the genetic basis of any agronomically important traits in rice with having higher mapping resolution of biparental QTLs and can be used as an alternative approach to classical linkage mapping using biparental recombinant population.

Among the four parameters of rice grain qualities, milling and nutritional quality have been studied less comparatively in terms of genetic understanding. A limited number of QTLs using linkage mapping for milling quality have been reported due to a smaller number of studies. Recently, only two GWAS studies on milling quality have been published using different sets of germplasm accessions. Therefore, more genetic studies need to be conducted to dissect its complex genetic control. Likewise, in the area of nutritional quality, few QTL mapping reports using linkage mapping on resistant starch, and only one GWAS analysis, have been published. Many QTL studies are required to fully elucidate the molecular basis of resistant starch content in cooked rice.

Regarding mineral and protein content, while a good number of QTLs have been found using the linkage mapping approach, few GWAS studies have been published to date. Due to constraints of biparental QTL mapping, it can be assumed that QTLs affecting mineral and protein content could still be unknown and needing to be discovered using more GWAS studies. The USDA mini-core germplasm, which somewhat overlaps the accessions used in the current study, was previously used to detect QTLs for mineral content with GWAS and 155 SSR markers. But using few markers made the study’s resolution capacity low, potentially missed key genetic regions affecting the mineral content. In addition to the low marker issue, this current study will

be different from the former one in terms of the methodology used for mineral determination and selected germplasm, suggested that different QTLs may be detected from the previous studies.

The molecular basis of appearance and eating and cooking quality are relatively well established because many QTL studies have been conducted in the past, resulting in the cloning of some major genes and their functional characterization, along with fine mapping of major QTLs. But novel QTLs still have been identified in recent other GWAS studies using different natural populations, suggesting the existence of concealed minor QTLs in different genetic populations. It is well known that rice has excellent germplasm resources, with 773,948 rice accessions stored in different rice gene bank collections around the world (Song et al., 2018). As this study uses diverse germplasm resources collected from across the world, along with the 7k SNP Chip, it is expected that novel QTLs could be unveiled, along with validation of previously identified QTLs from past studies.

Considering the above-mentioned background information and importance of traits and justification of the approach, the objectives of this study are:

1. Identifying the genetic basis of micronutrient content (Cu, Fe, Mn, Mg, K, Zn) in rice grains using a genome-wide association mapping study (GWAS);
2. Discovering candidate genomic regions for milling and head rice yield, grain size and shape, chalkiness, apparent amylose content, protein content and gelatinization temperature by using GWAS;

3. Genome-wide association mapping for resistant starch of cooked rice using diverse germplasm and characterizing the relationship of resistant starch with apparent amylose content, protein content, chalkiness, and gelatinization temperature in rice.

### 1.1. References


Bao, J. (2014). Genes and QTLs for Rice Grain Quality Improvement. In *Rice - Germplasm, Genetics and Improvement*.

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

Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., . . . McCouch, S. R. (2015). Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (Oryza sativa). *PLoS One, 10*(3), e0119873. doi:10.1371/journal.pone.0119873

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., . . . Dong, H. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature genetics, 46*(7), 714.

Global Rice Science Partnership. (2013). *Rice almanac* (4th ed.). Los Baños, Philippines: International Rice Research Institute.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., . . . Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet, 42*(11), 961-967. doi:10.1038/ng.695

Huang, X., Zhao, Y., Li, C., Wang, A., Zhao, Q., Li, W., . . . Fan, D. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature genetics, 44*(1), 32.

Jackson, S. A. (2016). Rice: The First Crop Genome. *Rice (N Y), 9*(1), 14. doi:10.1186/s12284-016-0087-4

Li, Y., Xiao, J., Chen, L., Huang, X., Cheng, Z., Han, B., . . . Wu, C. (2018). Rice Functional Genomics Research: Past Decade and Future. *Mol Plant, 11*(3), 359-380. doi:10.1016/j.molp.2018.01.007

Magwa, R. A., Zhao, H., & Xing, Y. (2016). Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (Oryza sativa L.). *BMC genetics, 17*(1), 28.

Mahender, A., Anandan, A., Pradhan, S. K., & Pandit, E. (2016). Rice grain nutritional traits and their enhancement using relevant genes and QTLs through advanced approaches. *Springerplus, 5*(1), 2086. doi:10.1186/s40064-016-3744-6

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J. i., Ebana, K., . . . Saito, K. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *The Plant Journal, 81*(1), 13-23.

Mezmouk, S., Dubreuil, P., Bosio, M., Decousset, L., Charcosset, A., Praud, S., & Mangin, B. (2011). Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theor Appl Genet, 122*(6), 1149-1160. doi:10.1007/s00122-010-1519-y

Mogga, M., Sibiya, J., Shimelis, H., Lamo, J., & Yao, N. (2018). Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (Oryza sativa L.) using DArT markers. *PLoS One, 13*(6), e0198012. doi:10.1371/journal.pone.0198012

Nawaz, Z., Kakar, K. U., Li, X. B., Li, S., Zhang, B., Shou, H. X., & Shu, Q. Y. (2015). Genome-wide Association Mapping of Quantitative Trait Loci (QTLs) for Contents of Eight Elements in Brown Rice (Oryza sativa L.). *J Agric Food Chem, 63*(36), 8008-8016. doi:10.1021/acs.jafc.5b01191

Norton, G. J., Douglas, A., Lahner, B., Yakubova, E., Guerinot, M. L., Pinson, S. R., . . . Price, A. H. (2014). Genome wide association mapping of grain arsenic, copper, molybdenum and zinc in rice (Oryza sativa L.) grown at four international field sites. *PLoS One, 9*(2), e89685. doi:10.1371/journal.pone.0089685

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., . . . Ye, G. (2015). Genome-Wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of Indica Rice Germplasm. *PLoS One, 10*(12), e0145577. doi:10.1371/journal.pone.0145577

Song, S., Tian, D., Zhang, Z., Hu, S., & Yu, J. (2018). Rice Genomics: over the Past Two Decades and into the Future. *Genomics Proteomics Bioinformatics, 16*(6), 397-404. doi:10.1016/j.gpb.2019.01.001

Wang, Q., Xie, W., Xing, H., Yan, J., Meng, X., Li, X., . . . Yu, S. (2015). Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. *Molecular plant, 8*(6), 946-957.

Wang, Q., Zhao, H., Jiang, J., Xu, J., Xie, W., Fu, X., . . . Wang, G. (2017). Genetic architecture of natural variation in rice nonphotochemical quenching capacity revealed by genome-wide association study. *Frontiers in plant science, 8*, 1773.

Wang, X., Pang, Y., Wang, C., Chen, K., Zhu, Y., Shen, C., . . . Li, Z. (2016). New Candidate Genes Affecting Rice Grain Appearance and Milling Quality Detected by Genome-Wide and Gene-Based Association Analyses. *Front Plant Sci, 7*, 1998. doi:10.3389/fpls.2016.01998

Wang, X., Pang, Y., Zhang, J., Wu, Z., Chen, K., Ali, J., . . . Li, Z. (2017). Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. *Sci Rep, 7*(1), 17203. doi:10.1038/s41598-017-17347-5

Yang, M., Lu, K., Zhao, F.-J., Xie, W., Ramakrishna, P., Wang, G., . . . Zhao, H. (2018). Genome-wide association studies reveal the genetic basis of ionomic variation in rice. *The Plant Cell, 30*(11), 2720-2740.

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., . . . McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat Commun, 2*, 467. doi:10.1038/ncomms1467

# 2. IDENTIFYING THE GENETIC BASIS OF MINERAL ELEMENTS (CU, FE, MN, MG, K, AND ZN) IN RICE GRAIN USING A GENOME-WIDE ASSOCIATION MAPPING STUDY (GWAS)

## 2.1. Introduction

Being a staple food for half of the world's population, the nutritional quality of rice can have a large impact on human nutrition. The lack of nutritional quality mainly affects the society or country where rice is eaten primarily as a staple food. Mineral malnutrition is one of the more serious problems for rice-eating societies, specifically in Asian countries (Y. Huang et al., 2015). More than 60% and 30% of the world's population have iron (Fe) and zinc (Zn) deficiency, respectively, because of low mineral content availability in their staple foods, including rice (Graham et al., 2007; Nawaz et al., 2015). To solve the issue, biofortification, improving the nutrition of foods through either conventional plant breeding or genetic engineering has been suggested as a possible way to address malnutrition by developing new varieties with high concentrations of minerals (Y. Huang et al., 2015). Towards that end, genetic mechanisms controlling the accumulation of the mineral elements in rice grain need to be studied thoroughly.

So far, linkage mapping has been applied to find QTLs controlling mineral uptake in rice grain. There are 217 QTLs identified for micro (Fe, Zn, Mn, Cu) and macro (Ca, Mg, P and K) nutrients elements (Mahender et al., 2016). As biparental QTL mapping only has the ability to detect two alleles per locus while multiple alleles per locus can segregate, and, due to the limited resolving power due to the limited number of recombination events, the identified QTLs are often found in relatively large genomic regions, making it very difficult to pinpoint the causal genes (Gong et al., 2017; Qiu et al., 2015).

Association mapping based on linkage disequilibrium (LD), as an alternative approach of linkage mapping, is a powerful method for dissecting the genetic basis of plant traits (Mezmouk et al., 2011). This approach has several advantages, including: (1) it permits the use of existing populations instead of cross-controlled populations that take time and money to develop; (2) it can detect more than two alleles per locus and (3) it enables a high resolution of mapping. Though it is a promising technique, there are still some drawbacks; for example, rare alleles necessitate large population sizes to provide statistical power to detect rare alleles; likewise, a large number of markers is required to provide high resolution, and population structure and/or relatedness between accessions need to be controlled (Mezmouk et al., 2011). Initially, genome-wide associate studies (GWAS) were applied in human genetics and then successfully introduced in various plant species (Mezmouk et al., 2011). This technique has also been used widely in rice, starting in 2010 by X. Huang et al. (2010) using GWAS to detect QTLs for 14 agronomic traits.

To conduct GWAS, several statistical models have been widely used, such as the general linear model (GLM) and the mixed linear model (MLM) (Bradbury et al., 2007). MLM was the most popular because of having the ability to account for population structure and family relatedness. The Efficient Mixed-Model Association eXpedited (EMMAX), Population Parameters Previously Determined (P3D), and Genome-wide Efficient Mixed Model Association (GEMMA) have been built up based on MLM, helping to reduce the computational time for analysis (Yang Xu et al., 2018). However, these methods are unidirectional, testing one locus at a time, resulting in failure to capture the multiple loci controlling complex traits simultaneously. Moreover, multiple test corrections for threshold values are required to control the false positive rate. The Bonferroni correction is often used; however, it is too conservative, resulting in many

important loci being ignored because they do not fulfill the significance threshold level (C. Li, Fu, Sun, Wang, & Wang, 2018; Y. Xu et al., 2018).

The multi-locus models have been proposed as an alternative to overcome the issues with the single-locus model GWAS. These multivariate models consider all the loci simultaneously; as a result, multiple test corrections are not needed. So far, several multi-locus GWAS models have been developed and used to study GWAS, such as MLMM (multi-locus mixed-model), FarmCPU (Fixed and random model Circulating Probability Unification), mrMLM (multi-locus random-SNP-effect MLM), FASTmrMLM (fast mrMLM), FASTmrEMMA (fast multi-locus random-SNP-effect efficient mixed model analysis), pLARmEB (polygenic background-control-based least angle regression plus empirical Bayes), pKWmEB (integration of Kruskal-Wallis test with empirical Bayes), ISIS EM-BLASSO (iterative modified-sure independence screening expectation-maximization-Bayesian least absolute shrinkage and selection operator), and GPWAS (Genome-Phenome Wide Association Study) (Liu et al., 2020). All the multi-locus models follow the two-step principle during analysis. In the first stage, all the potentially associated SNPs are identified or scanned in the whole genome. During the second step, all the identified SNPs are included into one model, then their effects are estimated by empirical Bayes, and finally all the non-zero effects are further evaluated using the likelihood ratio test. A less stringent critical p-value such as 0.01, is used to select the SNPs in the first step. Each of these multi-locus modes is different from each other in terms of algorithms utilized in the two steps (Cui, Zhang, & Zhou, 2018; C. Li et al., 2018; Y. Xu et al., 2018).

The main objectives of this study are: 1) to identify loci that are significantly associated with six mineral elements (Cu, Fe, K, Mg, Mn and Zn) by using single-locus and multi-locus GWAS methods; and 2) to compare the performance of these methods in terms of SNPs detection. The findings of the study will help us gain insight into the molecular mechanisms underlying mineral accumulation in rice grain and the identification of potential candidate genes, accelerating the development of new mineral-rich rice varieties.

## 2.2. Materials And Methods

### 2.2.1. Plant Materials

A total of 174 accessions, including 151 diverse accessions from the USDA-GRIN germplasm collection and 23 US-released varieties, having a similar heading date to avoid any effect of flowering time on rice grain quality, were used in the study. These accessions originated from 31 countries, where the highest number of accessions were from Bangladesh (19) followed by Russia (18), Uzbekistan (16), India (14) (Figure 2.1; Appendix A).

Table 2.1. Parameters used during microwave digestion.

| Power (W) | % Max | Time (min) to raise temperature | Temperature | Running time (min) |
|---|---|---|---|---|
| 1600 | 50 | 20.5 | 160 | 4.5 |

### 2.2.2. Sample Preparation For Phenotyping

The field experiment was conducted at the Texas A&M AgriLife Research Center, Beaumont, Texas (30.0802° N, 94.1266° W), during the growing season from late April to

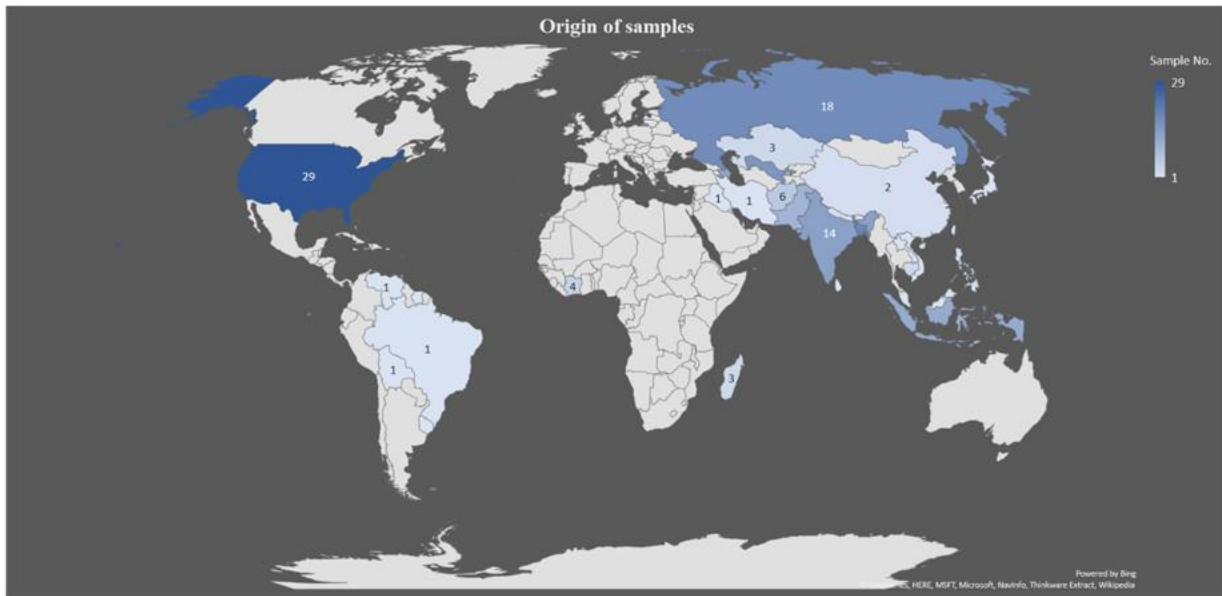



Figure 2.1. Origin of rice accessions used in the study.

September in 2018. A randomized complete block design with two replications with a two-row plot for each replication for each accession was used in field experiments. After reaching the maturity stage, plants in the middle of each plot were bulk harvested and air-dried for 3 months in the drying room. Then, around 120 g rough seeds were dehulled with electrical dehuller to make brown rice, followed by milling by using PAZ-1 DTA (Zaccaria USA, Anna, TX).

Table 2.2. Phenotypic variation in whole, *Indica* and *Japonica* rice accessions

| **Traits** | **Whole Panel** | | | | | | | **Indica** | | **Japonica** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | CV | R2 | P-value | Mean | SD | Mean | SD |
| Cu | 2.99 | 0.66 | 1.33 | 6.63 | 22.06 | 0.02 | 0.04 | 2.91* | 0.70 | 3.06 | 0.62 |
| Fe | 17.21 | 3.45 | 6.37 | 30.77 | 20.04 | 0.10 | 0.00 | 16.71* | 3.43 | 17.64 | 3.42 |
| K | 2492.84 | 422.67 | 793.67 | 4788.57 | 16.96 | 0.01 | 0.34 | 2471.84 | 425.71 | 2510.32 | 420.47 |
| Mg | 1419.99 | 253.32 | 435.10 | 2585.49 | 17.84 | 0.01 | 0.11 | 1434.59 | 250.75 | 1407.83 | 255.47 |
| Mn | 21.78 | 5.38 | 6.85 | 48.99 | 24.71 | 0.02 | 0.02 | 20.88* | 5.04 | 22.54 | 5.55 |
| Zn | 26.53 | 8.77 | 4.75 | 52.64 | 33.05 | 0.04 | 0.00 | 27.96* | 6.81 | 25.34 | 9.98 |

N.B: * indicate the significant difference at α= 0.05 level

### 2.2.3. Phenotypic Measurements

We used brown rice for mineral content determination for this study. At first, brown rice of all the samples were dried for 72hrs at 65ºC, followed by sterilization with 70% ethanol to remove contaminants and/or debris from the surface. Then, seeds were grounded into fine powder by mortar and pistol and kept in the airtight plastic zip lock bags or small container or tubes till the sample digestion was started.

About 0.5 g of rice sample was weighed accurately and poured directly into PFA vessel followed by adding reagents consisting of 6 ml HNO3 and 3 ml of 30% (v/v) H2O2 (Nardi et al., 2009). The digestion vessels were then closed and heated in the CEM MarXpress Microwave Accelerated Reaction System (CEM Corporation, NC, USA) using the parameters shown in the Table 2.1. After digestion, the obtained solutions were allowed to cool down to room temperature, and then were filtered through Whatman No. 1 (11 µm pores size) filter paper into a 25 volumetric flask. The volume was made up to the mark with ultrapure water. Next, two sets of

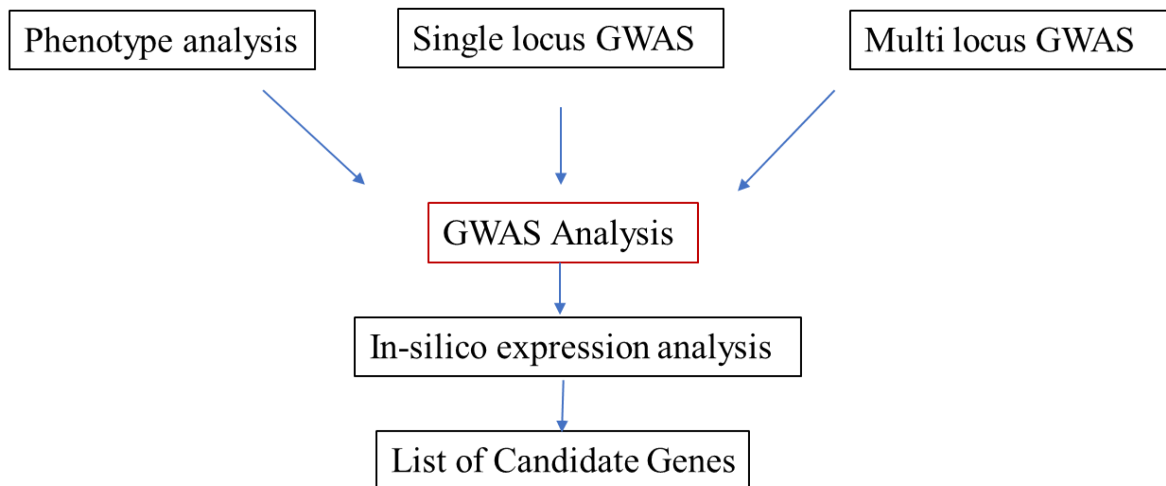



Figure 2.2. Outline of the GWAS analysis of the study

diluted samples were prepared from the solution for 100X and 1000X dilution for determining Cu, Fe, Mg and Mn content and K and Zn content, respectively.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS) (Agilent Technologies) was used to quantify the Cu, Fe, K, Mg, Mn and Zn content. Samples were run per batch and 3 blanks without samples were included in each batch during analyzing the elements. Five technical replications data were generated for each sample and were averaged. The average elemental

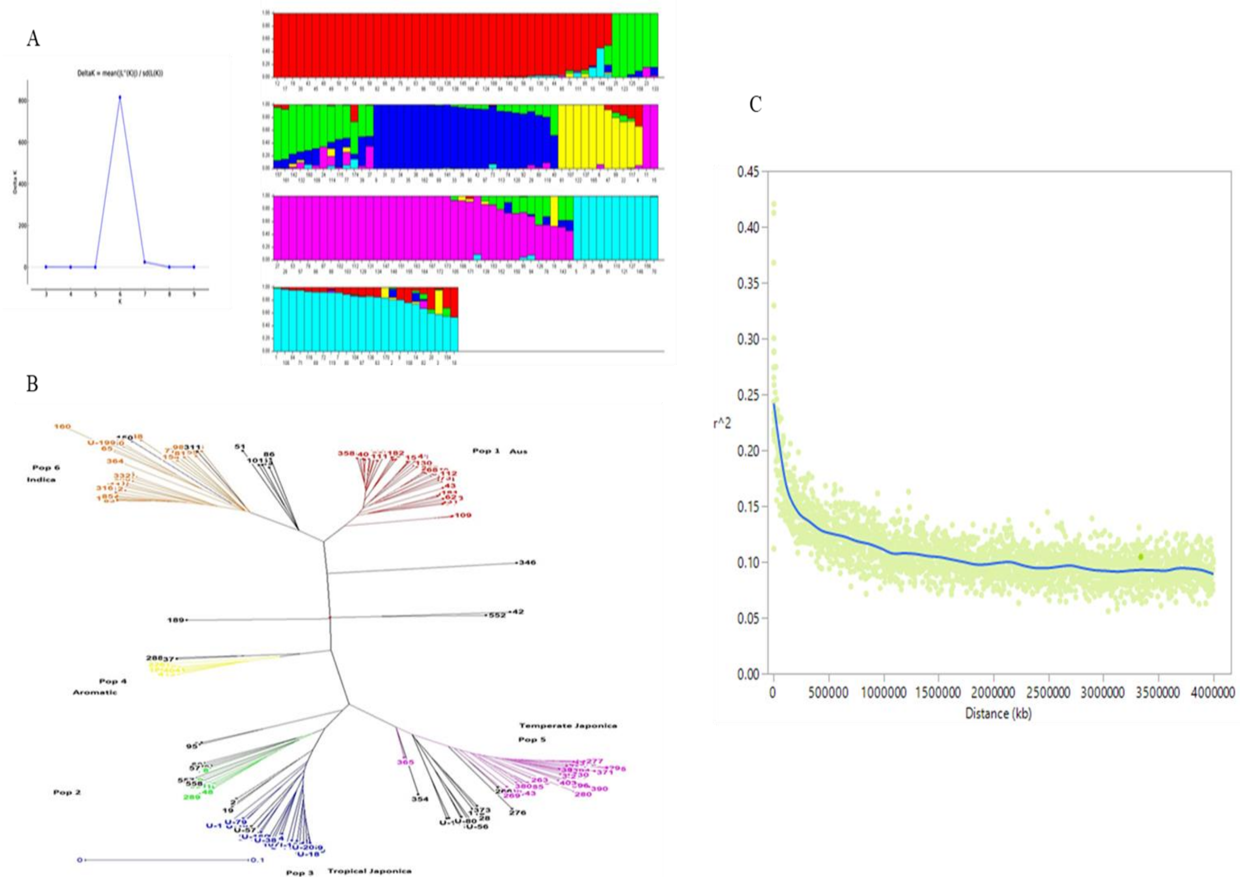



Figure 2.3. (A) Population structure analysis result. Δk is highest at 6, indicating six groups existence in the rice germplasm used in the study. (B) Result of the Cluster analysis, showed similar result of Structure analysis. The branches with different colors showed different sub-population of the rice germplasm. (C) LD decay pattern for the whole genome of rice.

concentration of two biological/ field replications of each accession was used during GWAS analysis.

### 2.2.4. Analysis Of Phenotypic Data

Basic statistics, including mean, standard deviation, coefficient of variation (CV), analysis of variance (ANOVA) was conducted on the whole accession as well as on two subpopulations accessions, Indica and Japonica, to determine the phenotypic variation. To find out if the population structure has an effect on the phenotypic variation, we used ANOVA using the general linear model, where population structure was set as the fixed independent variable. In addition, correlation analysis among the minerals was done. All the analysis was conducted by using JMP pro15.

### 2.2.5. Genotype Data Preparation

#### 2.2.5.1. Genotype Marker

We used 6,565 high quality SNPs from the 7K SNP array data (Morales et al., 2020) as our molecular markers for GWAS analysis. To impute the missing genotypes, MACH 1.0 was used, which is a Markov Chain based haplotyper that infers the missing genotypes by comparing the available genotypes to those in other accessions that have been typed at a higher density (Y. Li, Willer, Ding, Scheet, & Abecasis, 2010).

#### 2.2.5.2. Analysis Of Population Structure And Kinship Coefficient

Population structure and kinship analysis were conducted to control the false positive results in the GWAS analysis. STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) using Bayesian clustering analysis method was used for determining population structure with the

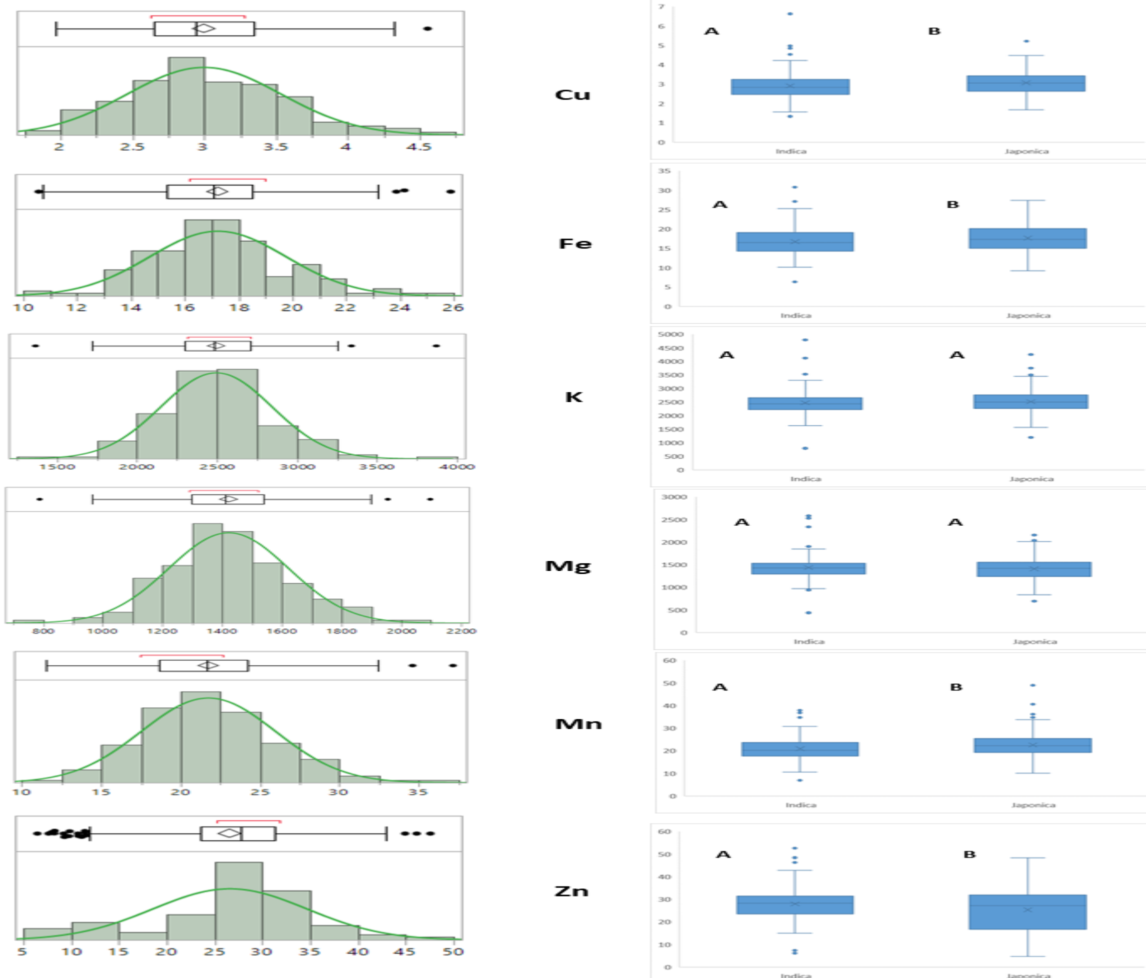



Figure 2.4. Phenotypic variation of the six mineral elements of rice grain. Different capital letters in the same box plot indicate Indica and Japonica rice accessions are significantly different at α= 0.05 for mean value of the six mineral elements.

following profile: k, the number of genetic clusters in the panel ranging from 2 to 10 with 10 runs for each K value; burn-in time for each run was 10,000 followed by 50,000 MCMC (Markov Chain Monte Carlo) iterations. The Structure Harvester program was used (http://taylor0.biology.ucla.edu) to determine the best k value using the method of Evanno, Regnaut, and Goudet (2005) by submitting the results for each K, and determining log(k)2 and Δk values. Within-population membership probability (Q) threshold was fixed at 0.80, so an

individual with higher Q was assigned to a population, whereas an individual having lower Q was considered admixed. For calculating kinship coefficient matrix (K), four methods implemented in four different software packages were utilized. The TASSEL 5 uses the scaled_IBS method, as a default method, to calculate kinship, whereas the VanRaden method is used in GAPIT. For GEMMA software, a centered relatedness matrix system was used to calculate the kinship in this study. The default method was used during running mrMLM for GWAS analysis.

#### 2.2.5.3. Linkage Disequilibrium (LD) Analysis

The LD decay distance across the whole genome was measured by squared allele frequency correlations ($r^2$) values between the pairs of markers of 6,565 SNPs calculated by PopLDdecay 3.41 (C. Zhang, Dong, Xu, He, & Yang, 2019). Marker pairs were discretized into bins of 1.5 kb and the average $r^2$ value was used as the estimate of $r^2$ of a bin. The LD decay was calculated as the chromosomal distance at which the average $r^2$ dropped to half of its maximum value (X. Huang et al., 2010).

### 2.2.6. Genome-Wide Association Analysis (GWAS)

The GWAS was conducted using nine models that can be divided broadly into two groups; single-locus models: CMLM (compressed mixed linear model), ECMLM (Enriched CMLM) and GEMMA (Genome-wide Efficient Mixed Model Association algorithm), and multi-locus models: mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO. All six multi-locus models are implemented in mrMLM R package (Y.-W. Zhang et al., 2020). SNPs with $p < 10^{-3}$ were considered significant for the single-locus models and LOD= 3.0 was used as a cut-off value to declare a significant QTN in all of the multi-locus models (Figure 2.2).

#### 2.2.7. In-Silico Gene Expression Analysis

We mined the genes within the LD decay distance on either side of the significant SNPs by using RAP-DB database (https://rapdb.dna.affrc.go.jp/). To check the in-silico expression levels of the mined genes, Nipponbare gene expression data was downloaded from the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/expression.shtml). A heatmap of the gene expression for each trait was created with LDheatmap R package.

### 2.3. Results

#### 2.3.1. Population Structure And Linkage Disequilibrium (LD) Pattern

According to the value of Δk from the Structure analysis result, there were six groups or sub-populations in our study sample panel. These six sub-populations were used for the Q-matrix as a covariate during the GWAS analysis to account for the population structure (Figure 2.3). It is well known that rice has two major sub-species, Indica and Japonica. Studies of global rice germplasm have shown that the Indica subspecies consists of the *aus* and *indica* subgroups and the *Japonica* subspecies consists of the temperate japonica, tropical japonica, and aromatic subgroups. To determine the population structure effect on the phenotypic variations, we just considered the two primary sub-populations to analyze the phenotypic variation, which was also observed during cluster analysis as two distinct clusters with some other sub-groups under the two main clusters (Figure 2.3). Therefore, 78 and 93 accessions were considered as Indica and Japonica groups in the panel: in total, 171 samples were analyzed during phenotypic analysis. Three accessions were removed due to admixture.

The linkage disequilibrium (LD) decay across all chromosomes was estimated to be 250 kb, with half the maximum of mean r2 values (Figure 2.3).

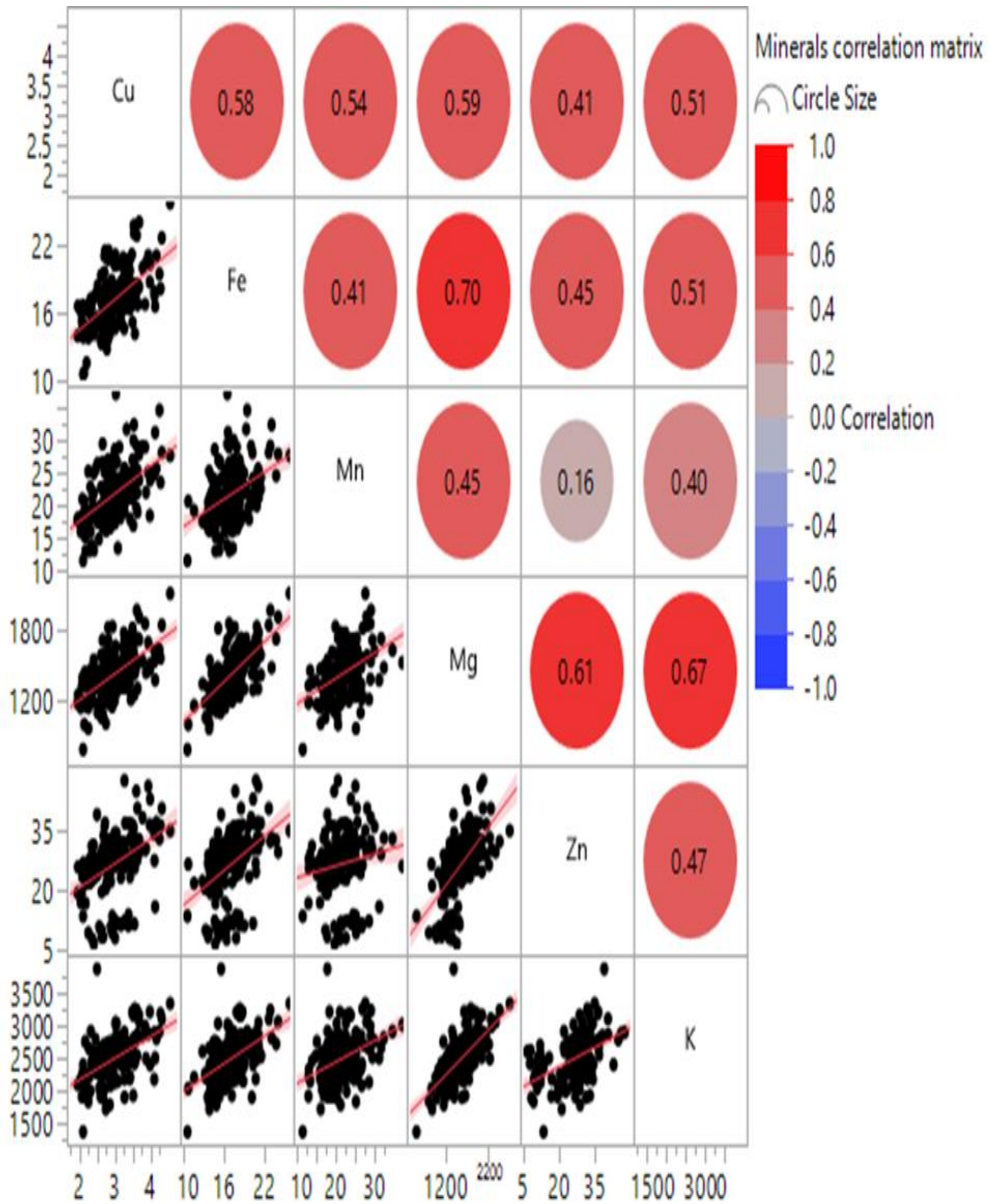



Figure 2.5. Correlation matrix for the six mineral elements. The value inside the circles shows correlation value between two minerals. Size of the circle indicate the magnitude of significant level at α= 0.05.

### 2.3.2. Phenotypic Variation Analysis

The phenotypic evaluation shows a broad variation among accessions. Overall, most of the traits appeared to be normally distributed, but Zn showed slightly skewed distribution (Figure 2.4). Given that population structure is the main factor affecting GWAS, population structure is explained from 1% (K, Mg) to 10% (Fe) of the phenotypic variation in the whole panel. Mean differences between the *indica* and *japonica* sub-group panels were found significant for Cu, Fe, Mn and Zn except for K and Mg (Figure 2.4; Table 2.2).

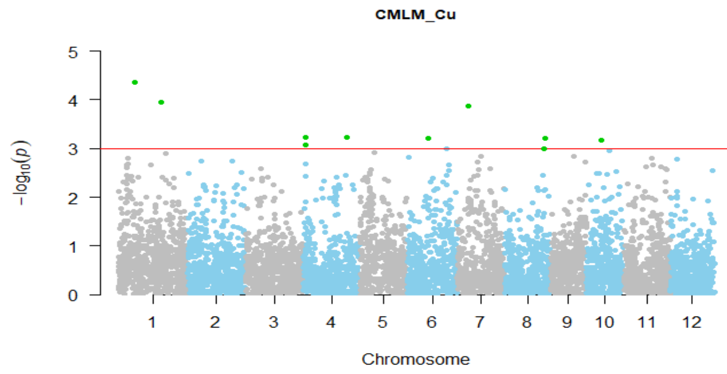
To determine the correlation among the six mineral elements, the Pearson's correlation coefficients were calculated. Figure 2.5 shows that all the pairwise correlations between any two minerals are significantly positive. Apart from Mn and Zn, all other r2 values were ≥ 40.0.

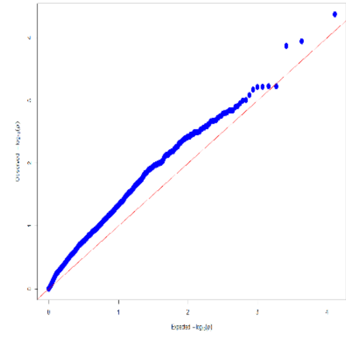
### 2.3.3. GWAS analysis

We declared a SNP as significant when it's –log10P ≥ 3.0 and LOD ≥ 3.0 for all single and multi-locus models, respectively. SNPs with MAF < 0.05 were not considered as significant. Multiple SNPs with physical distance less than 250 kb were regarded as the same significant locus (i.e., significant SNP-trait association). Based on these criteria, a total of 147 significant SNPs for six mineral elements were identified using nine models.

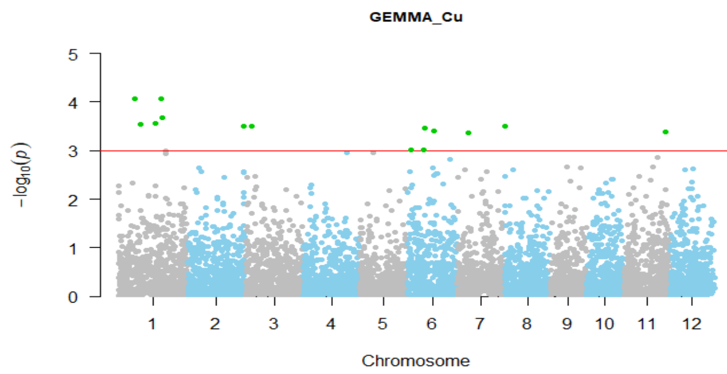
For Cu elements, 16 significant SNPs were detected by only single-locus GWAS models and explained 7.20-14.46% of the phenotypic variation. One SNP was found by only a multi-locus model and it explained 8.69% of the phenotypic variation. Both single and multi-locus models found two additional SNPs that explained 10.38-17.62% of the phenotypic variation. Only single-

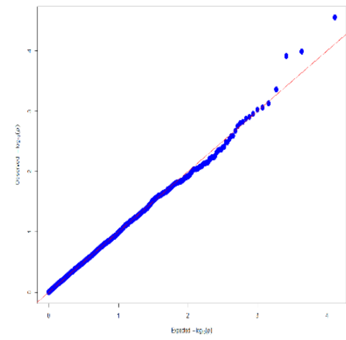
(A)



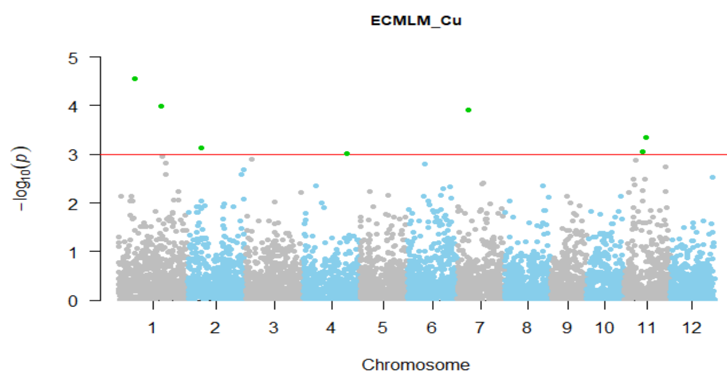



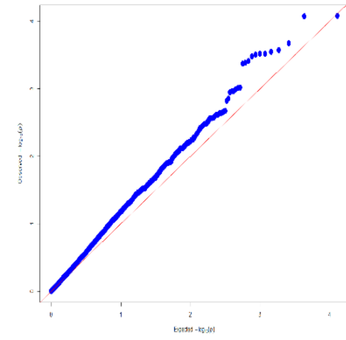
1



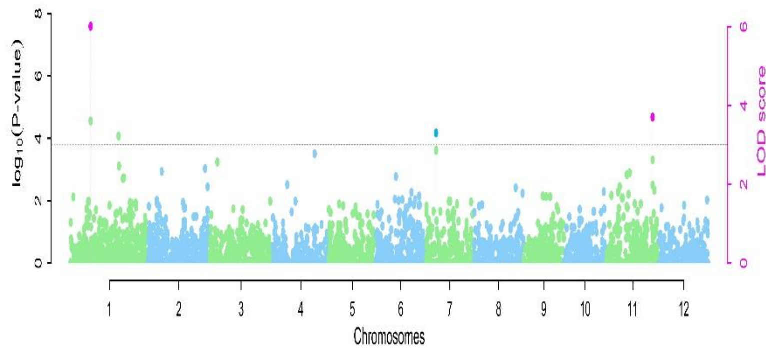



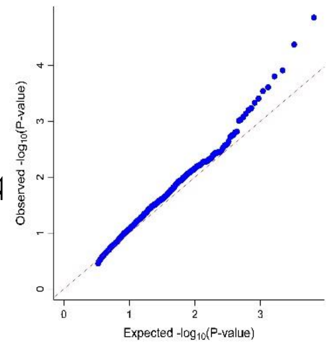
2



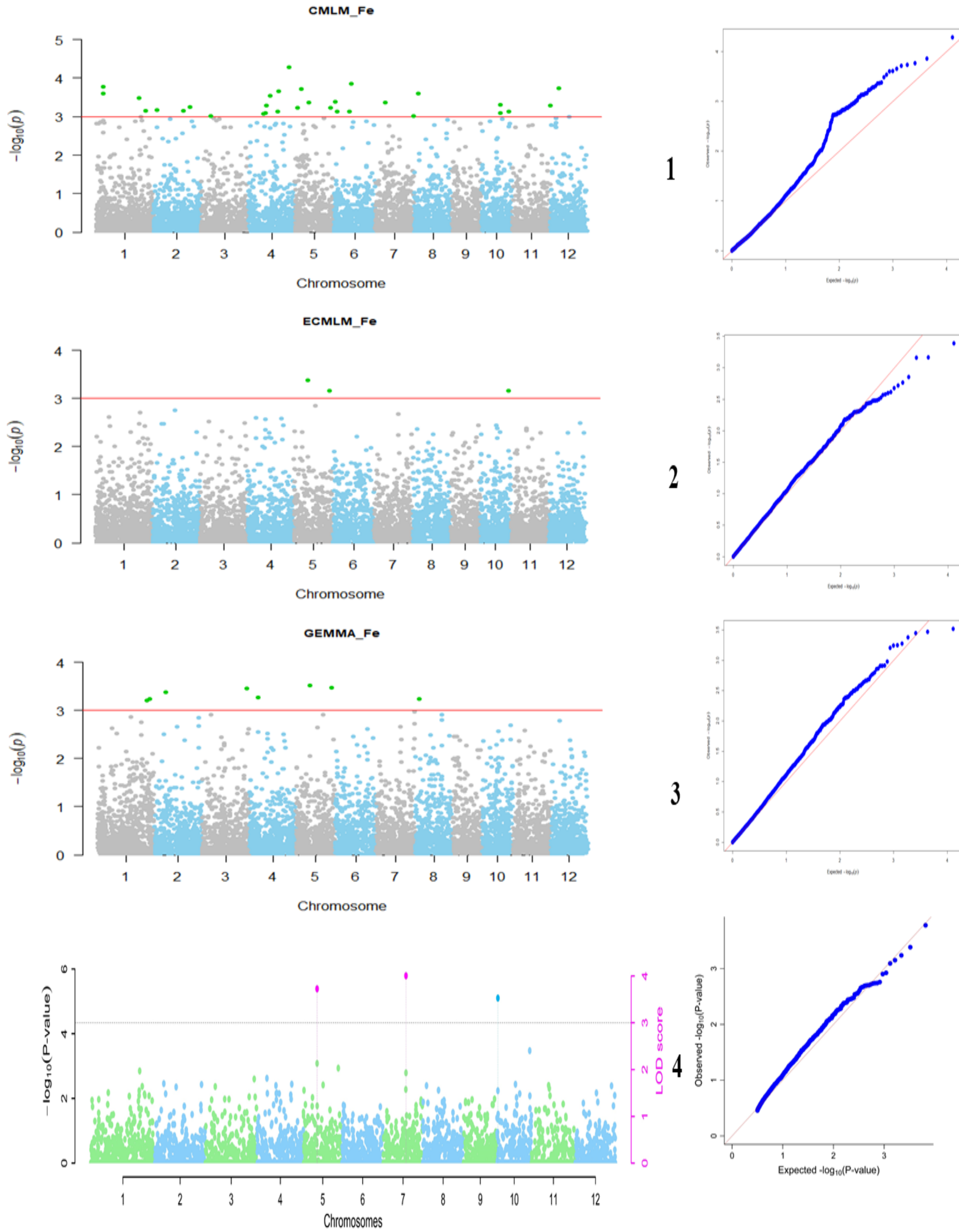
3



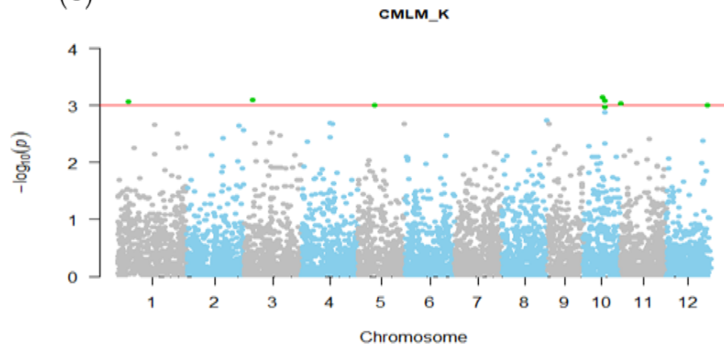




4

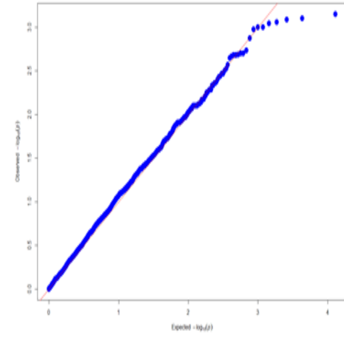
(B)

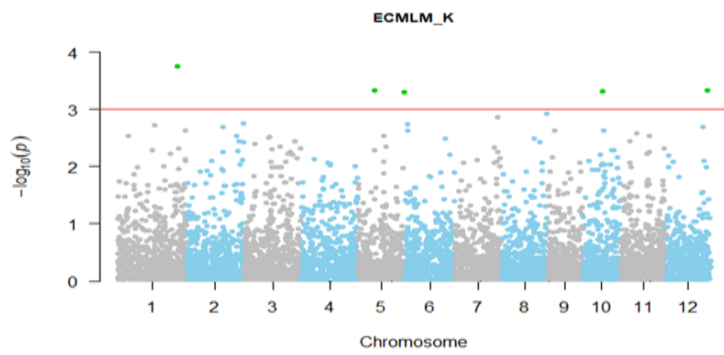


(C)

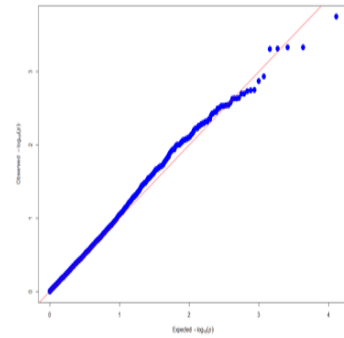


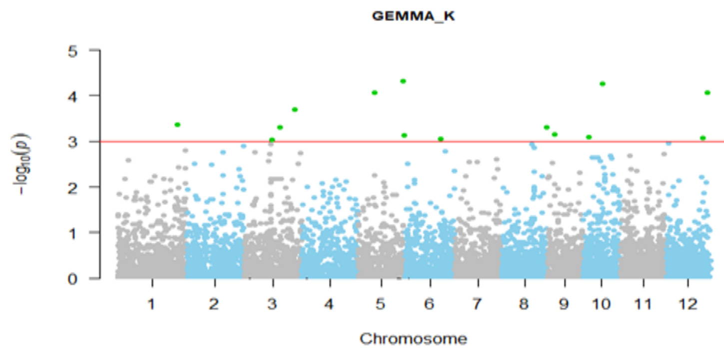

1



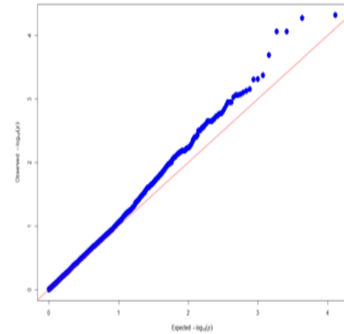


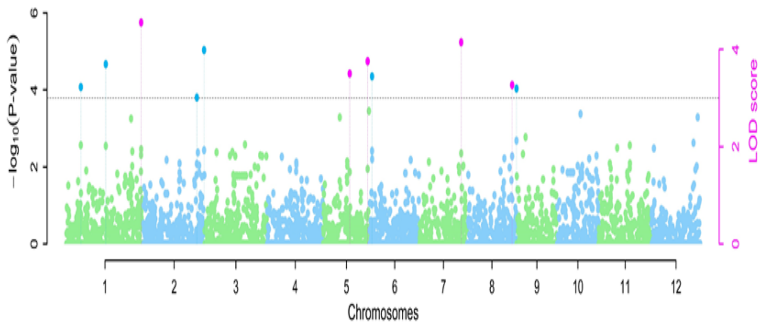

2



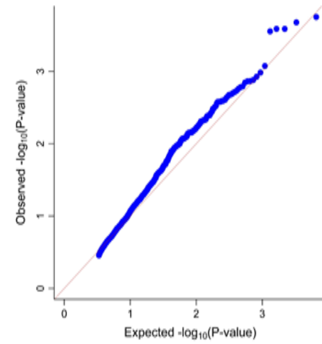


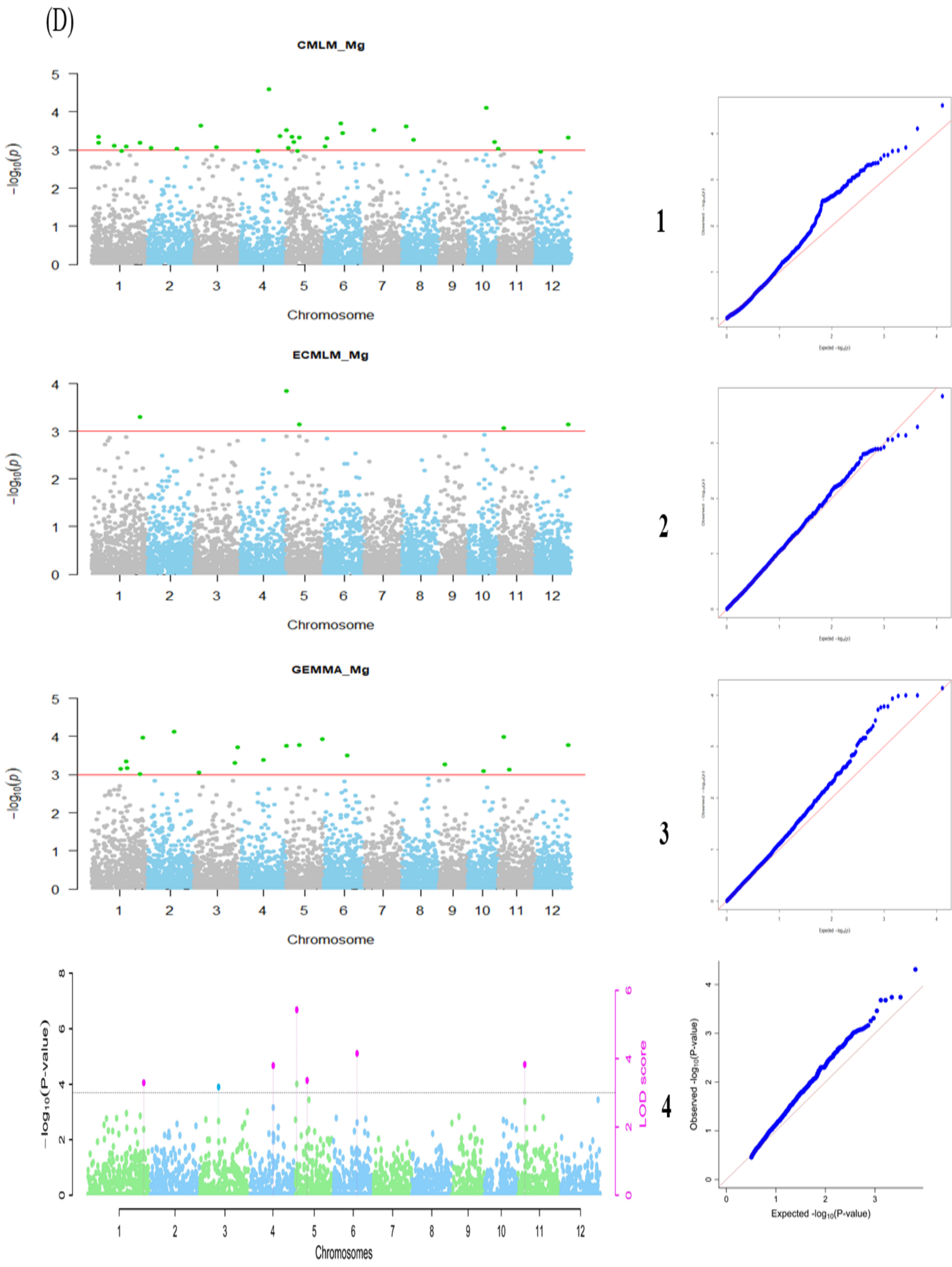

3



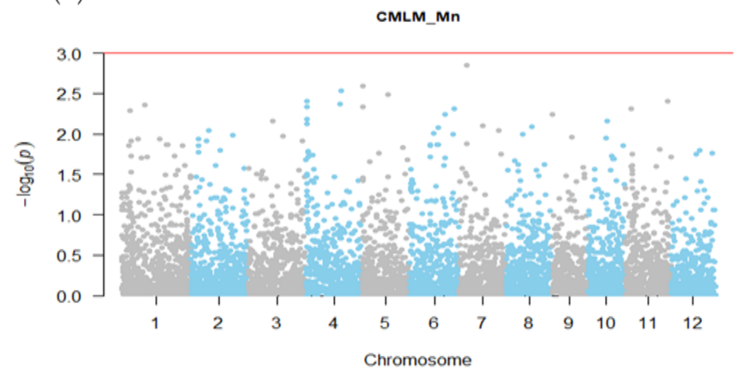


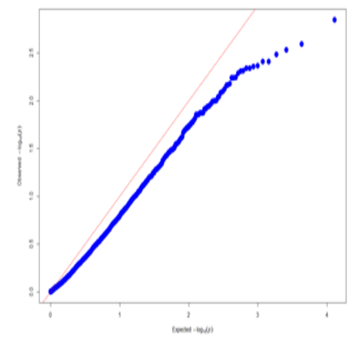

4



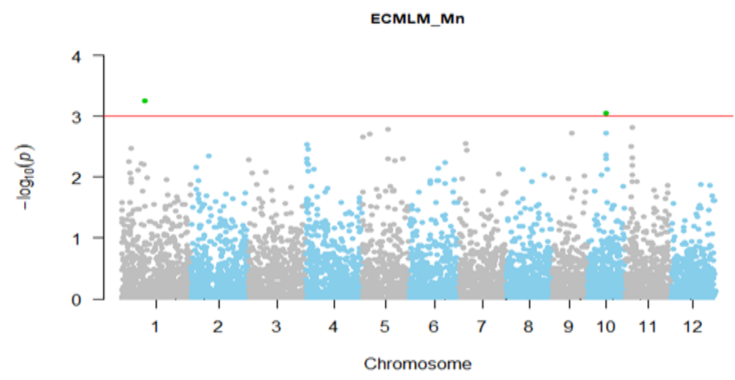

(D)
CMLM_Mg
-log10(p)
Chromosome
1
ECMLM_Mg
2
GEMMA_Mg
3
-log10(P-value)
LOD score
Chromosomes
4
Observed -log10(P-value)
Expected -log10(P-value)

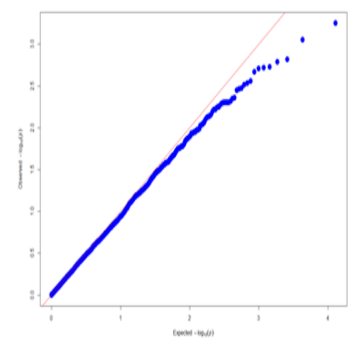
(E)



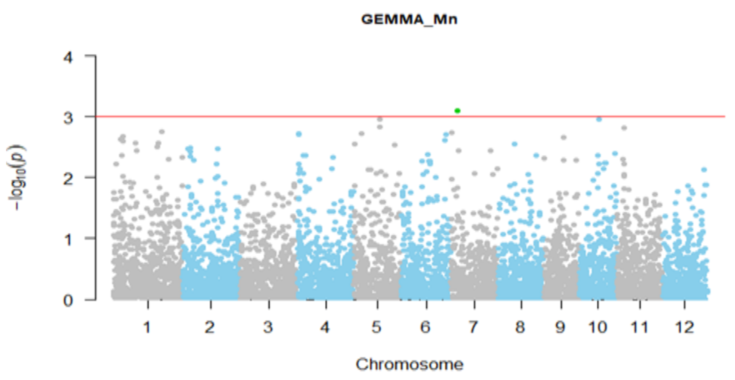



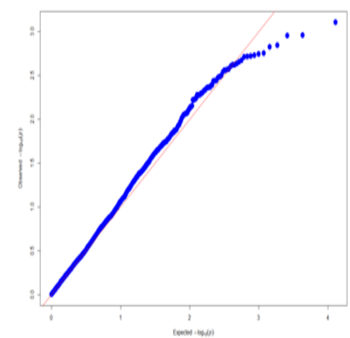
1



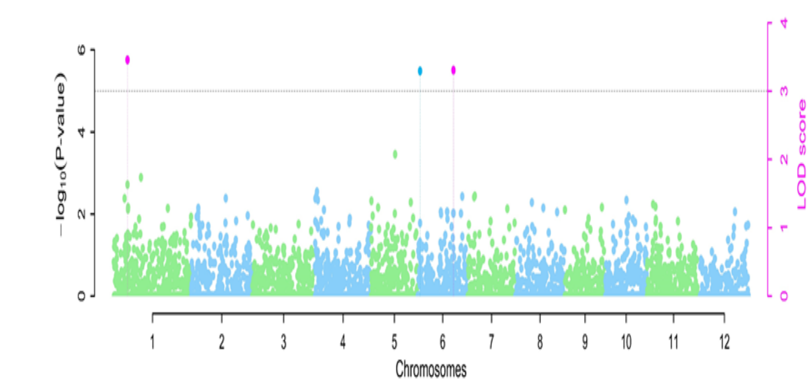



2

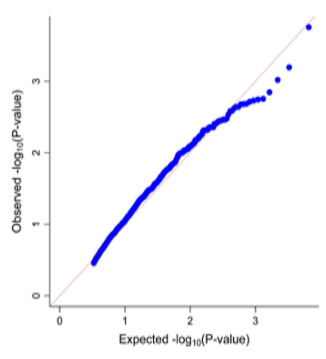


3




4



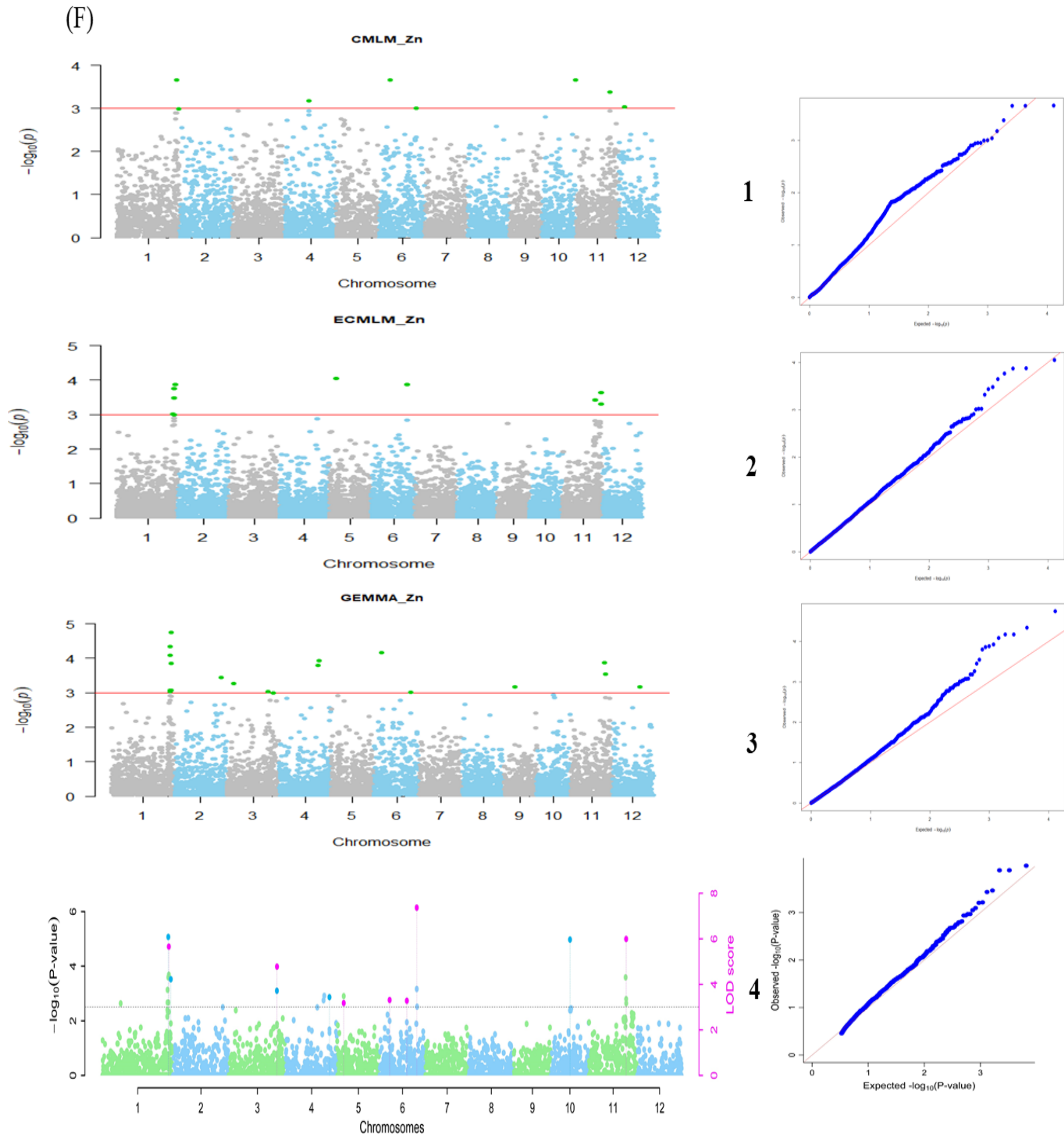



Figure 2.6. Manhattan plots of GWAS for six minerals. (1). Manhattan plot CMLM model. (2). Manhattan plot for ECMLM model. (3). Manhattan plot for GEMMA model. (4). Manhattan plot for multi-locus models, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, ISIS EM-BLASSO. The red horizontal line in the 1, 2, and 3 models is the threshold significant level used in the study to declare a SNP as being significant for measured traits. The green circles above the red line depict the significant SNPs. For 4 model, purple circles above the dashed horizontal line are the significant SNPs identified by all six multi-locus models, whereas green circles show only those SNPs identified by any two models of six models of multi-locus method. (A) Cu. (B) Fe, (C) K. (D) Mg. (E) Mn. (F) Zn.

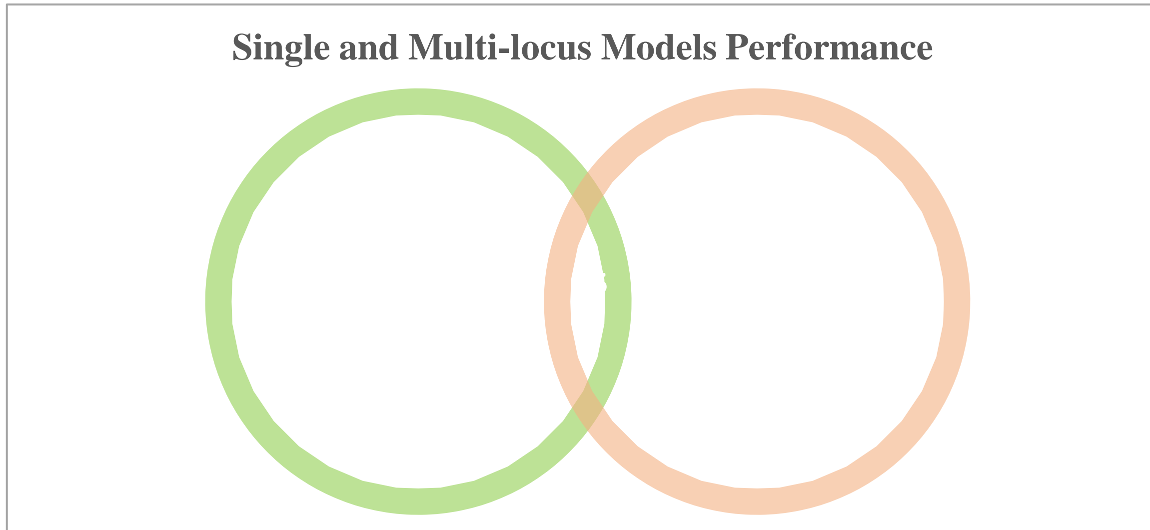



Figure 2.7. Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for the six mineral elements.

locus models found 32 SNPs for Fe minerals explaining 8.38-17.88% of the phenotypic variation, whereas only multi-locus models detected two SNPs with explaining 3.56-5.33% of the phenotypic variation. Both models identified one SNPs for Fe, thus explaining 0.00-26.83% of the phenotypic variations. For K, 13, 7 and 3 SNPs were identified by only single-locus, only multi-locus and both models together, explaining 6.57-23.80%, 3.85-34.37% and 5.86-14.41% of the phenotypic variations, respectively. For Mg, 35 SNPs were found by only single-locus models, explaining 6.71-13% of the phenotypic variation. Only two SNPs were detected by only multi-locus and this explained 8.79-16.4% of the phenotypic variation. Five SNPs were identified by both models that explained 3.62-24.99% of the phenotypic variation. Both only single and multi-locus models detected three SNPs separately, in total six SNPs, were found for Mn that explained 13.66-14.16% and 6.54-10.92% of the phenotypic variations, respectively. For Zn, single-locus, multi-locus and both models identified 11, 7 and 4, in total 22 SNPs, explaining 8.39-49.9%, 5.32-24.44% and 10.19-50.06% of the phenotypic variations (Figure 2.6; Table 2.3).

Among the 147 significant SNPs, thirty-two SNPs appear to control more than one trait, suggesting a pleiotropic effect. Among the six SNPs that affect Cu, two SNPs (SNP-6.2196821., 6285634) are associated with Fe, three SNPs (907175, SNP-6.2196821., 6285634) with Mg and one SNP (4572241) is also found for Zn. Similarly, 15 SNPs have an effect on both Fe and Mg and one SNP on both Fe and K. In addition, seven and one SNPs are associated with controlling both K and Mg and K and Zn, respectively. Two SNPs influence both Mg and Zn elements. Moreover, SNP-6.2196821 and 6285634 SNPs are involved in affecting Cu, Fe and Mg, whereas the 1202195 SNP has an effect on the Fe, K and Mg elements (Table 2.3; Figure 2.8)

In terms of SNP detection ability of the different models used in this study across the six mineral elements, the single-locus models outperformed the multi-locus models. Single-locus models identified 148 SNPs (CMLM= 75, ECMLM= 21 and GEMMA= 52) including SNPs identified with multi-locus models, whereas multi-locus models were able to detect 71 SNPs (MrMLM=18, FASTMrMLM= 17, FASTmrEMMA= 3, PKWmEB= 11, PLARmEB= 10 and ISIS-EB-BL= 12) including SNPs found with single-locus models. Moreover, the R2 (proportion of phenotypic variance explained by the QTL) of the co-identified SNPs identified by different models are different, likely due to small differences in the underlying calculations in each model (Figure 2.7).

### 2.3.4. In-Silico Gene Expression Analysis

After mining the genes within 250-kb region of the significant SNPs found for all mineral elements using the RAP-DB database (https://rapdb.dna.affrc.go.jp/), we found 371, 732, 506, 853, 140 and 527 genes for Cu, Fe, K, Mg, Mn, and Zn, respectively. To investigate which genes are

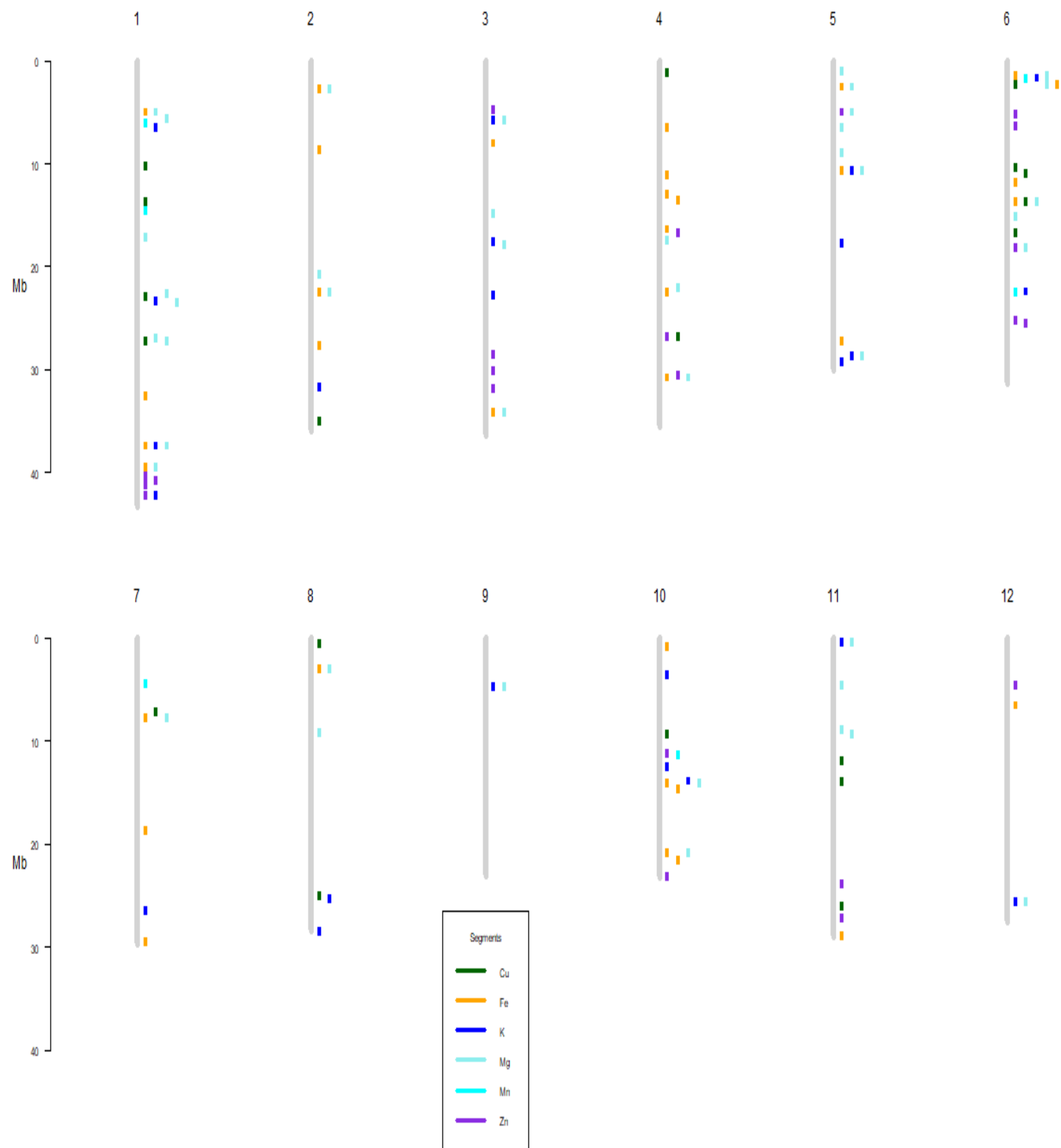



Figure 2.8. All the significant SNPs found in the study for the six mineral elements. SNPs positions are depicted by the rectangular box. Specific color shows the corresponding mineral elements. Rice chromosomes are displayed by vertical lines.

more likely to be responsible for mineral traits, we selected only those genes that were expressed both in the reproductive and vegetative stages of rice plants by using Nipponbare gene expression data in normalized FPKM values. After filtering the non-expressed genes in reproductive stages, 117, 101, 194, 193, 41 and 146 genes were found for Cu, Fe, K, Mg, Mn, and Zn, respectively, and will be used for further analysis (Figure 2.9).

## 2.4. Discussion

### 2.4.1. Population Structure, LD, And Phenotypic Variation

The rice germplasm used in this study has six sub-populations based on the Structure analysis, including two major sub-populations, *indica* and *japonica*, which is consistent with the previous studies using worldwide rice germplasm (Morales et al., 2020; Xu et al., 2016). The LD decay distance of this study was 250 kb for the whole panel, which is similar to the previous findings using different sub-populations with LD ranging from 100 kb to over 240 kb for cultivated rice (Qiu et al., 2015). Mather et al. (2007) found LD decay from >500 kb in Oryza sativa ssp. japonica, to ~75 kb in *O. sativa* subsp. *indica*, and down to merely ~40 kb or lower in O. rufipogon for different rice sub-populations. In our study sample, there was no *O. rufipogon* sub-populations. So, it can be said that LD decay of this study is well enough to conduct the association studies.

The phenotypic variation for all the mineral traits used in this study was abundant, suggesting that GWAS can be applied to this rice germplasm. Positive correlations with moderate levels were observed among the six mineral elements except between Mn and Zn, indicating that these minerals might share a common molecular pathway. This could be due to the pleiotropic effects of causal genes controlling these minerals in rice, which is supported by our

GWAS findings: 32 pleiotropic SNPs were found in our study, explaining the causal reasons for the correlation observed among the minerals. While no pleiotropic SNPs were found between Mn and Zn, a positive correlation with a low level was found in correlation studies, indicating SNPs with minor effects still might exist that our GWAS could not capture.

### 2.4.2. Performance Of Single And Multi-Locus GWAS Models

In terms of SNP detection ability, based on our GWAS studies, single-locus models found more significant SNPs than multi-locus models, when including SNPs shared by both models. Single-locus models detected 148 SNPs for the six mineral traits, where CMLM found 75 SNPs, followed by GEMMA (52) and ECMLM (21). Multi-locus models identified 71 SNPs across thesix traits. But the model's performance in our studies is opposite to the reports published previously, where the multi-locus approach was more powerful than the single-locus approach using both real and simulation datasets (C. Li et al., 2018; Y. Xu et al., 2018; Y. M. Zhang, Jia, & Dunwell, 2019). The main reason for this seemingly contradictory result is that multiple test corrections were not applied to the single-locus models that could have reduced the number of significant SNPs. Multiple test correction is not required for multi-locus methods, which is the obvious advantage, so the number of SNPs detected by these methods will be similar. Another possible explanation is that multi-locus approaches are more robust against loci with small effects that explain less than six percent of phenotypic variance (Y. Xu et al., 2018). It could be possible that most of the QTLs affecting mineral traits have a large effect that could hinder the performance of multi-locus models for mineral traits. Similar results were also found by Liu et al. (2020) where more SNPs were detected for the mineral content of rice by univariate models (GLM and MLM) than multivariate models (mrMLM and FarmCPU)

Table 2.3. List of loci detected in the study.

| Trait | SNP | Alleles | Chr. | Pos.(bp) | Single-locus GWAS | | | Multi-locus GWAS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cu | | | | | -log10P | R2 | Model | LOD | R2 | Model |
| | 304900 | G/A | 1 | 10100605 | 4.07-4.55 | 10.4-17.6 | C, EC, G | 4.54 | 9.4 | M, FM, PK, PL, I |
| | SNP-1.13478728. | G/A | 1 | 13479755 | 3.55 | | G | | | |
| | 760644 | A/C | 1 | 22723681 | 3.57 | | G | | | |
| | 907175 | G/A | 1 | 27047209 | 3.67 | | G | | | |
| | id2015767 | A/G | 2 | 34868096 | 3.5 | | G | | | |
| | id4000574 | C/A | 4 | 972749 | 3.22 | 9.1 | C | | | |
| | 4572241 | G/A | 4 | 26688183 | 3.02-3.23 | 7.2-13.6 | C, EC | | | |
| | SNP-6.2196821. | A/T | 6 | 2197821 | 3.01 | | G | | | |
| | 6147112 | G/A | 6 | 10199497 | 3.01 | | G | | | |
| | SNP-6_10761128 | A/G | 6 | 10762128 | 3.47 | | G | | | |
| | 6285634 | G/A | 6 | 13487635 | 3.21 | 9.1 | C | | | |
| | 6427131 | A/G | 6 | 16508748 | 3.4 | | G | | | |
| | id7001155 | G/A | 7 | 6987625 | 3.37-3.91 | 11.1-15.9 | C, EC, G | 3.3 | 42.0 | M |
| | 7993541 | C/A | 8 | 416250 | 3.51 | | G | | | |
| | id8006885 | G/A | 8 | 24753844 | 3.21 | 9.1 | C | | | |
| | SNP-10.9068762. | G/A | 10 | 9139902 | 3.17 | 9.0 | C | | | |
| | 11233430 | G/A | 11 | 11715177 | 3.06 | 13.7 | EC | | | |
| | SNP-11.13313880. | G/A | 11 | 13777560 | 3.35 | 14.5 | EC | | | |
| | SNP-11.25392640. | A/C | 11 | 25858722 | | | | 3.25 | 8.7 | I, PK |
| Fe | 153297 | G/A | 1 | 4823701 | 3.77 | 10.55 | C | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-1.32376151. | G/A | 1 | 32377196 | 3.48 | 9.7 | C | | | |
| | 1202195 | A/C | 1 | 37230818 | 3.16-3.21 | 8.8 | C, G | | | |
| | 1257104 | A/G | 1 | 39282883 | 3.24 | | G | | | |
| | SNP-2.2575985. | G/A | 2 | 2575988 | 3.17 | 8.8 | C | | | |
| | SNP-2.8455563. | G/C | 2 | 8455565 | 3.38 | | G | | | |
| | 2087054 | A/G | 2 | 22324900 | 3.15 | 8.8 | C | | | |
| | 2267750 | A/C | 2 | 27548893 | 3.25 | 9.1 | C | | | |
| | id3004190 | A/G | 3 | 7849199 | 3.02 | 8.4 | C | | | |
| | 3501392 | G/A | 3 | 33987612 | 3.45 | | G | | | |
| | SNP-4.6317262. | G/A | 4 | 6321823 | 3.28 | | G | | | |
| | SNP-4.10930754. | A/G | 4 | 10940054 | 3.08 | 8.5 | C | | | |
| | 4128471 | C/A | 4 | 12879859 | 3.1 | 8.6 | C | | | |
| | SNP-4.13348501. | C/A | 4 | 13357791 | 3.28 | 9.1 | C | | | |
| | 4241771 | G/A | 4 | 16199670 | 3.54 | 9.9 | C | | | |
| | SNP-4.22128339. | G/A | 4 | 22313458 | 3.66 | 10.2 | C | | | |
| | 4678550 | G/A | 4 | 30601123 | 4.28 | 12.1 | C | | | |
| | 4882140 | A/C | 5 | 2378143 | 3.23 | 9.0 | C | | | |
| | 5196119 | A/G | 5 | 10528231 | 3.36-3.52 | 7.4-18.4 | C, EC, G | 3.99 | 0.00-26.8 | FM, PK. PL |
| | 5735083 | A/G | 5 | 27094485 | 3.17-3.47 | 7.1-17.9 | C, EC, G | | | |
| | SNP-6.1343132. | A/G | 6 | 1344132 | 3.38 | 9.4 | C | | | |
| | SNP-6.2196821. | A/T | 6 | 2197821 | 3.14 | 8.7 | C | | | |
| | id6007260 | A/G | 6 | 11618178 | 3.13 | 8.7 | C | | | |
| | 6285634 | G/A | 6 | 13487635 | 3.86 | 10.8 | C | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | 7179219 | A/G | 7 | 7619494 | 3.37 | 9.4 | C | | | |
| | 7643802 | G/A | 7 | 18573822 | | | | 4.34 | 5.3 | FM, I |
| | SNP-7.29385457. | A/G | 7 | 29386450 | 3.02 | 8.4 | C | | | |
| | 8067129 | G/A | 8 | 2887584 | 3.25-3.60 | 10.1 | C, G | | | |
| | 9921984 | A/G | 10 | 650031 | | | | 3.53 | 3.6 | FM |
| | SNP-10.13843768. | T/A | 10 | 13915001 | 3.31 | 9.2 | C | | | |
| | 10555828 | G/A | 10 | 14438582 | 3.09 | 8.6 | C | | | |
| | SNP-10.20587837. | G/A | 10 | 20659359 | 3.13 | 8.7 | C | | | |
| | 10778744 | A/G | 10 | 21397933 | 3.16 | 17.9 | EC | | | |
| | SNP-11.28200021. | C/G | 11 | 28723243 | 3.29 | 9.1 | C | | | |
| | SNP-12.6356528. | C/A | 12 | 6357639 | 3.73 | 10.4 | C | | | |
| K | SNP-1.6382810. | G/A | 1 | 6383811 | 3.06 | 8.6 | C | 3.22 | 14.4 | M |
| | SNP-1.23170758. | G/A | 1 | 23171803 | | | | 3.7 | 34.4 | PK |
| | 1202195 | A/C | 1 | 37230818 | 3.37-3.75 | 23.8 | EC, G | | | |
| | SNP-1.41998191. | T/A | 1 | 41999235 | | | | 5.05 | 7.5 | FM, I, M |
| | 2375486 | G/A | 2 | 31547627 | | | | 3.01 | 10.6 | M |
| | SNP-3.5666296. | G/A | 3 | 5667297 | 3.1 | 8.7 | C | | | |
| | 2964807 | G/A | 3 | 17453008 | 3.03 | | G | | | |
| | 3173191 | A/G | 3 | 22657915 | 3.31 | | G | | | |
| | id5004837 | G/A | 5 | 10539124 | 3 | | C | | | |
| | 5452087 | A/G | 5 | 17598723 | | | | 3.67 | 7.1 | FM, M |
| | SNP-5.28500625. | C/G | 5 | 28563271 | 4.32 | | G | 4.4 | 12.0 | FM, M |
| | 5787299 | A/G | 5 | 29150826 | 3.12-3.30 | 22.8 | EC, G | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-6_1500959 | A/G | 6 | 1501961 | | | | 3.44 | 7.7 | PK |
| | 6674186 | C/A | 6 | 22249886 | 3.06 | | G | | | |
| | 7892688 | G/A | 7 | 26282546 | | | | 4.51 | 8.1 | FM, M |
| | 8966923 | C/A | 8 | 25213697 | | | | 3.3 | 3.9 | I, M |
| | id8007977 | A/G | 8 | 28377609 | 3.31 | | G | 3.19 | 5.9 | FM |
| | c9p4565514 | C/A | 9 | 4565515 | 3.15 | | G | | | |
| | 10063204 | A/G | 10 | 3400212 | 3.1 | | G | | | |
| | 10480545 | G/A | 10 | 12310115 | 3.15 | 7.0 | C | | | |
| | id10003608 | G/A | 10 | 13711367 | 3.09 | 8.7 | C | | | |
| | SNP-11.235195. | G/A | 11 | 236194 | 3.04 | 8.6 | C | | | |
| | 13022382 | A/C | 12 | 25490919 | 3 | 6.6 | C | | | |
| Mg | 153297 | G/A | 1 | 4823701 | 3.19 | 8.8 | C | | | |
| | 170435 | G/A | 1 | 5408523 | 3.34 | 9.3 | C | | | |
| | 572891 | G/A | 1 | 17008280 | 3.11 | 8.6 | C | | | |
| | id1012746 | A/G | 1 | 22494508 | 3.16 | | G | | | |
| | SNP-1.23342685. | G/A | 1 | 23343730 | 2.98 | 8.2 | C | | | |
| | 899561 | A/G | 1 | 26762494 | 3.1 | 6.7 | C | | | |
| | 907175 | G/A | 1 | 27047209 | 3.16 | | G | | | |
| | 1202195 | A/C | 1 | 37230818 | 3.02-3.29 | 8.9-25.5 | C, EC, G | | | |
| | 1257104 | A/G | 1 | 39282883 | 3.98 | | G | 3.42 | 3.6 | I, PL |
| | SNP-2.2575985. | G/A | 2 | 2575988 | 3.05 | 8.4 | C | | | |
| | 2031305 | G/A | 2 | 20616529 | 4.13 | | G | | | |
| | 2087054 | A/G | 2 | 22324900 | 3.03 | 8.4 | C | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-3.5666296. | G/A | 3 | 5667297 | 3.63 | 10.1 | C | | | |
| | 2853978 | G/A | 3 | 14652096 | | | | 3.17 | 16.4 | M |
| | 2972375 | G/A | 3 | 17674269 | 3.08 | 8.5 | C | | | |
| | 3501392 | G/A | 3 | 33987612 | 3.71 | | G | | | |
| | 4288833 | A/G | 4 | 17316219 | 3.39 | | G | 3.8 | 7.1 | I, PL |
| | 4448877 | G/A | 4 | 21864875 | 4.6 | 13.0 | C | | | |
| | 4678550 | G/A | 4 | 30601123 | 3.36 | 9.3 | C | | | |
| | 4833352 | A/G | 5 | 922530 | 3.52-3.85 | 9.8-26.8 | C, EC, G | 7.43 | 13.3 | FM, I, M, PL |
| | 4882140 | A/C | 5 | 2378143 | 3.05 | 8.4 | C | | | |
| | id5002528 | C/A | 5 | 4819475 | 3.35 | 9.3 | C | | | |
| | 5011602 | G/A | 5 | 6374926 | 3.21 | 8.9 | C | | | |
| | 5121882 | A/G | 5 | 8842405 | | | | 4.02 | 8.8 | FM, M, PL |
| | id5004837 | G/A | 5 | 10539124 | 3.32 | 7.3 | C | | | |
| | SNP-5.28500625. | C/G | 5 | 28563271 | 3.93 | | G | | | |
| | SNP-6.1343132. | A/G | 6 | 1344132 | 3.09 | 8.6 | C | | | |
| | SNP-6.2196821. | A/T | 6 | 2197821 | 3.31 | 9.2 | C | | | |
| | 6285634 | G/A | 6 | 13487635 | 3.69 | 10.3 | C | | | |
| | 6351040 | G/A | 6 | 14937819 | 3.45 | 9.6 | C | | | |
| | 6496457 | C/A | 6 | 17969922 | 3.5 | | G | 4.8 | 8.3 | FM, M |
| | 7179219 | A/G | 7 | 7619494 | 3.53 | 9.8 | C | | | |
| | 8067129 | G/A | 8 | 2887584 | 3.62 | 10.1 | C | | | |
| | 8322255 | G/A | 8 | 9019202 | 3.28 | 9.1 | C | | | |
| | c9p4565514 | C/A | 9 | 4565515 | 3.27 | | G | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-10.13843768. | T/A | 10 | 13915001 | 4.1 | 11.5 | C | | | |
| | SNP-10.20587837. | G/A | 10 | 20659359 | 3.21 | 8.9 | C | | | |
| | SNP-11.235195. | G/A | 11 | 236194 | 3.04 | 8.4 | C | | | |
| | 10943015 | A/G | 11 | 4419880 | 3.06-3.99 | 25.0 | EC, G | 4.04 | 7.65 | I, PL |
| | 11112426 | G/A | 11 | 8788201 | 3.12 | | G | | | |
| | 11130199 | G/A | 11 | 9217367 | 3.12 | | G | | | |
| | 13022382 | A/C | 12 | 25490919 | 3.32 | 7.3 | C | | | |
| Mn | SNP-1.5867020. | A/G | 1 | 5868021 | | | | 3.46 | 10.9 | FE, I, PK |
| | SNP-1.14460354. | G/A | 1 | 14461381 | 3.25 | 14.2 | EC | | | |
| | 5868825 | A/C | 6 | 1521855 | | | | 4.01 | 6.5 | PK |
| | SNP-6.22337184. | G/A | 6 | 22338182 | | | | 3.31 | 7.9 | FM, M |
| | 7066952 | G/A | 7 | 4232489 | 3.1 | | G | | | |
| | id10002943 | C/A | 10 | 11195773 | 3.05 | 13.7 | EC | | | |
| Zn | 1280193 | A/G | 1 | 40154802 | 3.76-4.34 | 49.9 | EC, G | | | |
| | SNP-1.40596823. | T/A | 1 | 40597867 | 3.48-3.86 | | EC, G | 6.08 | 23.9 | FE |
| | 1305247 | G/A | 1 | 41042727 | 3.66-4.74 | 10.2-50.1 | C, EC, G | 6.73 | 11.0 | I, PK, PL |
| | SNP-1.41998191. | T/A | 1 | 41999235 | | | | 4.22 | 6.0 | PL |
| | SNP-3.4621271. | G/A | 3 | 4622270 | 3.26 | | G | | | |
| | SNP-3.28426789. | G/A | 3 | 28433737 | 3.04 | | G | | | |
| | 3405830 | G/A | 3 | 29987079 | | | | 4.78 | 24.4 | FM, M |
| | rd3001044 | A/G | 3 | 31627459 | 3 | | G | | | |
| | id4004654 | G/A | 4 | 16559384 | 3.17 | 8.8 | C | | | |
| | 4572241 | G/A | 4 | 26688183 | 3.92 | | G | | | |

| Table 2.3. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | id4010220 | A/G | 4 | 30330971 | | | | 3.44 | 7.4 | I |
| | id5002528 | C/A | 5 | 4819475 | 4.05 | 50.3 | EC | 3.18 | 13.0 | FM, M |
| | rd6001756 | G/A | 6 | 5007776 | 4.17 | | G | | | |
| | SNP-6.6241072. | A/G | 6 | 6242072 | | | | 3.31 | 11.4 | FM, M |
| | 6496457 | C/A | 6 | 17969922 | | | | 3.47 | 5.3 | PK, PL |
| | SNP-6.25063527. | G/A | 6 | 25064525 | 3.01 | | G | 7.36 | 14.4 | FM, M |
| | 6783797 | G/A | 6 | 25387111 | 3.00-3.88 | 6.4-50.1 | C, EC | | | |
| | 10430775 | A/G | 10 | 11051662 | | | | 5.96 | 12.6 | PK |
| | id10007301 | A/G | 10 | 23033344 | 3.66 | 10.2 | C | | | |
| | 11769276 | G/A | 11 | 23660957 | | | | 8.13 | 14.0 | FE, PK |
| | 11915122 | G/A | 11 | 27015384 | 3.65 | 49.7 | EC | | | |
| | 12134336 | G/A | 12 | 4433511 | 3.04 | 8.4 | C | | | |

N.B: Single-locus models: C-CMLM, EC-ECMLM, G- GEMMA; Multi-locus models: M- MrMLM, FM- FASTmrMLM, FE- FASTmrEMMA, PK- pKWmEB, PL- pLARmEB, I- ISIS EM-BLASSO

As our main goal of this study is to find as many related SNPs as possible without losing any potential causal SNPs, but with reliable, genuine SNPs, we applied both models, which were also recommended by others (Liu et al., 2020). This will lead to the validation of the significance of the underlying target region. Thus, combining single-locus and multi-locus GWAS could improve the power and robustness of association analyses.

### 2.4.3. Comparison And Reliability Of Our GWAS Studies

We compared our detected SNPs for the six minerals with the genes/QTLs and markers related to mineral content identified from linkage mapping and association mapping in previous studies. To compare with the QTLs/SSR and RFLP markers of the previous studies, the surrounding 250 kb of our associated SNPs were regarded as potentially the same loci for any particular trait when this region was found between the borders of QTLs/SSR and RFLP markers. In the case of SNP marker comparisons, the markers of past studies that were located within the 250 kb region of our significant associated SNPs were considered as the same loci for the particular trait. Thus, with these parameters, 43 (~ 29%) of the 147 of the significant SNPs of this study coincided with the previously reported genes/QTLs and/or markers for the six minerals and the remaining 104 (70%) SNPs were considered as novel. Out of the shared 43 SNPs, 17 were found for Fe, followed by eight for Mg, seven for K, five for both Cu and Zn and the remaining one was found for Mn (Table 2.4).

The molecular mechanisms of uptake, transport and accumulation of the mineral elements used in this study are well established (Norton et al., 2014). So far, several genes and gene families have been found as being involved in the acquisition and transport of copper and zinc in

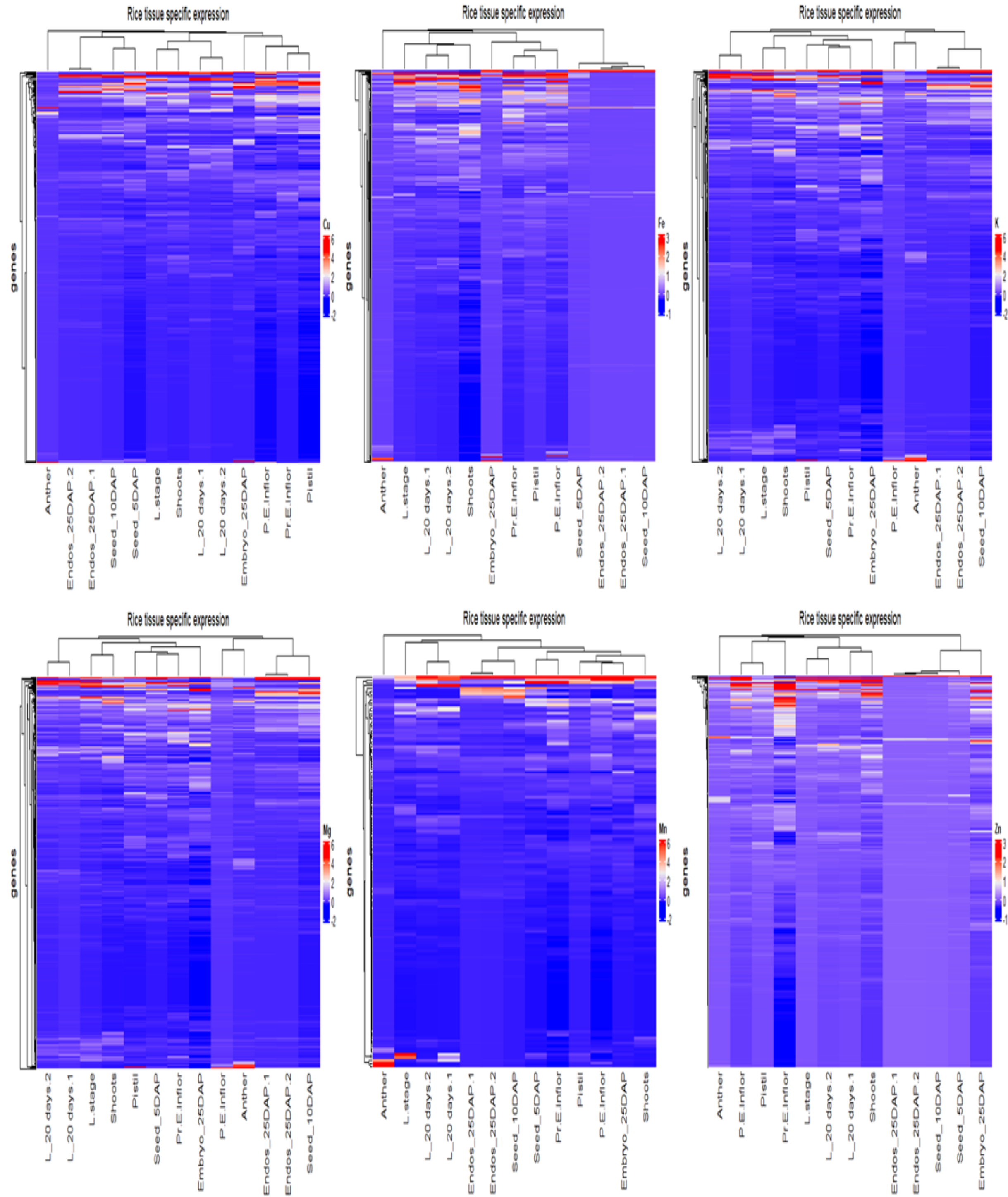



Figure 2.9. Heatmap of In-silico gene expression analysis results for the six mineral elements of the study.

rice seeds. These include, but are not limited to, the ZIP (Zinc-regulated transporter (ZRT)) gene families, Iron-regulated transporter (IRT-like protein) gene family, YSL (yellow stripe-like) protein, MTPs (metal tolerance proteins), COPT (COPper Transporter) family, and NRAMPs gene family. Some of them are also involved in the pathway of uptake, transport, and accumulation of iron, magnesium and cadmium (White & Broadley, 2009). ZIP3 of the six ZIP genes (ZIP1, 3, 4, 5, 7a and 8) was identified within a 250 kb region at id4010220 SNP in chromosome 4 associated with Zn in our GWAS study. Two SSRs and five SNPs were also reported in the same chromosomal position previously by Zhang et al. (2014), Norton et al. (2014) and Bollinedi et al. (2020). Similarly, the NAS gene family is involved in the accumulation of Fe, Zn and Cu in rice endosperm (Lee et al., 2009). The current study found OsNAS3 at SNP-7.29385457. in chromosome 7 associated with Fe, where two SSR markers were also reported before (Nawaz et al., 2015; M. Zhang et al., 2014). The COPT transporter gene family for Cu was not identified by our GWAS analysis. Interestingly, our study identified OsIRO2, an iron-related bHLH transcription factor 2 that regulates Fe uptake from soil, transport during germination and translocation to the grain, at 1305247 SNP in chromosome 1 associated with Zn, where Bollinedi et al. (2020) also found a SNP almost at the same position, supporting the fact that a single gene may control the molecular mechanism of multiple elements simultaneously (Table 2.4).

Since 43 QTLs with three known genes were rediscovered by this study, we can confirm the accuracy of our GWAS studies. More importantly, these 43 QTLs regulating six mineral elements, simultaneously detected in various populations with different genetic backgrounds, eventually can be further validated and used to conduct marker assisted selection in future biofortification programs.

Table 2.4. Comparison of the GWAS result with the previous studies.

| Trait | SNP | Position (bp) | Chr. | Markers linked/ associated | Position (bp) | Types of Markers | Known genes | References |
|---|---|---|---|---|---|---|---|---|
| Cu | SNP-1.13478728. | 13479755 | 1 | ud1000606 | 13429671 | SNP | | Norton et al. (2014) |
| | 6147112 | 10199497 | 6 | id6006288 | 10090472 | SNP | | Norton et al. (2014) |
| | id7001155 | 6987625 | 7 | RM214 | 5Mb - 20Mb | SSR | | Zhang et al. (2014) |
| | id8006885 | 24881549 | 8 | RM3155, id8007452 | 23Mb – 28Mb, 27210520 | SSR, SNP | | Zhang et al. (2014), Norton et al. (2014) |
| | SNP-11.25392640. | 25858722 | 11 | id11010366, id11010372, id11010373 | 25850373, 25851231, 25851251 | SNP | | Norton et al. (2014) |
| Fe | 153297 | 4823701 | 1 | RM283 | 4886944 - 4886983 | SSR | | Nawaz et al. (2015) |
| | SNP-1.32376151. | 32377196 | 1 | RM5 | 23Mb - 35Mb | SSR | | Zhang et al. (2014) |
| | SNP-2.8455563. | 8455565 | 2 | RG437, RM452, RM145/Os02ssr0079000 | 7Mb – 1Mb0, 7706972 - 7707033 | RFLP, SSR, SSR | | Zhang et al. (2014), Nawaz et al. (2015) |
| | 2267750 | 27548893 | 2 | RM6933 | 24Mb - 31Mb | SSR | | Zhang et al. (2014) |
| | 3501392 | 33987612 | 3 | RM514, AX-95935621, AX-95950999, AX-95935460, AX-95924055, AX-95923317, AX-95923159 | 29Mb – 36Mb, 32326592, 32335075, 32374286, 32380341, 32380432, 32380964 | SSR, SNP | | Zhang et al. (2014), Bollinedi et al. (2020), |
| | 4241771 | 16199670 | 4 | RM3317, RM471 | 5Mb – 20Mb, 18996850 - 18996873 | SSR, SSR | | Zhang et al. (2014), Nawaz et al. (2015) |
| | SNP-4.6317262. | 6321823 | 4 | RM3317 | 5Mb - 20Mb | SSR | | Zhang et al. (2014) |
| | SNP-4.10930754. | 10940054 | 4 | RM3317 | 5Mb - 20Mb | SSR | | Zhang et al. (2014) |
| | 4128471 | 12879859 | 4 | RM3317 | 5Mb - 20Mb | SSR | | Zhang et al. (2014) |

| Table 2.4. Continued | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Position (bp)** | **Chr.** | **Markers linked/ associated** | **Position (bp)** | **Types of Markers** | **Known genes** | **References** |
| | SNP-4.13348501. | 13357791 | 4 | RM3317 | 5Mb - 20Mb | SSR | | Zhang et al. (2014) |
| | 4882140 | 2378143 | 5 | RM13 | 0 - 5500000 | SSR | | Zhang et al. (2014) |
| | SNP-6.1343132. | 1344132 | 6 | RM190, RM190 | 0 – 4Mb, 1765669 - 1765704 | SSR, SSR | | Zhang et al. (2014), Nawaz et al.(2015) |
| | SNP-6.2196821. | 2197821 | 6 | RM190 | 0 - 4Mb | SSR | | Zhang et al. (2014) |
| | SNP-7.29385457. | 29386450 | 7 | RM248, RM1335 | 26Mb – 30Mb, 28299722 - 28299763 | SSR, SSR | | Zhang et al. (2014), Nawaz et al.(2015) |
| | 9921984 | 650031 | 10 | RM222 | 0 - 12Mb | SSR | | Zhang et al. (2014) |
| | SNP-10.20587837. | 20659359 | 10 | RM1108 | 18Mb - 21Mb | SSR | | Zhang et al. (2014) |
| | 10778744 | 21397933 | 10 | RM1108 | 18Mb - 21Mb | SSR | | Zhang et al. (2014) |
| K | SNP-1.6382810. | 6383811 | 1 | RG532a, RM1 | 3Mb - 7Mb | RFLP, SSR | | Zhang et al. (2014) |
| | SNP-1.23170758. | 23171803 | 1 | CDO455, RM5501, RM5 | 23Mb – 35Mb, 23972466 - 23972495 | RFLP, SSR | | Zhang et al. (2014), Garcia-Oliveira et al (2009) |
| | 2375486 | 31547627 | 2 | RM3732-RM492 | 7290000 - 44080000 | SSR | | Du et al. (2013) |
| | 2964807 | 17453008 | 3 | RM282-RM6266/Os03ssr0099800-Os03ssr0183400 | 12408722 - 23823856 | SSR | | Du et al. (2013) |
| | 3173191 | 22657915 | 3 | RM282-RM6266/Os03ssr0099800-Os03ssr0183400 | 12408722 - 23823856 | SSR | | Du et al. (2013) |
| | SNP-5.28500625. | 28563271 | 5 | RM188 | 21Mb - 27Mb | SSR | | Zhang et al. (2014) |
| | 7892688 | 26282546 | 7 | RM248, RM505/RM21926 | 26Mb – 30Mb, 24527811 - 24527834 | SSR | | Zhang et al. (2014), Nawaz et al.(2015) |
| Mg | 4882140 | 2378143 | 5 | RM13 | 0 - 5500000 | SSR | | Zhang et al. (2014) |
| | id5002528 | 4819475 | 5 | RM13 | 0 - 5500000 | SSR | | Zhang et al. (2014) |

| Table 2.4. Continued | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Position (bp)** | **Chr.** | **Markers linked/ associated** | **Position (bp)** | **Types of Markers** | **Known genes** | **References** |
| | 6285634 | 13487635 | 6 | RM527-RM3 | 9862309 - 28130383 | SSR | | Du et al. (2013) |
| | 6351040 | 14937819 | 6 | RM527-RM3 | 9862309 - 28130383 | SSR | | Du et al. (2013) |
| | 6496457 | 17969922 | 6 | RM527-RM3 | 9862309 - 28130383 | SSR | | Du et al. (2013) |
| | SNP-10.20587837. | 20659359 | 10 | RM5494 | 18Mb - 21Mb | SSR | | Zhang et al. (2014) |
| | SNP-11.235195. | 236194 | 10 | RM5494 | 18Mb - 21Mb | SSR | | Zhang et al. (2014) |
| | 10943015 | 4419880 | 11 | RZ781, RM332 | 0 - 5Mb | RFLP, SSR | | Zhang et al. (2014) |
| Mn | SNP-6.22337184. | 22338182 | 6 | RM527-RM3 | 9862309 - 28130383 | SSR | | Du et al. (2013) |
| Zn | 1305247 | 41042727 | 1 | AX-95918225 | 41121295 | SNP | OsIRO2 | Bollinedi et al. (2020) |
| | SNP-3.28426789. | 28433737 | 3 | id3013232 | 28309539 | SNP | | Norton et al. (2014) |
| | id4010220 | 30330971 | 4 | RM317, id4010984, wd4003179, id4011016, id4011022, AX-95951158, RM317/Os04ssr0174300 | 28Mb – 33Mb, 31853012, 31953475, 31953961, 31957137, 32811874, 29246223 - 29246242 | SSR, SNP | ZIP3 | Zhang et al. (2014), Norton et al. (2014), Bollinedi et al. (2020), Huang et al.(2015) |
| | 10430775 | 11051662 | 10 | RM222, id10003681, AX-95932094 | 0 – 12Mb, 13867606, 12685215 | SSR, SNP | | Zhang et al. (2014), Norton et al. (2014), Bollinedi et al. (2020) |

## 2.5. Conclusion

This study reported the GWAS of six mineral elements using 174 rice accessions across the world using 7k SNP array genotype data. A total of 147 SNPs affecting mineral elements were identified by single-locus and multi-locus methods. Of these SNPs, 110 SNPs were detected by only single-locus methods, whereas the multi-locus methods detected 22 SNPs, and 15 SNPs were co-detected by both methods. While 43 SNPs were matched with the previously reported genes/QTLs and markers, 104 SNPs were novel. After mining genes within a 250 kb region of these SNPs, a total of 3,129 genes were found. Among these genes, 792 genes could be identified as candidate genes for controlling accumulation in rice grain. These shortlisted genes could be used for future studies to further investigate the gene expression levels, followed by functional gene characterization, to better understand the complex molecular mechanisms controlling grain concentration of these six mineral elements.

## 2.6. References


Bollinedi, H., Yadav, A. K., Vinod, K. K., Gopala Krishnan, S., Bhowmick, P. K., Nagarajan, M., . . . Singh, A. K. (2020). Genome-Wide Association Study Reveals Novel Marker-Trait Associations (MTAs) Governing the Localization of Fe and Zn in the Rice Grain. *Front Genet, 11*, 213. doi:10.3389/fgene.2020.00213

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308

Cui, Y., Zhang, F., & Zhou, Y. (2018). The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice. *Front Plant Sci, 9*, 1464. doi:10.3389/fpls.2018.01464

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology, 14*(8), 2611-2620.

Graham, R. D., Welch, R. M., Saunders, D. A., Ortiz-Monasterio, I., Bouis, H. E., Bonierbale, M., . . . Liria, R. (2007). Nutritious subsistence food systems. *Advances in agronomy, 92*, 1-74.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., . . . Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet, 42*(11), 961-967. doi:10.1038/ng.695

Huang, Y., Sun, C., Min, J., Chen, Y., Tong, C., & Bao, J. (2015). Association Mapping of Quantitative Trait Loci for Mineral Element Contents in Whole Grain Rice (Oryza sativa L.). *J Agric Food Chem, 63*(50), 10885-10892. doi:10.1021/acs.jafc.5b04932

Li, C., Fu, Y., Sun, R., Wang, Y., & Wang, Q. (2018). Single-Locus and Multi-Locus Genome-Wide Association Studies in the Genetic Dissection of Fiber Quality Traits in Upland Cotton (Gossypium hirsutum L.). *Front Plant Sci, 9*, 1083. doi:10.3389/fpls.2018.01083

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology, 34*(8), 816-834.

Liu, S., Zhong, H., Meng, X., Sun, T., Li, Y., Pinson, S. R. M., . . . Peng, Z. (2020). Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. *BMC Plant Biol, 20*(1), 441. doi:10.1186/s12870-020-02603-0

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., & Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (Oryza sativa L.). *Genetics, 177*(4), 2223-2232. doi:10.1534/genetics.107.079616

Morales, K. Y., Singh, N., Perez, F. A., Ignacio, J. C., Thapa, R., Arbelaez, J. D., . . . Thomson, M. J. (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS One, 15*(5), e0232479. doi:10.1371/journal.pone.0232479

Nardi, E. P., Evangelista, F. S., Tormen, L., Saint´Pierre, T. D., Curtius, A. J., Souza, S. S. d., & Barbosa, F. (2009). The use of inductively coupled plasma mass spectrometry (ICP-MS) for the determination of toxic and essential elements in different types of food samples. *Food Chemistry, 112*(3), 727-732. doi:10.1016/j.foodchem.2008.06.010

Nawaz, Z., Kakar, K. U., Li, X. B., Li, S., Zhang, B., Shou, H. X., & Shu, Q. Y. (2015). Genome-wide Association Mapping of Quantitative Trait Loci (QTLs) for Contents of Eight Elements in Brown Rice (Oryza sativa L.). *J Agric Food Chem, 63*(36), 8008-8016. doi:10.1021/acs.jafc.5b01191

Norton, G. J., Douglas, A., Lahner, B., Yakubova, E., Guerinot, M. L., Pinson, S. R., . . . Price, A. H. (2014). Genome wide association mapping of grain arsenic, copper, molybdenum and zinc in rice (Oryza sativa L.) grown at four international field sites. *PLoS One, 9*(2), e89685. doi:10.1371/journal.pone.0089685

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., . . . Ye, G. (2015). Genome-Wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of Indica Rice Germplasm. *PLoS One, 10*(12), e0145577. doi:10.1371/journal.pone.0145577

White, P. J., & Broadley, M. R. (2009). Biofortification of crops with seven mineral elements often lacking in human diets–iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytologist, 182*(1), 49-84.

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

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., . . . Xu, C. (2018). Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus

Models. *Frontiers in plant science, 9*, 1311. Retrieved from https://www.frontiersin.org/article/10.3389/fpls.2018.01311
Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., . . . Xu, C. (2018). Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. *Front Plant Sci, 9*, 1311. doi:10.3389/fpls.2018.01311
Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. (1367-4811 (Electronic)).
Zhang, M., Pinson, S. R. M., Tarpley, L., Huang, X.-Y., Lahner, B., Yakubova, E., . . . Salt, D. E. (2014). Mapping and validation of quantitative trait loci associated with concentrations of 16 elements in unmilled rice grain. *Theoretical and Applied Genetics, 127*(1), 137-165.
Zhang, Y.-W., Lwaka Tamba, C., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., . . . Zhang, Y.-M. (2020). mrMLM v4.0: An R Platform for Multi-locus Genome-wide Association Studies. *Genomics, Proteomics & Bioinformatics*. doi:https://doi.org/10.1016/j.gpb.2020.06.006
Zhang, Y. M., Jia, Z., & Dunwell, J. M. (2019). Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Front Plant Sci, 10*, 100. doi:10.3389/fpls.2019.00100.

# 3. DISCOVERING CANDIDATE GENOMIC REGIONS FOR MILLING AND HEAD RICE YIELD, GRAIN SIZE AND SHAPE, CHALKINESS, APPARENT AMYLOSE CONTENT, PROTEIN CONTENT, AND GELATINIZATION TEMPERATURE BY USING GWAS

## 3.1. Introduction

Grain quality of rice can be defined as the overall features and characteristics of rice or/and its derivative products fulfilling the demands of the consumer (Bao, 2014). Grain appearance quality consists of grain size, shape, chalkiness and translucency. Grain size is expressed as grain length (GL), grain width (GW) and grain thickness (GT), while grain shape is the ratio of length to width, influencing the market values. Chalkiness is the opaque part of the endosperm. Chalky grains have a negative effect on milling and eating and cooking quality by reducing head milled rice rate (HMRR) and the palatability of the cooked rice, respectively. To measure chalkiness, two measurement systems are widely used, namely, degree of endosperm chalkiness (Mezmouk et al., 2011) and percentage of grains with chalkiness (PGWC). Milling qualities are determined by three parameters: brown rice rate (BRR), milled rice rate (MRR) and head milled rice rate (HMRR), representing various rice grades produced by different milling scales. Brown rice is the result of removing just the hull (consisting of the palea and lemma) and is increasingly consumed as a whole grain largely due to its status as a "health food" in Western countries (Bao, 2014). White milled rice is produced by getting rid of aleurone and pericarp and germ or embryo. HMRR is usually used as standard criteria for weighing whole milled grain having a length grain longer or equal to 3/4 full length of a kernel. It is the most determiner factor affecting market value. Cooking and eating quality (ECQ) measures the cooking flexibility as well as nature of the cooked rice, by using four major physicochemical properties such as apparent amylose content (AC), gel consistency

(GC) and gelatinization temperature (GT) and pasting viscosity. Nutritional quality basically includes protein and amino acid content, minerals and vitamins, fat content, phenolic and flavonoid content and resistant starch. Brown rice is the main source of minerals, vitamins, dietary fibers and phenols. White milled rice is produced by additionally removing the bran layer (consisting of the aleurone and pericarp) along with the germ (or embryo). HMRR is usually used as standard criteria for weighing whole milled grain having a length grain longer or equal to 3/4 full length of a kernel. It is the most important determining factor affecting market value. Cooking and eating quality (ECQ) measures the cooking flexibility as well as nature of the cooked rice by using four major physicochemical properties: apparent amylose content (AC), gel consistency (GC) and gelatinization temperature (GT) and pasting viscosity. Nutritional quality basically includes protein and amino acid content, minerals and vitamins, fat content, phenolic and flavonoid content and resistant starch. Brown rice has a higher nutritional value than white rice when considering the minerals, vitamins, dietary fibers and phenols.

QTL mapping has been widely used to dissect the genetic basis of rice grain appearance, milling, eating and nutritional quality traits. Due to the low-resolution power and limited ability to detect QTLs using biparental mapping populations, association mapping is increasingly being used to map key traits in rice. Initially, the genome-wide association study (GWAS) was applied in human genetics and then introduced in various plant species successfully (Mezmouk et al., 2011). This technique has also used widely in rice, starting in 2010 by Huang et al. (2010) to detect QTLs for 14 agronomic traits.

The general linear model (GLM) and mixed linear model (MLM) (Bradbury et al., 2007) methods were used initially to conduct GWAS. But due to their limitations, such as the requirement for multiple test correction and the constraint in detecting single loci at a time while multiple loci are involved for complex traits (C. Li, Fu, Sun, Wang, & Wang, 2018; Y. Xu et al., 2018), multi-locus models have been proposed. These multivariate models consider all the loci simultaneously; as a result, multiple test corrections are not needed. So far, several multi-locus GWAS models have been developed and used to study GWAS. All the multi-locus models follow the two-step principle during analysis. In the first stage, all the potentially associated SNPs are identified or scanned in the whole genome. During the second step, all the identified SNPs are included in one model, then their effects are estimated by empirical Bayes, and finally all the non-zero effects are further evaluated using the likelihood ratio test. A less stringent critical p-value, such as 0.01, is used to select the SNPs in the first step. Each of these multi-locus models is different from the other in terms of algorithms utilized in the two steps (Cui, Zhang, & Zhou, 2018; C. Li et al., 2018; Y. Xu et al., 2018).

In this study, the main goal is 1) to detect all possible loci controlling rice grain appearance qualities (grain length, width, GLWR, DEC, and PGWC), milling qualities (MRY, HRY), eating qualities (AAC, ASV) and nutritional quality (protein content), 2) to compare the detection ability of SNPs of single-locus and multi-locus methods of GWAS. The outcome of this study will be helpful to get information on the genetic basis of these traits, helping to accelerate new rice varieties development.

## 3.2. Materials And Methods

### 3.2.1. Plant Materials And Sample Preparation

We used the USDA-GRIN germplasm collection, including 151 diverse accessions collected from 31 countries across the world and 23 US-released varieties for our study. All accessions had similar heading dates to avoid the effect of flowering time on rice grain quality (Table. 1). The field experiment was conducted in Texas A&M AgriLife Research Center, Beaumont, Texas, in 2018. The details field experiment design and sample preparation for this study were described in Chapter 2.2.1 and 2.2.2.

### 3.2.2. Phenotypic Measurements

Grain appearance and milling quality traits

While milling the rice using the PAZ-1 DTA instrument (Zaccaria USA, Anna, TX), two milling quality related traits, MRR and HMRR, were measured according to GIPSA standard (GIPSA, 2009). Then, all full head milled rice kernels of each accession were used to measure grain length (GL, mm), grain width (GW, mm), grain length-width ratio (GLWR), degree of endosperm chalkiness (DEC, %), percentage of grain with chalkiness (PGWC, %) by using Winseedle Image analysis, Regent Instruments.

Alkali spreading value (ASV)

The alkali spreading value (ASV) method, used for predicting gelatinization temperature (GT), was determined according to Little (1958) using 1.7% KOH, for 23 h, at 25 °C. Six whole milled rice kernels were immersed in 10 ml KOH solution in each replication of ASV analysis.

AAC and protein content

Determination of AAC in rice flour was measured by following the iodine colorimetric method of AACC Method, except the automation of the color reaction part. The absorbance of the solution was measured at 620 nm with a microplate reader. AAC was calculated using a standard curve made from four different rice samples with known AACs that belongs to low, medium and high amylose class. Wx (0.8%), Bengal (13.1%), Cypress (21.5%) and DXBL (25.8%) varieties were used as standard samples in this study.

Protein content was measured from brown rice by using NIR (Near InfraRed spectroscopy). The average trait value of two replications of each accession was used during GWAS data analysis.

### 3.2.3. Analysis Of Phenotypic Data

Basic statistics were conducted to characterize the phenotypic variation in the panel. One-way analysis of variance (ANOVA) was used to determine the effect of population structure on the phenotypic variation. All the analyses were done in JMP Pro 15.

### 3.2.4. GWAS Analysis

A total of 6,565 high quality SNPs from the 7K SNP array data (Morales et al., 2020) were used for the GWAS analysis. Before conducting association studies, SNPs markers were imputed for missing genotypes by MACH 1.0. To control the false positive results, population structure and kinship analysis were also conducted. A detailed description of the procedures was mentioned in Chapter 2, section 2.5. The LD decay across the whole genome was measured by PopLDdecay 3.41. The LD decay was calculated as the chromosomal distance at which the average r2 is half of its maximum value (Huang et al., 2010). We used both single-locus and multi-locus models to conduct the GWAS analysis. CMLM, ECMLM and GEMMA models were used as single-locus

models, whereas mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO were the multi-locus models. TASSEL 5, GAPIT v.3 and GEMMA were utilized to conduct the CMLM, ECMLM and GEMMA model, respectively, in our study. All the multi-locus models were implemented in mrMLM R package (Y.-W. Zhang et al., 2020). To declare a SNP as significant, we used p <10-3 as the cut-off value in the single-locus model. For all the multi-locus models, the critical LOD scores for significance were set at 3.0 (Figure 2.2).

### 3.2.5. In-Silico Gene Expression Analysis

The genes within the LD decay distance on either side of the significant SNPs were mined by using RAP-DB database (https://rapdb.dna.affrc.go.jp/). We conducted, then, in-silico gene expression analysis using Nipponbare gene expression data in the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/expression.shtml).

## 3.3. Results

### 3.3.1. Phenotypic Variation And Correlation Analysis

Overall, most of the phenotypic traits showed skewed distributions, except grain length, width and GLWR. Bimodal distributions were observed for ASV and PC (Figure 3.1). As population structure was considered as the main affecting factor for GWAS, it explained a wide range (1%-26%) of the phenotypic variations. Mean differences between the *indica* and *japonica* panels were found significant for all traits, except GLWR, MRY and HRY traits (Table 3.1).

The phenotype pairwise correlation showed that strong positive and negative correlations were found among the traits of grain appearance. Overall, weak correlations were observed among grain appearance, milling, and eating quality traits. Protein content had a moderate level correlation with grain appearance and milling quality traits but a very weak correlation with eating

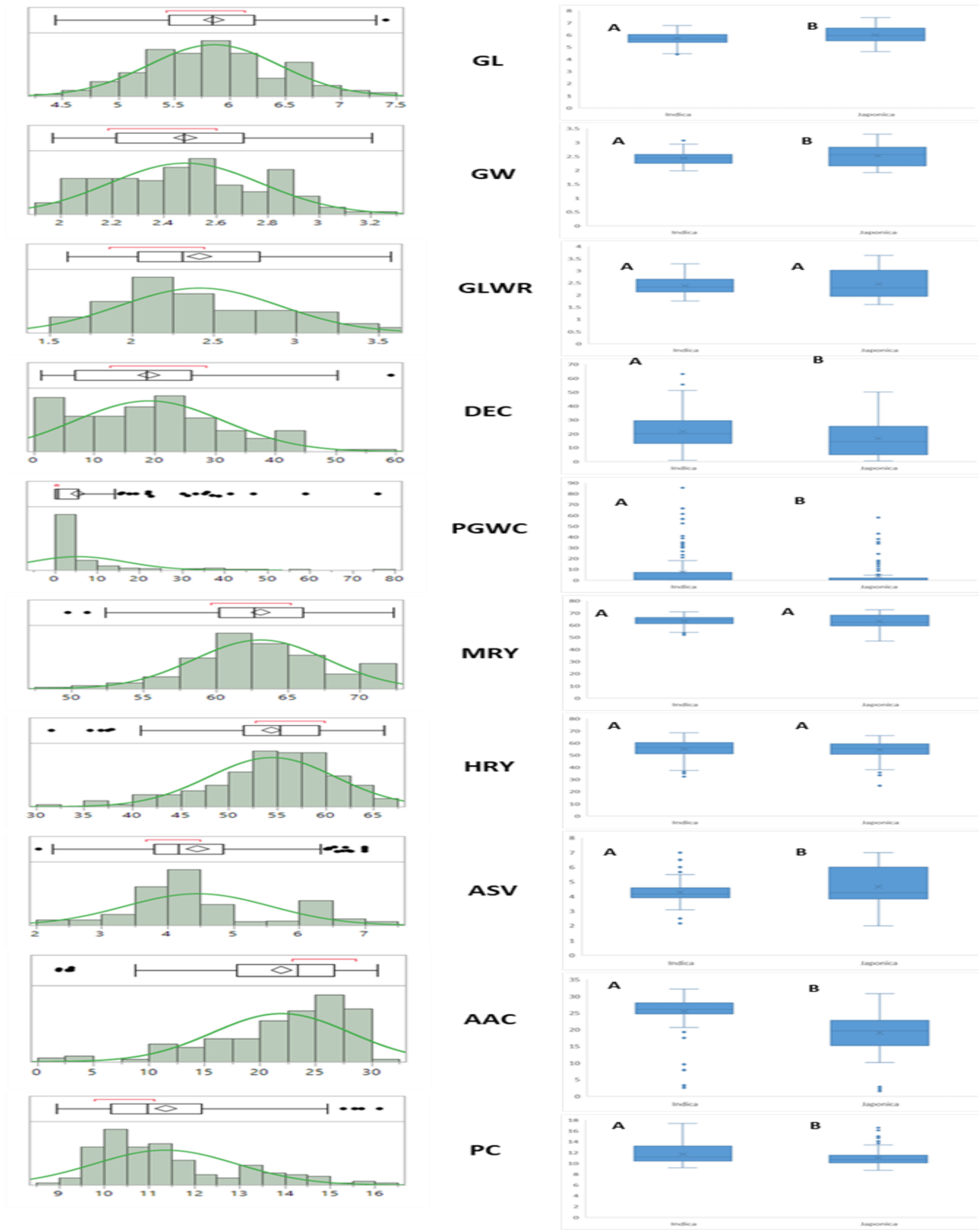



Figure 3.1. Phenotypic variation of rice grain appearance, milling, eating and cooking, and nutritional quality traits. Different capital letters in the same box plot indicate Indica and Japonica rice accessions are significantly different at α= 0.05 for mean value of the measured traits.

quality traits. GL and GLWR were positively correlated with each other, and negatively correlated with GW. DEC and PGWC were positively correlated with each other and they were negatively correlated with GL and GLWR but positively correlated with GW. MRY and HRY were positively correlated, and both were negatively correlated with all grain appearance quality traits except GL and GLWR. AAC and ASV had a low negative correlation with each other (Figure 3.2).

### 3.3.2. Population structure and GWAS analysis

STRUCTURE analysis shows that six sub-populations are present in our sample collections because Δk was at the highest peak at k=6. So, a six Q-matrix was used as a covariate during the GWAS analysis (Figure 2.3). It is well known that rice has two major sub-populations, Indica and Japonica, which themselves consist of smaller subgroups (Figure 2.3). To determine the population structure effect on the phenotypic variation, we considered the two sub-populations to analyze the phenotypic variation. Therefore, 78 and 93 accessions were identified in the Indica and Japonica panels, respectively; in total, 171 samples were analyzed during phenotypic analysis. Three accessions were removed due to admixture.

The LD decay of all the chromosomes was estimated to 250 kb, with half the maximum of mean r2 values (Figure 2.3).

SNPs with MAF $< 0.05$ only were considered as significant SNPs when their significant levels exceeded $-\log10P \geq 3.0$ and LOD $\geq 3.0$ for all single and multi-locus models, respectively. SNPs with a physical distance of less than 250 kb were regarded as the same significant SNP-trait association locus, also referred to as a GWAS QTL. This study found a total of 216 GWAS QTLs for seven grain appearance and milling quality traits, one eating quality and two nutritional qualities traits across all nine models (Figure 3.3; Figure 3.5; Table 3.2).

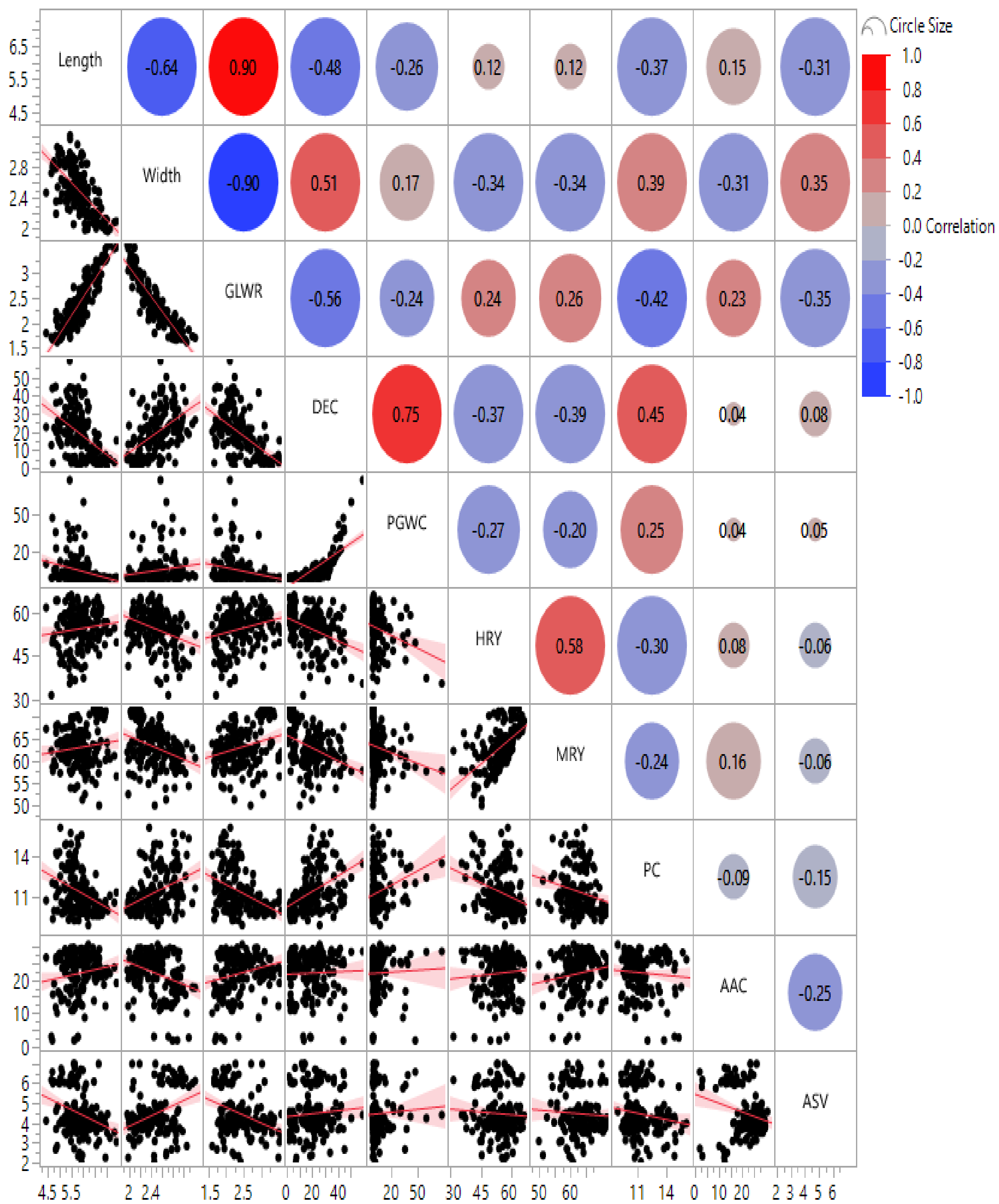



Figure 3.2. Correlation matrix for ten measured traits belong to rice grain appearance, milling, eating, and cooking, and nutritional quality. The value inside the circles shows correlation value between two minerals. Size of the circle indicate the magnitude of significant level at α= 0.05.

Among 216 significant SNPs, 17, 42, 33, 15, 21, 19 and 18 SNPs were identified for length, width, GLWR, DEC, PGWC, MRY and HRY, respectively. For length traits, multi-locus models detected 10 SNPs, followed by both single and multi-locus modes (4), and only single-locus models (3). The phenotypic variation explained by the only multi-locus and both models were 2.1-11.54% and 15.99-67.97%, respectively. In width, 25, 11 and 6 SNPs were detected by the only single-locus model, only multi-locus model and both models, respectively with explaining 8.19-70.83%, 0.93-17.48% and 10.68-72.18% of the phenotypic variation, respectively. In GLWR, only single-locus model found 11 significant SNPs explaining 6.85-74.76a% of the phenotypic

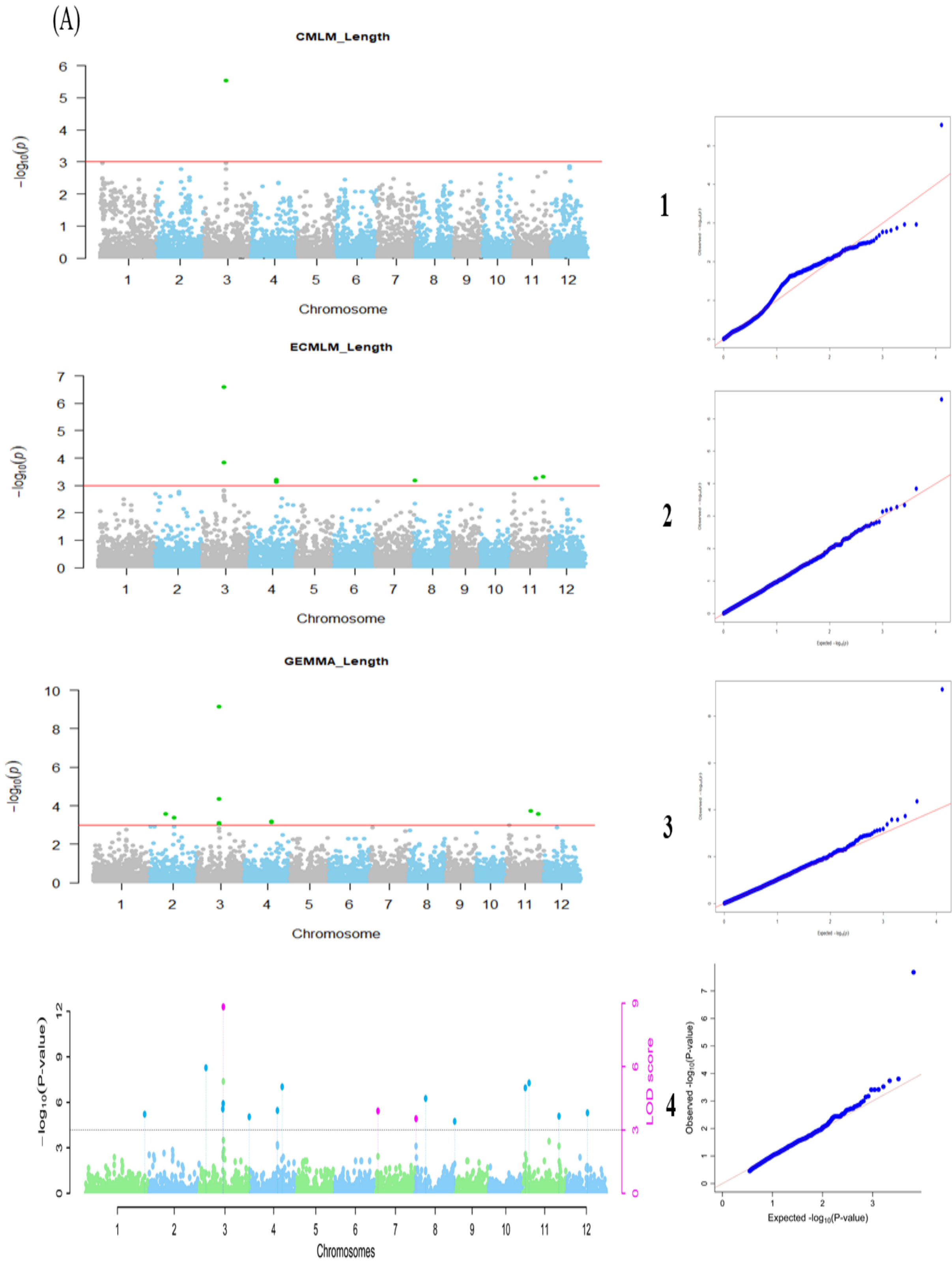
Table 3.1. Phenotypic variation in whole, *Indica* and *Japonica* rice

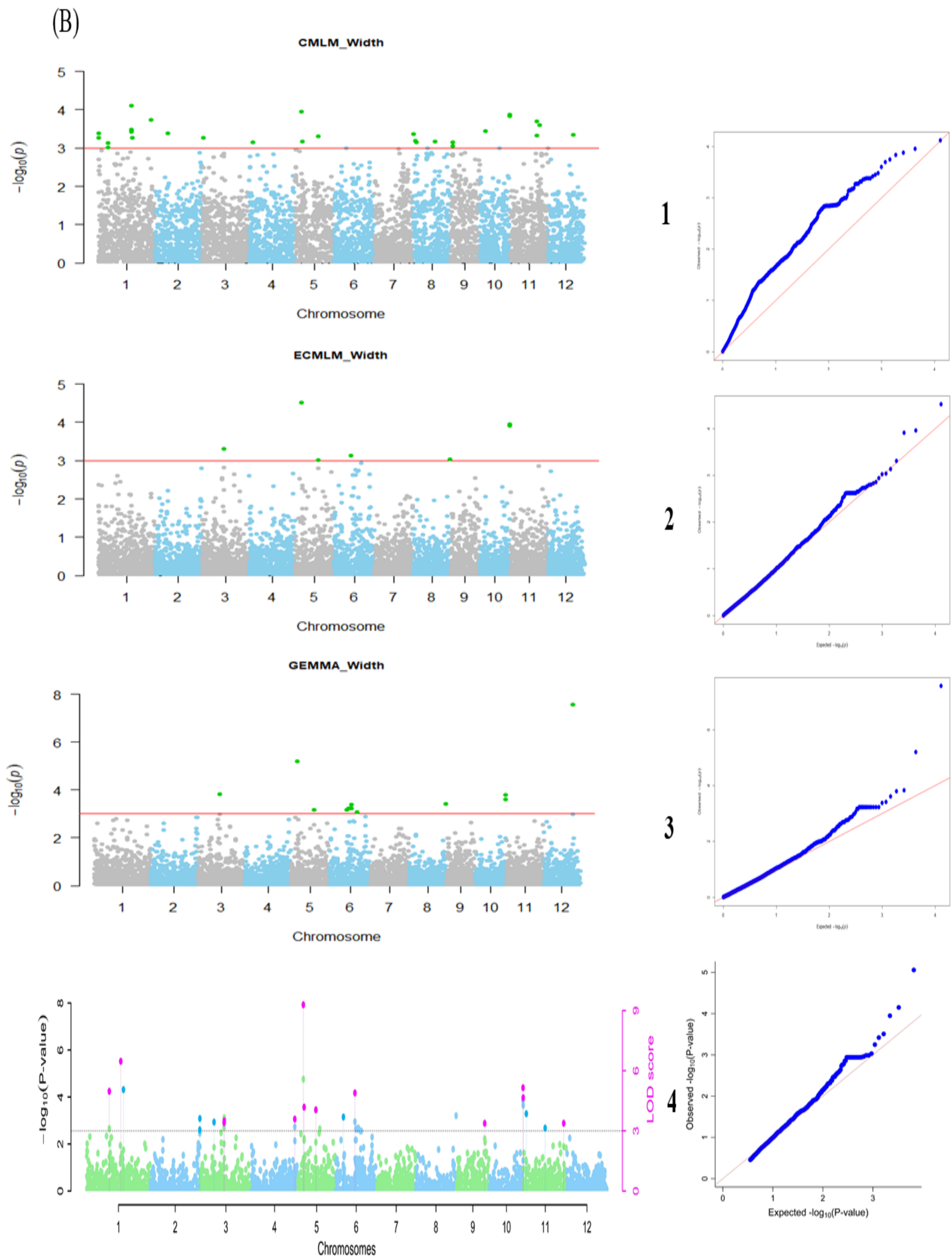
| **Traits** | **Whole Panel** | | | | | | | **Indica** | | **Japonica** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | CV | R2 | P-value | Mean | SD | Mean | SD |
| Length | 5.86 | 0.59 | 4.41 | 7.44 | 10.12 | 0.06 | 0.00 | 5.72* | 0.46 | 5.99 | 0.66 |
| Width | 2.48 | 0.30 | 1.92 | 3.30 | 11.97 | 0.03 | 0.02 | 2.43* | 0.22 | 2.52 | 0.34 |
| GLWR | 2.42 | 0.49 | 1.61 | 3.63 | 20.37 | 0.01 | 0.44 | 2.38 | 0.35 | 2.45 | 0.59 |
| DEC | 18.89 | 12.81 | 0.42 | 62.98 | 67.80 | 0.04 | 0.00 | 21.58* | 12.52 | 16.51 | 12.62 |
| PGWC | 5.58 | 11.84 | 0.00 | 85.62 | 212.31 | 0.02 | 0.02 | 7.47* | 14.08 | 3.90 | 9.14 |
| MRY | 63.23 | 5.04 | 47.05 | 72.66 | 7.97 | 0.00 | 0.87 | 63.33 | 4.36 | 63.14 | 5.57 |
| HRY | 54.60 | 7.24 | 24.93 | 68.55 | 13.26 | 0.01 | 0.30 | 54.94 | 7.65 | 54.31 | 6.88 |
| ASV | 4.48 | 1.13 | 2.00 | 7.00 | 25.12 | 0.03 | 0.00 | 4.28* | 0.83 | 4.66 | 1.31 |
| AAC | 22.02 | 6.25 | 1.52 | 32.15 | 28.39 | 0.26 | 0.00 | 25.44* | 4.87 | 19.03 | 5.78 |
| PC | 11.35 | 1.63 | 8.70 | 17.40 | 14.34 | 0.05 | 0.00 | 11.71* | 1.78 | 11.03 | 1.41 |

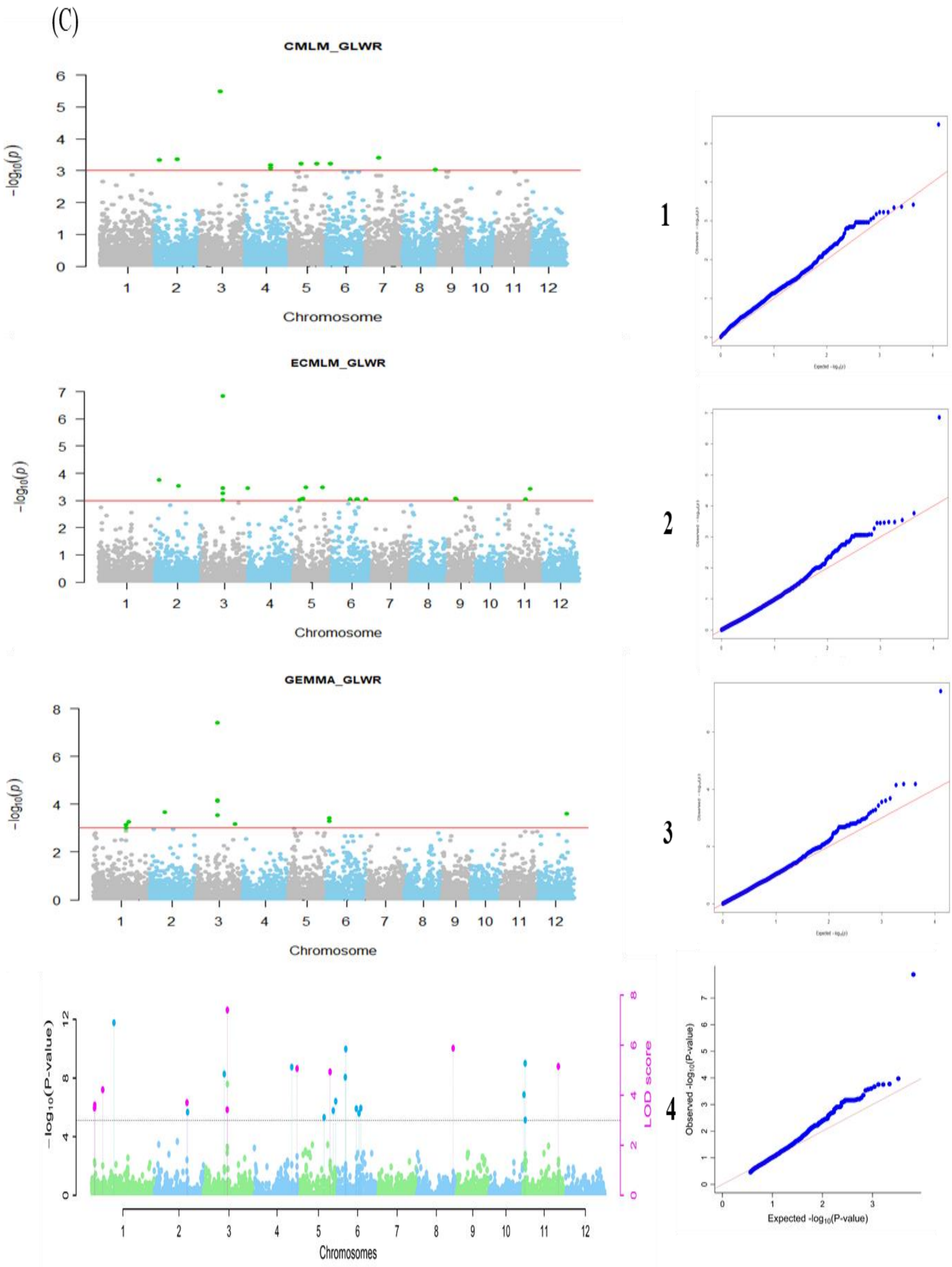
variation, followed by 21 SNPs identified by only multi-locus models that explained 1.44-12.84% of the phenotypic variations. Only one SNP was co-detected by both models that explained 15.58-77.63% of the phenotypic variation. For DEC, the highest number of SNPs (14) were detected by only single-locus models, followed by four SNPs by only multi-locus models and three SNPs by both models. The phenotypic variation explained was 8.54-40.7%, 4.7-15.09% and 9.44-43, respectively. For PGWC traits, 14, three and four significant SNPs were found by only single-locus, only multi-locus and both models that explained 6.89-18.35%, 4.14-8.12% and 11.32-23.38% of the phenotypic variation. For MRY traits, 19 SNPs were detected, including seven SNPs by only single-locus, six by only multi-locus and the remaining six SNPs were co-detected by both models, with 52.01-52.12%, 2.57-9.97% and 3.44-53.14% of the phenotypic variation explained by only single-locus, only multi-locus, and both models, respectively. In the case of HRY, only single-locus and both models detected an equal number of SNPs (7) with explaining 7.06-22.16% and 1.56-22.69% of the phenotypic variation, respectively, and only multi-locus models identified four SNPs that explained 1.1-9.21% of the phenotypic variation (Figure 3.3; Table 3.2).

For ASV, 22 significant SNPs were detected, where 14, 3 and 5 SNPs were identified by only single-locus models, only multi-locus models, and both models, respectively. Only single-locus models explained 6.69-54.76%, followed by 0-0.88% and 2.91-41.47% of the phenotypic variation that was explained by only multi-locus and both models, respectively (Figure 3.3; Table 3.2).

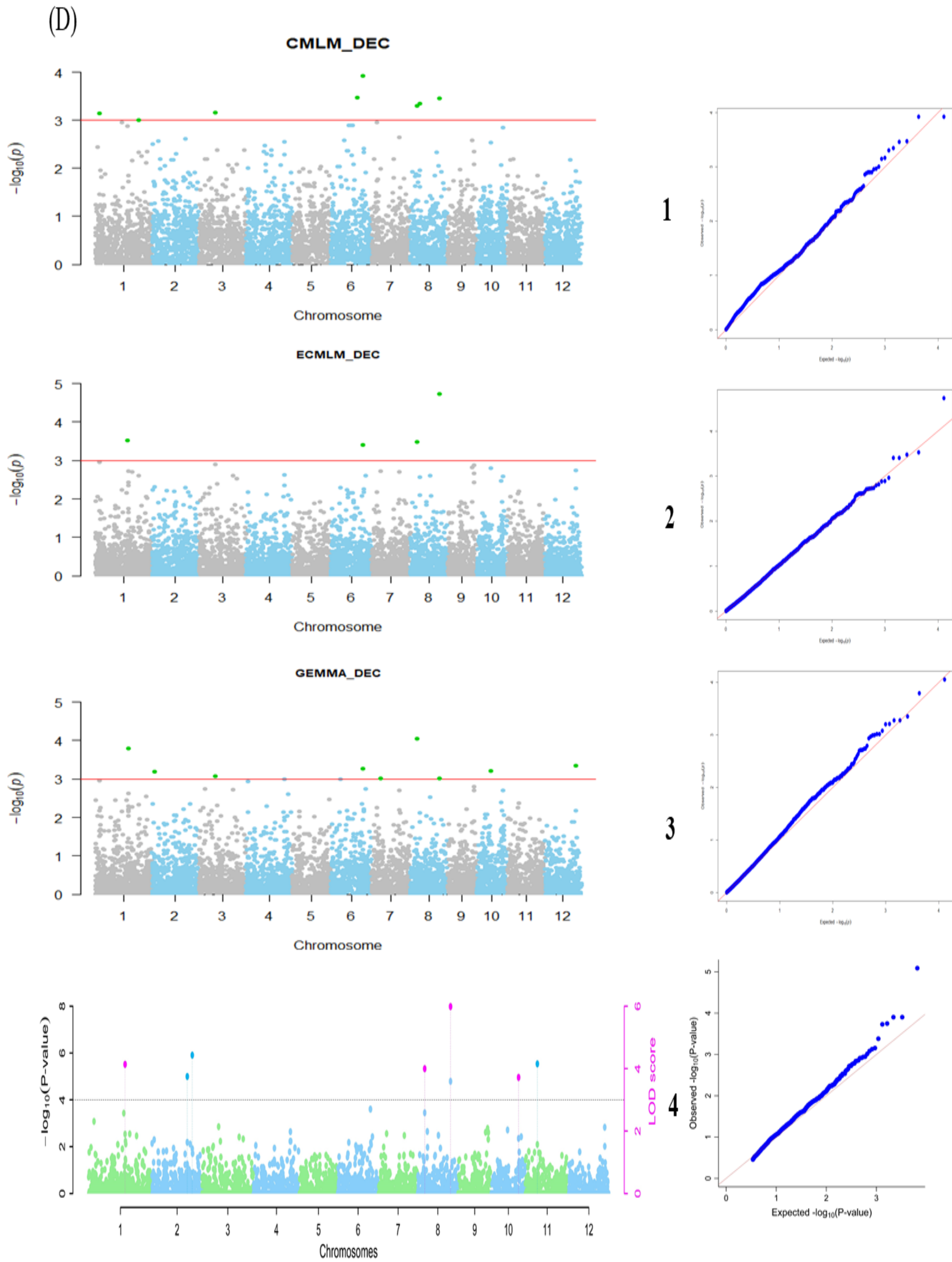
For nutritional quality traits, 16 and 13 significant SNPs were identified for AAC and PC, respectively. For AAC, among the 16 SNPs, nine SNPs were found by only a single-locus model that explained 8.74-50.09% of the phenotypic variation. Six SNPs were detected by both models with explaining 3.53-61.38% of the phenotypic variation and the remaining SNP was identified by

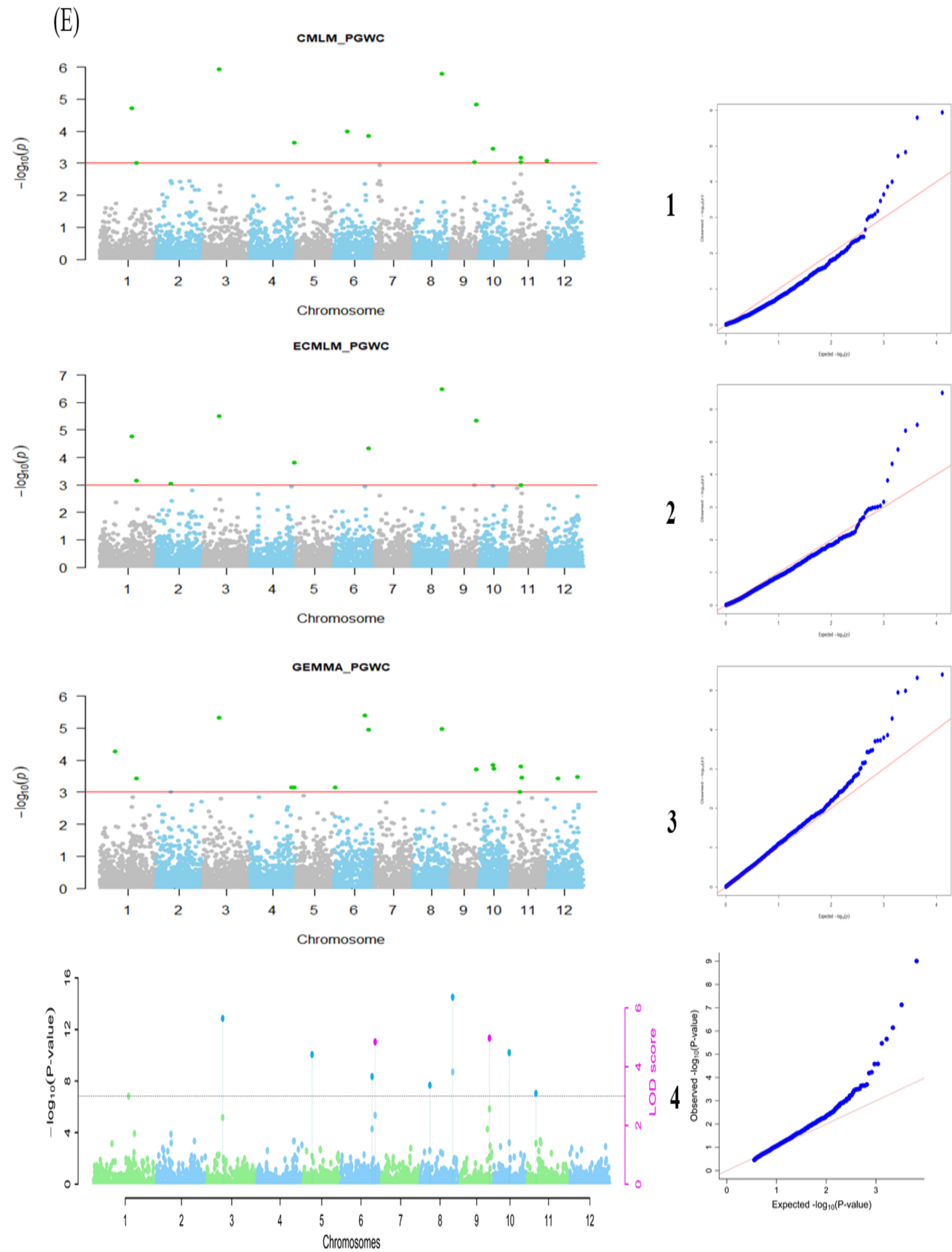

(A)
CMLM_Length
ECMLM_Length
GEMMA_Length
Chromosome
Chromosomes
LOD score
1
2
3
4

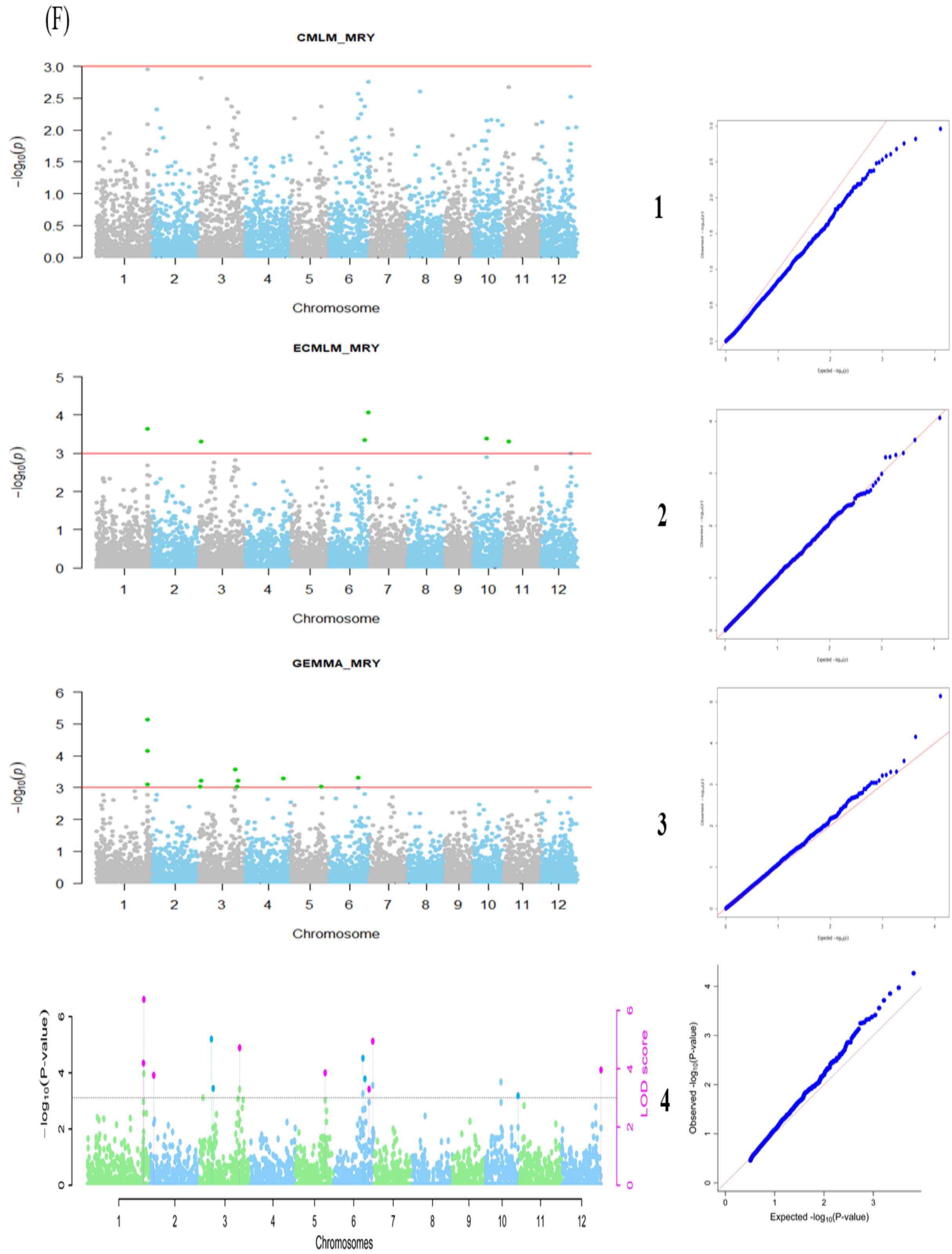

(B)
CMLM_Width
ECMLM_Width
GEMMA_Width
Chromosome
Chromosomes
1
2
3
4
LOD score
Expected -log10(P-value)
Observed -log10(P-value)


(C)
CMLM_GLWR
-log10(p)
Chromosome
ECMLM_GLWR
GEMMA_GLWR
-log10(P-value)
LOD score
Chromosomes
Observed -log10(P-value)
Expected -log10(P-value)
1
2
3
4

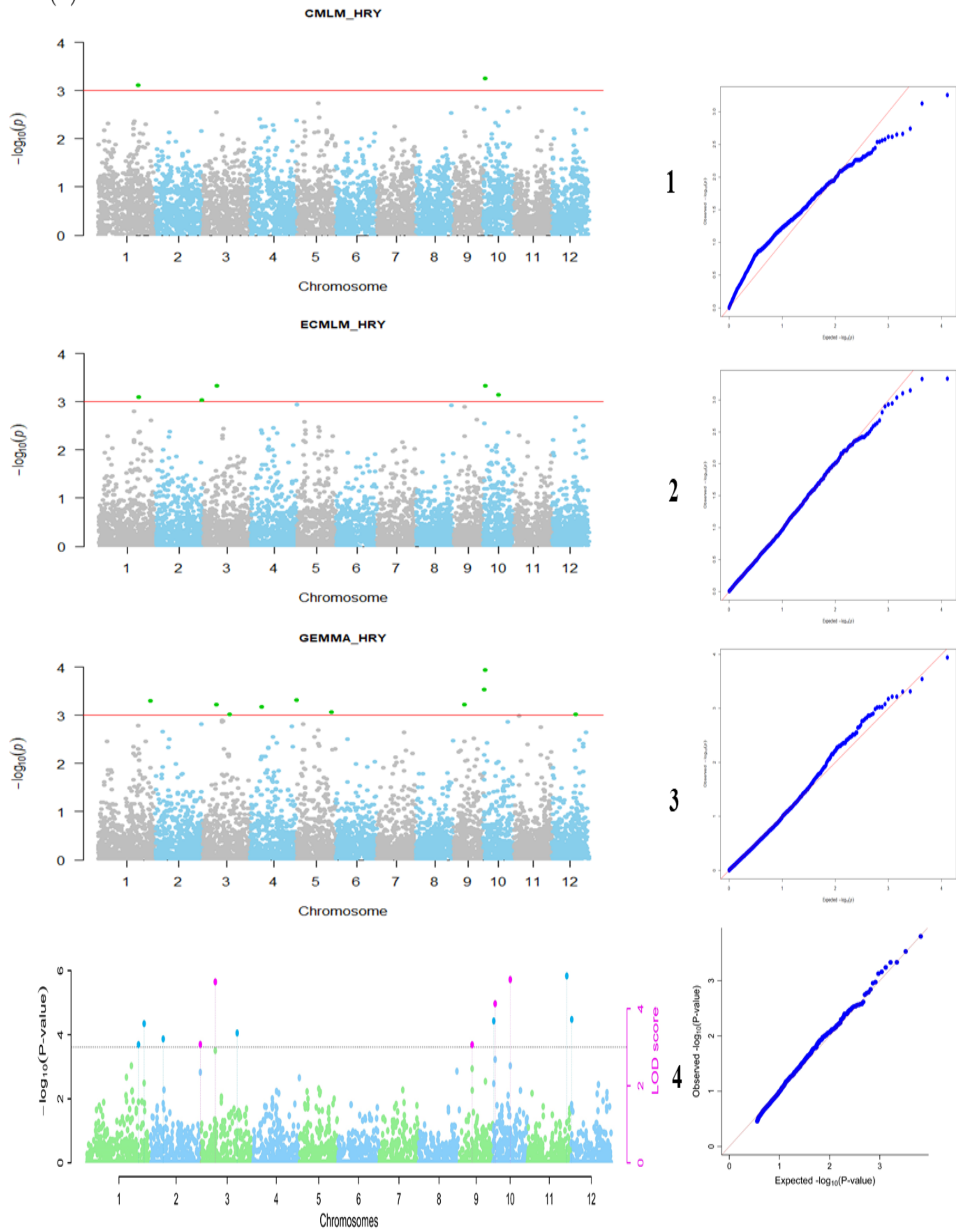
(D)

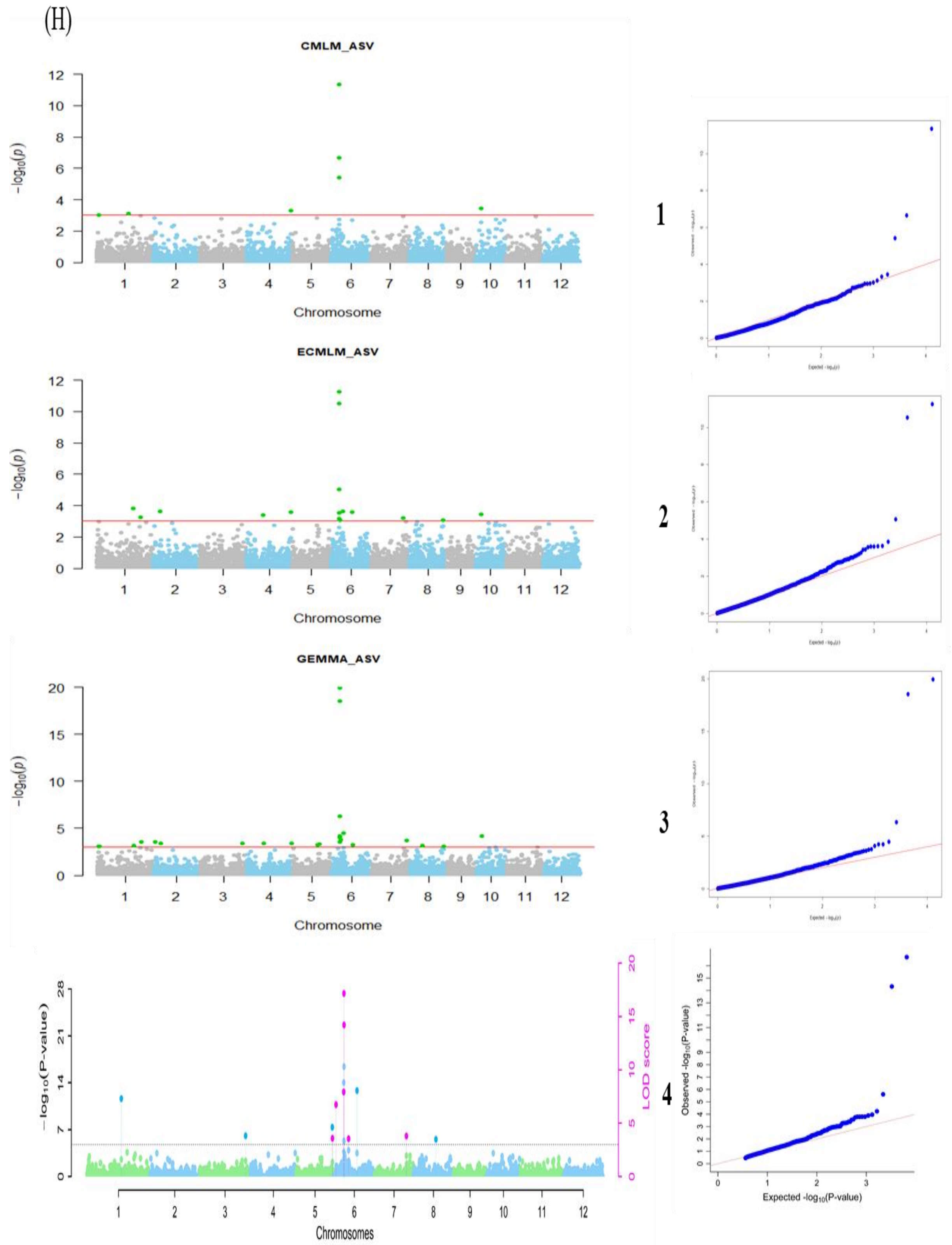


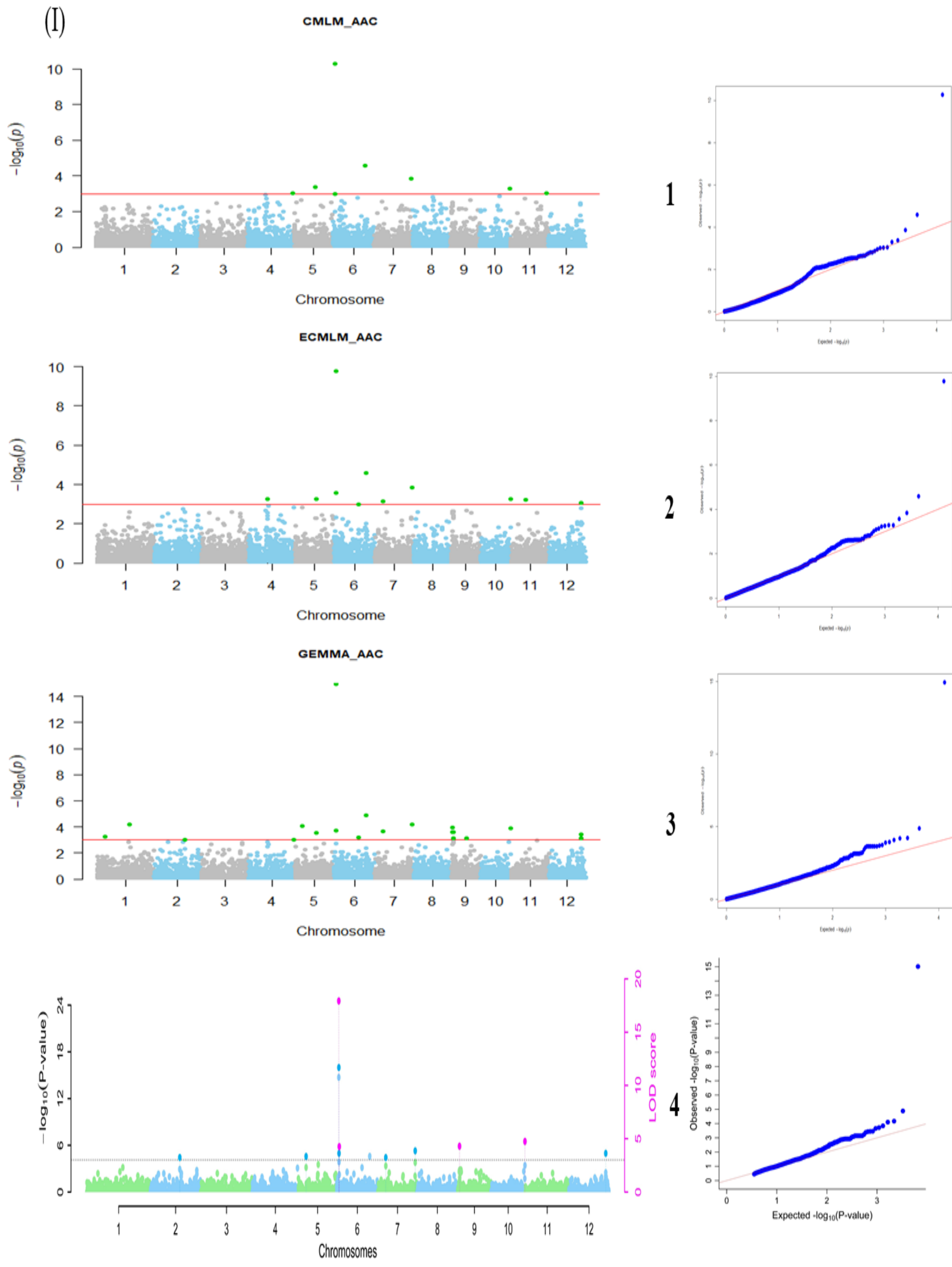

(E)
CMLM_PGWC
ECMLM_PGWC
GEMMA_PGWC
Chromosome
Chromosomes
LOD score
1
2
3
4


(F)
CMLM_MRY
ECMLM_MRY
GEMMA_MRY
Chromosome
Chromosomes
LOD score
1
2
3
4

(G)




(H)
CMLM_ASV
ECMLM_ASV
GEMMA_ASV
Chromosome
Chromosomes
LOD score
1
2
3
4
Expected -log10(P-value)
Observed -log10(P-value)



(I)
CMLM_AAC
ECMLM_AAC
GEMMA_AAC
Chromosome
Chromosomes
LOD score
1
2
3
4

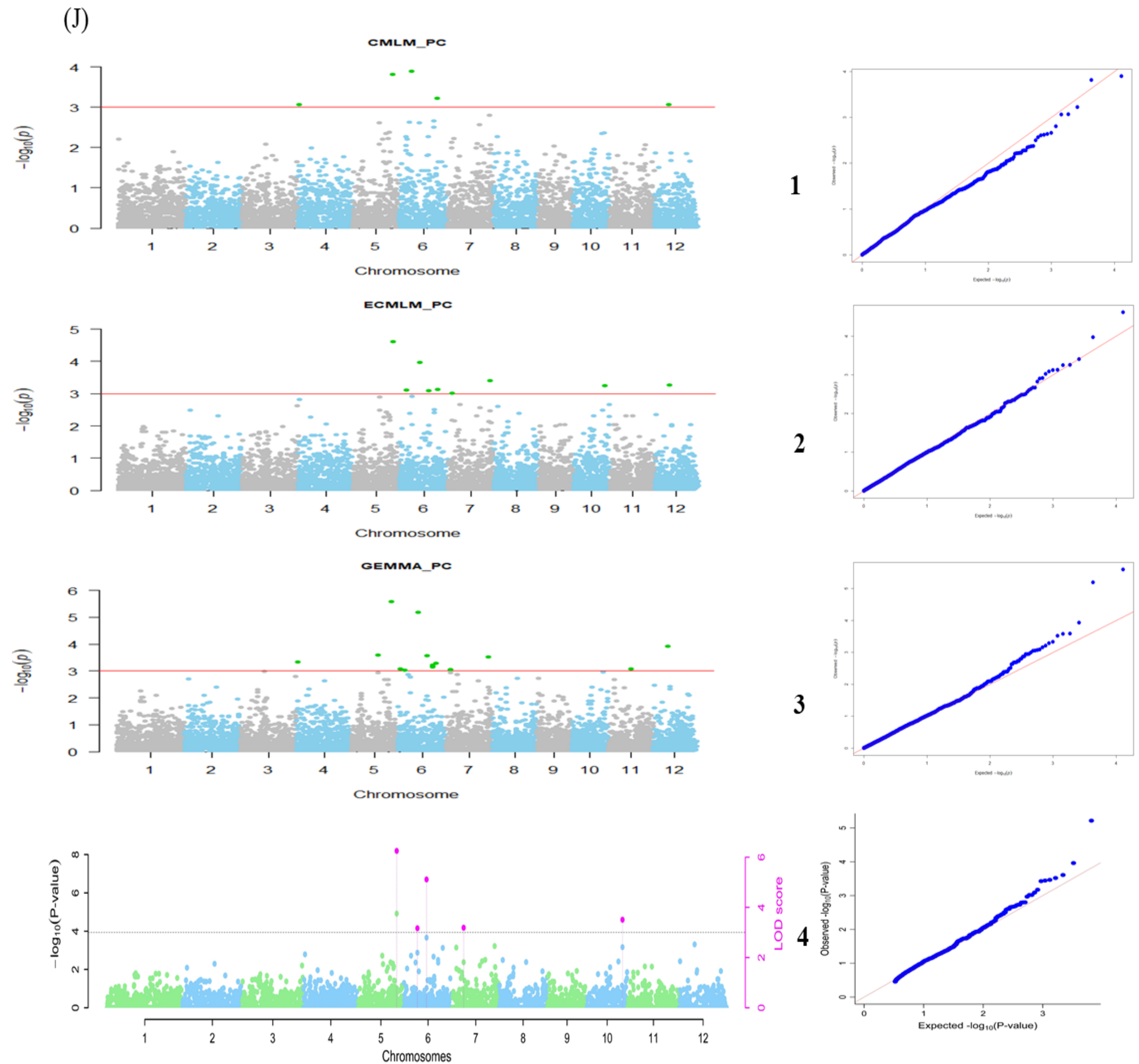



Figure 3.3. Manhattan plots of GWAS for ten measured traits used in the GWAS study. (1). Manhattan plot CMLM model. (2). Manhattan plot for ECMLM model. (3). Manhattan plot for GEMMA model. (4). Manhattan plot for multi-locus models, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, ISIS EM-BLASSO. The red horizontal line in the 1, 2, and 3 models is the threshold significant level used in the study to declare a SNP as being significant for measured traits. The green circles above the red line depict the significant SNPs. For 4 model, purple circles above the dashed horizontal line are the significant SNPs identified by all six multi-locus models, whereas green circles show only those SNPs identified by any two models of six models of multi-locus method. (A) Grain Length (GL). (B) Grain Width (GW). (C) Grain length-width ratio (GLWR). (D) Degree of endosperm chalkiness (DEC). (E) Percentage of grains with chalkiness (PGWC). (F) Milled rice yield (MRY). (G) Head rice yield (HRY). (H) Alkali spreading value (ASV). (I) Apparent amylose content (AAC). (J) Protein content (PC).

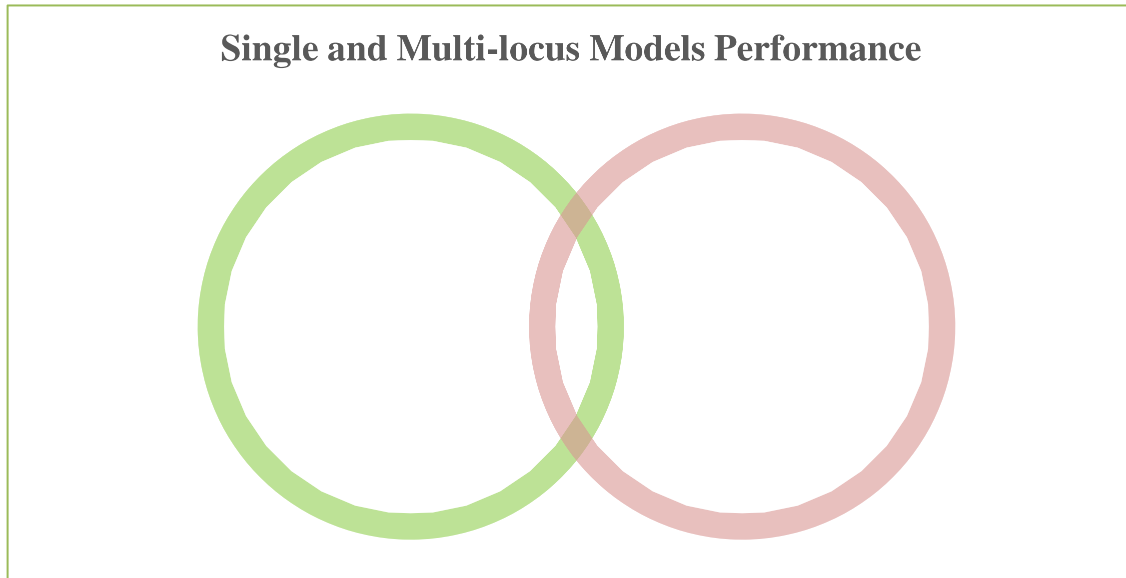



Figure 3.4. Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for ten traits of rice grain appearance, milling, eating and cooking, and nutritional quality.

only multi-locus model that explained 4.48% of the phenotypic variation. For PC, eight SNPs were identified by only a single-locus model, explaining 8.99-30.46% of the phenotypic variation. Only multi-locus models and both models detected one SNP and four SNPs, respectively, explaining 6.6% and 2.94-42% of the phenotypic variation, respectively (Figure 3.3; Table 3.2).

Across the 216 significant SNPs, 23 SNPs were found that had an effect on more than one trait (i.e., a pleiotropic effect). For example, c1p40455715 and GS3 SNPs have effects on length and width traits. Similarly, four SNPs had effects on length and GLWR, five on width and GLWR, one on width and MRY, one on width and HRY, one on width and PC, one on GLWR and DEC, one on GLWR and ASV, one on GLWR and PC, four on DEC and PGWC, one on MRY and HRY and one on ASV and AAC. Three SNPs were found affecting multiple traits, including SNP GS3

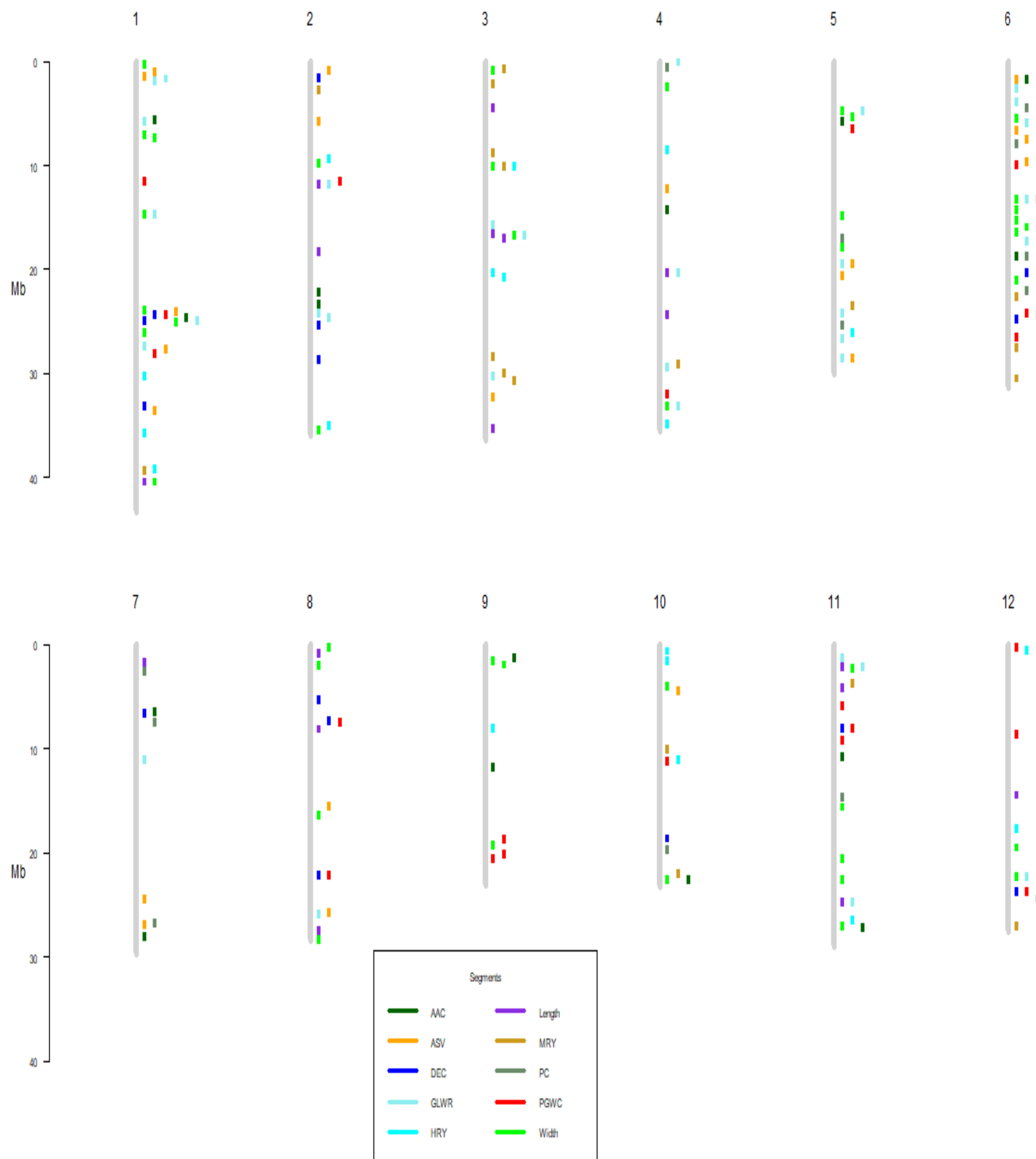



Figure 3.5. All the significant SNPs found in the study for ten rice grain appearance, milling, eating and cooking, and nutritional quality. SNPs positions are depicted by the rectangular box. Specific color shows the corresponding traits. Rice chromosomes are displayed by vertical lines.

Table 3.2. List of loci detected in the study.

| Trait | SNP | Alleles | Chr. | Pos.(bp) | Single-locus GWAS | | | Multi-locus GWAS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Length | | | | | -log10P | R2 | Model | LOD | R2 | Model |
| | c1p40455715 | G/A | 1 | 40455716 | | | | 3.74 | 2.3 | I |
| | 1739007 | A/G | 2 | 11861014 | 3.57 | | G | | | |
| | 1955025 | G/A | 2 | 18370293 | 3.37 | | G | | | |
| | SNP-3.4541545. | G/A | 3 | 4542544 | | | | 5.94 | 4.3 | I |
| | GS3 | G/A | 3 | 16733441 | 5.53-9.14 | 15.9-67.9 | C, EC, G | 8.83 | 27.3 | FE, FM, I, M, PL |
| | id3008418 | A/G | 3 | 17019705 | 3.06 | | G | | | |
| | 3538410 | C/A | 3 | 35295694 | | | | 3.62 | 2.1 | PL |
| | id4006172 | G/A | 4 | 20396474 | 3.17-3.22 | 64.0 | EC, G | 3.92 | 9.4 | PK |
| | 4518584 | G/A | 4 | 24397585 | | | | 5.04 | 10.3 | 4. I |
| | 7006027 | A/C | 7 | 1762346 | | | | 4.23 | 5.3 | FE, PL |
| | 8004315 | G/A | 8 | 811201 | 3.17 | 63.9 | EC | 3.37 | 10.0 | FM, M |
| | SNP-8.8136656. | A/G | 8 | 8137653 | | | | 4.49 | 7.7 | PL |
| | 9037775 | G/A | 8 | 27517812 | | | | 3.41 | 10.8 | M |
| | 10877755 | C/A | 11 | 2208347 | | | | 4.99 | 11.5 | PK |
| | SNP-11.4140934. | G/A | 11 | 4145033 | | | | 5.22 | 6.3 | PL |
| | SNP-11.24316754. | C/A | 11 | 24782915 | 3.33-3.58 | 64.1 | EC, G | 3.65 | 2.1 | PL |
| | id12005205 | C/A | 12 | 14489210 | | | | 3.81 | 4.8 | PK |
| Width | SNP-1.303375. | A/G | 1 | 304376 | 3.38 | 9.2 | C | | | |
| | SNP-1.7150499. | A/T | 1 | 7151500 | 3.13 | 8.6 | C | | | |
| | 232009 | C/A | 1 | 7421797 | 3.01 | 8.2 | C | | | |
| | 482027 | A/G | 1 | 14736394 | | | | 3.3 | 3.1 | FE, I, PK, PL |
| | SNP-1.23956743. | A/G | 1 | 23957788 | | | | 7.8 | 6.5 | PK, PL |
| | id1015006 | A/G | 1 | 25139879 | 4.12 | 11.4 | C | 5.05 | 7.0 | PL |
| | 882462 | A/C | 1 | 26095782 | 3.26 | 8.9 | C | | | |
| | c1p40455715 | G/A | 1 | 40455716 | 3.74 | 10.3 | C | | | |

| Table 3.2. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | id2004711 | G/A | 2 | 9880575 | 3.38 | 9.3 | C | | | |
| | SNP-2.35507276. | C/A | 2 | 35513146 | | | | 3.61 | 3.6 | PK |
| | SNP-3.876244. | C/A | 3 | 877247 | 3.28 | 9.0 | C | | | |
| | 2724361 | A/G | 3 | 10184356 | | | | 3.44 | 6.1 | PK |
| | GS3 | G/A | 3 | 16733441 | 3.31-3.83 | 71.1 | EC, G | 3.11 | 4.7 | FE, FM, PL |
| | SNP-4.2481854. | A/G | 4 | 2486257 | 3.15 | 8.6 | C | | | |
| | 4754818 | C/A | 4 | 33157662 | | | | 3.17 | 3.1 | I, PK |
| | 4953842 | A/G | 5 | 4799269 | 3.95-5.19 | 10.9-72.2 | C, EC, G | 9.62 | 20.1 | FM, M, PL |
| | 4970110 | A/G | 5 | 5338205 | 3.17 | 8.7 | C | 4.65 | 10.2 | FM, M |
| | 5344637 | G/A | 5 | 14835431 | | | | 3.04 | 0.9 | FM, PK, PL |
| | 5463395 | A/C | 5 | 17952485 | 3.02- 3.32 | 9.1-70.8 | C, EC, G | | | |
| | id6003591 | C/G | 6 | 5579211 | | | | 3.69 | 1.1 | I |
| | 6274558 | C/A | 6 | 13250266 | 3.13-3.17 | 70.9 | EC, G | 5.22 | 5.6 | FM, G, I, M |
| | SNP-6.14278613. | G/A | 6 | 14279613 | 3.22 | | G | | | |
| | 6368871 | G/A | 6 | 15315368 | 3.22 | | G | | | |
| | SNP-6.15968673. | G/A | 6 | 15969672 | 3.22 | | G | | | |
| | 6424219 | G/A | 6 | 16465801 | 3.38 | | G | | | |
| | 6623174 | G/A | 6 | 21059263 | 3.06 | | G | | | |
| | 7988967 | G/A | 8 | 250632 | 3.36 | 9.2 | C | | | |
| | id8000555 | A/G | 8 | 1995144 | 3.19 | 8.7 | C | | | |
| | 8640728 | A/G | 8 | 16443521 | 3.17 | 8.6 | C | | | |
| | id8007977 | A/G | 8 | 28377609 | 3.03-3.41 | 70.8 | EC, G | | | |
| | 9125104 | G/A | 9 | 1636524 | 3.05 | 8.3 | C | | | |
| | SNP-9.1950655. | G/A | 9 | 1951656 | 3.14 | 8.6 | C | | | |
| | 9796547 | G/A | 9 | 19322095 | | | | 3.35 | 3.9 | FM. M |
| | SNP-10.4004044. | A/G | 10 | 4005069 | 3.45 | 9.5 | C | | | |
| | SNP-10.22576326. | A/C | 10 | 22647853 | 3.79- 3.96 | 10.7-71.7 | C, EC, G | 4.62 | 10.8 | M, PL |

| Table 3.2. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-11.2356543. | A/G | 11 | 2360685 | | | | 3.85 | 11.3 | I |
| | 11387335 | A/G | 11 | 15638505 | | | | 3.14 | 17.5 | PK |
| | 11618220 | C/A | 11 | 20653107 | 3.69 | 10.2 | C | | | |
| | 11720074 | G/A | 11 | 22562351 | 3.6 | 9.9 | C | | | |
| | SNP-11.26673214. | A/G | 11 | 27144828 | | | | 3.18 | 10.1 | FM, M |
| | id12006560 | G/A | 12 | 19522102 | 3.35 | 9.2 | C | | | |
| | 12915589 | A/G | 12 | 22277106 | 7.57 | | G | | | |
| GLWR | SNP-1.1700426. | G/A | 1 | 1701427 | | | | 3.09 | 3.5 | FM, M |
| | SNP-1.1981466. | T/A | 1 | 1982467 | | | | 3.14 | 2.9 | I, PK |
| | SNP-1.5867020. | A/G | 1 | 5868021 | | | | 5.06 | 4.8 | M, PL |
| | 482027 | A/G | 1 | 14736394 | | | | 6.89 | 4.4 | PL |
| | SNP-1.25001551. | A/G | 1 | 25002596 | 3.02 | | G | | | |
| | 918735 | G/A | 1 | 27471641 | 3.25 | | G | | | |
| | 1739007 | A/G | 2 | 11861014 | 3.67 | | G | | | |
| | SNP-2.24270214. | G/A | 2 | 24276084 | | | | 3.95 | 1.9 | FM, I |
| | SNP-2.24721576. | G/A | 2 | 24727446 | | | | 3.32 | 1.4 | I |
| | 2894214 | G/A | 3 | 15709424 | | | | 4.85 | 2.7 | PL |
| | GS3 | G/A | 3 | 16733441 | 5.48-6.85 | 15.6-77.6 | C, EC, G | 11.35 | 14.3 | FE, FM, I, M, PK, PL |
| | rd3001030 | C/A | 3 | 30253323 | 3.19 | | G | | | |
| | id4000001 | A/G | 4 | 59946 | 3.45 | 74.8 | EC | | | |
| | SNP-4.20233911. | T/A | 4 | 20405868 | 3.17 | 6.9 | C | | | |
| | SNP-4.29251127. | A/C | 4 | 29436261 | | | | 5.12 | 5.8 | M |
| | SNP-4.32979048. | G/A | 4 | 33164161 | | | | 5.55 | 3.6 | I, M |
| | 4953842 | A/G | 5 | 4799269 | 3.02 | 74.4 | EC | | | |
| | SNP-5.19483749. | G/A | 5 | 19546266 | | | | 3.11 | 4.0 | PK |
| | 5661708 | G/A | 5 | 24298176 | | | | 4.93 | 4.3 | FM, I, M |

Table 3.2. Continued

| Trait | SNP | Alleles | Chr. | Pos.(bp) | -log10P | R2 | Model | LOD | R2 | Model |
|---|---|---|---|---|---|---|---|---|---|---|
| | 5724509 | G/A | 5 | 26692950 | | | | 3.38 | 10.9 | PK |
| | SNP-5.28500625. | C/G | 5 | 28563271 | | | | 3.76 | 2.9 | PL |
| | rd6001732 | G/A | 6 | 2641425 | 3.43 | | G | | | |
| | 5931130 | G/A | 6 | 3951547 | 3.22 | 8.9 | C | | | |
| | SNP-6.5975349. | G/A | 6 | 5976349 | | | | 5.84 | 12.8 | M |
| | 6274558 | C/A | 6 | 13250266 | | | | 3.46 | 3.2 | FE |
| | id6009207 | C/A | 6 | 16017986 | | | | 3.27 | 3.6 | PK |
| | 6461495 | C/A | 6 | 17273260 | | | | 3.48 | 1.8 | I |
| | 7343834 | G/A | 7 | 11160163 | 3.42 | 7.5 | C | | | |
| | id8007067 | G/C | 8 | 25902075 | | | | 8.19 | 3.9 | I, PL |
| | c11p1238204 | G/A | 11 | 1238205 | | | | 4.02 | 8.0 | M |
| | 10877755 | C/A | 11 | 2208347 | | | | 5.27 | 2.3 | I |
| | SNP-11.24316754. | C/A | 11 | 24782915 | | | | 3.01 | 1.7 | I, M, PL |
| | 12915589 | A/G | 12 | 22277106 | 3.6 | | G | | | |
| DEC | 827062 | A/G | 1 | 24357077 | 3.52 | 40.7 | EC | | | |
| | SNP-1.25001551. | A/G | 1 | 25002596 | 3.79 | | G | 3.4 | 10.6 | FM, PL |
| | 1085563 | A/G | 1 | 33145570 | 3 | 8.5 | C | | | |
| | 1407860 | A/G | 2 | 1660713 | 3.2 | | G | | | |
| | 2204731 | G/A | 2 | 25339669 | | | | 3.75 | 5.4 | M |
| | id2012493 | G/A | 2 | 28733338 | | | | 4.43 | 15.1 | M |
| | 6593105 | G/A | 6 | 20366418 | 3.47 | 9.9 | C | | | |
| | 6764049 | A/G | 6 | 24787268 | 3.27- 3.92 | 9.2-40.5 | C, EC, G | | | |
| | 7142172 | C/A | 7 | 6693263 | 3.01 | | G | | | |
| | 8149888 | A/G | 8 | 5284717 | 3.30-4.05 | 9.4-40.6 | C, EC, G | 4.68 | 7.4 | FE, FM, I, M, |
| | 8244660 | G/A | 8 | 7347361 | 3.34 | 9.6 | C | | | |
| | 8886338 | A/G | 8 | 22200719 | 3.02- 3.46 | 9.9-43.0 | C, EC, G | 5.25 | 41.6 | FE, M |
| | SNP-10.18588829. | G/A | 10 | 18660272 | | | | 3.24 | 4.7 | FM, PK |

Table 3.2. Continued

| Trait | SNP | Alleles | Chr. | Pos.(bp) | -log10P | R2 | Model | LOD | R2 | Model |
|---|---|---|---|---|---|---|---|---|---|---|
| | 11083523 | A/G | 11 | 8083891 | | | | 4.15 | 12.6 | PK |
| | 12970597 | G/A | 12 | 23736085 | 3.35 | | G | | | |
| PGWC | 353599 | C/A | 1 | 11576126 | 4.28 | | G | | | |
| | 827062 | A/G | 1 | 24357077 | 4.71-4.76 | 14.0-18.4 | C, EC | | | |
| | 939001 | A/G | 1 | 28183638 | 3.00-3.43 | 8.7-13.9 | C, EC, G | | | |
| | SNP-2.11519045. | A/C | 2 | 11519050 | 3.03 | 13.7 | EC | | | |
| | id4010985 | A/G | 4 | 32038496 | 3.15 | | G | | | |
| | 5020568 | A/G | 5 | 6593347 | | | | 4.41 | 7.6 | M |
| | 6137646 | G/A | 6 | 10049864 | 3.99 | 11.8 | C | | | |
| | 6746522 | A/G | 6 | 24198300 | 5.4 | | G | 3.66 | 12.0 | I |
| | 6817862 | G/A | 6 | 26566949 | 3.85-4.95 | 11.3-17.1 | C, EC, G | 5.35 | 15.0 | FM, M |
| | id8002314 | C/A | 8 | 7549599 | | | | 3.37 | 4.1 | FE |
| | 8886338 | A/G | 8 | 22200719 | 4.98- 6.49 | 17.5-23.4 | C, EC, G | 6.36 | 54.2 | FE |
| | 9776646 | G/A | 9 | 18670151 | 3.04 | 6.9 | C | | | |
| | 9819278 | C/A | 9 | 20138665 | 3.70- 5.34 | 14.4-20.0 | C, EC, G | 7.22 | 27.5 | FM, M, PL |
| | 9834082 | C/A | 9 | 20597071 | 3.72 | | G | | | |
| | id10002943 | C/A | 10 | 11195773 | 3.73 | | G | | | |
| | 10996639 | A/G | 11 | 5909358 | | | | 3.09 | 8.1 | M |
| | 11083523 | A/G | 11 | 8083891 | 3.02 | | G | | | |
| | 11129011 | G/A | 11 | 9183139 | 3.00-3.80 | 13.6 | EC, G | | | |
| | 12010072 | G/A | 12 | 246313 | 3.09 | 9.0 | C | | | |
| | 12311678 | G/A | 12 | 8609984 | 3.43 | | G | | | |
| | 12970597 | G/A | 12 | 23736085 | 3.48 | | G | | | |
| MRY | SNP-1.39395295. | G/A | 1 | 39396339 | 3.64-5.13 | | EC, G | 7.86 | 12.9 | FM, I, M, PK, PL |
| | 1434943 | A/G | 2 | 2796173 | | | | 3.21 | 2.7 | I, PL |
| | 2499175 | C/A | 3 | 815419 | 3.04 | | G | | | |
| | SNP-3.2199211. | C/A | 3 | 2200216 | 3.21- 3.31 | 52.0 | EC, G | | | |

| Table 3.2. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | id3004633 | G/A | 3 | 8803255 | | | | 5.01 | 5.9 | I |
| | 2724361 | A/G | 3 | 10184356 | | | | 3.32 | 2.6 | FM |
| | SNP-3.28390394. | G/A | 3 | 28397342 | 3.57 | | G | 4.71 | 22.1 | FM, M, PL |
| | 3405830 | G/A | 3 | 29987079 | 3.04 | | G | | | |
| | id3014633 | G/A | 3 | 30751779 | 3.23 | | G | | | |
| | SNP-4.28964173. | G/A | 4 | 29149314 | 3.3 | | G | | | |
| | SNP-5.23488784. | G/A | 5 | 23551364 | 3.04 | | G | 4.22 | 9.0 | FE, FM, I, M, PL |
| | SNP-6.22696571. | G/A | 6 | 22697569 | 3.3 | | G | 4.36 | 3.4 | PL |
| | 6749451 | G/A | 6 | 24331932 | | | | 3.65 | 10.0 | FE |
| | 6855107 | G/A | 6 | 27609767 | 3.35 | 52.1 | EC | 3.12 | 4.6 | I, M, PL |
| | SNP-6.30518734. | C/A | 6 | 30519733 | 4.06 | 53.1 | EC | 3.42 | 4.0 | FE, I, M, PK |
| | SNP-10.10050447. | A/G | 10 | 10121627 | 3.38 | 52.1 | EC | | | |
| | 10794778 | A/C | 10 | 22078159 | | | | 3.07 | 5.0 | M |
| | SNP-11.3756217. | A/G | 11 | 3760316 | 3.31 | 52.0 | EC | | | |
| | 13069784 | C/A | 12 | 27023424 | | | | 4.87 | 7.7 | I, PK |
| HRY | 1006459 | A/G | 1 | 30318336 | 3.10- 3.12 | 7.2-22.2 | C, EC | | | |
| | 1163456 | C/A | 1 | 35825579 | | | | 3.07 | 1.1 | PL |
| | 1257104 | A/G | 1 | 39282883 | 3.31 | | G | 3.61 | 7.7 | PK |
| | id2004534 | C/A | 2 | 9441925 | | | | 3.21 | 9.2 | PK |
| | 2462399 | A/G | 2 | 35081420 | 3.04 | 22.0 | EC | 3.07 | 6.2 | FM, M, PL |
| | 2724361 | A/G | 3 | 10184356 | 3.22- 3.33 | 22.7 | EC, G | 5.09 | 9.0 | FE, FM, I, M, PK, PL |
| | id3009515 | G/A | 3 | 20366430 | 3.02 | | G | | | |
| | 3096758 | G/A | 3 | 20774372 | 3.02 | | G | | | |
| | SNP-4.8510513. | C/G | 4 | 8515241 | 3.17 | | G | | | |
| | 4794120 | G/A | 4 | 34935052 | 3.31 | | G | | | |
| | 5711540 | A/G | 5 | 26158316 | 3.07 | | G | | | |
| | | | | | | | | | | |

| Table 3.2. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | id9002494 | G/A | 9 | 8040762 | 3.22 | | G | 3.02 | 9.4 | FE, FM, I, M, PK, PL |
| | 9921984 | A/G | 10 | 650031 | 3.54 | | G | 3.68 | 1.6 | PL |
| | SNP-10.1590493. | G/A | 10 | 1591517 | 3.25-3.94 | 9.4-22.7 | C, EC, G | 4.13 | 16.0 | FE, FM, I, M, PL |
| | 10432540 | G/A | 10 | 11087428 | 3.14 | 22.3 | EC | 4.57 | 12.1 | FM, I, M, PL |
| | 11893268 | C/A | 11 | 26488851 | | | | 4.85 | 7.0 | PL |
| | 12018243 | A/G | 12 | 614213 | | | | 3.72 | 1.9 | PL |
| | id12005992 | G/A | 12 | 17691781 | 3.01 | | G | | | |
| ASV | 31587 | C/A | 1 | 1015503 | 3.12 | | G | | | |
| | 46141 | G/A | 1 | 1483319 | 3.00-3.06 | 6.7 | C, G | | | |
| | 817625 | A/G | 1 | 24108184 | 3.11 | 8.9 | C | 7.3 | 9.2 | M |
| | 924930 | G/A | 1 | 27710094 | 3.16- 3.84 | 39.2 | EC, G | | | |
| | 1099618 | G/A | 1 | 33630310 | 3.26-3.59 | 39.9 | EC, G | | | |
| | 1391852 | A/G | 2 | 855183 | 3.52 | | G | | | |
| | 1517351 | G/A | 2 | 5836334 | 3.38- 3.61 | 39.5 | EC, G | | | |
| | 3460782 | C/A | 3 | 32380470 | 3.36 | | G | 3.82 | 2.9 | M |
| | 4111683 | A/G | 4 | 12333059 | 3.42-3.43 | 54.8 | EC, G | | | |
| | SNP-5.19418236. | A/C | 5 | 19480753 | 3.21 | | G | | | |
| | 5552255 | A/C | 5 | 20612374 | 3.31 | | G | | | |
| | SNP-5.28500625. | C/G | 5 | 28563271 | | | | 4.62 | 0.0 | PL |
| | Waxy-Intron1 | C/A | 6 | 1765761 | | | | 5.44 | 0.9 | PK, PL |
| | ALK-SNP4_FWD | C/A | 6 | 6752887 | 11.25-18.52 | 36.7-38.9 | C, EC, G | 6.48 | 41.5 | FM, M, PK, PL |
| | 6047367 | G/A | 6 | 7490489 | 3.06-3.75 | 39.9 | EC, G | | | |
| | id6006147 | T/A | 6 | 9684814 | 3.63-4.46 | 39.8 | EC, G | 3.85 | 3.4 | FM, M, PK, PL |
| | 6440144 | G/A | 6 | 16837502 | 3.24- 3.58 | 39.1 | EC, G | 8.05 | 6.6 | PK |
| | 7839126 | G/A | 7 | 24485103 | 3.19 | 38.9 | EC | | | |
| | 7912213 | G/A | 7 | 26983927 | 3.69 | | G | | | |

Table 3.2. Continued

| Trait | SNP | Alleles | Chr. | Pos.(bp) | -log10P | R2 | Model | LOD | R2 | Model |
|---|---|---|---|---|---|---|---|---|---|---|
| | SNP-8.15619850. | C/A | 8 | 15622565 | | | | 3.48 | 0.6 | PK |
| | 8983572 | A/C | 8 | 25757847 | 3.08-3.10 | 39.5 | EC, G | | | |
| | 10110880 | C/A | 10 | 4521396 | 3.43-4.21 | 7.9-39.5 | C, EC, G | | | |
| AAC | 176306 | G/A | 1 | 5608892 | 3.23 | | G | | | |
| | 839855 | G/A | 1 | 24733982 | 4.2 | | G | | | |
| | 2084926 | G/A | 2 | 22256108 | | | | 3.2476 | 4.5 | PK |
| | 2120738 | G/A | 2 | 23400020 | 3.01 | | G | | | |
| | 4189864 | G/A | 4 | 14321993 | 3.27 | 50.1 | EC | | | |
| | 4987236 | A/G | 5 | 5789766 | 4.07 | | G | 3.34 | 5.3 | PK |
| | Waxy-Intron1 | C/A | 6 | 1765761 | 9.77-14.93 | 29.8-61.4 | C, EC, G | 21.11 | 26.9 | FE, FM, I, M, |
| | 6527310 | G/A | 6 | 18809305 | 3.00-3.18 | 49.7 | EC, G | | | |
| | 7133760 | A/G | 7 | 6434783 | 3.13-3.68 | 49.9 | EC, G | 3.24 | 3.5 | M |
| | SNP-7.28073407. | C/A | 7 | 28074401 | 3.83-4.18 | 9.2-50.9 | C, EC, G | 3.87 | 28.7 | M |
| | SNP-9.1275036. | G/A | 9 | 1276037 | 3.94 | | G | 5.55 | 9.6 | M, PL |
| | id9003198 | C/A | 9 | 11871485 | 3.11 | | G | | | |
| | SNP-10.22600875. | C/A | 10 | 22672402 | 3.25-3.89 | 7.6-50.1 | C, EC, G | 4.75 | 6.6 | FM, M, PL |
| | SNP-11.10843784. | G/A | 11 | 10849468 | 3.22 | 50.0 | EC | | | |
| | SNP-11.26742377. | T/A | 11 | 27213991 | 3.01 | 8.7 | C | | | |
| | 12993236 | G/A | 12 | 24417433 | 3.16 | | G | | | |
| PC | 3598944 | G/A | 4 | 648669 | 3.06-3.33 | 9.0 | C, G | | | |
| | 5432007 | A/G | 5 | 17078792 | 3.59 | | G | | | |
| | SNP-5.25374745. | A/G | 5 | 25437325 | 3.82-5.59 | 11.3-33.8 | C, EC, G | 5.37 | 42.0 | FE, FM, I, M, PL |
| | 5946748 | A/C | 6 | 4499092 | 3.04-3.12 | 30.5 | EC, G | | | |
| | SNP-6.7907666. | A/G | 6 | 7908666 | 3.9 | 11.6 | C | 3.06 | 2.9 | FE, FM |
| | 6274558 | C/A | 6 | 13250266 | 3.97-5.20 | 32.3 | EC, G | 4.82 | 17.8 | FE, FM, I, M, PK, PL |
| | 6526839 | G/A | 6 | 18798208 | 3.09-3.58 | 30.4 | EC, G | | | |

| Table 3.2. Continued | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Alleles** | **Chr.** | **Pos.(bp)** | **-log10P** | **R2** | **Model** | **LOD** | **R2** | **Model** |
| | SNP-6.22145120. | A/G | 6 | 22146118 | 3.21 | | G | | | |
| | SNP-7.2636123. | A/G | 7 | 2637123 | 3.02-3.05 | 30.3 | EC, G | | | |
| | 7173448 | G/A | 7 | 7503778 | | | | 4.29 | 6.6 | FM, I, M |
| | 7909791 | A/G | 7 | 26860306 | 3.40-3.52 | 31.1 | EC, G | | | |
| | 10732011 | A/G | 10 | 19780519 | 3.25 | 30.7 | EC | 4.11 | 9.0 | FE, FM, PK |
| | 11345455 | G/A | 11 | 14741103 | 3.09 | | G | | | |

N.B: Single-locus models: C-CMLM, EC-ECMLM, G- GEMMA; Multi-locus models: M- MrMLM, FM- FASTmrMLM, FE- FASTmrEMMA, PK- pKWmEB, PL- pLARmEB, I- ISIS EM-BLASSO

for length, width and GLWR, SNP 6274558 for width, GLWR and PC and SNP 2724361 for width, MRY and HRY (Table 3.2).

To compare the performance of the models in terms SNP detection, single-locus models detected (shared SNPs were included) a total of 215 SNPs including 53, 60 and 102 SNPs were identified by CMLM, ECMLM and GEMMA, respectively. Total multi-locus models found 226 SNPs (MrMLM=52, FASTMrMLM= 36, FASTmrEMMA= 21, PKWmEB= 35, PLARmEB= 43 and ISIS-EB-BL= 39) including SNPs also found with single-locus models (Figure 3.4).

### 3.3.3. In-Silico Gene Expression Analysis

After mining the genes within 250-kb region of the significant SNPs for all the traits related to grain appearance and milling quality, eating quality and nutritional quality, using RAP-DB database (https://rapdb.dna.affrc.go.jp/), we found 363 genes for length, 849 genes for width, 740 genes for GLWR, 324 genes for DEC, 445 genes for PGWC, 479 genes for MRY, 396 genes for HRY, 498 genes for ASV, 289 genes for AAC, and 226 genes for PC, respectively. To investigate which genes are responsible for the traits, we selected only those genes that are expressed at the reproductive stage of rice plant while being expressed in other vegetative stages by using Nipponbare gene expression data in normalized FPKM values. After filtering the non-expressed genes in reproductive stages, 124, 287, 229, 39, 132, 128, 108, 140, 61 and 81 genes were found for length, width, GLWR, DEC, PGWC, MRY, HRY, ASV, AAC and PC, respectively, and will be used for further analysis (Figure 3.6).

Figure 3.6. Heatmap of In-silico gene expression analysis results for the ten measured traits of the study.

### 3.4. Discussion

#### 3.4.1. Population Structure, LD, And Phenotypic Variation

The structure analysis revealed six sub-populations that belong to two major sub-populations, *indica* and *japonica*. This finding is consistent with the previous studies using worldwide rice germplasm (Morales et al., 2020; F. Xu et al., 2016). The LD decay pattern of this study with having 250 kb is also supported by the previous studies (Mather et al., 2007; Qiu et al., 2015).

Great variation was observed for all traits, indicating the possibility of applying association studies in these rice accessions. This study used an NIR machine to measure protein content from brown rice and the range of our protein content (8.70-17.40%) is consistent with the previous studies with 4.3–18.2% for diverse rice germplasm where a chemical assay was used (Bryant et al., 2013). Analysis of variance (ANOVA) analysis shows that seven of 10 traits are strongly affected by population structure and sufficiently explain the phenotypic variation of the affected traits.

It is well known fact that a complex relationship exists among grain appearance traits. This study showed a strong positive correlation between GL and GLWR, but both have a negative correlation with GW, which was consistent with the previous studies (Qiu et al., 2015; X. Wang et al., 2016). We found DEC and PGWC were correlated with each other strongly and both have a positive correlation with GW but negative with GL and GLWR. These findings were also supported by the previous studies (Qiu et al., 2015; X. Wang et al., 2016). Both DEC and PGWC were found to be negatively correlated with MRY and HRY and similar results were found in the

past studies (Qiu et al., 2015; Zheng, Xu, Li, Zhai, & Wan, 2007). These findings support the fact that chalky grains are more prone to breakage at the chalky area than the non-chalky part. A low negative correlation was found between AAC and GT in our study, but X. Wang et al. (2017) reported a medium negative correlation between AAC and GT for two environments. We reported a weak, negative correlation for PC with both AAC and GT, which is, to some extent, different from the result of X. Wang et al. (2017), where they found a low positive correlation between PC and GT and low positive for one environment and low negative for another environment correlation was found between PC and AAC. This discrepancy may be owing to using different germplasm in the respective study and methods used to determine the AAC content and GT because they used near-infrared spectroscopy (NIRS) where this study used colorimetric and ASV methods to determine AAC and GT, respectively.

From the previous studies, it is well established that grain appearance traits are complex, and they are correlated to each other with having either positive or negative correlation (Qiu et al., 2015). This fact is supported by our SNP detection results, where pleiotropic SNPs have been observed for grain appearance traits. Grain length, width and GLWR are interrelated (Qiu et al., 2015; X. Wang et al., 2016). We found one SNP (GS3) that is common in three traits, along with eight SNPs were observed for among the SNPs controlling these genes. Similarly, the previous studies found the correlation among DEC, PGWC and GLWR and this is also supported by our results with finding seven pleiotropic SNPs (X. Wang et al., 2016). HRR was found to be correlated with DEC and PGWC by the past studies but we have not found any common SNPs for these traits in our GWAS study (Qiu et al., 2015; Zheng et al., 2007). We found one common SNP for MRY

Table 3.3. Comparison of the GWAS result with the previous studies.

| Trait | SNP | Position (bp) | Chr. | Markers linked/associated | Types of Markers | Known genes/QTLs | Position (bp) | References |
|---|---|---|---|---|---|---|---|---|
| | c1p40455715 | 40330716 | 1 | | | grl1-1 | 38895210-40167979 | Amarawathi et al. (2007) |
| Length | GS3 | 16614917 | 3 | S3_16785761 | SNP | GS3, qGL-3 | 16785761 | Wang et al. (2017), Wan et al. (2005) |
| | | | 3 | S03_16663793 | SNP | GS3, qGL-3 | 16663793 | Qiu et al. (2015), Wan et al. (2005) |
| | | | 3 | S03_16731182 | SNP | GS3, qGL-3 | 16731182 | Qiu et al. (2015), Wan et al. (2005) |
| | id3008418 | 17019705 | 3 | S3_16883926 | SNP | qGL-3a, qGL-3 | 16883926 | Wang et al. (2017), Wan et al. (2006), Wan et al. (2005) |
| | | | 3 | S03_16996600 | SNP | qGL-3a, qGL-3 | 16996600 | Qiu et al. (2015), Wan et al. (2006), Wan et al. (2005) |
| | | | 3 | S03_17000111 | SNP | qGL-3a, qGL-3 | 17000111 | Qiu et al. (2015), Wan et al. (2006), Wan et al. (2005) |
| | id4006172 | 20396474 | 4 | S4_20297417 | SNP | GIF1 | 20297417 | Wang et al. (2016), Wang et al. (2008) |
| | 10877755 | 2208347 | 11 | S11_2576141 | SNP | CycT1;3 | 2576141 | Qiu et al. (2015) |
| Width | GS3 | 16733441 | 3 | | | GS3 | 16729501-16735109 | Fan et al. (2006) |
| | 4970110 | 5338205 | 5 | S5_5369802 | SNP | GW5 | 5369802 | Wang et al. (2016) |
| | | | 5 | S5_5459847 | SNP | GW5 | 5459847 | Wang et al. (2016) |
| | | | 5 | S05_5368086 | SNP | GW5 | 5368086 | Qiu et al. (2015) |
| | | | 5 | S05_5368151 | SNP | GW5 | 5368151 | Qiu et al. (2015) |
| | | | 5 | S05_5369527 | SNP | GW5 | 5369527 | Qiu et al. (2015) |
| | id8000555 | 1995144 | 8 | | | OsFIE1 | 2095644-2100604 | Folsom et al. (2014) |
| | | | 8 | | | OsFIE2 | 2077234-2083272 | Nallamilli et al. (2013) |

Table 3.3. Continued

| Trait | SNP | Position (bp) | Chr. | Markers linked/associated | Types of Markers | Known genes/QTLs | Position (bp) | References |
|---|---|---|---|---|---|---|---|---|
| GLWR | SNP-1.5867020. | 5868021 | 1 | S01_5811755 | SNP | | 5811755 | Qiu et al. (2015) |
| | SNP-1.25001551. | 25002596 | 1 | | | DIF1 | 25009514-25012233 | Cai et al. (2014) |
| | GS3 | 16733441 | 3 | S3_16785761 | SNP | GS3, qLWR-3 | 16785761 | Wang et al. (2016), Wan et al. (2005) |
| | | | 3 | S03_16858510 | SNP | GS3, qLWR-3 | 16858510 | Qiu et al. (2015), Wan et al. (2005) |
| | 2894214 | 15709424 | 3 | | | qLWR-3 | 15645609-24600376 | Wan et al. (2005) |
| | SNP-4.20233911. | 20405868 | 4 | | | GIF1 | 20422171-20426921 | Wang et al. (2008) |
| | 10877755 | 2208347 | 11 | S11_2576141 | SNP | CycT1;3 | 2576141 | Qiu et al. (2015) |
| AAC | Waxy-Intron1 | 1765761 | 6 | No name | SNP | Wx, qAC-6 | 1765761 | Xu et al. (2016), Li et al. (2003) |
| | | | 6 | No name | SNP | | 1529682 | Xu et al. (2016) |
| | | | 6 | No name | SNP | | 1585864 | Xu et al. (2016) |
| | | | 6 | S6_1746440 | SNP | Wx | 1746440 | Wang et al. (2017) |
| | SNP-9.1275036. | 1276037 | 9 | RM4413/Os09ssr0006900*, RM5122/Os09ssr0098400* | SSR | | 1173589-15249042 | Wada et al. (2006) |
| GT | ALK-SNP4_FWD | 6752887 | 6 | S6_6752888 | SNP | ALK, SSIIa, alk6-1, asv6-1 | 6752888 | Wang et al. (2017), Zhang et al. (2011), Aluko et al. (2004), Amarawathi et al. (2007) |
| | 7912213 | 26983927 | 7 | S7_27788464 | SNP | | 27788464 | Wang et al. (2017) |
| | SNP-8.15619850. | 15622565 | 8 | | | OsAGPS2b | 15666336-15672583 | Tuncel et al. (2014) |
| | 1391852 | 855183 | 2 | | | BiP1 | 838743-842672 | Wakasa et al. (2011) |

| Table 3.3. Continued | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Trait** | **SNP** | **Position (bp)** | **Chr.** | **Markers linked/ associated** | **Position (bp)** | **Types of Markers** | **Known genes** | **References** |
| | Waxy-Intron1 | 1765761 | 6 | | | Wx, alk6-1 | 1765622-1770653 | Wang et al. (1995), Aluko et al. (2004) |
| | 6047367 | 7490489 | 6 | | | asv6-1, qGT-6 | 4235101-27253355 | Amarawathi et al. (2007), Tian et al. (2005) |
| | id6006147 | 9684814 | 6 | | | asv6-1, qGT-6 | 4235101-27253355 | Amarawathi et al. (2007), Tian et al. (2005) |
| | 6440144 | 16837502 | 6 | | | asv6-1 | 4235101-27253355 | Amarawathi et al. (2007) |
| PC | SNP-7.2636123. | 2637123 | 7 | RM427 | SSR | | 2711812-2712005 | Bryant et al. (2013) |
| RS | Waxy-Intron1 | 1765761 | 6 | chr06_1765761 | SNP | Wx | 1765761 | Bao et al. (2017) |
| DEC | 8149888 | 5284717 | 8 | | | SSIIIa | 5352105-5363276 | Zhang et al. (2011) |
| MRY | SNP-4.28964173. | 29149314 | 4 | | | xiao | 29084153-29086526 | Jiang et al. (2012) |
| | 13069784 | 27023424 | 12 | | | gpa1 | 27022013-27025203 | Wang et al. (2010) |

and HRY that confirms the correlation between them found by the previous studies (X. Wang et al., 2016).

### 3.4.2. Performance Of Single And Multi-Locus GWAS Models

To detect SNPs for controlling ten traits, the multivariate or multi-locus models found more SNPs (226) than single-locus or univariate GWAS models (215). This performance is expected with respect to the reports published in the past, mentioning multivariate or multi-locus models are more powerful and robust than the single-locus methods (C. Li et al., 2018; Y. Xu et al., 2018; Y. M. Zhang, Jia, & Dunwell, 2019). Within the single-locus models, GEMMA found the highest number of SNPs (102), followed by ECMLM (60) and CMLM (53), and this performance is consistent with the previous report (M. Li et al., 2014). Though multivariate models performed well in comparison to single-locus models, none of these methods identified all the previously known SNPs or QTLs. So, the previous studies recommended combining the single-locus methods and/or multi-locus methods to improve the detection power and robustness of GWAS (He et al., 2019; C. Li et al., 2018; M. Li et al., 2014; Liu et al., 2020).

### 3.4.3. Comparison And Reliability Of Our GWAS Studies

The current GWAS analysis identified 216 significant SNPs for ten traits related to rice grain appearance, milling, eating and nutritional quality, which were then compared with the genes/QTLs and markers for the same traits identified using linkage mapping and association mapping in previous studies. The surrounding 250 kb of significant associated SNPs were regarded as potentially co-located loci for any particular trait when this region was between the borders of mapping QTLs, known genes, or flanking SSR, RFLP, and SNP markers of previous studies.

Therefore, 29 (~ 13%) of the 216 significant SNPs of this study co-located with 26 significant SNP, SSR and RFLP markers, 14 known genes, and 9 QTLs reported in the previous studies, and the remaining 187 (87%) significant SNPs could be considered as novel loci detected only in the current study. Out of 29 coincided SNPs, the highest number (8) of SNPs were associated with gelatinization temperature trait, followed by six SNPs for GLWR, five for length trait. SNPs from MRY (2), DEC (1), and PC (1) traits shared fewer matches with the previously reported markers. For PGWC and HRY, no common SNP was found with the past studies (Table 3.3).

Since grain length, width, and length-to-width ratio (grain shape) have high stability and are highly heritable; many QTLs have been previously identified for those traits. Some of them already have been fine mapped and cloned; some of them have pleiotropic effects, controlling multiple grain related traits simultaneously. For grain length, chromosome 3 harbors more QTLs than other chromosomes (Bao, 2014). The current study also identified more SNPs in chromosome 3 than other chromosomes. GS3 is the first major QTL for grain length that has been cloned. Our GWAS study identified the GS3 SNP at the position of GS3 gene, confirming its major effect on rice grain length. This is expected, as the gene-based GS3 SNP was previously designed to tag the GS3 gene (Morales et al., 2020). SNPs markers from the other studies were also reported at the same position (Qiu et al., 2015; X. Wang et al., 2016). GS3 is well known for its pleiotropic effect, having a major effect on length and weight and a minor role for width and thickness. This fact is also confirmed by our study with detecting the GS3 SNP at the GS3 gene location on chromosome 3 for both width and GLWR traits. Similarly, qGL3 on chromosome 3 is another major QTL for grain length, encoding a putative protein phosphatase with Kelch-like repeat domain (OsPPKL1) (Bao, 2014; Wan et al., 2005). The current study also identified the id3008418 SNP at the location

of qGL3 region. Beside these major gene/QTLs for length, this study identified several SNPs on other chromosomes close to known gene positions regulating grain shape. For example, SNP id4006172 on chromosome 4 in our study was mapped near to the GIF1 gene position, whose function is to regulate grain filling and grain size. This gene is also partially responsible for chalkiness (Bao, 2014; E. Wang et al., 2008). X. Wang et al. (2016) also reported a SNP at the same gene location. Similarly, SNP 10877755 on chromosome 11 of this study was positioned at the CycT1;3 gene location that also affects grain size. A SNP was also found close to this gene by Qiu et al. (2015). Since the functions of both GIF1 and CycT1;3 genes are related to grain size, both were identified for GLWR traits, as expected, by identifying SNP-4.20233911. and 10877755 SNP on chromosome 4 and 11, respectively, that were located close to those genes. For grain width, our GWAS analysis found GS3, 4970110, and id8000555 SNPs that were closed to GS3, GW5, OsFIE1 and OsFIE2 genes, respectively. GW5 is the major gene for grain width on chromosome 5 that influences grain width and weight negatively (Weng et al., 2008). Several SNPs markers from the past studies were reported at this gene location, confirming its major effect on grain width (Qiu et al., 2015; X. Wang et al., 2016). On chromosome 8, OsFIE1 and OsFIE2 genes were identified within a 250 kb region of id8000555 SNP whose functions are related to grain size, width, weight, grain filling rate, and seed dormancy. No markers have been reported yet at this location, indicating a novel QTL of grain width. In the case of GLWR trait, along with GS3, GIF1, and CycT1;3 (described earlier), the 2894214 SNP was found in this study where previously qLWR-3 QTL was reported by Wan et al. (2005). Chalkiness can be caused by both environment and genetics. Genetically, genes involved in starch biosynthesis, starch granule structure, and grain filling, including but are not limited to starch branching enzyme IIb (BEIIb), branching enzyme IIb (BEIIb), starch synthase IIIa (SSIIIa), floury and sugary genes, have been

reported as being a genetic reason for rice grain chalkiness (Bao, 2014). Our study observed a single SNP (8149888) on chromosome 8 associated with the DEC trait that was close to the FLO5 gene, which has been implicated for white-core floury endosperm (Ryoo et al., 2007). qPGWC-8 and qPGWC-7 QTL were previously reported as being a major QTL for PGWC trait. Unfortunately, we did not identify any significant SNP on chromosome 7 in our study for PGWC. But we identified two SNPs on chromosome 8 on two different positions for PGWC trait, indicating the novel QTLs for PGWC (Table 3.3).

The genetic control for milling quality is still comparatively less understood. No map-based cloning and fine mapping of milling quality related genes have been reported yet. Many QTLs were found in past studies without consistent results (Bao, 2014). However, our study results for MRY and HRY were consistent with the previous studies in terms of finding associated markers across all twelve chromosomes. Most importantly, we detected SNP-4.28964173. and 13069784 SNP markers within a 250 kb boundary of the xiao and gpa1 genes on chromosome 4 and 12, respectively. While the xiao gene was reported to be involved in dwarfism, grain size, leaf angle, fertility, and cell division, the gpa1 gene was reported as being involved in pro-glutelin content in seed and floury endosperm (Jiang et al., 2012; Y. Wang et al., 2010). These findings are consistent with the fact that rice milling yield is largely influenced by grain size and chalkiness, which could be the reason for being associated with MRY trait (Table 3.3).

The eating and cooking quality of rice is largely influenced by starch properties which are controlled by genes involved in the starch synthesis pathway. Among these, Wx and ALK are the most influential genes governing AAC and GT, respectively. Other starch synthesis related genes,

such as AGPlar, BEI, GBSSII, GPT1, ISA2, PUL, SSI, SSIIb, SSIIc, SSIIIa, SSIIIb and SSIVa have minor effects on eating and cooking quality (Bao, 2014; X. Wang et al., 2017). Wx gene encoding GBSSI, which is the major enzyme responsible for amylose synthesis, ALK or SSIIa gene is mainly for GT, thermal properties and amylopectin structure (Bao, 2014). Our GWAS study result is consistent with these facts by identifying significant SNPs close to the position of these genes. For example, for the AAC trait, the gene-based Waxy-Intron1 SNP on chromosome 6 was found that is exactly located in the Wx gene. Similarly, in the case of GT, we found a SNP (ALK-SNP4_FWD) on chromosome 6 located close to ALK gene. Besides this gene, we also detected some other genes that may have a minor effect on GT. OsAGPS2b and BiP1 genes were found close to SNP-815619850 and 1391852 on chromosome 8 and 2, respectively. OsAGPS2b and BiP1 genes have been demonstrated to be involved in starch synthesis in the middle to late stage of developing endosperm, and seed storage protein and starch content, respectively (Tuncel et al., 2014; Wakasa et al., 2011). The current study also identified new QTL positions for AAC and GT, where the past studies found markers at the same location without having any known genes. For instance, on chromosome 9, SNP-9.1275036. was associated with AAC where previously an SSR marker was reported by Wada, Uchimura, Ogata, Tsubone, and Matsue (2006). Similarly, in the case of GT, 7912213 SNP on chromosome 7 was found where X. Wang et al. (2017) also reported a SNP, indicating a possible new QTL for GT (Table 3.3).

For protein content, the past studies reported more QTLs on chromosome 1, 2 and 7. This study did not find any SNP on chromosome 1 and 2 but detected a significant SNP on chromosome 7. On chromosome 7, SNP-7.2636123. was identified where an SSR marker was also reported by Bryant et al. (2013) (Table 3.3).

Overall, the current study mapped many major genes/QTLs for the measured traits, although some major genes/QTLs related to grain appearance qualities were not identified in our study. For example, SNPs co-localizing with GW2, GS5, GW8 were not detected for grain size. Similarly, SNPs co-localizing with Chalk5, qPGWC-8, and qPGWC-7 genes/QTLs were also not found in our study. This may be due to the genetic makeup of our diversity panel or the relatively small sample size of 174 samples. Since 14 known major genes with known functions controlling the measured traits used in this study have been rediscovered by our GWAS study, we can confirm the accuracy of our GWAS results. We also found 187 new loci, implying that many more genes are yet to be discovered that may have an effect on rice grain appearance, milling, eating and nutritional qualities. Most importantly, these common 28 loci found across different populations with different genetic backgrounds can be used for future studies for gene functional characterization and validation.

### 3.5. Conclusion

This association study used a global diversity panel consisting of 174 rice accessions on 6,565 SNPs to detect significant SNPs controlling rice grain appearance, milling, eating and nutritional quality traits. To do so, univariate and multivariate GWAS methods were used, and a total of 216 significant SNPs, or GWAS QTLs, were detected. Among these SNPs, single-locus methods alone and multi-locus methods alone detected 106 and 64 SNPs, respectively, while 46 SNPs were identified by both methods simultaneously. While our analysis got 29 verified SNPs with the previously reported genes/QTLs and marker, 188 novel QTLs were discovered by this study. A total of 4,609 genes were mined within a 250 kb region of these SNPs and 1,329 genes, including

previously known genes, were expressed during in-silico analysis, indicating their candidacy for these ten traits. This information could be useful for future studies for expression analysis using NGS technology, followed by functional gene characterization using CRISPR/Cas9 gene editing, to help characterize the molecular mechanisms underlying these traits and ultimately to accelerate rice breeding efforts.

### 3.6. References


Bao, J. (2014). Genes and QTLs for Rice Grain Quality Improvement. In *Rice - Germplasm, Genetics and Improvement*.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308

Bryant, R. J., Jackson, A. K., Yeater, K. M., Yan, W. G., McClung, A. M., & Fjellstrom, R. G. (2013). Genetic variation and association mapping of protein concentration in brown rice using a diverse rice germplasm collection. *Cereal chemistry, 90*(5), 445-452.

Cui, Y., Zhang, F., & Zhou, Y. (2018). The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice. *Front Plant Sci, 9*, 1464. doi:10.3389/fpls.2018.01464

GIPSA. (2009). United States standards for rice.

He, L., Xiao, J., Rashid, K. Y., Yao, Z., Li, P., Jia, G., . . . You, F. M. (2019). Genome-Wide Association Studies for Pasmo Resistance in Flax (Linum usitatissimum L.). *Frontiers in plant science, 9*, 1982. Retrieved from https://www.frontiersin.org/article/10.3389/fpls.2018.01982

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., . . . Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet, 42*(11), 961-967. doi:10.1038/ng.695

Jiang, Y., Bao, L., Jeong, S. Y., Kim, S. K., Xu, C., Li, X., & Zhang, Q. (2012). XIAO is involved in the control of organ size by contributing to the regulation of signaling and homeostasis of brassinosteroids and cell cycling in rice. *The Plant Journal, 70*(3), 398-408.

Li, C., Fu, Y., Sun, R., Wang, Y., & Wang, Q. (2018). Single-Locus and Multi-Locus Genome-Wide Association Studies in the Genetic Dissection of Fiber Quality Traits in Upland Cotton (Gossypium hirsutum L.). *Front Plant Sci, 9*, 1083. doi:10.3389/fpls.2018.01083

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y.-M., Todhunter, R. J., . . . Zhang, Z. (2014). Enrichment of statistical power for genome-wide association studies. *BMC biology, 12*(1), 1-10.

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

Liu, S., Zhong, H., Meng, X., Sun, T., Li, Y., Pinson, S. R. M., . . . Peng, Z. (2020). Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice

and comparative analyses of different mapping methods. *BMC Plant Biol, 20*(1), 441. doi:10.1186/s12870-020-02603-0

Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., & Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (Oryza sativa L.). *Genetics, 177*(4), 2223-2232. doi:10.1534/genetics.107.079616

Mezmouk, S., Dubreuil, P., Bosio, M., Decousset, L., Charcosset, A., Praud, S., & Mangin, B. (2011). Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theor Appl Genet, 122*(6), 1149-1160. doi:10.1007/s00122-010-1519-y

Morales, K. Y., Singh, N., Perez, F. A., Ignacio, J. C., Thapa, R., Arbelaez, J. D., . . . Thomson, M. J. (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS One, 15*(5), e0232479. doi:10.1371/journal.pone.0232479

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., . . . Ye, G. (2015). Genome-Wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of Indica Rice Germplasm. *PLoS One, 10*(12), e0145577. doi:10.1371/journal.pone.0145577

Ryoo, N., Yu, C., Park, C.-S., Baik, M.-Y., Park, I. M., Cho, M.-H., . . . Jeon, J.-S. (2007). Knockout of a starch synthase gene OsSSIIIa/Flo5 causes white-core floury endosperm in rice (Oryza sativa L.). *Plant cell reports, 26*(7), 1083-1095.

Tuncel, A., Kawaguchi, J., Ihara, Y., Matsusaka, H., Nishi, A., Nakamura, T., . . . Cakir, B. (2014). The rice endosperm ADP-glucose pyrophosphorylase large subunit is essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. *Plant and Cell Physiology, 55*(6), 1169-1183.

Wada, T., Uchimura, Y., Ogata, T., Tsubone, M., & Matsue, Y. (2006). Mapping of QTLs for physicochemical properties in japonica rice. *Breeding Science, 56*(3), 253-260.

Wakasa, Y., Yasuda, H., Oono, Y., Kawakatsu, T., Hirose, S., Takahashi, H., . . . Takaiwa, F. (2011). Expression of ER quality control-related genes in response to changes in BiP1 levels in developing rice endosperm. *The Plant Journal, 65*(5), 675-689.

Wan, X. Y., Wan, J. M., Weng, J. F., Jiang, L., Bi, J. C., Wang, C. M., & Zhai, H. Q. (2005). Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theoretical and Applied Genetics, 110*(7), 1334-1346.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., . . . Lin, H. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nature genetics, 40*(11), 1370-1374.

Wang, X., Pang, Y., Wang, C., Chen, K., Zhu, Y., Shen, C., . . . Li, Z. (2016). New Candidate Genes Affecting Rice Grain Appearance and Milling Quality Detected by Genome-Wide and Gene-Based Association Analyses. *Front Plant Sci, 7*, 1998. doi:10.3389/fpls.2016.01998

Wang, X., Pang, Y., Zhang, J., Wu, Z., Chen, K., Ali, J., . . . Li, Z. (2017). Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. *Sci Rep, 7*(1), 17203. doi:10.1038/s41598-017-17347-5

Wang, Y., Ren, Y., Liu, X., Jiang, L., Chen, L., Han, X., . . . Lv, J. (2010). OsRab5a regulates endomembrane organization and storage protein trafficking in rice endosperm cells. *The Plant Journal, 64*(5), 812-824.

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., . . . Guo, X. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell research, 18*(12), 1199-1209.

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

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., . . . Xu, C. (2018). Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. *Front Plant Sci, 9*, 1311. doi:10.3389/fpls.2018.01311

Zhang, Y.-W., Lwaka Tamba, C., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., . . . Zhang, Y.-M. (2020). mrMLM v4.0: An R Platform for Multi-locus Genome-wide Association Studies. *Genomics, Proteomics & Bioinformatics*. doi:https://doi.org/10.1016/j.gpb.2020.06.006

Zhang, Y. M., Jia, Z., & Dunwell, J. M. (2019). Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Front Plant Sci, 10*, 100. doi:10.3389/fpls.2019.00100

Zheng, T. Q., Xu, J. L., Li, Z. K., Zhai, H. Q., & Wan, J. M. (2007). Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (Oryza sativa L.). *Plant breeding, 126*(2), 158-163.

# 4. GENOME-WIDE ASSOCIATION MAPPING FOR RESISTANT STARCH OF COOKED RICE USING DIVERSE GERMPLASM AND CHARACTERIZING THE RELATIONSHIP OF RESISTANT STARCH WITH APPARENT AMYLOSE CONTENT, PROTEIN CONTENT, CHALKINESS, AND GELATINIZATION TEMPERATURE

## 4.1. Introduction

Being a staple food for half of the world's population, the nutritional quality of rice has a direct impact on human health. For many areas in the world, rice is the main source of dietary carbohydrates, including starch. Nutritionists have explored the digestibility of starch in terms of human health (Englyst, Kingman, & Cummings, 1992). Three types of starch have been found in terms of digestion, namely rapidly digestible starch, slowly digestible starch (SDS), and resistant starch (RS) (Bao, Zhou, Xu, He, & Park, 2017). RS is the part of starch that is not digestible by pancreatic amylase of the human body, making it available for fermentation by microbiota in the colon (Chen, Bergman, McClung, Everette, & Tabien, 2017). It has multiple health benefits, such as increasing adiposity and insulin resistance, gut health, decreasing cardiovascular disease risk factors, and lowering the risk of colon cancer.

In general, rice has a low amount of resistant starch. From the health point of view, further improvement of RS content in cooked rice would be beneficial (Chen et al., 2017). So far, 0.6-1.21% of RS content has been reported in wild rice accessions and 2.33- 4.46% of RS content from cultivated rice varieties has been reported using the AOAC Method 2002.02 (Butardo et al., 2012; Chen et al., 2017). While most of the studies focus on the starch structure and its digestive properties, little attention has been paid to the genetic basis of RS content. Wx, starch synthase

IIIa (SSIIIa), isoamylase 2 (ISA2) have been reported in past studies as influencing RS content. However, few studies have been reported using GWAS to detect QTLs controlling RS. So, it is important to perform a genome-wide association study to identify SNPs that control RS content in rice.

Association mapping is a powerful approach to dissect the genetic basis of complex traits. In contrast to the linkage mapping, association studies can associate genotypes with phenotypes in natural populations and detect many natural allelic variants in a single study (C. Li, Fu, Sun, Wang, & Wang, 2018). GWAS has been applied to detect QTLs for many traits in rice so far (Huang et al., 2010; Qiu et al., 2015; Yang et al., 2018).

Initially, single-locus-based GWAS models, such as GLM and MLM (Bradbury et al., 2007) were used in most GWAS publications. But because of multiple test correction issues, along with the lack of ability to detect multiple loci simultaneously, as an alternative approach, multiple-locus models have been developed and applied. These multivariate models consider all loci simultaneously; as a result, multiple test corrections are not needed. So far, several multi-locus GWAS models have been developed and used to study GWAS. All the multi-locus models follow the two-step principle during analysis. In the first stage, all the potentially associated SNPs are identified or scanned in the whole genome. During the second step, all the identified SNPs are included in one model, then their effects are estimated by empirical Bayes, and finally all the non-zero effects are further evaluated using the likelihood ratio test. A less stringent critical p-value, such as 0.01, is used to select the SNPs in the first step. Each of these multi-locus modes is different

from each other in terms of algorithms utilized in the two steps (Cui, Zhang, & Zhou, 2018; C. Li et al., 2018; Y. Xu et al., 2018).

The main objective of this study is 1) to identify all possible SNPs controlling RS content in rice grains by using single-locus and multi-locus GWAS methods, 2) to compare the performance of the two methods in terms of SNP detection ability, 3) to determine the relationship between RS content and other traits related to rice appearance, eating and nutritional quality.

## 4.2. Materials And Methods

### 4.2.1. Plant Materials And Sample Preparation

A worldwide collection of 151 diverse accessions, obtained from the USDA GRIN collection, along with 23 US-released varieties, in total 174 accessions, were used in this study. All of the accessions were selected from a larger panel to have a similar heading date to avoid the effect of flowering time on grain quality. The field experiment was conducted in Texas A&M AgriLife Research Center, Beaumont, Texas, in 2018. The details of the field experiment design and sample preparation for this study were described in Chapter 2.2.1 and 2.2.2.

### 4.2.2. Phenotypic Measurements

The RS concentration in cooked rice was measured based on the AOAC Method 2002.02 (Horwitz & Latimer, 2005) by using the RS Assay kit from Megazyme (K-RSTAR 08/11, Wicklow, Ireland). Sample preparation to the determination of RS was performed according to Chen et al. (2017) protocol.

### 4.2.3. Analysis Of Phenotypic Data

Analysis of variance (ANOVA) was conducted to determine the population structure effect on RS phenotypic variation. Also, the correlation between RS and AAC was evaluated. Multiple regression analysis was performed to know the relationship among RS, AAC, length, width, GLWR, DEC, PGWC and PC. The backward and forward stepwise selection method was used with the criteria using P-value thresholds of 0.25 to enter the model and 0.15 to leave the model. All the predictors in both models had $p < 0.05$. All the analyses were conducted in JMP Pro 15.

### 4.2.4. GWAS Analysis

A total of 6,565 high quality SNPs from the 7K SNP array data (Morales et al., 2020) were used for the GWAS analysis. Imputation was carried out to infer the untyped markers by MACH 1.0. population structure (Q), kinship analysis (K), and LD analysis were conducted for these SNP markers. The detailed procedure was described in Chapter 2, section 2.5. Single-locus models including CMLM, ECMLM and GEMMA and multi-locus models including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO were used to study the genome-wide associations. CMLM, ECMLM and GEMMA models were implemented in TASSEL 5, GAPIT v.3 and GEMMA, respectively, whereas all the multi-locus models were implemented in mrMLM R package (Y.-W. Zhang et al., 2020). p <10-3 and LOD= 3.0 were set as critical values to declare significant SNPs and QTNs in the single-locus and multi-locus models, respectively (Figure 2.2).

### 4.2.5. *In-Silico* Gene Expression Analysis

*In-silico* gene expression analysis was conducted for the mined genes around the significant SNPs. Nipponbare gene expression data downloaded from the MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/expression.shtml) was used for this analysis.

## 4.3. Results

### 4.3.1. Phenotypic Variation Analysis

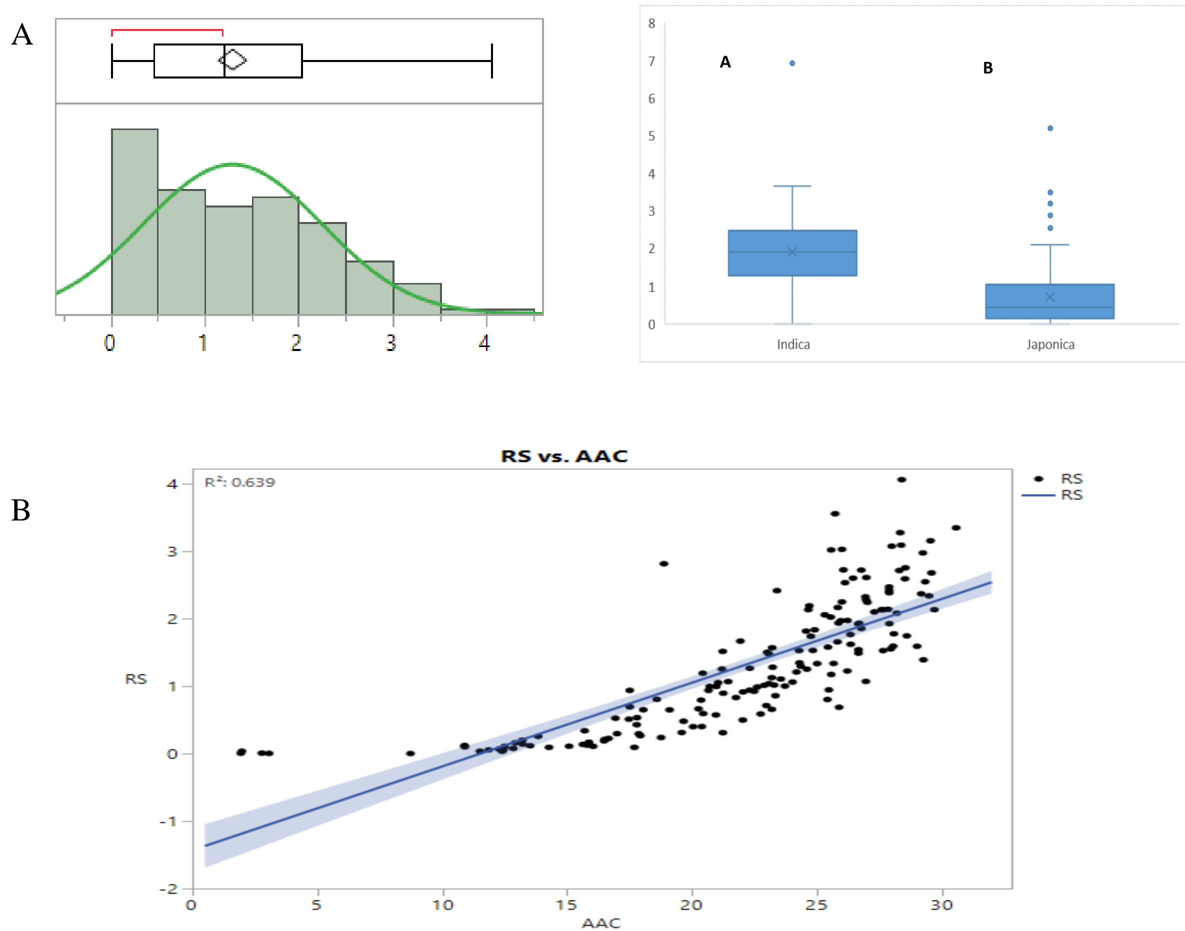



Figure 4.1. (A) Phenotypic variation of resistant starch (RS) of rice grain. Different capital letters in the same box plot indicate Indica and Japonica rice accessions are significantly different at α= 0.05 for mean value of the six mineral elements. (B) Correlation between resistant starch (RS) and apparent amylose content (AAC).

The frequency distribution of RS was normally distributed. Population structure explained 33% of the phenotypic variation and the mean difference between *indica* and *japonica* panel was found significant (Figure 4.1; Table 4.1).

Table 4.1. Phenotypic variation in whole, *Indica* and *Japonica* rice accessions

| **Traits** | **Whole Panel** | | | | | | | **Indica** | | **Japonica** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Mean** | **SD** | Min | Max | **CV** | **R2** | **P** | Mean | SD | Mean | SD |
| RS | 1.36 | 1.07 | 0 | 6.92 | 79.17 | 0.33 | <.0001 | 1.92* | 0.95 | 0.70 | 0.81 |

N.B: * indicate the significant at α= 0.05 level

Pairwise correlation revealed a strong positive correlation (64%) between AAC and RS. So, we conducted a simple and multiple regression analysis to predict RS. AAC explained 64% of the variance in RS. Adding other predictors improved the RS prediction to 67%. DEC, PGWC, PC and AAC were found in both backward and forward stepwise methods except length and GLWR was found only in forward and backward methods, respectively (Table 4.2).

Table 4.2. Regression models for resistant starch (RS) content

| | **Regression** | **R2** | **R2_Adj** | **P-value of the model** | **Intercep t** | **Predictors** |
|---|---|---|---|---|---|---|
| Whole Panel | Simple | 0.639 | | <.0001 | -1.430 | AAC (0.124)** |
| | Multiple/forward stepwise | | 0.674 | <.0001 | -1.272 | Length (-0.181)**, DEC (-0.021)**, PGWC (0.018)**, PC (0.092)**, AAC (0.131)** |
| | Multiple/backward stepwise | | 0.673 | <.0001 | -1.808 | GLWR (-0.219), DEC (-0.023)**, PGWC(0.020)**, PC(0.091)**, AAC(0.133)** |

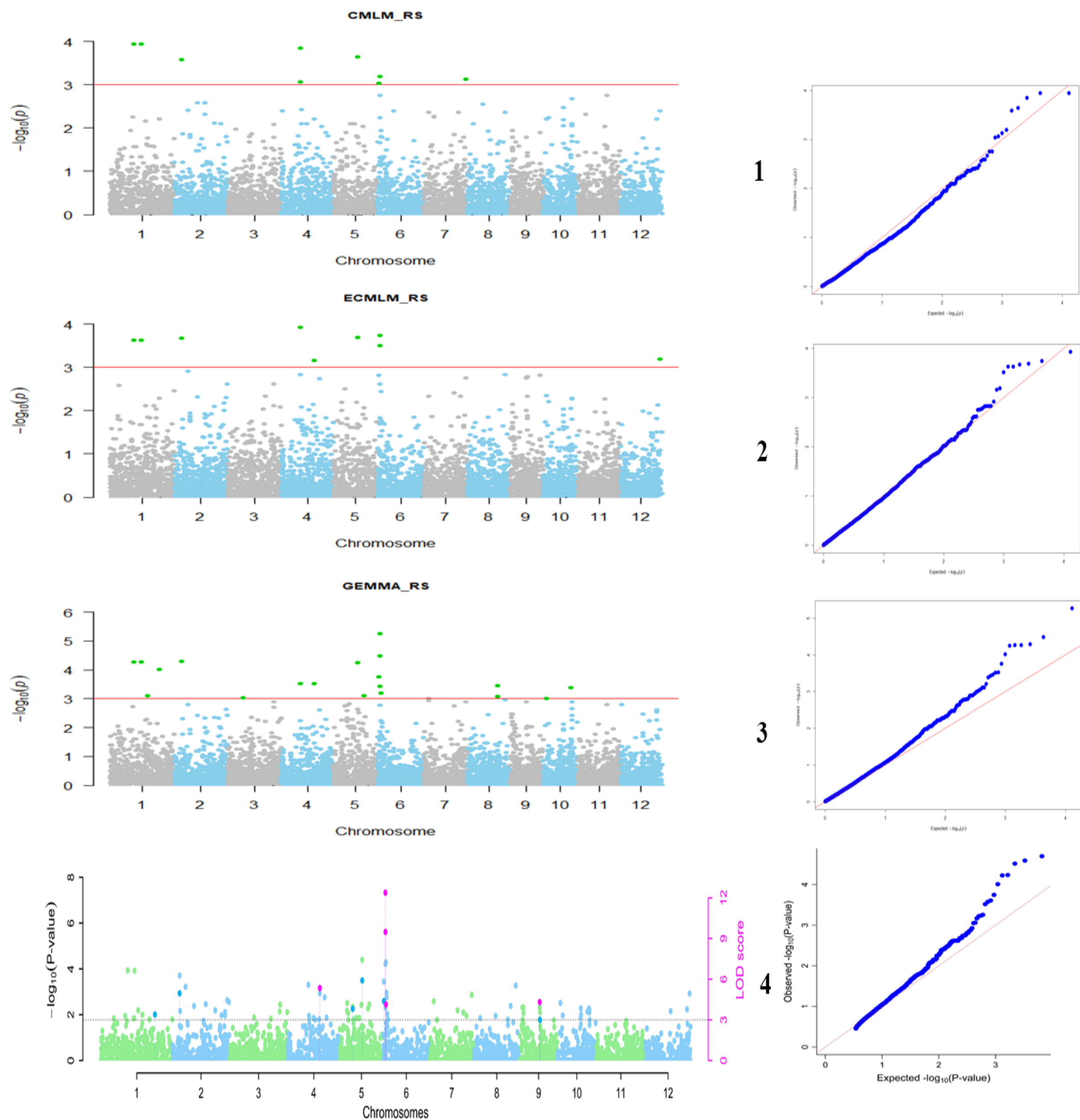



Figure 4.2. Manhattan plots of GWAS for resistant starch (RS). (1). Manhattan plot CMLM model. (2). Manhattan plot for ECMLM model. (3). Manhattan plot for GEMMA model. (4). Manhattan plot for multi-locus models, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, ISIS EM-BLASSO. The red horizontal line in the 1, 2, and 3 models is the threshold significant level used in the study to declare a SNP as being significant for measured traits. The green circles above the red line depict the significant SNPs. For 4 model, purple circles above the dashed horizontal line are the significant SNPs identified by all six multi-locus models, whereas green circles show only those SNPs identified by any two models of six models of multi-locus method.

### 4.3.2. Population structure and GWAS analysis

According to the STRUCTURE result, based on Δk value, there were six groups or sub-populations in our study sample panel. So, a six Q-matrix was used as a covariate during GWAS analysis (Figure 2.3). It is well known that rice has two major sub-populations, Indica and Japonica. Indica and Japonica each are divided into more sub-groups, such as temperate and tropical japonica. To determine the population structure effect on the phenotypic variations, we just considered the two primary sub-populations to analyze the phenotypic variation (Figure 2.3). Therefore, 78 and 93 accessions were considered as the Indica and Japonica panel; in total, 171 samples were analyzed during phenotypic analysis. Three accessions were removed due to admixture.

The LD decay of all the chromosomes was estimated to 250 kb, with half the maximum of mean r2 values (Figure 2.3).

SNPs with $-\log10P \geq 3.0$ and LOD $\geq 3.0$ for single and multi-locus models, respectively, was declared as significant for marker-trait association except those having MAF $< 0.05$. For RS, a total 17 significant SNPs were identified by nine models. Among them, only single-locus modes detected nine SNPs that explained 8.99-49.57% of the phenotypic variation. Only multi-locus

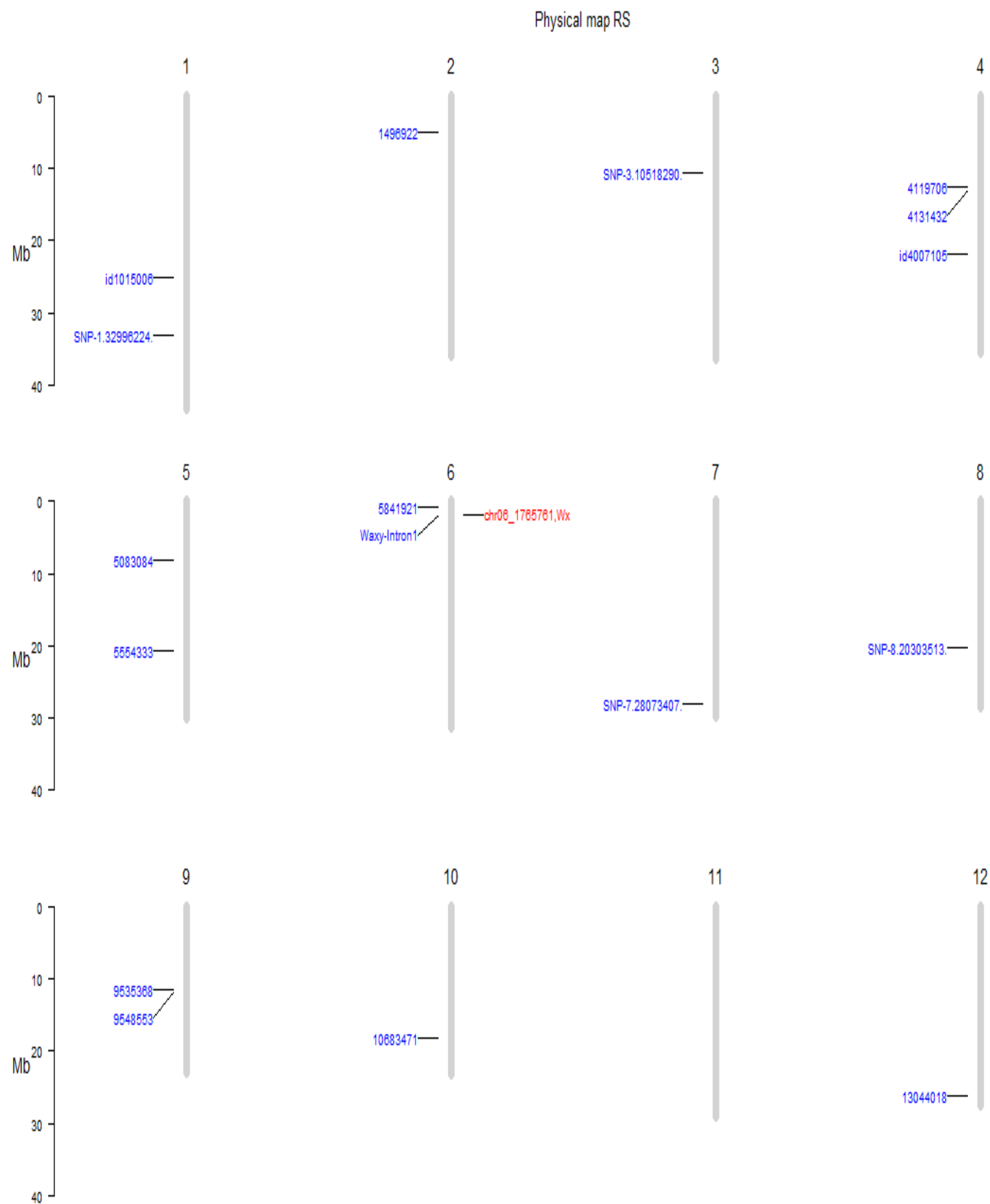



Figure 4.3. Physical map of the significant SNPs associated with resistant starch (RS). Rice chromosomes are displayed by vertical lines. The significant SNPs detected in this study are marked as blue color on the left side of the chromosomes. The previously reported genes/QTLs and markers are marked as red color on the right of the chromosomes.

models identified three SNPs with explaining 5.16-8.35% of the phenotypic variation. Five SNPs were co-detected by both models, explaining 3.10-50.35% of the phenotypic variation (Figure 4.2; Figure 4.4; Table 4.3). Among them, four SNPs had a pleiotropic effect (i.e., were significant for multiple traits). Waxy-Intron1 and SNP-7.28073407. SNPs have an effect on RS, ASV and AAC. Similarly, the id1015006 SNP was found to be significant for both RS and Width traits (Table 3.2; Table 4.3).

Three single-locus models individually identified a total 21 SNPs-where CMLM, ECMLM and GEMMA contributed 5, 5 and 11 SNPs. Multi-locus models found 14 SNPs (Table 4.3; Figure 4.3).

### 4.3.3. In-Silico Gene Expression Analysis

381 genes were found after mining within 250-kb region of the significant SNPs for RS using RAP-DB database (https://rapdb.dna.affrc.go.jp/). Then, we filtered out the non-expressed genes in the reproductive stage of rice plant by using Nipponbare gene expression data and 122 genes were selected for further analysis.

## 4.4. Discussion

### 4.4.1. Phenotypic Analysis

RS content can vary depending on the methods used to measure this trait. So far, three methods have been reported to measure rice RS content (Chen et al., 2017). We used the AOAC Method 2002.02 to measure the RS for our studies. The average RS content of the rice accessions is 1.36%, with a range between 0-6.92%. The similar result was found by Chen et al. (2017) with average ranging from 2.33-4.46%, 2.0% and 0.27% for high, intermediate, and low amylose type rice samples, respectively. Bao et al. (2017) reported average RS content ranging

from 0.3-2.42% using 105 rice accessions; however, they used rice flour, whereas we used cooked rice samples to measure RS. Similarly, they found a higher correlation between AAC and RS ($r^2$ =0.7529), which is a little bit higher than our study ($r^2$ =0.639). RS contents of the *indica* and japonica panels were significantly different because of population structure. Based on regression analysis, AAC alone explained 63% of the variance of RS, which is similar to Chen et al. (2017) where AAC explained 61%. To check if RS can be predicted by other variables along with AAC,

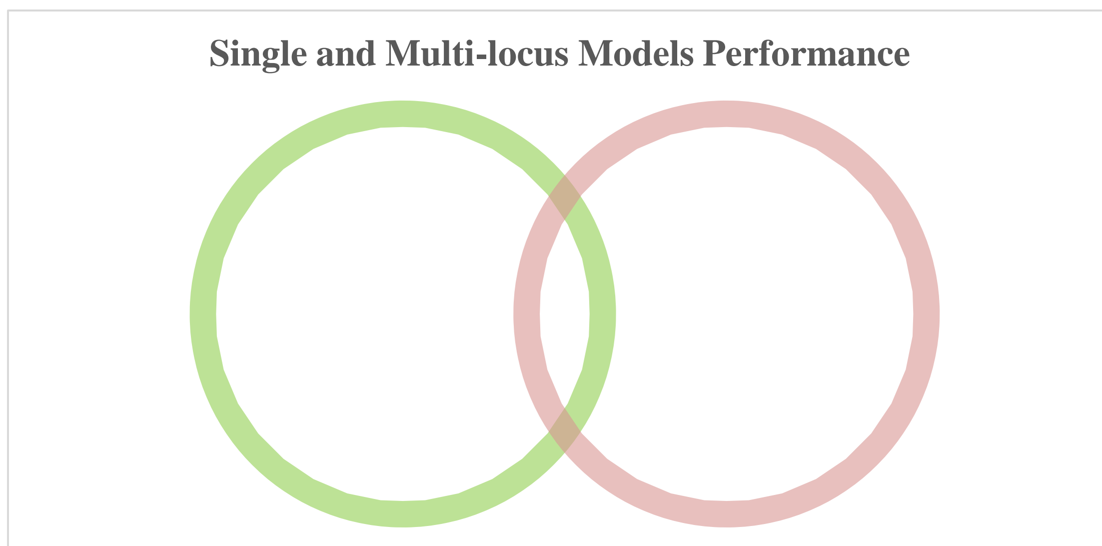



Figure 4.4. Venn diagram showed the number of SNPs identified by single-locus and multi-locus methods for resistant starch (RS).

we conducted multiple regression with eight variables (length, width, GLWR, DEC, PGWC, ASV, PC, AAC) and found that variance of RS could be explained by a combination of length, GLWR, DEC, PGWC, PC and AAC, explaining about 67% overall. Surprisingly, width and ASV did not appear to be significant predictors in the model, while id1015006 and Waxy-Intron1 SNPs were found for RS and width and RS and ASV, respectively, in our GWAS studies.

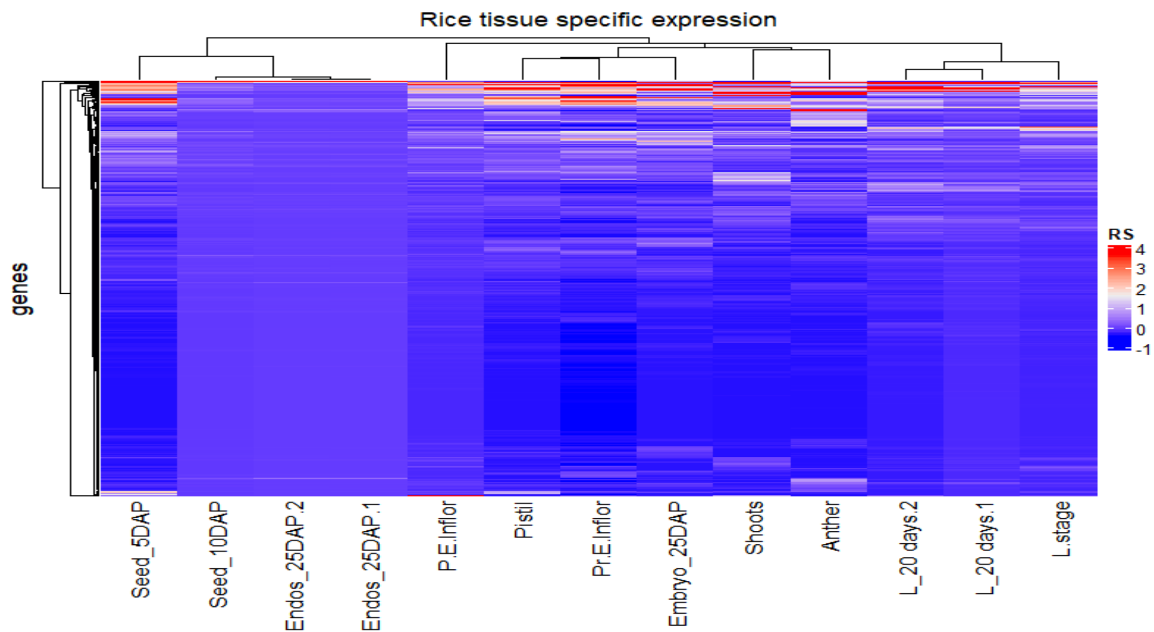



Figure 4.5. Heatmap of In-silico gene expression analysis result of resistant starch (RS).

### 4.4.2. Performance Of Single And Multi-Locus GWAS Models And Comparison GWAS Result

In terms of SNP detection ability, single-locus models found more than multi-locus models, including SNPs shared by both models, which is similar to our studies for mineral content. The same number of SNPs were detected by both CMLM and ECMLM, which is different from the previous report (Li et al., 2014). This may be due to not applying multi-test corrections for a threshold value, along with the possible reason of having SNPs with higher effect, in which situation the multivariate model performs poorly (Xu et al., 2018). As our goal is to identify the maximum number of causal SNPs controlling RS, but with reliable and genuine SNPs, we combined the single and multi-locus models to be complementary, as was recommended by the previous studies as well (Zhang et al., 2019).

Among 17 identified SNPs for the RS trait, a single SNP, Waxy-Intron1, was positioned close to the Wx gene, a gene encoding GBSSI which synthesizes amylose, affecting RS content in the rice grain. Similar results were also reported previously, confirming as being a major influential gene regulating RS content, as well as supporting the fact that RS is positively correlated with AAC (Bao et al., 2017; Fitzgerald et al., 2011; Kong et al., 2015). Additionally, Bao et al. (2017) also found three more known genes (SSIIa, ISA1, and AGPS1) on chromosome 6, 8, and 9 regulating RS content that we have not observed in our result. Instead, we identified significant SNPs at new locations on those chromosomes. While this study identified a single SNP out of 17 SNP (5%) that rediscovered a major gene (Wx) affecting RS content, the remaining 16 SNPs (95%) could be regarded as new loci for RS content. This result confirms that there are more genes that still need to be discovered that contribute to RS content.

Table 4.3. List identified QTLs for resistant starch (RS) content.

| Trait | SNP | Alleles | Chr. | Pos.(bp) | Single-locus GWAS | | | Multi-locus GWAS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| RS | | | | | -log10P | R2 | Model | LOD | R2 | Model |
| | id1015006 | A/G | 1 | 25139879 | 3.11 | | G | | | |
| | SNP-1.32996224. | G/A | 1 | 32997269 | 4.01 | | G | 3.39 | 15.3 | FM |
| | 1496922 | A/C | 2 | 5092779 | 3.59-4.29 | 10.6-50.4 | C, EC, G | 4.96 | 10.7 | M |
| | SNP-3.10518290. | G/A | 3 | 10519375 | 3.04 | | G | | | |
| | 4119706 | G/A | 4 | 12589729 | 3.52-3.93 | 11.4-50.8 | C, EC, G | | | |
| | 4131432 | G/A | 4 | 12954360 | 3.06 | 9.0 | C | | | |
| | id4007105 | A/G | 4 | 21815986 | 3.16-3.52 | 49.5 | EC, G | 4.24 | 3.1 | FM, PK |
| | 5083084 | A/G | 5 | 8080666 | | | | 3.81 | 7.7 | PK |
| | 5554333 | A/C | 5 | 20669935 | 3.11 | | G | | | |
| | 5841921 | G/A | 6 | 777271 | 3.04-3.75 | 8.9 | C, G | 4.37 | 26.6 | PK |
| | Waxy-Intron1 | C/A | 6 | 1765761 | 3.74-4.48 | 50.5 | EC, G | 3.85 | 10.8 | FE, FM, I, M, PL |
| | SNP-7.28073407. | C/A | 7 | 28074401 | 3.13 | 7.2 | C | | | |
| | SNP-8.20303513. | G/A | 8 | 20306227 | 3.46 | | G | | | |
| | 9535368 | C/A | 9 | 11435014 | | | | 5.47 | 8.4 | M, PL |
| | 9548553 | A/G | 9 | 11746684 | | | | 3 | 5.2 | PK |
| | 10683471 | A/G | 10 | 18138508 | 3.38 | | G | | | |
| | 13044018 | A/C | 12 | 26193413 | 3.19 | 49.6 | EC | | | |

N.B: Single-locus models: C-CMLM, EC-ECMLM, G- GEMMA; Multi-locus models: M- MrMLM, FM- FASTmrMLM, FE- FASTmrEMMA, PK- pKWmEB, PL- pLARmEB, I- ISIS EM-BLASSO

Table 4.4. Comparison of the GWAS result with the previous studies.

| Trait | SNP | Position (bp) | Chr. | Markers linked/associated | Types of Markers | Known genes/QTLs | Position (bp) | References |
|---|---|---|---|---|---|---|---|---|
| RS | Waxy-Intron 1 | 1765761 | 6 | chr06_1765761 | SNP | Wx | 1765761 | Bao et al. (2017) |

## 4.5. Conclusion

This study reported a GWAS and phenotypic analysis for RS content in cooked rice using 174 global rice accessions with 7k SNP array genotype data. A total of 17 significant SNPs affecting RS were identified by single-locus and multi-locus methods. Of these SNPs, 9 SNPs were detected by only single-locus methods, whereas the multi-locus methods detected 3 SNPs, and 5 SNPs were co-detected by both methods. After mining genes within the 250 kb region of these SNPs, 381 genes were found. Among these genes, 122 genes, including known genes, could be the candidate genes for controlling RS content in rice grain. These shortlisted genes could be used for future study to explore the gene expression levels, followed by functional gene characterization, helping to understand the complex molecular mechanisms of RS content in cooked rice. In addition, multiple regression analysis was conducted to predict the RS content by other rice grain quality-related traits that could be useful for rice breeding to select rice varieties with high RS content in any rice crop improvement program.

### 4.6. References


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

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308

Butardo, V. M., Daygon, V. D., Colgrave, M. L., Campbell, P. M., Resurreccion, A., Cuevas, R. P., . . . Fitzgerald, M. (2012). Biomolecular Analyses of Starch and Starch Granule Proteins in the High-Amylose Rice Mutant Goami 2. *Journal of Agricultural and Food Chemistry, 60*(46), 11576-11585. doi:10.1021/jf303205p

Chen, M. H., Bergman, C. J., McClung, A. M., Everette, J. D., & Tabien, R. E. (2017). Resistant starch: Variation among high amylose rice varieties and its relationship with apparent amylose content, pasting properties and cooking methods. *Food Chem, 234*, 180-189. doi:10.1016/j.foodchem.2017.04.170

Cui, Y., Zhang, F., & Zhou, Y. (2018). The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice. *Front Plant Sci, 9*, 1464. doi:10.3389/fpls.2018.01464

Englyst, H. N., Kingman, S. M., & Cummings, J. H. (1992). Classification and measurement of nutritionally important starch fractions. *European journal of clinical nutrition, 46*, S33-50.

Fitzgerald, M. A., Rahman, S., Resurreccion, A. P., Concepcion, J., Daygon, V. D., Dipti, S. S., . . . Bird, A. R. (2011). Identification of a major genetic determinant of glycaemic index in rice. *Rice, 4*(2), 66-74.

Horwitz, W., & Latimer, G. W. (2005). Official methods of analysis of AOAC International. 18th edGaithersburg. *Md: AOAC International*.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., . . . Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet, 42*(11), 961-967. doi:10.1038/ng.695

Kong, X., Chen, Y., Zhu, P., Sui, Z., Corke, H., & Bao, J. (2015). Relationships among genetic, structural, and functional properties of rice starch. *Journal of Agricultural and Food Chemistry, 63*(27), 6241-6248.

Li, C., Fu, Y., Sun, R., Wang, Y., & Wang, Q. (2018). Single-Locus and Multi-Locus Genome-Wide Association Studies in the Genetic Dissection of Fiber Quality Traits in Upland Cotton (Gossypium hirsutum L.). *Front Plant Sci, 9*, 1083. doi:10.3389/fpls.2018.01083

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y.-M., Todhunter, R. J., . . . Zhang, Z. (2014). Enrichment of statistical power for genome-wide association studies. *BMC biology, 12*(1), 1-10.

Morales, K. Y., Singh, N., Perez, F. A., Ignacio, J. C., Thapa, R., Arbelaez, J. D., . . . Thomson, M. J. (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS One, 15*(5), e0232479. doi:10.1371/journal.pone.0232479

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., . . . Ye, G. (2015). Genome-Wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of Indica Rice Germplasm. *PLoS One, 10*(12), e0145577. doi:10.1371/journal.pone.0145577

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., . . . Xu, C. (2018). Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. *Front Plant Sci, 9*, 1311. doi:10.3389/fpls.2018.01311

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., . . . Xu, C. (2018). Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. *Frontiers in plant science, 9*, 1311. Retrieved from https://www.frontiersin.org/article/10.3389/fpls.2018.01311

Yang, M., Lu, K., Zhao, F.-J., Xie, W., Ramakrishna, P., Wang, G., . . . Zhao, H. (2018). Genome-wide association studies reveal the genetic basis of ionomic variation in rice. *The Plant Cell, 30*(11), 2720-2740.

Zhang, Y.-W., Lwaka Tamba, C., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., . . . Zhang, Y.-M. (2020). mrMLM v4.0: An R Platform for Multi-locus Genome-wide Association Studies. *Genomics, Proteomics & Bioinformatics*. doi:https://doi.org/10.1016/j.gpb.2020.06.006

Zhang, Y. M., Jia, Z., & Dunwell, J. M. (2019). Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Front Plant Sci, 10*, 100. doi:10.3389/fpls.2019.00100

# 5. CONCLUSIONS

The present study was conducted to further characterize the genetic basis of rice grain quality, milling, eating and nutritional quality traits by using GWAS approach. A total of 174 rice accessions were used for these purposes, which have the diverse genetic backgrounds required for association studies. A total of 6,565 SNP markers were utilized for the marker-trait association studies.

Phenotypic analysis was conducted to explore the phenotypic variation existing in the germplasm and a wide variation across the traits was found. Similarly, population structure and kinship analysis were performed to reduce the type 1 error rate in the study. Six populations were found in our rice germplasm that affect the phenotypic variation for all the traits except GLWR, MRY, HRY, K, Mg traits.

We used several single-locus and multi-locus GWAS methods to identify as many significant SNPs as possible, controlling the corresponding traits of the study. In this regard, our study was successful because both methods performed well in terms of possible SNP identification. Single-locus methods alone detected 110, 106 and 9 SNPs for minerals, appearance and nutritional and RS content, respectively. Similarly, 22, 64 and 3 SNPs were found by multi-locus methods alone for minerals, appearance, and nutritional and RS content, respectively. Both methods have their own advantages and disadvantages, so this study recommends using both methods as a complementary approach, which was also recommended by past studies, to identify all possible SNPs affecting the corresponding traits.

We conducted in silico gene expression analysis to identify the candidate genes for the respective traits. We found that 3129, 4609 and 381 genes are located within 250 kb region of the corresponding SNPs for minerals, appearance and nutritional traits RS content, respectively, and among them, 792, 1329 and 122 genes were found being expressed in the rice reproductive stage, indicating their candidacy for the corresponding traits. Among these expressed genes, the known genes, which is functionally characterized before, were included too.

The findings of the current study will be useful for future studies to narrow down the genes by conducting further gene expression analysis. Also, it could be helpful to target genes identified by both methods that validate the reliability of particular SNPs, for functional characterization using CRISPR/Cas9 gene editing.

## APPENDIX A

## DESCRIPTION OF THE RICE GERMPLASM USED IN THE STUDY

| Sample | Plant_ID | Name | Origin |
|---|---|---|---|
| RG-2 | CIor 12153 | Quinimpol | Philippines |
| RG-4 | CIor 12234 | Long Gnar Jim | US |
| RG-6 | CIor 12244 | Creole Bred | US |
| RG-9 | CIor 2490 | Karang Serang | Indonesia |
| RG-11 | CIor 7404 | Kin Shan Zim | China |
| RG-14 | CIor 9403 | Century Patna Original | US |
| RG-15 | PI 127076 | Spin Mere | Afghanistan |
| RG-18 | PI 160530 | Pan Ju | China |
| RG-19 | PI 161567 | Criollo Chivacoa 2 | Venezuela |
| RG-21 | PI 180060 | Dhala Shaitta | Bangladesh |
| RG-23 | PI 199553 | Secano do Brazil | El Salvador |
| RG-24 | PI 208447 | Early No. 1 | Nepal |
| RG-28 | PI 223612 | Sel. No. 388 | Uruguay |
| RG-29 | PI 224605 | Sigadis | Indonesia |
| RG-30 | PI 226204 | SHIMIZU MOCHI | Japan |
| RG-31 | PI 229262 | N 32 | India |
| RG-33 | PI 238190 | Charmarumi | India |
| RG-34 | PI 240638 | Dular | India |

| | | | |
|---|---|---|---|
| RG-37 | PI 277414 | Red Khosha Cerma | Afghanistan |
| RG-39 | PI 283681 | Hashikalmi Aus | Bangladesh |
| RG-40 | PI 283682 | Kataktara Aus | Bangladesh |
| RG-42 | PI 291608 | WC 4443 | Bolivia |
| RG-43 | PI 294423 | GHRAIBA | Iraq |
| RG-44 | PI 297569 | Dharial | Bangladesh |
| RG-48 | PI 369813 | Samanis | Suriname |
| RG-49 | PI 373053 | Bala | India |
| RG-51 | PI 373232 | Khao Phoi | Laos |
| RG-52 | PI 373347 | Karayal | Sri Lanka |
| RG-53 | PI 373403 | ARC 6578 | India |
| RG-55 | PI 373537 | ARC 10638 | India |
| RG-56 | PI 373777 | C 8447 | Indonesia |
| RG-57 | PI 373779 | Tia Heret | Indonesia |
| RG-60 | PI 373816 | Padi Pohon Batu | Malaysia |
| RG-62 | PI 376252 | Pelu | India |
| RG-65 | PI 385344 | Ratna | India |
| RG-67 | PI 385529 | Jhona | Pakistan |
| RG-68 | PI 385578 | Sufaida | Pakistan |
| RG-70 | PI 385621 | Mahlar | Pakistan |
| RG-73 | PI 385849 | Ziri | Pakistan |

| | | | |
|---|---|---|---|
| RG-74 | PI 385888 | Sathra | Pakistan |
| RG-75 | PI 389037 | Ai Chueh Ta Pai Ku | Taiwan |
| RG-77 | PI 389267 | Heo Trang | Vietnam |
| RG-78 | PI 389876 | Sipirasikkam | Indonesia |
| RG-79 | PI 389879 | Sigoendaba | Indonesia |
| RG-80 | PI 389945 | Angkrang | Cambodia |
| RG-81 | PI 389960 | Srav Prapay | Cambodia |
| RG-83 | PI 391827 | Lantjang | Indonesia |
| RG-84 | PI 391936 | Ali Combo | Madagascar |
| RG-85 | PI 391943 | Sabharaj | Bangladesh |
| RG-86 | PI 392170 | Torh | Pakistan |
| RG-87 | PI 392217 | Sugdasi | Pakistan |
| RG-90 | PI 392677 | ASWINA 330 | Bangladesh |
| RG-92 | PI 393114 | DNJ 151 | Bangladesh |
| RG-94 | PI 400042 | AS 46 | India |
| RG-95 | PI 400586 | Putih Montor | Indonesia |
| RG-96 | PI 400587 | Gendjah Banten | Indonesia |
| RG-98 | PI 400662 | Janeri | Nepal |
| RG-101 | PI 400773 | Vary Vato 275 | Madagascar |
| RG-105 | PI 400782 | Bengaly Morino 120 | Madagascar |
| RG-107 | PI 401750 | Kuning Tinggi | Indonesia |

| RG-109 | PI 402634 | Koi Murali | India |
|---|---|---|---|
| RG-111 | PI 402691 | Trandeup Kandir | Cambodia |
| RG-112 | PI 402720 | Angana | India |
| RG-113 | PI 402747 | Banajira | Bangladesh |
| RG-114 | PI 402804 | Brondol | Indonesia |
| RG-117 | PI 403091 | DJ 53 | Bangladesh |
| RG-118 | PI 403109 | DJ 90 | Bangladesh |
| RG-119 | PI 403114 | DJ 102 | Bangladesh |
| RG-120 | PI 403160 | DM 55 | Bangladesh |
| RG-121 | PI 403287 | DV 85 | Bangladesh |
| RG-123 | PI 403310 | DV 132 | Bangladesh |
| RG-130 | PI 412790 | Daudzai Field Mix | Pakistan |
| RG-137 | PI 413802 | Bengawan | Indonesia |
| RG-138 | PI 413989 | Sug | India |
| RG-148 | PI 431292 | Akabona | Pakistan |
| RG-150 | PI 433833 | Aus 8 | Bangladesh |
| RG-154 | PI 439078 | Ngoba | India |
| RG-159 | PI 494105 | M202 | US |
| RG-160 | PI 497682 | IR64 | Philippines |
| RG-178 | PI 575134 | Bak Tushi | Bangladesh |
| RG-179 | PI 575201 | Gambir | Bangladesh |

| RG-181 | PI 575212 | Ghorbhai | Bangladesh |
|---|---|---|---|
| RG-182 | PI 575217 | Shoni | Bangladesh |
| RG-186 | PI 584569 | FIROOZ | Iran |
| RG-189 | PI 584625 | Ak Tokhum | Azerbaijan |
| RG-192 | PI 585042 | EMBRAPA 1200 | Brazil |
| RG-193 | PI 593892 | Jefferson | US |
| RG-226 | PI 67150 | Mushkan | India |
| RG-237 | PI 231642 | Caucasica | Russia |
| RG-257 | PI 277417 | Shevkati Kundry | Azerbaijan |
| RG-259 | PI 282171 | ARPA SHALI | Uzbekistan |
| RG-263 | PI 282208 | Uz Rosz 17 | Uzbekistan |
| RG-264 | PI 282210 | UZ ROSZ 269 | Uzbekistan |
| RG-266 | PI 282212 | Uz Rosz 2832 | Uzbekistan |
| RG-269 | PI 291427 | Uz Rosz M9 | Uzbekistan |
| RG-276 | PI 346926 | Nahodka | NA |
| RG-277 | PI 346927 | VILKID ZIRE | NA |
| RG-279 | PI 346932 | KUBAN 3 | Russia |
| RG-280 | PI 348904 | Uz Ros 275 | Uzbekistan |
| RG-283 | PI 348909 | SADRI MASALINSKIJ | Azerbaijan |
| RG-284 | PI 348910 | Ambarby White | Azerbaijan |
| RG-288 | PI 373900 | Besudi Long-Grain | Afghanistan |

| | | | |
|---|---|---|---|
| RG-289 | PI 373901 | Besudi Short-Grain | Afghanistan |
| RG-305 | PI 431000 | Dera Wadi 1/43 | Afghanistan |
| RG-310 | PI 431024 | Cat 1747 | Russia |
| RG-311 | PI 431031 | P 817 | Russia |
| RG-316 | PI 431195 | Vulgaris Ko Ch Azpasaly | Uzbekistan |
| RG-319 | PI 431201 | UZ ROS 59 | Uzbekistan |
| RG-332 | PI 431235 | P 1041 | Russia |
| RG-334 | PI 431242 | P 1048 | Russia |
| RG-336 | PI 431267 | HZ ROS 637 | Uzbekistan |
| RG-338 | PI 439621 | Azerbaidjanica | Azerbaijan |
| RG-340 | PI 439624 | Kasakstanica | Kazakhstan |
| RG-343 | PI 439629 | Nigrescens | Russia |
| RG-346 | PI 439633 | Ak Tohum | Azerbaijan |
| RG-348 | PI 439637 | Dicolorata | Azerbaijan |
| RG-354 | PI 439650 | Bak Saly Mestnyj | Azerbaijan |
| RG-355 | PI 439661 | DONSKOJ 2 | Russia |
| RG-357 | PI 439664 | Dv Ros 0219 | Russia |
| RG-358 | PI 439665 | Dv Ros 2568 | Russia |
| RG-363 | PI 439677 | Kasaki Shala Mestnyj | Uzbekistan |
| RG-364 | PI 439679 | Kesa | Azerbaijan |
| RG-365 | PI 439683 | KUBANETS 508 | Russia |

| RG-371 | PI 439724 | Severnyj | Russia |
|---|---|---|---|
| RG-372 | PI 439730 | UZBEKSKIJ 2 | Uzbekistan |
| RG-373 | PI 439733 | Uz Ros 421 | Uzbekistan |
| RG-375 | PI 446913 | Pioner 320 | Uzbekistan |
| RG-376 | PI 458444 | Krasnodarski | Russia |
| RG-379 | PI 584584 | LUK TAKHAR | Afghanistan |
| RG-380 | PI 584615 | WIR 623 | Uzbekistan |
| RG-383 | PI 584618 | WIR 1528 | Azerbaijan |
| RG-384 | PI 584620 | Hi Muke | Kazakhstan |
| RG-385 | PI 584622 | WIR 2623 | Russia |
| RG-389 | PI 584629 | Celiaj | Azerbaijan |
| RG-390 | PI 584633 | UZ ROS 2759 | Uzbekistan |
| RG-393 | PI 584637 | KROS 358 | Kazakhstan |
| RG-394 | PI 584640 | NF-1 | Russia |
| RG-395 | PI 584642 | NF-9 | Russia |
| RG-396 | PI 584644 | SPALCIK | Russia |
| RG-400 | PI 584649 | INTENSIVNYJ | Uzbekistan |
| RG-403 | PI 584652 | ZEMCYZNYJ | Russia |
| RG-404 | PI 596813 | WIR 3419 | Azerbaijan |
| RG-412 | PI 61718 | Shala | Turkistan |
| RG-413 | PI 65884 | Styk | Azerbaijan |

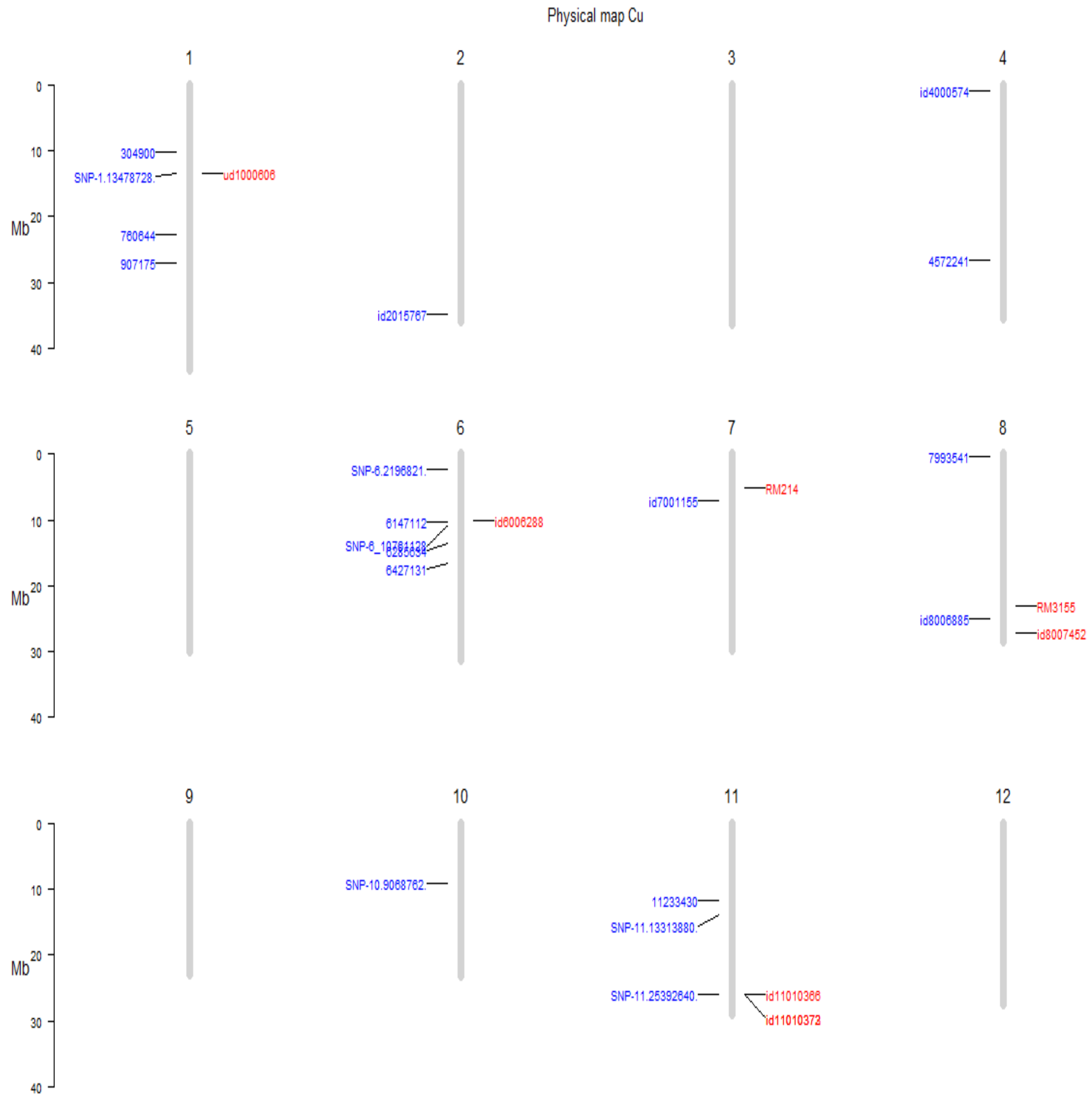
| | | | |
|---|---|---|---|
| RG-550 | PI 636839 | WAB450-11-1-3-P40-HB | Cote D'Ivoire |
| RG-551 | PI 636840 | WAB450-24-2-3-P33-HB | Cote D'Ivoire |
| RG-552 | PI 636841 | WAB450-24-3-P38-1-HB | Cote D'Ivoire |
| RG-557 | PI 636846 | WAB450-I-B-P-38-HB | Cote D'Ivoire |
| RG-558 | PI 636847 | WAB450-I-B-P-62-HB | NA |
| RG-575 | PI 636465 | Presidio | US |
| RG-578 | PI 385751 | Kharsu | Pakistan |
| U-117 | NA | JAZZMAN 2 | US |
| U-118 | NA | CL 172 | US |
| U-119 | NA | M206 | US |
| U-120 | NA | CL 163 | US |
| U-158 | NA | DELLA 2 | US |
| U-159 | NA | ANTONIO | US |
| U-160 | NA | THAD | US |
| U-17 | NA | CL 111 | US |
| U-18 | NA | CL 153 | US |
| U-199 | NA | RONDO | US |
| U-20 | NA | MERMENTAU | US |
| U-200 | NA | CL 151 | US |
| U-37 | NA | JUPITER | US |
| U-38 | NA | WELLS | US |

| U-39 | NA | LAKAST | US |
|---|---|---|---|
| U-40 | NA | DIAMOND | US |
| U-56 | NA | MM-14 | US |
| U-57 | NA | REX | US |
| U-58 | NA | CHENIERE | US |
| U-59 | NA | COCODRIE | US |
| U-60 | NA | CL272 | US |
| U-79 | NA | ROY J | US |
| U-80 | NA | TITAN | US |

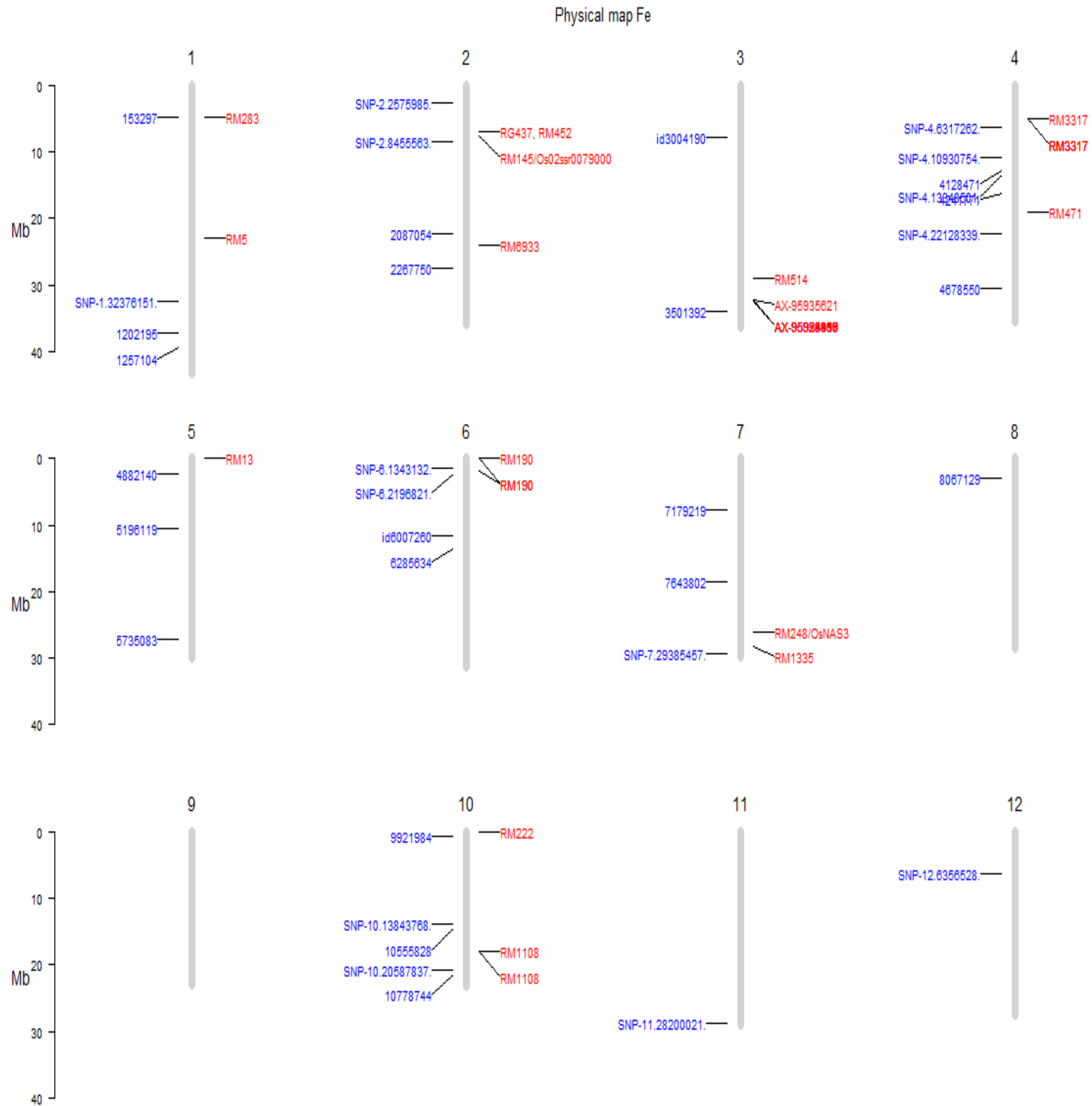
# APPENDIX B

## PHYSICAL MAP FOR SIX MINERAL ELEMENTS

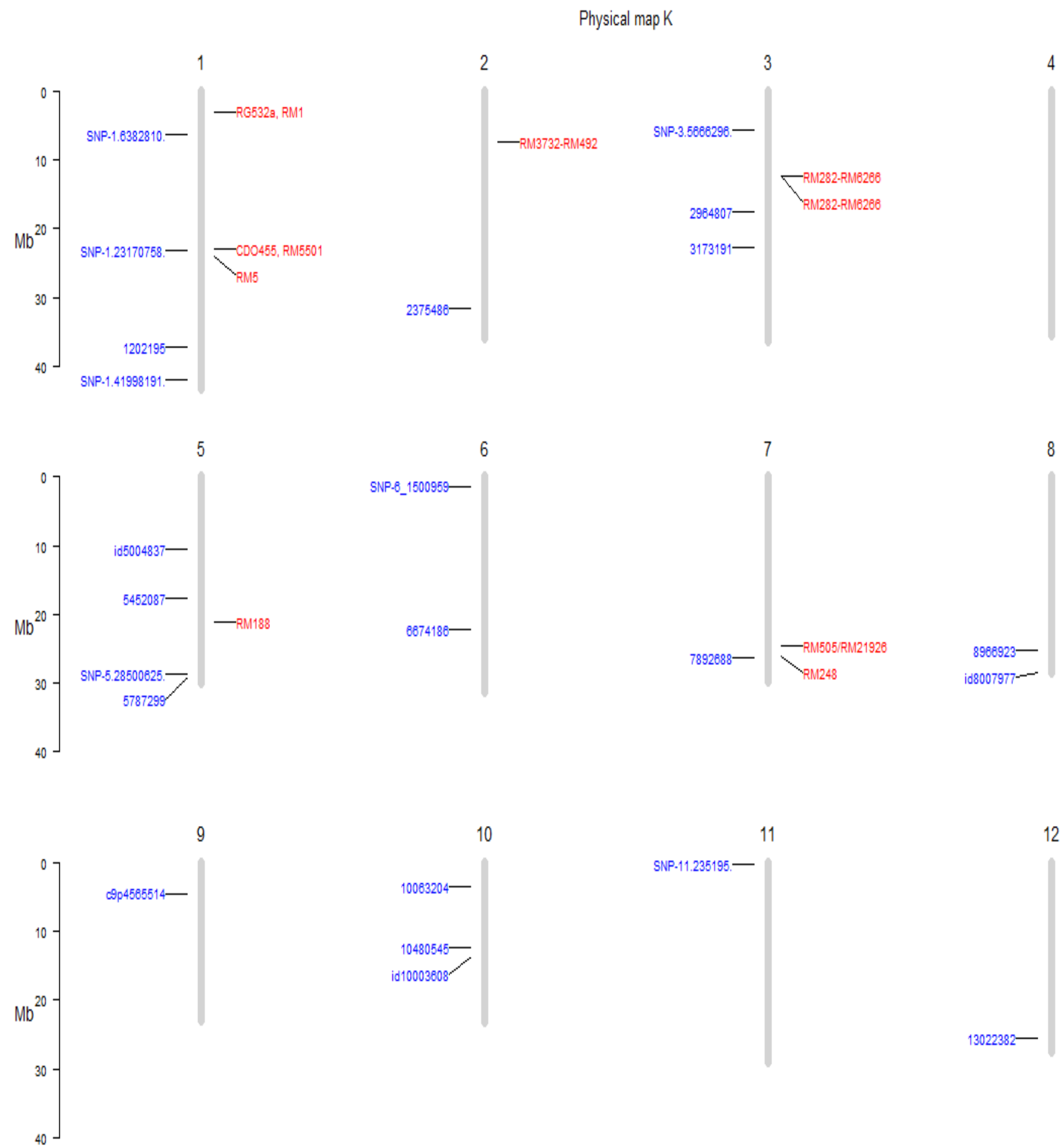
1. Cu



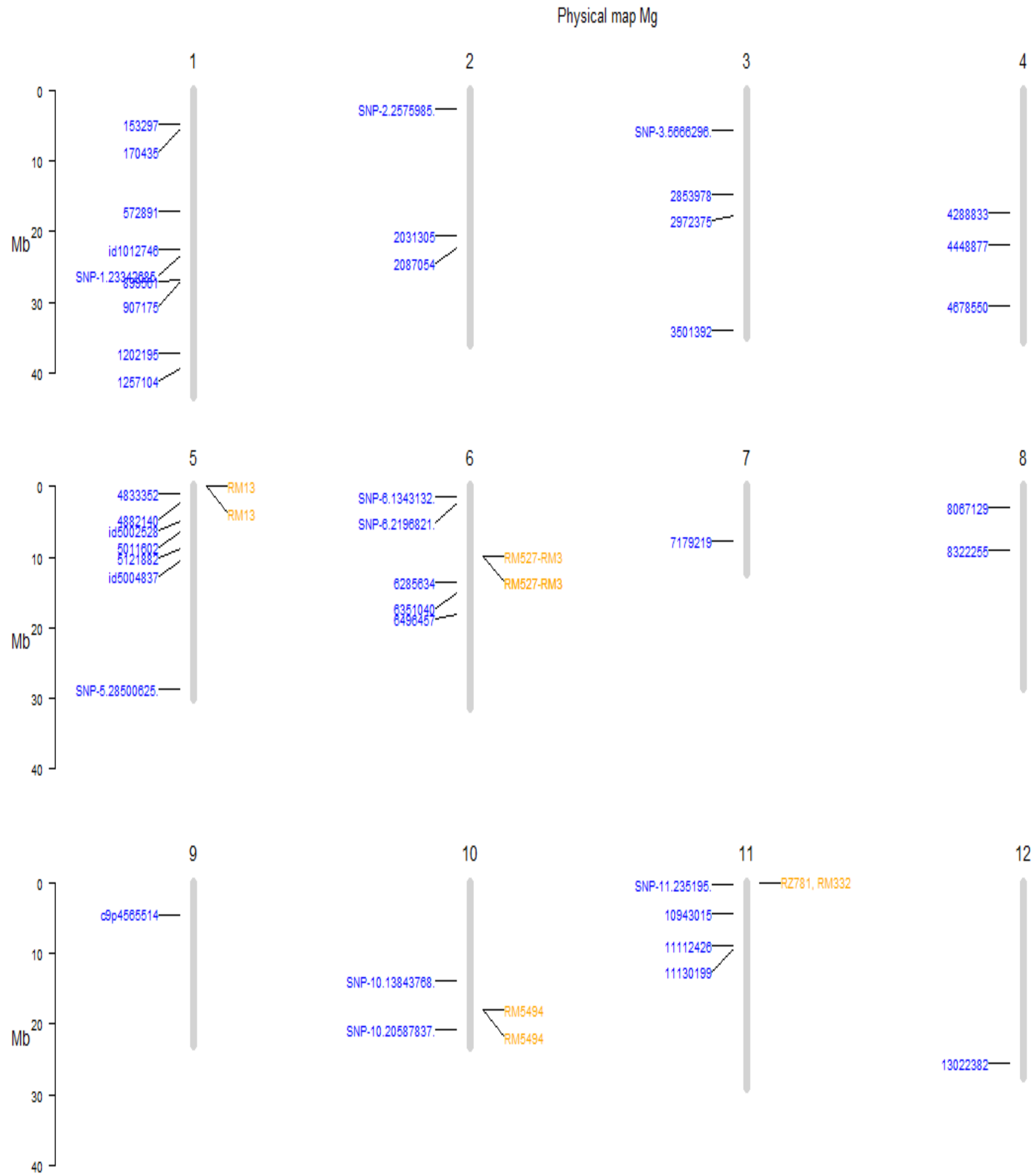
2. Fe



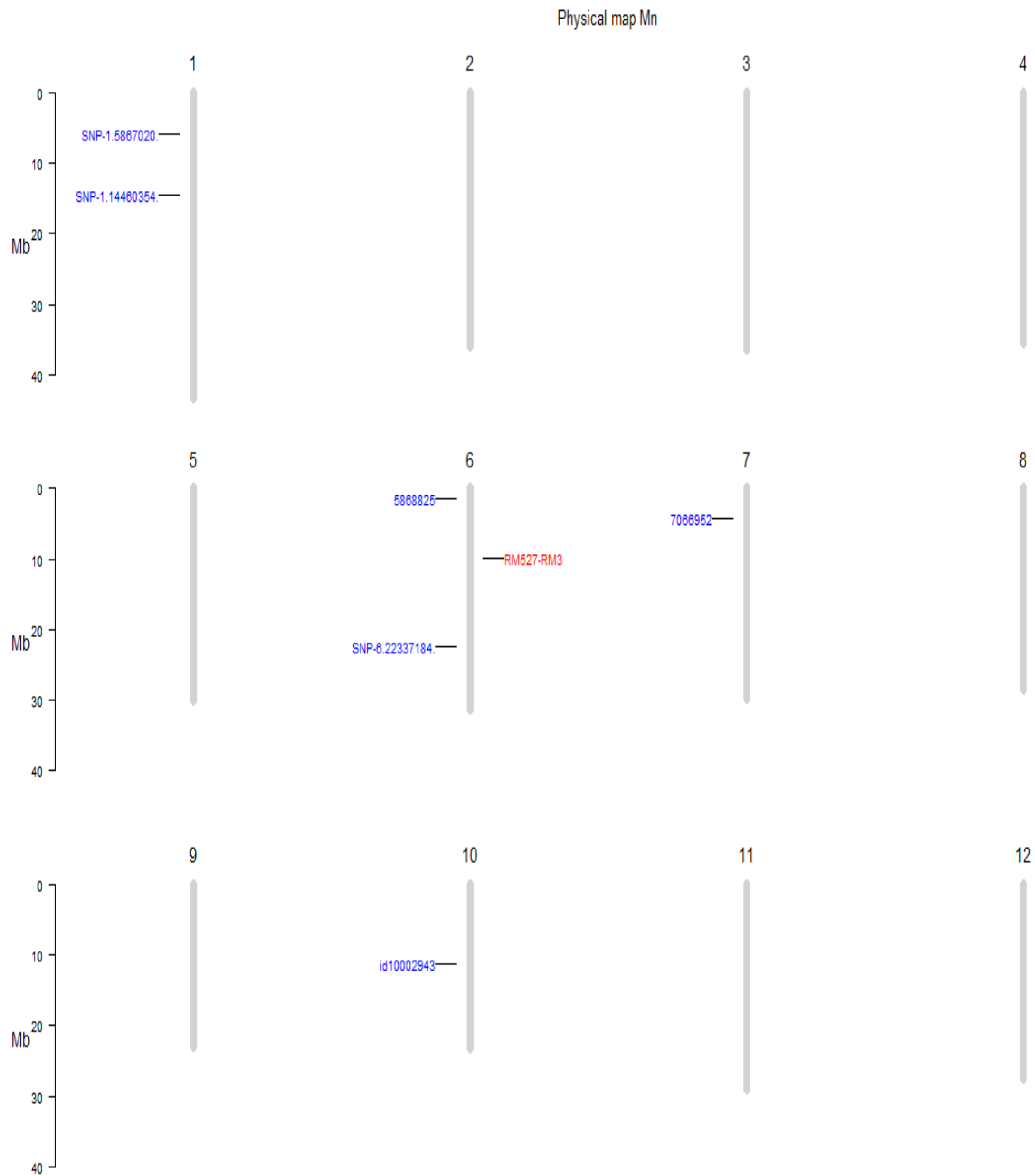
3. K



4. Mg



5. Mn



6. Zn

Physical map Zn

1
1280193
AX-95918225/OsIRO2

3
SNP-3.4621271.
SNP-3.28426789.
id3013232
3405830
id3001044

4
id4004654
4572241
RM317
id4010220
RM317/Os04ssr0174300
id4010984

5
id5002528

6
id6001756
SNP-6.6241072.
6496457
SNP-6.25063527.
6783797

7

8

9

10
RM222
10430775
AX-95932094
id10003681
id10007301

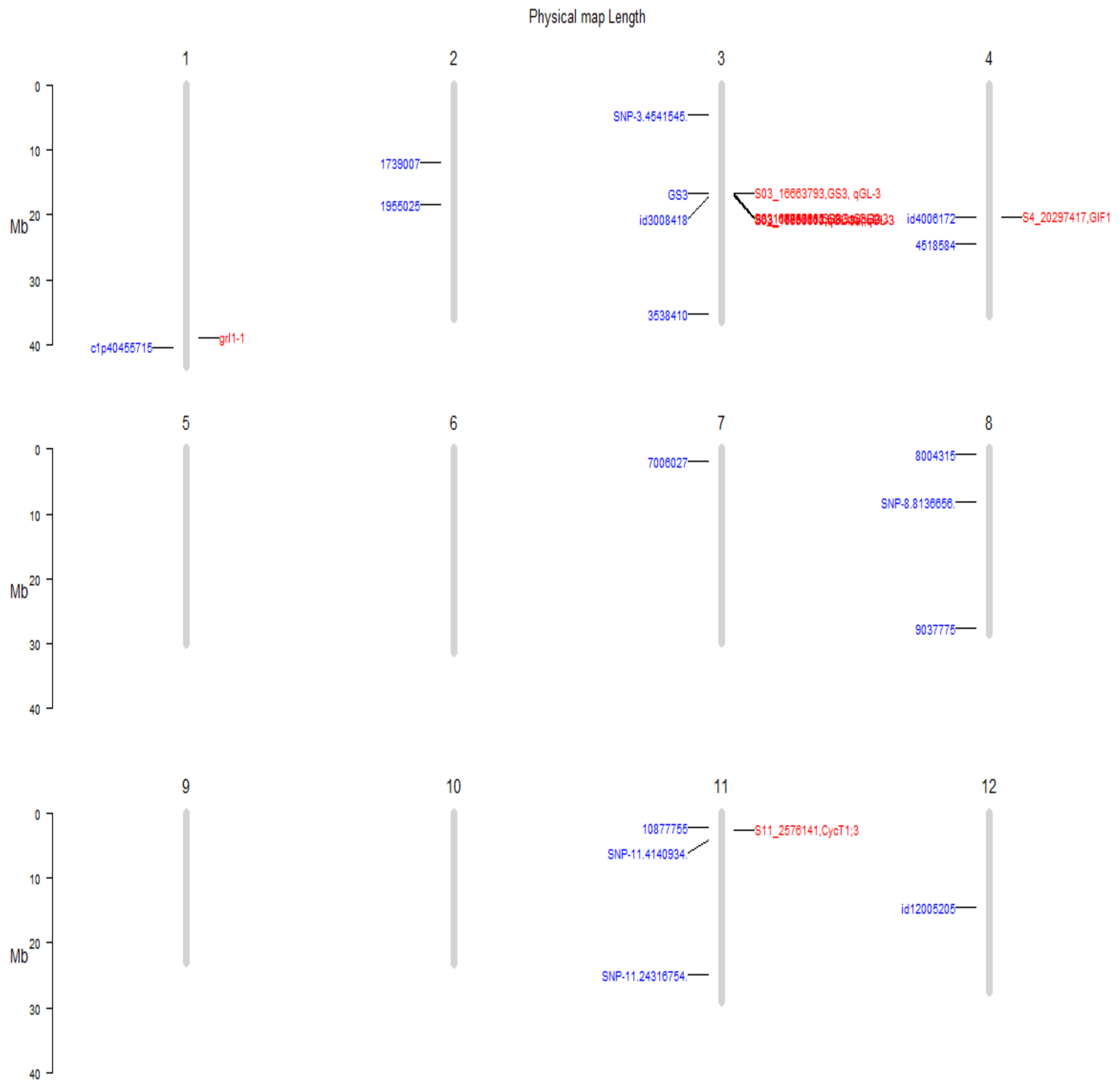
11
11769276
11915122

12
12134336

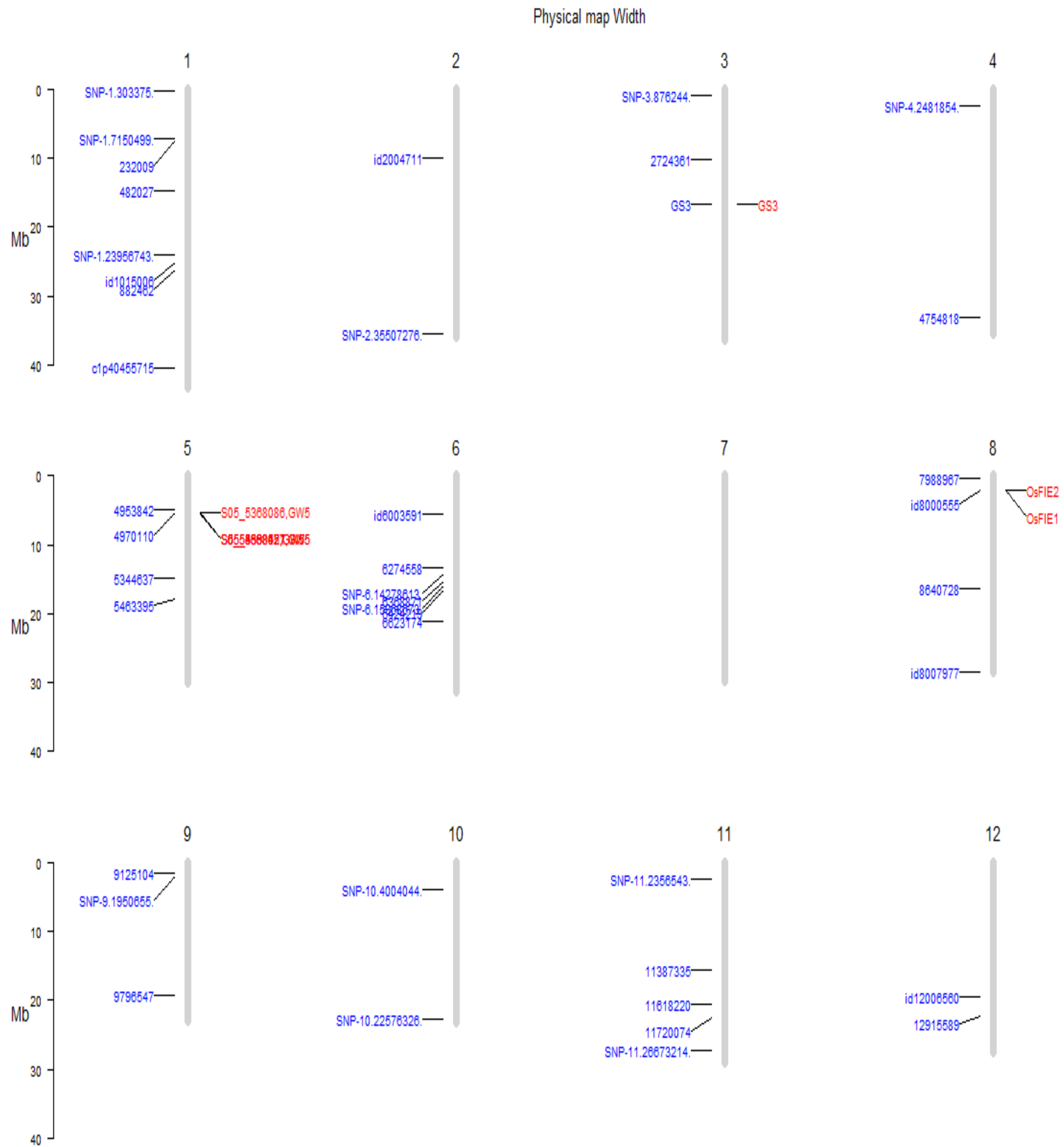
# APPENDIX C

## PHYSICAL MAP FOR RICE GRAIN APPEARANCE, MILLING, EATING AND COOKING, AND NUTRITIONAL TRAITS.

### Grain Length (GL)



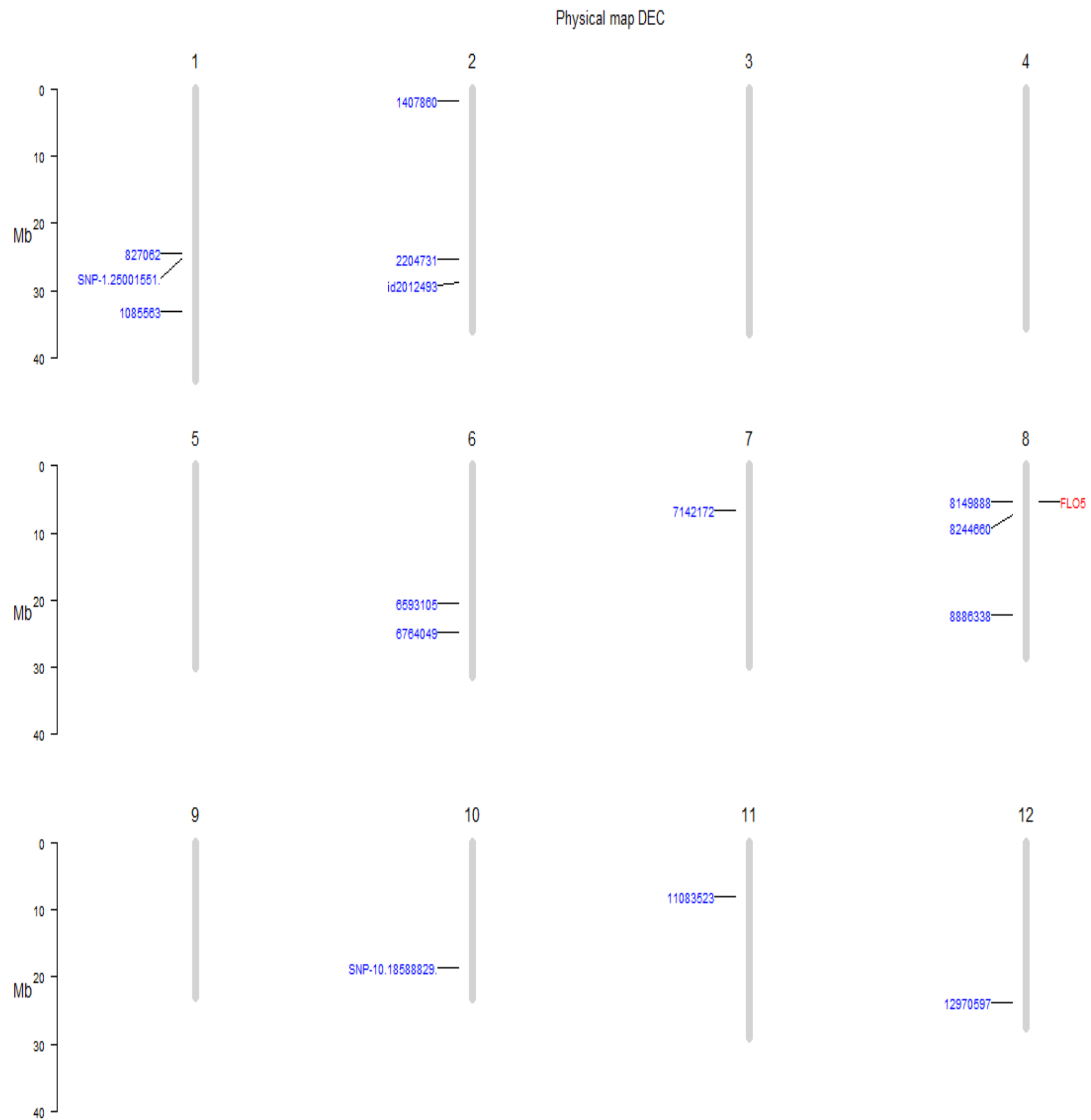
2. Grain Width (GW)



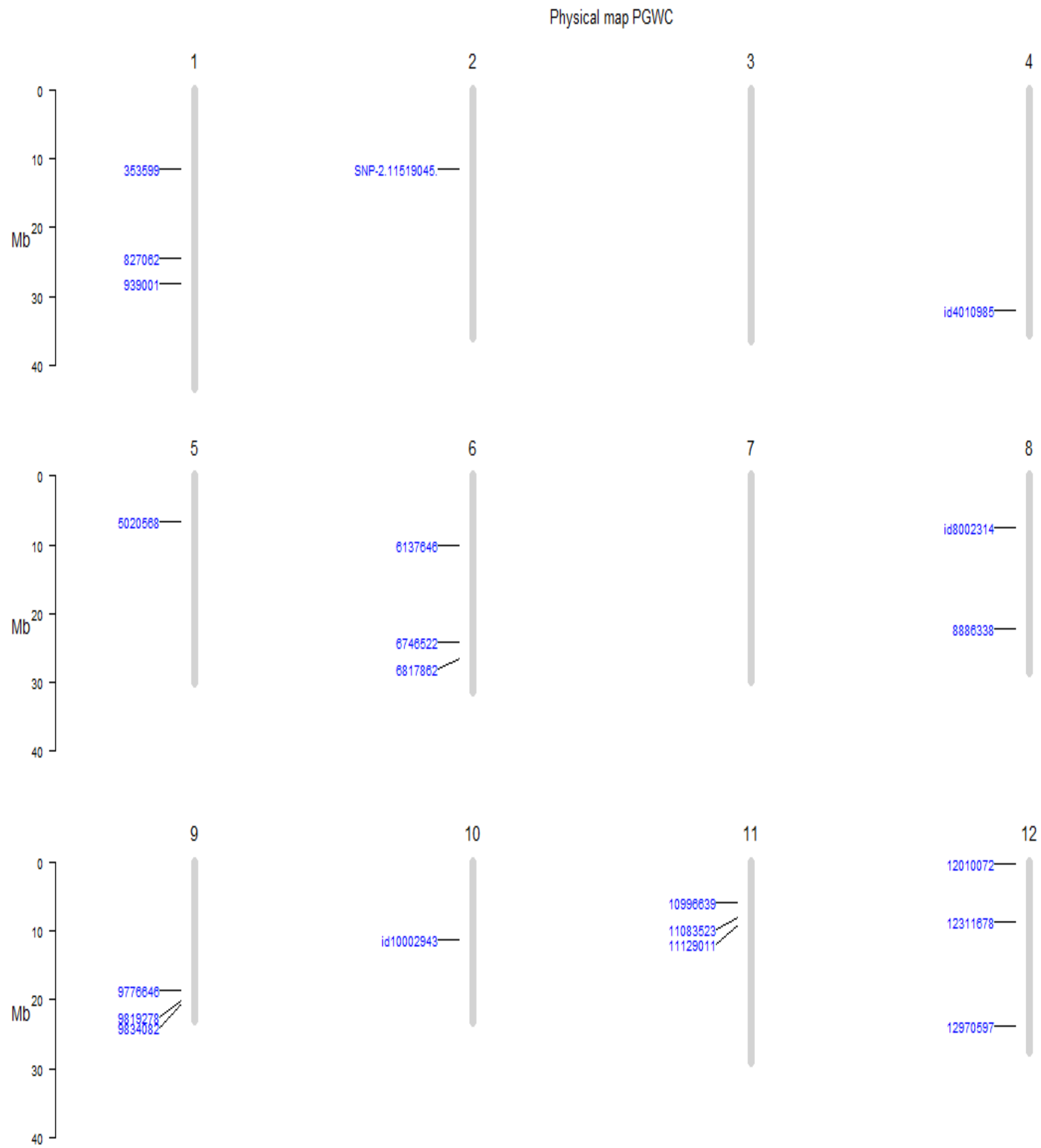
## 3. GLWR

Physical map GLWR

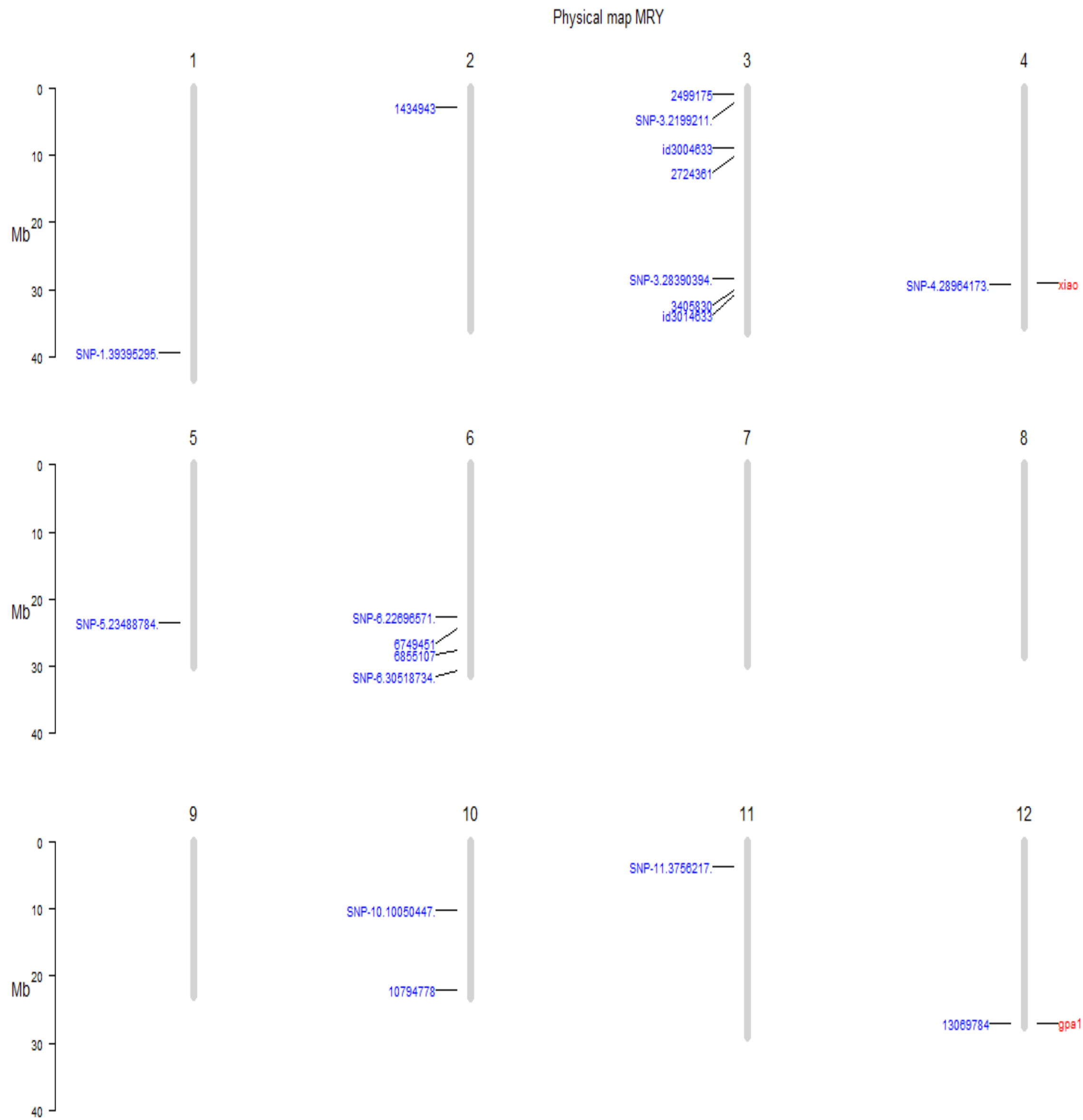
4. DEC



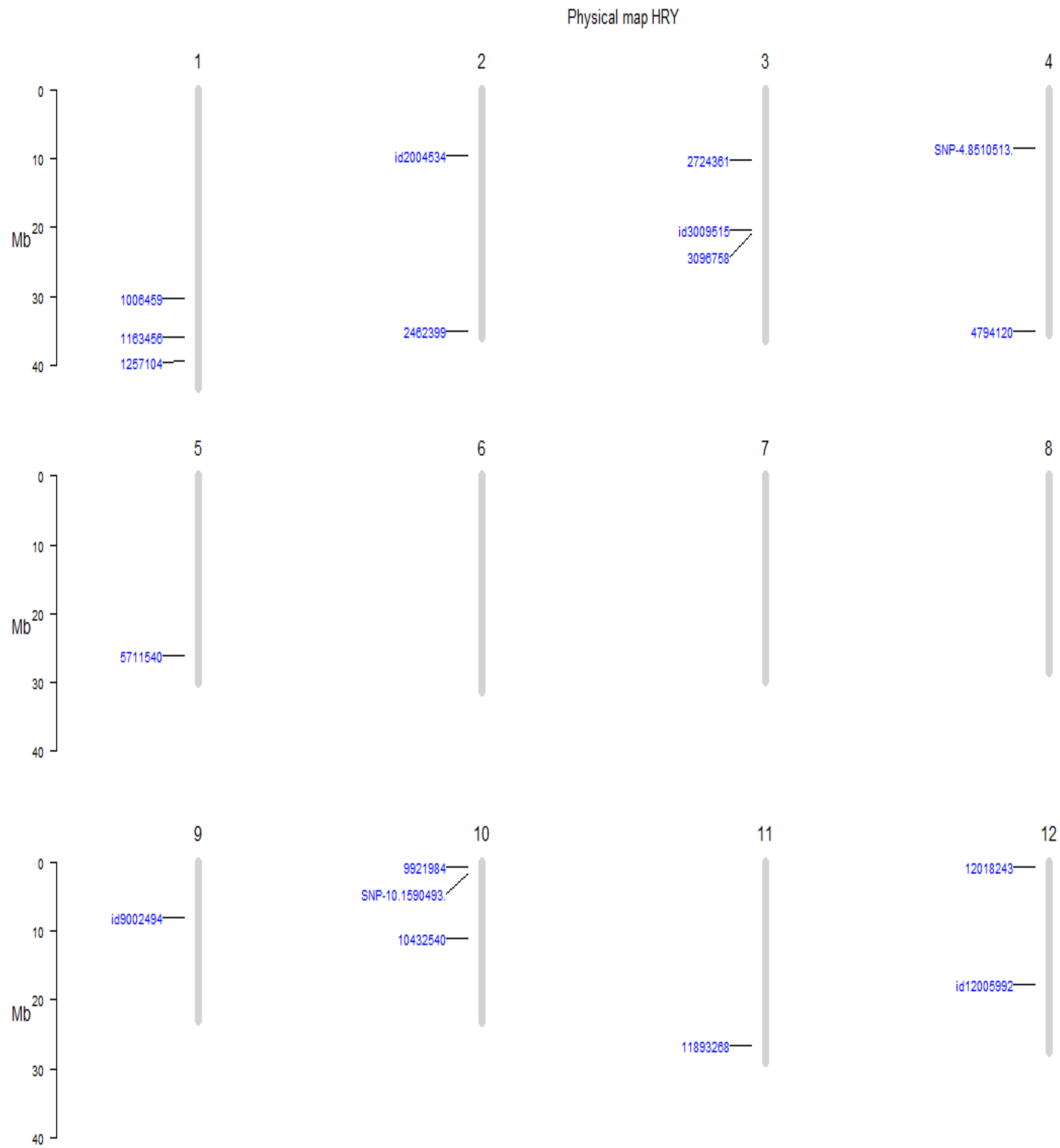
5. PGWC



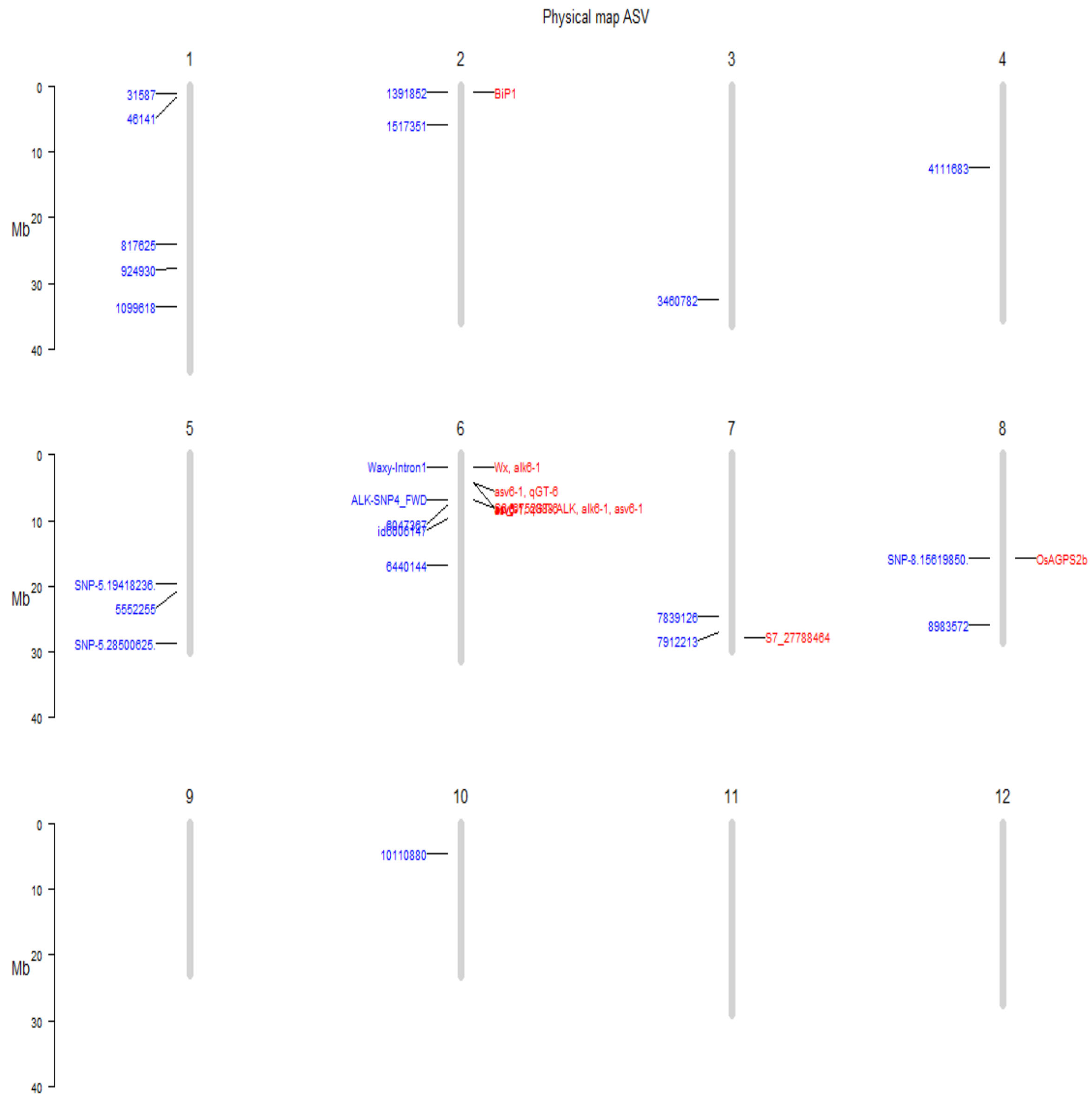
## 6. MRY



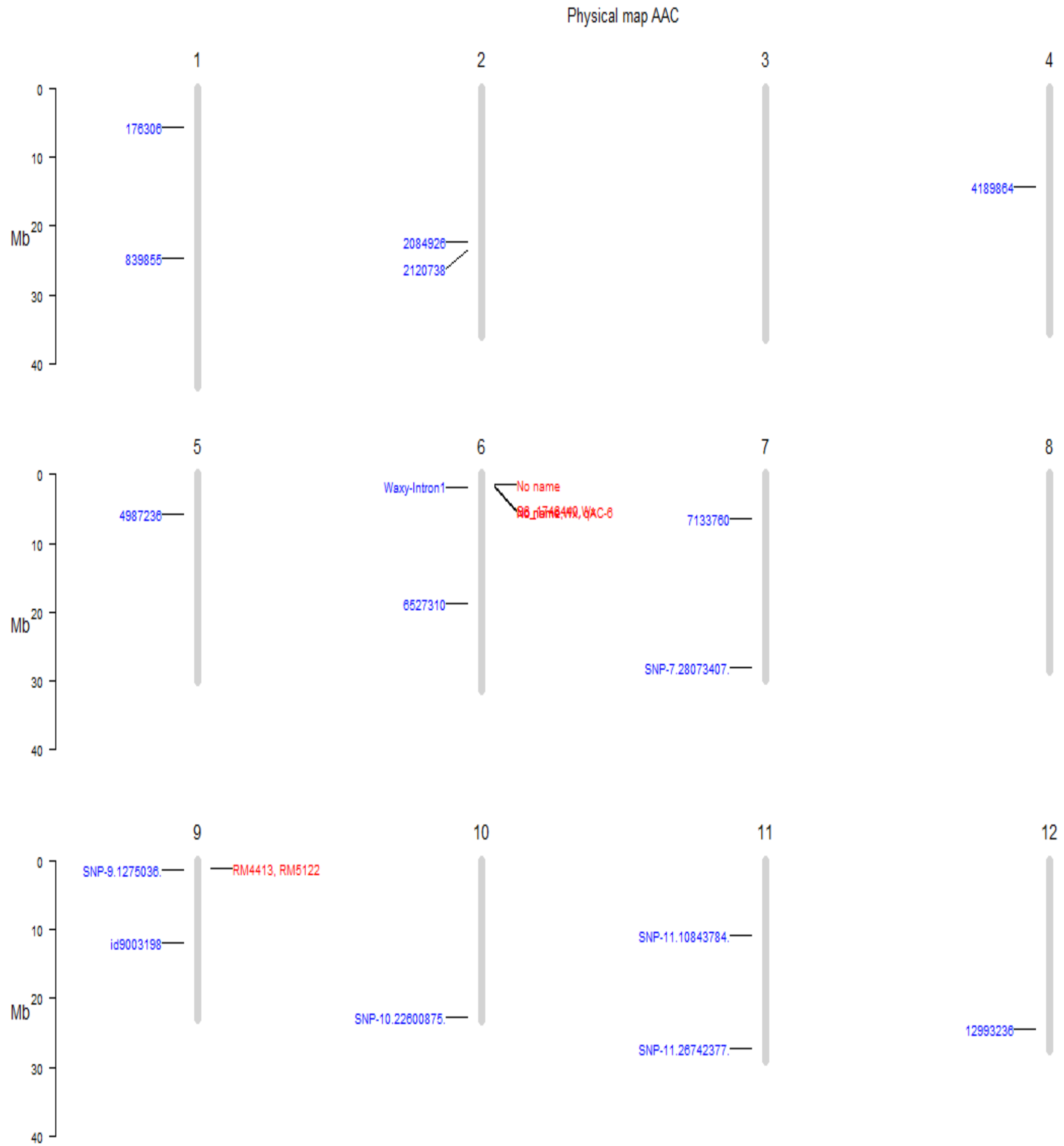
7. HRY



8. GT



9. AAC



10. PC