J. Trop. Agric. and Fd. Sc. 51 (2)(2023): 35 – 44

Comparative single nucleotide polymorphisms (SNPs) analysis of MR 297 and MRQ 76 rice varieties reveals potential SNPs in coding genes

Rabiatul Adawiah, Z. A.[1*], Norliza, A. B.[1], Sew, Y. S.[1], Shahril Firdaus, A. R.[1] and Sanimah, S.[2]

[1]*Biotechnology and Nanotechnology Research Centre, MARDI Headquarters, 43400 Serdang, Selangor, Malaysia*
[2]*Strategic Planning and Innovation Management Centre, MARDI Headquarters, 43400 Serdang, Selangor, Malaysia*

Abstract

MR 297 and MRQ 76 are commercial rice varieties in Malaysia, but the genomic information of these varieties are scarcely reported. This study aims to unravel the genomic variations in these two rice varieties, annotate the single nucleotide polymorphisms (SNP) and generate a list of potential SNPs in coding genes related to agronomically important genes. In this study, the genomes of MR 297 and MRQ 76 were sequenced using Illumina HiSeq 4000. Bioinformatics analysis was carried out to identify potential SNPs in both varieties. A total of 2,424,521 and 2,354,097 SNPs were identified in MR 297 and MRQ 76, respectively. Annotation of SNPs in coding genes identified 153,058 SNPs in MR 297 and 147,431 SNPs in MRQ 76. Deleterious SNPs analysis using SIFT4G found 12 deleterious SNPs in genes encoding seed metabolism, salinity, cell-wall biosynthesis, and disease resistance. Identification and annotation of SNPs in MR 297 and MRQ 76 suggested potential SNPs that valuable for PCR-based marker development or whole-genome genotyping to facilitate marker-assisted rice breeding (MAB) and marker-assisted selection (MAS) in rice breeding programme.

Keywords: MR 297, MRQ 76, *Oryza sativa*, rice, single nucleotide polymorphism

Introduction

High-throughput sequencing plays an essential role for exploring the genome of crop species. The genome sequence of crops provides the rapid identification of agronomically important genes and the discovery of molecular markers (i.e., microsatellite, single nucleotide polymorphism (SNP), insertion-deletion (InDel)). For example, the first large rice re-sequencing project has sequenced 3000 rice genomes from various rice sub-species and varieties to discover millions of SNPs in coding genes (Consortium et al. 2016). Previous studies have performed genome re-sequencing of extensive collections of rice germplasms (Tanaka et al. 2020) within rice sub-species (Lv et al. 2020) and pigmented rice varieties (Lachagari et al. 2019) to discover potential SNPs in coding genes that contribute to agronomic traits. These efforts indicate the usefulness and effectiveness of a high-throughput sequencing platform for researchers to exploit large-scale genomic variation in rice. Consequently, this genomic variation can be used as molecular markers to aid in the development of quantitative trait loci (QTL), marker-assisted selection (MAS) and marker-assisted breeding (MAB).

Genome-wide identification of single nucleotide polymorphism (SNP) in the coding genes has been widely used in investigating the genomic basis of phenotypic differences within and intra sub-species, genetic diversity study, and evolutionary relationships (Manrique-Carpintero et al. 2013; Rezaei et al. 2016; Vasconcelos et al. 2016). SNP is highly abundant in the genome, codominance inheritance and amenable for high-throughput genotyping (Edwards et al. 2005). In addition, the SNP in the coding region is crucial as the substition of amino acid may affect the protein function that subsequently may lead to differential phenotypic expression (Martinez-Garcia et al. 2013).

To date, information on polymorphisms in the coding genes of Malaysian white rice varieties remains limited. In this study, we report the genome re-sequencing of two Malaysian white rice varieties, namely MR 297 and MRQ 76. Genome re-sequencing is an approach for which

Article history
Received: 27.2.23
Accepted: 12.6.23

Authors' full names: Rabiatul Adawiah Zainal Abidin, Norliza Abu Bakar, Sew Yun Shin, Shahril Firdaus Ab Razak and Sanimah Simoh
Corresponding author: rabiatul@mardi.gov.my

a reference genome of species exists, and several rice genomes from *the japonica* and *indica* sub-species have been established (Guo et al. 2014). Hence, we performed the re-sequencing of MR 297 and MRQ 76. Additionally, the genome re-sequencing approach has been performed in many crop species and varieties to identify the variation in different species and cultivars (Guo et al. 2013; Thudi et al. 2016; Gramazio et al. 2019).

MR 297 and MRQ 76 are white, cultivated and modern rice varieties from the *indica* sub-species. MR 297 is a high-yielding variety and widely planted in Malaysia whereas MRQ 76 is fragrant rice variety and low amylose content. In addition, the demand for fragrant rice in Malaysia has increased due to it taste and aroma (Hashim et al. 2016). Therefore, it is of importance to gain insight into the genomics variation in these two rice varieties, which could be used in the future rice breeding programme. *Table 1* shows the morpho-physiological features of MR 297 and MRQ 76.

This study aims to unravel the genomic variations in these two rice varieties, focusing on the polymorphism in the coding genes that are responsible for expression of agronomically important traits, such as nutritional, aromatic, amylose content and stress response. We also identified unique SNPs between both rice varieties and annotated the SNPs into biological processes, molecular functions and pathways. The results from this study will facilitate us in developing more QTLs and molecular markers for application in in rice breeding programme.

Materials and methods

Plant materials

Seeds of MR 297 and MRQ 76 were obtained from MARDI Seberang Perai, Penang, Malaysia. Seeds were sterilised, incubated at 42 °C overnight and soaked in water for two days before being placed onto wet tissues or directly sowed into the soil.

DNA isolation and genome sequencing

Total DNA of each variety was extracted from leaves of two weeks old germinated seedlings using Murray and Motou protocol (Murray and Thompson 1980), and Sigma DNA extraction kit. Analysis of the DNA quality and quantity was carried out using a nanodrop spectrophotometer. The integrity of the DNA samples was determined using 0.8% agarose gel. The DNA samples were sent to the sequencing service provider and sequencing was performed using Illumina HiSeq4000 sequencing (Illumina, Inc. San Diego, CA, USA). The sequencing procedure was performed by following the standard Illumina protocol. The cleaned paired-end sequence reads from MR 297 and MRQ 76 have been deposited in the ENA database (www.ebi.ac.uk/ena) under the acession number PRJEB32344.

Table 1. Morpho-physiological features of MR 297 and MRQ 76

	MR 297	MRQ 76
Grain length (mm)	7.07	6.54
Width grain (mm)	2.07	1.96
Maturity day	110 days	117 days
Amylose content	23.3 (hard)	16.0 (intermediate)
Plant height (mm)	6.4 – 7.0	7.8
Stem length (mm)	2.1 – 2.7	2.2
Weight 1,000 grains (g)	27.8 – 29.2	25.2

Reads mapping and identification of SNPs

The pair-end sequencing reads from MR 297 and MRQ 76 with the read length of 150bp at each end were aligned against Nipponbare genome sequences using Burrows-Wheeler Aligner (BWA) (H. Li and Durbin, 2009) software with default parameters except for "mem -m 10000 -o 1 -e 10 -t 4". The mapped reads were merged and indexed as BAM files. All reads from each variety were cleaned by removing PCR duplicate reads using PICARD version 0.7.12. Haplotype Caller package in (GATK) version 3.6 was used for SNP calling with default parameters except a minimum phred-scaled confidence threshold of 50, a minimum phred-scaled confidence threshold for emitting variants at 10, every site in a genome were set at >30 for mapping quality, >50 for variant quality and >10 for the number of supporting reads for every base.

Annotation of SNPs

SnpEff version 4.1 was used to annotate the SNPs into three different categories: i) intergenic, ii) genic and iii) non-coding (Cingolani et al. 2012). The genic SNPs were classified into four different genomic regions: i) exon, ii) 3'prime untranslated region (UTR), iii) 5'prime UTR and iv) intron. The SNPs in the exon region were divided into synonymous and non-synonymous amino acid substitutions. Filtering of annotated SNPs according to the above criteria was performed using R packages (dataframe plyr (dplyr), structured query language data frame (sqldf) and tidy R language (tidyr). Unique SNPs were identified and compared among the rice cultivars using R scripts. The SNPs from different rice cultivars were retrieved from Ensembl Plants Variation database (http://plants.ensembl.org/Oryza_indica/Info/Index).

Functional enrichment of non-synonymous SNPs (nsSNPs)

Gene ontology (GO) and pathway analysis were performed using ClueGO plugin (Bindea et al. 2009) in Cytoscape version 3.6.1 (Cline et al. 2007) to decipher the biological interpretation of these non-synonymous SNPs (nsSNPs). A two-sided hypergeometric test and multiple testing by Bonferroni were selected to identify the GO terms and pathways. Two-sided hypergeometric

test represents enrichment or depletion of GOs and pathways. It is the most appropriate and standard parameter for GO and pathway annotation analyses in ClueGO (Rivals et al. 2007).

Prediction of deleterious and tolerated non-synonymous SNPs

Deleterious and tolerated of non-synonymous SNPs were predicted using the SIFT4G (Sorting Intolerant from Tolerant) (Kumar et al. 2009). In SIFT4G, the SNPs are classified as deleterious if tolerance index ≤0.05 whereas tolerated if tolerance index ≥0.05. We used list of agronomically important genes from The Rice Annotation Project database (RAP-DB) (https://rapdb.dna.affrc.go.jp/agri_genes/agri_gene_list.html) to screen the potential genes containing deleterious SNPs in MR 297 and MRQ 76 rice varieties *(Supplementary Table 1)***.**

Results and discussion

Genome re-sequencing of MR 297 and MRQ 76 and identification of SNPs

Using Illumina HiSeq 4000, 96.8 million reads (14.52 Gb) and 96.5 million reads (14.48 Gb) were generated in MR 297 and MRQ 76, respectively (*Table 1*). The sequencing depth was 30X coverage. 30X coverage is sufficient for variants discovery (Sims et al. 2014). This genome re-sequencing has covered 89.3% and 88.50% genome of MR 297 and MRQ 76, respectively. After removing adapter and low quality base, clean reads achieved 99.26% and 99.36% for MR 297 and MRQ 76, respectively. The clean reads were mapped on the *Oryza sativa japonica* cultivar Nipponbare genome, which resulted in 96.54% and 96.64% mapping rates in MR 297 and MRQ 76, respectively (*Table 1*). The difference in mapping rates could be related to the divergence between *indica* and *japonica* sub-species (Xu et al. 2012) and the difference in genomic features that could not be shared between the genomes of *indica* and *japonica* sub-species (Schatz et al. 2014). For instance, structural variants (i.e., insertions, deletions, inversions) and genes that contributed to domestication can be identified as divergent between these two genomes of sub-species (Wang et al. 2014; Kou et al. 2020).

We mapped the genomes of MR 297 and MRQ 76 onto *O. sativa cv*. Nipponbare. Nipponbare is a high-quality and well-annotated reference genome. Although several *indica* rice genomes are available and well-assembled (i.e., 93-11, HR-12, R498, IR64), they need to be well-annotated as compared to Nipponbare. For instance, many hypothetical and unknown genes annotation in 93-11 requires the user to annotate the gene sequences for further description. In addition, the gene identifier of Nipponbare is widely used in many bioinformatics tools and databases, which ease the user to analyse the genes into biologically meaningful information. Previous studies have used the Nipponbare to map onto *indica* rice variety (Jain et al. 2014 ; Srivasta et al. 2014; Rathinasabapathi et al. 2015). According to a previous report, the difference between the genomes of *indica* and *japonica* sub-species should not be too different (Campbell et al. 2020).

A total of 2,425,462 and 2,355,126 raw putative SNPs were identified in MR 297 and MRQ 76, respectively (approximately six SNPs for every one kilobase of MR 297 and MRQ 76 rice genomes). In addition, SNP filtering identified 479,341 high-quality SNPs in MR 297, whereas 468,254 high-quality SNPs in MRQ 76 (*Table 1*). This finding shows a higher frequency of SNPs was found in MR 297 as compared to MRQ 76. The frequency of SNPs indicates the measurements of genetic diversity between two individuals or within the population (Edwards et al. 2005). The different SNPs frequency between the individuals was likely due to the combination of mutation rate, generating new polymorphism either through natural or artificial selection in the breeding activity (Edwards et al. 2005). High SNPs frequency in MR 297 could indicate MR 297 has received new genes during the artificial breeding selection.

SNPs annotation to identify the functional effects of SNPs

A total of 153,058 SNPs in the MR 297 and 147,431 SNPs in the MRQ 76, were annotated in the coding regions of rice genes (*Table 1*). 330,682 and 324,961 SNPs were located in the intergenic regions of MR 297 and MRQ 76, respectively. High-number of SNPs in intergenic regions has been commonly observed in rice (Ebana et al. 2010; Mehra et al. 2015; Tatarinova et al. 2016).

A total of 76,213 genic SNPs were identified in the intron, followed by 44,347 SNPs in the exon and 32, 498 in the UTRs of MR 297. In MRQ 76, 73,581 genic SNPs were found in the intron, followed by 42,461 SNPs in the exon and 32,498 SNPs in the UTR (*Figure 1*). SNP in the intron region resides at the splice junction, intron splice enhancer or silence elements (Coulombe-Huntington et al. 2009). Furthermore, SNPs in the intron region have the potential to alter the DNA splicing and regulate the transcript level by changing the binding sites of miRNAs, which reside in the intron region (Wang and Cooper, 2007).

Low-frequency of SNPs were identified in the 3' prime UTR and 5' prime UTR. Fewer SNPs in this region due to the amino acid conservation (Tatarinova et al. 2016). Mutation in the UTR is capable of changing the local mRNA structure near the 3' and 5' cap and potentially affect the translation process (Tatarinova et al. 2016).

Identification of unique SNPs

Unique SNP shows the presence of alleles only in one variety and with different alleles in multiple varieties or individuals *(Figure 2)*. Hence, unique SNPs can be used to investigate the relationship and differentiate between accessions and varieties. In this study, 76,645 unique SNPs were identified between MR 297 and MRQ 76 (*Figure 3*)

(Supplementary Table 2). The proportion of unique SNPs between MR 297 and MRQ 76 is low (42.68%), showing that these two varieties tend to share the alleles. This finding is consistent with previous report that MRQ 76 is one of the parental lines to MR 297 (Sunian et al. 2022).

Of these, 135 unique SNPs were identified after comparing with SNP data from Ensembl Plants Variation, consisting the variation data from six large-scale SNP studies from various cultivars *(Supplementary Table 2)*. This finding indicates that many SNPs have been discovered from various rice cultivars by rice genome-sequencing efforts from time to time. The application of unique SNPs has been demonstrated in seed purity testing and DNA fingerprinting analysis (Rhomdane et al. 2018; Ooi et al. 2019; Josia et al. 2021).

Functional annotation on non-synonymous and deleterious SNPs

A total of 23,636 nsSNPs in the MR 297 and 22,640 nsSNPs in the MRQ 76 were further analysed into pathway, biological process and molecular function terms to gain better insight into the biological function of the potential genes and proteins (*Table 2*). The biological process (BP) terms, such as peptide transport (GO:0015833) and positive regulation of transcription DNA-template (GO:0045893) were enriched in genes containing nsSNPs of MR 297. For molecular function (MF) terms, such primary active transmembrane transporter activity (GO:0015399), ether lipid metabolism (GO:0046485), ion transmembrane transporter activity (GO:0046873), and histone binding (GO:0042393) were enriched in genes containing nsSNPs of MRQ 76 *(Figure 4a)*. GO terms, such as metal ion transmembrane transporter activity, primary active transmembrane transporter activity and peptide transport are likely related to the catalysis and transport of substrates throughout the cell membrane (Miyadate et al. 2011; Ueno et al. 2011).

In MRQ 76, BP terms, such as oxylipin metabolic process (BP) (GO:0031407) and molecular function of coenzyme binding (MF) (GO:0050662) were enriched in genes containing nsSNPs *(Figure 4b)*. The GO term coenzyme binding is likely associated with plant growth and development (Lu et al. 2013) whereas the oxylipin metabolic process is likely related to the defence response in rice (Chehan et al. 2007). A previous study has reported that repression of metabolic genes during stress might be able to defend the plant from pathogen attack (Zeier 2013).

Five pathways, including diterpenoid biosynthesis (osa00904), glycerophospholipid metabolism (osa00564), starch and sucrose metabolism (osa00500) and amino sugar and nucleotide sugar metabolism (osa00520), and aminoacyl-tRNA biosynthesis (osa00970) were enriched

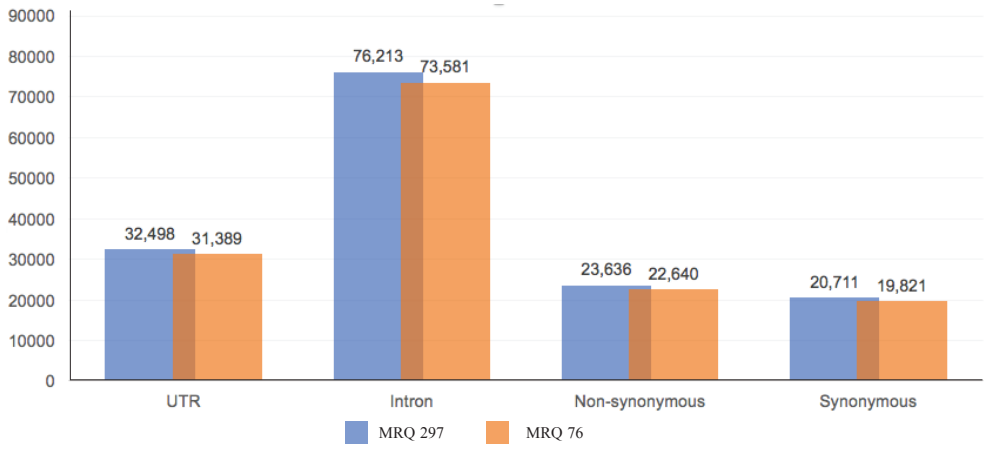

Figure 1. Frequency of SNPs in each functional effect of MR 297 and MRQ 76. The functional effects were classified into UTR, intron, non-synonymous and synonymous

	MR 297	MRQ 76	Acc 1	
snp001	C	T	T	unique SNPs
snp002	A	G	A	
snp003	T	A	G	

Figure 2. Unique SNPs show the allele presence in one variety and with different alleles combinaBon in other varieBes

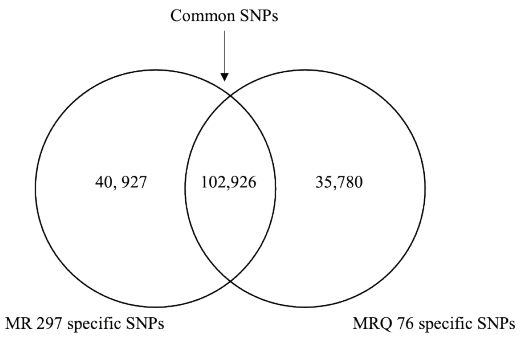

Figure 3. Unique and common SNPs present in MR 297 and MRQ 76

Table 2. Summary of genome mapping and SNPs discovery from MR 297 and MRQ 76

	MR 297	MRQ 76
Total reads (bp)	96,838,269	96,537,127
Total data (Gb)	14.52	14.48
Total clean reads (bp)	96,121,666 (99.26%)	95,919,290 (99.36%)
Total mapped reads	92,795,856 (96.54%)	92,696,401 (96.64%)
Total raw SNPs	2,424,521	2,354,097
Total high-quality SNPs	479,341	468,254
Total SNPs in coding genes	153,058	147,431
Total SNPs in intergenic region	330,682	324,961

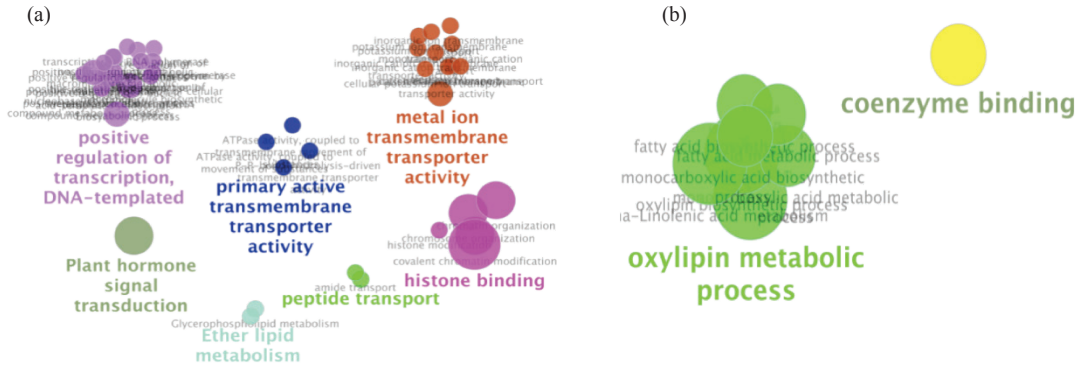

Figure 4. Gene ontology enrichment analysis of rice gene families that were contained non-synonymous SNPs. The gene ontology term was grouped into biological process, molecular function and cellular component. a) The most significant biological process and molecular function in MR 297. b) The most significant biological process and molecular function in MR 297 in MRQ 76

in genes containing nsSNPs of MR 297 and MRQ 76. Starch and sucrose metabolism has been associated with plant abiotic stress, such as high salinity, drought resistance (Thalmann and Santelia 2017) and submergence (Lee et al. 2019).

The nsSNP or missense SNP are categorised into deleterious or tolerated SNPs. Deleterious SNPs may alter the conserved regions, affect the protein stability and cause phenotypic changes (Kono et al. 2016). In this study, SIFT4G predicted 775 nsSNPs were deleterious in MR 297 while 677 nSNPs were predicted as deleterious in MRQ 76 (*Table 2*) *(Supplementary Table 3)*. Annotation of deleterious SNPs identified 30 potential deleterious SNPs in agronomically important genes, such as seed metabolism, disease resistance, cell wall biosynthesis, aromatic, salinity, yield, flowering and quality (*Table 3*).

Deleterious SNP (chr06_4799710) predicted in the sucrose synthase 2 (*RSUS2*) (*Os*06g0194900), a gene which is involved in the carbohydrate metabolism and sucrose biosynthesis of rice. *RSUS2* has been suggested to alter the sugar metabolism under submerged conditions (Fukuda et al. 2008). Expression of *RSUS2* in the germinated rice seedlings and grown in the submerged conditions suggests its role in increasing sucrose metabolism. In another study, *RSUS2* was found to be involved in the starch biosynthetic pathway, which is contribute for developing the rice grain and vegetative tissues (Chen et al. 2012). This finding may indicate the involvement of *RSUS2* in grain quality and response to submerged condition in rice. To date, no further study has been carried out that investigates the loss-of-function in the *RSUS2* gene. Hence, *RSUS2* could be suggested to be studied in the future.

Deleterious SNP (chr06_662221) in MRQ 76 was found in the cellulose synthase-like protein D2 (*CSLD2*) (Os06g0111800). *CSLD2* has been reported as a plasma membrane protein, which is responsible for the biosynthesis of cell wall polysaccharides (Luan et al. 2011; Marcotuli et al. 2018; Nater et al. 2015). As *CSLD2* is one of the members in the CSL superfamily, thus it is a promising candidates for encoding the glycosyl synthases that make the hemicellulose backbones of plant cell walls (Li et al. 2009). Previous studies have revealed the involvement of cell wall biosynthesis in several agronomic traits in rice, such as resistance to disease (Hu et al. 2017) and grain yield (Xu et al. 2017)

Although MRQ 76 is an aromatic rice, only one nsSNP (chr08_20382857) was identified in the *BADH2* (Os08g0424500) gene encoding aromatic. This nsSNPs was not predicted as deleterious SNP.

Two nsSNPs were identified in gene encoding GTPCHI. GTPCHI is a gene in folate biosynthetic pathway and responsible for seed metabolism in rice (Dong et al., 2014). Previous study has identified GTPCHI in the QTL region of weedy rice, indicating this gene could be potential to be responsible for weedy rice domestication (Islam et al. 2020). Therefore, these two SNPs in GTPCHI are suggested for further validation using a low-

Table 3. Summary of non-synonymous and deleterious SNPs in selected genes that are responsible to agronomically important traits

Traits	Gene name/gene ID	Description	MR 297 allele	MRQ 76 allele	SNP ID	Chromosome	SIFT4G score	AA change
Seed metabolism	*RSUS2* (Os06g0194900)	Sucrose synthase 2	T	C	chr06_4799710	Chr 6	0.008	T82A
	GTPCHI (Os04g0662700)	Similar to GTP cyclohydrolase 1 isoform TaA	C	C	chr04_33815518	Chr 4	-	-
Disease resistance	*OsGLP3-4* (Os03g0693800)	Oxalate oxidase 2, Positive regulation of panicle blast resistance.	G	A	chr03_27784328	Chr 3	0.032	T85M
	GLP8 (Os08g0188900)	Germin-like protein 8-1, disease resistance	C	T	chr08_5187250	Chr 8	0.016	T71M
	BAG6 (Os11g0506800)	Gene regulation and stress responses	G	A	chr11_18071614	Chr 11	0.024	R46W
	OsCAF1CL2 (Os09g0540600)	Similar to WD-40 repeat protein MSI1	T	G	chr09_21276714	Chr 9	0.005	F64V
	OsCAX4 (Os02g0138900)	calcium:cation antiporter/	T	G	chr02_2074088	Chr 2	0.002	H58Q
	OsHyPRP09 (Os10g0349300)	Similar to root-specific protein	T	C	chr10_10563484	Chr 10	0.029	T93A
	OsPP2C22 (Os02g0607500)	Putative protein phosphatase 2C 22	G	A	chr02_23823651	Chr 2	0.02	A250V
Cell wall biosynthesis	*CSLD2* (Os06g0111800)	Similar to CSLD2	G	A	chr06_662221	Chr 6	0.01	S273G
	OsIDD11 (Os01g0572300)	Zinc finger, C2H2-type domain containing protein	A	T	chr01_21993686	Chr 1	0.036	N40K
Aromatic	*BADH2* (Os08g0424500)	Betaine aldehyde dehydrogenase, Rice fragrance	A	C	chr08_20382857	Chr 8	-	-
Salinity	*OsHAK21* (Os03g0576200)	High affinity K^+ transporter	G	A	Chr03_21060651	Chr 3	0.029	S796L
	OsGA2ox5 (Os07g0103500)	2OG-Fe(II) oxygenase domain containing protein	T	G	Chr07_218893	Chr 7	0.017	D227G

(Cont.)

Table 3. (*Cont.*)

Traits	Gene name/gene ID	Description	MR 297 allele	MRQ 76 allele	SNP ID	Chromosome	SIFT4G score	AA change
Yield	*TGW6* Os06g0623700	Protein with indole-3-acetic acid (IAA)-glucose hydrolase activity, regulation of grain length and weight	T	G	chr06_25094225	Chr 6	0.048	E328D
	OsSPL4 Os02g0174100	Unknown phosphatidylethanolamine-binding protein (PEBP) like domain protein, 426-amino-acid protein, homologous to the keratin-associated protein (KAP) 5-4 family, panicle architecture, panicle erectness	C	T	chr02_4075194	Chr 2	0	R139W
	OsDLT Os06g0127800	GAI-RGA-SCR (GRAS) family protein, brassinosteroid signaling	C	G	chr06_1466091	Chr 6	0.032	G575R
Flowering	*OsCRY2* Os02g0625000	Cryptochrome 2, blue light photoreceptor, promotion of flowering time	G	A	chr02_24916475	Chr 2	0.045	D18N
	OsPHYB Os03g0309200	Similar to phytochrome B	T	C	chr03_11021139	Chr 3	0.018	Y219H
	OsPHD36 Os08g0105000	Homeodomain (PHD) transcriptional regulator, flowering promoter	A	G	chr08_276669	Chr 8	0.047	P8L
Quality	*OsSBEI* Os06g0726400	Starch branching enzyme	T	C	chr06_30899986	Chr 6	0.015	G607D

throughput genotyping platform (i.e., KASPar), which could serve as a potential functional marker for weedy rice identification.

Seven deleterious SNPs were identified in several genes related to disease resistance traits (*Table 3*). For instance, germin-like proteins (*OsGLP3-4* and *GLP8*) are the genes that play a role in the regulation of disease resistance and plant height (Banerjee and Maiti, 2010). In addition, a previous study has revealed that the Bcl-2-associated athanogene (*BAG*) genes exhibited diverse expression patterns, such as response to stress (Zhou et al. 2021). Hence, the nsSNPs in genes related to disease resistance could be suggested to be studied in the future.

OsHAK21 is a family member of the High-Affinity (K+) transporter that plays a crucial role in the rice salt tolerance mechanism (Shen et al. 2015). In this study, one deleterious SNP was identified in *OsHAK21* (*Table 3*). Therefore, validating this potential nsSNPs using the genotyping platform could serve as a functional marker for salt tolerance traits in rice.

Identification of deleterious SNPs in the MR 297 and MRQ 76 is essential in prioritising the candidate SNPs and genes that responsible for agronomically important traits in rice. Prediction of deleterious SNP is based on sequence conservation approach, showing conserved region is functionally important region. In addition, conserved regions in the *O. sativa* could be due to the selective sweeps (Duitama et al. 2015). The potential SNPs discussed in this study could be prioritised and suggested for genotyping work, which could be applied in the rice breeding programme.

Conclusion

Genome re-sequencing of SNPs in MR 297 and MRQ 76 rice genomes highlights the wealth of genomic variations present in MR 297 and MRQ 76. The short reads sequence from MR 297 and MRQ 76 were mapped against Nipponbare rice genome sequences. This mapping has resulted in the identification of SNPs between MR 297 and MRQ 76. The functional effect of SNPs was observed to investigate their potential effects in several biosynthetic genes related to essential traits, such as resistance to disease, response to abiotic stress (i.e., salinity) and seed metabolism. This effort could be important in exploiting the DNA variations that are responsible for traits of interest.The identified SNPs in local rice varieties will serve as a useful resources, which in return will accelerate molecular genetics in rice breeding programme.

Acknowledgement

This work was supported by Malaysian Agricultural Research and Development Institute (MARDI) Development Fund (PRB401). The authors would like to thank Malaysian Agricultural Research and Development Institute (MARDI) for financial support. The authors are grateful to colleagues from Paddy and Rice Research Centre (RI) and Rice GeneBank, MARDI for their contributions to the germplasm collections and suggestions.

References

Banerjee, J., & Maiti, M. K. (2010). Functional role of rice germin-like protein1 in regulation of plant height and disease resistance. *Biochemical and biophysical research communications*, *394(1)*, 178 – 183. https://doi.org/10.1016/j.bbrc.2010.02.142

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W. H., Pages, F., Trajanoski, Z., & Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091 – 1093. https://doi.org/10.1093/bioinformatics/btp101

Campbell, M. T., Campbell, M. T., Du, Q., Liu, K., Sharma, S., Sharma, S., Zhang, C., & Walia, H. (2020). Characterization of the transcriptional divergence between the subspecies of cultivated rice (*Oryza sativa*). *BMC Genomics*, *21*(1), 1 – 16. https://doi.org/10.1186/s12864-020-06786-6

Chehan, E. W., V.Perea, J., Gopalan, B., Theg, S., & Dehesh, K. (2007). Oxylipin Pathway in Rice and Arabidopsis. *Journal of Integrative Plant Biology*, *49*(1), 43 – 51. https://doi.org/10.1111/j.1672-9072.2007.00405.x

Chen, Y., Wang, M., & Ouwerkerk, P. B. F. (2012). Molecular and environmental factors determining grain quality in rice. *Food and Energy Security*, *1*(2), 111 – 132. https://doi.org/10.1002/fes3.11

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*, *6*(2), 80 – 92. https://doi.org/10.4161/fly.19695

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Bader, G. D. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc*, *2*(10), 2366–2382. https://doi.org/10.1038/nprot.2007.324

Consortium, I. R. P., Hao, L., Zhang, H., Zhang, Z., Hu, S., & Xue, Y. (2016). Information Commons for Rice (IC4R). *Nucleic Acids Res*, *44*(D1), D1172-80. https://doi.org/10.1093/nar/gkv1141

Coulombe-Huntington, J., Lam, K. C., Dias, C., & Majewski, J. (2009). Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS Genet*, *5*(12), e1000766. https://doi.org/10.1371/journal.pgen.1000766

Dong, W., Cheng, Z. jun, Lei, C. lin, Wang, X. le, Wang, J. J. lin, Wang, J. J. lin, Wu, F. qing, Zhang, X., Guo, X. ping, Zhai, H. qu, & Wan, J. min. (2014). Overexpression of Folate Biosynthesis Genes in Rice (Oryza sativa L .) and Evaluation of Their Impact on Seed Folate Content. *Plant Foods Human Nutrition*, *69*(4), 379 – 385. https://doi.org/10.1007/s11130-014-0450-9

Duitama, J., Silva, A., Sanabria, Y., Cruz, D. F., Quintero, C., Ballen, C., Lorieux, M., Scheffler, B., Farmer, A., Torres, E., Oard, J., & Tohme, J. (2015). Whole genome sequencing of elite rice cultivars as a comprehensive information resource for marker assisted selection. *PLoS One*, *10*(4), e0124617. https://doi.org/10.1371/journal.pone.0124617

Ebana, K., Yonemaru, J., Fukuoka, S., Iwata, H., Kanamori, H., Namiki, N., Nagasaki, H., & Yano, M. (2010). Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (Oryza sativa L.). *Breeding Science*, *60*(4), 390 – 397. https://doi.org/10.1270/jsbbs.60.390

Edwards, D., Forster, J., Chagné, D., Batley, J., & p, li white-space: pre-wrap and. (2005). WHAT ARE SNPs? In C. R. E. H. A. G. S. E. de S. H. N. Oraguzie Nnadozie (Ed.), *Association mapping in plants* (Vol. 1, pp. 41 – 52). Springer-Verlag New York.

Fukuda, A., Yoshinaga, S., Nagata, K., & Shiratsuchi, H. (2008). Rice Cultivars with Higher Sucrose Synthase Activity Develop Longer Coleoptiles under Submerged Conditions. *Plant Production Science*, *11*(1), 67 – 75. https://doi.org/10.1626/pps.11.67

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J., & Bombarely, A. (2019). Whole-Genome Resequencing of Seven Eggplant (Solanum melongena) and One Wild Relative (S. incanum) Accessions Provides New Insights and Breeding Tools for Eggplant Enhancement. *Frontiers in Plant Science*, *10*(October), 1 – 17.

Guo, L., Gao, Z., & Qian, Q. (2014). Application of resequencing to rice genomics, functional genomics and evolutionary analysis. *Rice*,7(4), 1 – 10. https://doi.org/10.1186/s12284-014-0004-7

Hashim, S., Norsuha Misman, S., Khairani Abu Bakar, N., Najib Mohd Yusof, M., Naim Fadzli Mohd Rani, M., Ramli, A., Mohd Yusob, S., Shahida, H., Siti Norsuha, M., Nur Khairani, A., Mohamad Najib, M., Muhammad Naim Fadzli, M., Asfaliza, R., & Shajarutulwardah, M. Y. (2016). Effect of organic fertiliser as a basal fertiliser on growth, yield and disease incidence of local fragrant rice varieties. *J. Trop. Agric. and Fd. Sc*, *44*(2), 167 – 178.

Hu, K., Cao, J., Zhang, J., Xia, F., Ke, Y., Zhang, H., Xie, W., Liu, H., Cui, Y., Cao, Y., Sun, X., Xiao, J., Li, X., Zhang, Q., & Wang, S. (2017). Improvement of multiple agronomic traits by a disease resistance gene via cell wall reinforcement. *Nat Plants*, *3*, 17009. https://doi.org/10.1038/nplants.2017.9

Islam, M. S., Coronejo, S., & Subudhi, P. K. (2020). Whole-genome sequencing reveals uniqueness of black-hulled and straw-hulled weedy rice genomes. *Theoretical and Applied Genetics*, *133*, 2461 – 2475. https://doi.org/10.1007/s00122-020-03611-2

Jain, M., Moharana, K. C., Shankar, R., Kumari, R., & Garg, R. (2014). Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant biotechnology journal*, *12*(2), 253 – 264. https://doi.org/10.1111/pbi.12133

Josia, C., Mashingaidze, K., Amelework, A. B., Kondwakwenda, A., Musvosvi, C., & Sibiya, J. (2021). SNP-based assessment of genetic purity and diversity in maize hybrid breeding. *PLoS ONE*, *16*(8 August), 1 – 14. https://doi.org/10.1371/journal.pone.0249505

Kono, T. J., Fu, F., Mohammadi, M., Hoffman, P. J., Liu, C., Stupar, R. M., Smith, K. P., Tiffin, P., Fay, J. C., & Morrell, P. L. (2016). The Role of Deleterious Substitutions in Crop Genomes. *Mol Biol Evol*, *33*(9), 2307 – 2317. https://doi.org/10.1093/molbev/msw102

Kou, Y., Liao, Y., Toivainen, T., Lv, Y., Tian, X., Emerson, J. J., Gaut, B. S., & Zhou, Y. (2020). Evolutionary genomics of structural variation in asian rice (Oryza sativa) domestication. *Molecular Biology and Evolution*, *37*(12), 3507 – 3524. https://doi.org/10.1093/molbev/msaa185

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073 – 1082.

Lachagari, V. B. R., Gupta, R., Lekkala, S. P., Mahadevan, L., Kuriakose, B., Chakravartty, N., Mohan Katta, A., Santhosh, S., Reddy, A. R., & Thomas, G. (2019). Whole Genome Sequencing and Comparative Genomic Analysis Reveal Allelic Variations Unique to a Purple Colored Rice Landrace (Oryza sativa ssp. indica cv. Purpleputtu). *Front Plant Sci*, *10*, 513. https://doi.org/10.3389/fpls.2019.00513

Lee, H. S., Hwang, W. H., Jeong, J. H., Ahn, S. H., Baek, J. S., Jeong, H. Y., Park, H. K., Ku, B. I., Yun, J. T., Lee, G. H., & Choi, K. J. (2019). Analysis of the distribution of assimilation products and the characteristics of transcriptomes in rice by submergence during the ripening stage. *BMC Genomics*, *20*(1), 18. https://doi.org/10.1186/s12864-018-5320-7

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754 – 1760. https://doi.org/10.1093/bioinformatics/btp324

Li, M., Xiong, G., Li, R., Cui, J., Tang, D., Zhang, B., Pauly, M., Cheng, Z., & Zhou, Y. (2009). Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *Plant J*, *60*(6), 1055 – 1069. https://doi.org/10.1111/j.1365-313X.2009.04022.x

Lu, Z., YU, H., Xiong, G., Wang, J., Jiao, Y., Liu, G., Jing, Y., Meng, X., Hu, X., Qian, Q., Fu, X., Wang, Y., & Li, J. (2013). Genome-Wide Binding Analysis of the Transcription Activator IDEAL PLANT ARCHITECTURE1 Reveals a Complex Network Regulating Rice Plant Architecture. *The Plant Cell*, *25*(10), 3743 – 3759. https://doi.org/10.1105/tpc.113.113639

Luan, W., Liu, Y., Zhang, F., Song, Y., Wang, Z., Peng, Y., & Sun, Z. (2011). OsCD1 encodes a putative member of the cellulose synthase-like D sub-family and is essential for rice plant architecture and growth. *Plant Biotechnol J*, *9*(4), 513 – 524. https://doi.org/10.1111/j.1467-7652.2010.00570.x

Lv, Q., Li, W., Sun, Z., Ouyang, N., Jing, X., He, Q., Wu, J., Zheng, J., Zheng, J., Tang, S., Zhu, R., Tian, Y., Duan, M., Tan, Y., Yu, D., Sheng, X., Sun, X., Jia, G., Gao, H., Yuan, D. (2020). Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. *Nat Commun*, *11*(1), 4778. https://doi.org/10.1038/s41467-020-18608-0

Manrique-Carpintero, N. C., Tokuhisa, J. G., Ginzberg, I., Holliday, J. A., & Veilleux, R. E. (2013). Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. *G3 (Bethesda)*, *3*(9), 1467 – 1479. https://doi.org/10.1534/g3.113.007146

Marcotuli, I., Colasuonno, P., Blanco, A., & Gadaleta, A. (2018). Expression analysis of cellulose synthase-like genes in durum wheat. *Sci Rep*, *8*(1), 15675. https://doi.org/10.1038/s41598-018-34013-6

Martinez-Garcia, P. J., Fresnedo-Ramirez, J., Parfitt, D. E., Gradziel, T. M., & Crisosto, C. H. (2013). Effect prediction of identified SNPs linked to fruit quality and chilling injury in peach [Prunus persica (L.) Batsch]. *Plant Mol Biol*, *81*(1 – 2), 161 – 174. https://doi.org/10.1007/s11103-012-9989-8

Mehra, P., Pandey, B. K., & Giri, J. (2015). Genome-wide DNA polymorphisms in low Phosphate tolerant and sensitive rice genotypes. *Sci Rep*, *5*, 13090. https://doi.org/10.1038/srep13090

Miyadate, H., Adachi, S., Hiraizumi, A., Tezuka, K., Nakazawa, N., Kawamoto, T., Katou, K., Kodama, I., Sakurai, K., Takahashi, H., Satoh-Nagasawa, N., Watanabe, A., Fujimura, T., & Akagi, H. (2011). OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles. *New Phytologist*, *189*(1), 190 – 199. https://doi.org/10.1111/j.1469-8137.2010.03459.x

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, *8*(19), 4321 – 4326. https://doi.org/10.1093/nar/8.19.4321

Nater, A., Burri, R., Kawakami, T., Smeds, L., & Ellegren, H. (2015). Resolving Evolutionary Relationships in Closely Related Species with Whole-Genome Sequencing Data. *Syst Biol*, *64*(6), 1000 – 1017. https://doi.org/10.1093/sysbio/syv045

Ooi, L. C. L., Low, E. T. L., Ordway, J., Marjuni, M., Yaakub, Z., Jiang, N., Smith, S., Bacher, B., Garner, P. A., Leinenger, M. T., Sander, N., Chan, P. L., Ong, P. W., Ongabdullah, M., Nookiah, R., Manaf, M. A. A., Lakey, N., Sambanthamurthi, R., & Singh, R. (2019). SureSawitTM TRUE-TO-TYPE - A high throughput universal single nucleotide polymorphism panel for DNA fingerprinting, purity testing and origin verification in oil palm. *Journal of Oil Palm Research*, *31*(4), 561 – 571. https://doi.org/10.21894/jopr.2019.0048

Rathinasabapathi, P., Purushothaman, N., Ramprasad, V. L., & Parani, M. (2015). Whole genome sequencing and analysis of Swarna, a widely cultivated indica rice variety with low glycemic index. *Scientific reports*, *5*, 11303. https://doi.org/10.1038/srep11303

Rezaei, M. K., Deokar, A., & Tar'an, B. (2016). Identification and Expression Analysis of Candidate Genes Involved in Carotenoid Biosynthesis in Chickpea Seeds. *Front Plant Sci*, *7*, 1867. https://doi.org/10.3389/fpls.2016.01867

Rivals, I., Personnaz, L., Taing, L., & Potier, M.-C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, *23*(4), 401 – 407. https://doi.org/10.1093/bioinformatics/btl633

Romdhane, B., M., Riahi, L., Jardak, R., Ghorbel, A., & Zoghlami, N. (2018). Fingerprinting and genetic purity assessment of F_1 barley hybrids and their salt-tolerant parental lines using nSSR molecular markers. *3 Biotech*, *8*(1), 57. https://doi.org/10.1007/s13205-017-1080-3

Schatz, M. C., Maron, L. G., Stein, J. C., Hernandez Wences, A., Gurtowski, J., Biggers, E., Lee, H., Kramer, M., Antoniou, E., Ghiban, E., Wright, M. H., Chia, J. ming, Ware, D., McCouch, S. R., & McCombie, W. R. (2014). Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. *Genome Biology*, *15*(11), 506. https://doi.org/10.1186/s13059-014-0506-z

Shen, Y., Shen, L., Shen, Z., Jing, W., Ge, H., Zhao, J., & Zhang, W. (2015). The potassium transporter OsHAK21 functions in the maintenance of ion homeostasis and tolerance to salt stress in rice. Plant, cell & environment, 38(12), 2766 – 2779. https://doi.org/10.1111/pce.12586

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121 – 132. https://doi.org/10.1038/nrg3642

Srivastava, S. K., Wolinski, P., & Pereira, A. (2014). A strategy for genome-wide identification of gene based polymorphisms in rice reveals non-synonymous variation and functional genotypic markers. *PloS one*, *9*(9), e105335. https://doi.org/10.1371/journal.pone.0105335

Sunian, E., Ramli, A., Jamal, M. S., & Amri, S. (2022). *Pembangunan varieti padi berhasil tinggi untuk kelestarian pengeluaran makanan*. *30*, 83 – 97.

Tanaka, N., Shenton, M., Kawahara, Y., Kumagai, M., Sakai, H., Kanamori, H., Yonemaru, J., Fukuoka, S., Sugimoto, K., Ishimoto, M., Wu, J., & Ebana, K. (2020). Whole-Genome Sequencing of the NARO World Rice Core Collection (WRC) as the Basis for Diversity and Association Studies. *Plant and Cell Physiology*, *61*(5), 922 – 932. https://doi.org/10.1093/pcp/pcaa019

Tatarinova, T. V, Chekalin, E., Nikolsky, Y., Bruskin, S., Chebotarov, D., McNally, K. L., & Alexandrov, N. (2016). Nucleotide diversity analysis highlights functionally important genomic regions. *Sci Rep*, *6*, 35730. https://doi.org/10.1038/srep35730

Thalmann, M., & Santelia, D. (2017). Starch as a determinant of plant fitness under abiotic stress. *New Phytol*, *214*(3), 943 – 951. https://doi.org/10.1111/nph.14491

Thudi, M., Palakurthi, R., Schnable, J. C., Chitikineni, A., Dreisigacker, S., Mace, E., Srivastava, R. K., Satyavathi, C. T., Odeny, D., Tiwari, V. K., Lam, H. M., Hong, Y. Bin, Singh, V. K., Li, G., Xu, Y., Chen, X., Kaila, S., Nguyen, H., Sivasankar, S., … Varshney, R. K. (2021). Genomic resources in plant breeding for sustainable agriculture. *Journal of Plant Physiology*, *257*, 153351. https://doi.org/10.1016/j.jplph.2020.153351

Ueno, D., Koyama, E., Yamaji, N., & Ma, J. F. (2011). Physiological, genetic, and molecular characterization of a high-Cd-accumulating rice cultivar, Jarjan. *Journal of Experimental Botany*, *62*(7), 2265 – 2272. https://doi.org/10.1093/jxb/erq383

Vasconcelos, L. M., Brito, A. C., Carmo, C. D., & Oliveira, E. J. (2016). Polymorphism of starch pathway genes in cassava. *Genetics and Molecular Research*, *15*(4), 1 – 15. https://doi.org/10.4238/gmr15049082

Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, *8*(10), 749 – 761. https://doi.org/10.1038/nrg2164

Wang, X., Kudrna, D. A., Pan, Y., Wang, H., Liu, L., Lin, H., Zhang, J., Song, X., Goicoechea, J. L., Wing, R. A., Zhang, Q., & Luo, M. (2014). Global genomic diversity of *Oryza sativa* varieties revealed by comparative physical mapping. *Genetics*, *196*(4), 937 – 949. https://doi.org/10.1534/genetics.113.159970

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., Dong, Y., Gutenkunst, R. N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, W. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, *30*(1), 105 – 111. https://doi.org/10.1038/nbt.2050

Xu, Z., Li, S., Zhang, C., Zhang, B., Zhu, K., Zhou, Y., & Liu, Q. (2017). Genetic connection between cell-wall composition and grain yield via parallel QTL analysis in indica and japonica subspecies. *Sci Rep*, *7*(1), 12561. https://doi.org/10.1038/s41598-017-12903-5

Zeier, J. (2013). New insights into the regulation of plant immunity by amino acid metabolic pathways. *Plant Cell Environ*, *36*(12), 2085 – 2103. https://doi.org/10.1111/pce.12122

Zhou, H., Li, J., Liu, X., Wei, X., He, Z., Hu, L., Wang, J., Duan, M., Xie, G., Wang, J., & Wang, L. (2021). The Divergent Roles of the Rice bcl-2 Associated Athanogene (BAG) Genes in Plant Development and Environmental Responses. *Plants*, 10(10), 2169. https://doi.org/10.3390/plants10102169