Vietnam Journal of Biotechnology **21**(2): 337-345, 2023

GENOME-WIDE ASSOCIATION STUDIES FOR IDENTIFICATION OF GENES AND QTLS CONTROLLING THE PALMITIC ACID CONTENT IN RICE BRAN OIL

Mai Thi Phuong Nga, Nguyen Thi Thuy Linh, To Thi Mai Huong✉

University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam

✉To whom correspondence should be addressed. E-mail: to-thi-mai.huong@usth.edu.vn

Received: 02.5.2023
Accepted: 26.6.2023

SUMMARY

Rice bran oil is becoming more and more popular plant oil source for human consumption. As one of the biggest producers of rice worldwide, the regulation of rice bran oil contained in Vietnam is becoming more necessary. In the present study, 161 rice accessions, which derive from diverse ecosystems in Vietnam, were used to investigate their palmitic acid (PA) content, the significant Single Nucleotide Polymorphisms (SNPs) as well as candidate genes relating to PA content using genome-wide association studies (GWAS). From the GWAS analysis, sixty-two noteworthy markers and nine quantitative trait loci were detected on chromosomes 1 and 11. Moreover, 187 candidate genes were identified. Notably, we identified the candidate genes including *OsFAD* and *GDSL-like lipase/acylhydrolase*, whose function relates to fatty acid (FA) content. Furthermore, the *LOC_Os11g27370.1* gene, which encodes UDP-glucuronosyltransferase domain-containing protein, located on *qC16.0.11.6*, chromosome 11, has been involved in the FA biosynthesis pathway in animals but has not been proven yet in plants. Therefore, this gene could serve as a potential candidate gene to validate its function in rice plants. Other candidate genes with diverse functions were identified from our study but the relation to FA biosynthesis pathway has not been proven yet in any other studies. Therefore, these candidate genes can also open a number of in-depth studies to validate their involvement in FA biosynthesis pathways. The results from this study not only give a deeper knowledge of PA in rice bran oil but also pave the way for future applications in breeding programs aimed at improving the oil profile in rice.

Keywords: fatty acid, gas chromatography, *Oryza sativa* L., palmitic acid, rice bran oil.

INTRODUCTION

Rice is an important staple food for more than 50% of the worldwide population. The rice grains are composed of husk, bran, starchy endosperm and embryo, among which bran contains around 65% of the nutrients of the whole rice grain (Limtrakul *et al*., 2019). One of the most common uses of rice bran is rice bran oil extracted from the outer bran of grains, in which oleic, linoleic, and PA are the three main FAs, accounting

for 42%, 32%, and 20%, respectively (Ghazani and Marangoni, 2016).

The PA contains a 16-carbon backbone with the molecular formula $C_{16}H_{32}O_2$, is one of the primary saturated FA in the human body and can be provided in the diet. It is an essential component of membrane, secretory, and transport lipids, with important roles in protein palmitoylation and signal molecules (Innis, 2016). PA is also known to link to a number of chronic diseases including reduction of autophagic flux and insulin sensitivity (Hernández-Cáceres *et al*., 2019), or association with genetic disorders. Therefore, controlling the PA amount in diet is important for human health.

The investigation of individual FA content and genes related to their synthesis in rice has been performed in several studies (Goffman *et al*., 2003; Kitta *et al*., 2005; Mai *et al*., 2023). In 2012, Ying and his colleagues used 145 rice varieties in India including the wild type and the hybrid varieties to discover the ketoacyl–ACP synthase I (*LOC_Os04g36800*), which regulates the C16:0 content, by using the GWAS approach (Ying *et al*., 2012).

GWAS is known as a promising method to discover the candidate genes relating to FA content, in particular PA, based on SNPs for high-throughput genotyping (Pearson and Manolio, 2008). Therefore, in the present study, 161 rice varieties out of 180 that have been sequenced (Phung *et al*., 2014) were chosen to study the correlation between the PA content and the significant SNPs as well as the candidate genes relating to the PA content. The results from the study may provide insights into the genetic basis of PA content in rice bran oil, which is important for genetic engineering to control PA content in rice plants for human consumption.

MATERIALS AND METHODS

Materials

A collection of 161 rice varieties, provided by the Plant Resources Center in Hanoi, Vietnam, was used for analyzing the PA content. Approximately 21623 markers have been yielded by sequencing using DArTseq™ and Illumina NGS technology (Phung *et al*., 2014).

PA extraction and gas chromatography analysis (GC)

The procedure for oil extraction and fatty acid methylation followed a previous study (Mai *et al*., 2023). Briefly, the rice grains were ground into a fine powder, and incubated with absolute chloroform: methanol (1:2 v/v) (Merck, Darmstadt, Germany) in a shaker for 1 h at room temperature before being mixed with chloroform (Merck, Darmstadt, Germany) and saturated NaCl (Thermo Scientific, MA, USA) and centrifuged at 3000 rpm for 5 min. The bottom phase containing the oil was transferred into a new glass vial. The lipid-containing fatty acid solution was dried using nitrogen. In the next step, the fatty acid was derivatized into methyl esters using BF3-methanol (Merck, Darmstadt, Germany), then mixed with a mixture of n-hexane (Thermo Scientific, MA, USA) and NaCl. After centrifugation, the upper phase was aspirated and filtered using 0.22-µm Sartorius membrane. The PA content was analyzed by using a TRACE 1300 Series Gas Chromatograph machine (Thermo Scientific, MA, USA).

The GC analysis for PA acid analysis was modified from Ren *et al*., (2013) and Mai *et al*., (2023). While hexane was the eluent, helium was used as carrier gas. The content of PA was calculated as a percentage of its area (%) in the chromatogram among the fatty acids in rice bran oil.

Genome-wide association studies and Linkage disequilibrium (LD) heatmap analysis

The PA content and the associated SNPs were analyzed using the mixed linear model, six principle components and Kinship analysis in TASSEL v5.0 (Mai *et al*. 2023). The SNPs having −log10(P) >3.5 was decided as the threshold for being considered as significant. The LD heatmap, which analyzes the linkage between the significant SNPs, was constructed using R studio software (To *et al*. 2020).

Candidate gene identification

The database from the Rice Genome Annotation Project Database was used to screen the candidate genes. The screening position was around 25 kb before and after the significant SNPs (Kawahara *et al*., 2013).

RESULTS

GWAS analysis reveals significant markers relating to PA content

The PA was extracted from the bran oil of 161 rice varieties. The FA profile was analyzed using GC analysis, and then the PA content was calculated. The Manhattan plot shows the distribution between SNPs relating to PA content in twelve chromosomes of rice. The SNPs having $-\log_{10}(p) > 3.5$ were considered to have strong linkages with PA content. A total of 62 noteworthy markers were recognized above the threshold of $p < 3.0E\text{-}4$. Most of the significant SNPs located on chromosomes 1 and 11, some other SNPs spread on chromosomes 2, 4, 5, 6, 7, and 9 (Figure 1).

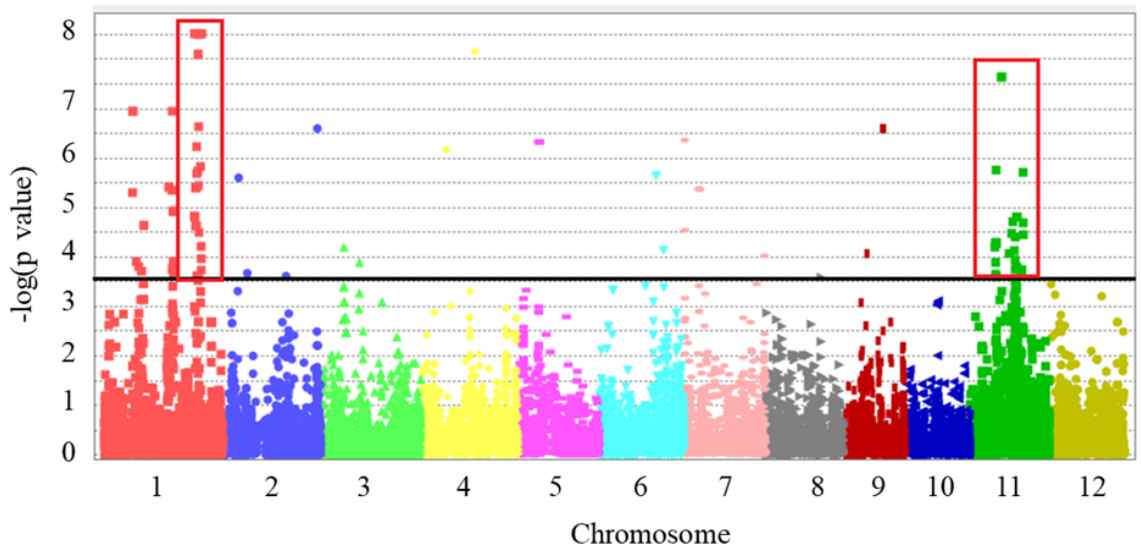

Figure 1. Manhattan plot of the association between SNPs and C16:0 fatty acid content in rice. The black line indicates the suggestive threshold of $-\log_{10}$(P value) equal 3.5. The red box indicate the potential SNPs link to PA content.

The quantile-quantile plot exhibits a well fit plot between theoretical and observed $-\log_{10}(P)$ value, which demonstrates our reliable data. An upper tail beginning at the value of 3.5 was spotted due to significant association surpassing the threshold established (Figure 2).

Promising quantitative trait locus (QTLs) associated to the PA content in rice bran oil

From the 62 significant SNPs, nine QTLs was identified, in which three and six QTLs locate in chromosome 1 and chromosome 11, respectively, with the length ranging from 30 kb to 335.5 kb (Table 1). QTLs *qC16.0.1.3* and *qC16.0.11.1* have the highest number of SNPs to be 6 SNPs, while each QTL *qC16.0.11.4, qC16.0.11.5, qC16.0.11.6* has only one SNPs.

LD heatmap analysis reveals the association between SNPs

Fine mapping by implementing linkage disequilibrium analysis can strengthen the association between a region of SNPs and the interesting trait and thus increase the reliability of the results. It is very interesting that the linkage between SNPs in the two QTLs *qC16.0.1.3* and *qC16.0.11.1* (having a high number of SNPs) strongly link together which illustrate by the red blocks. This result indicates that these SNPs could have a chance to bond in the cell division and transmit to the offspring together (Figure 3).

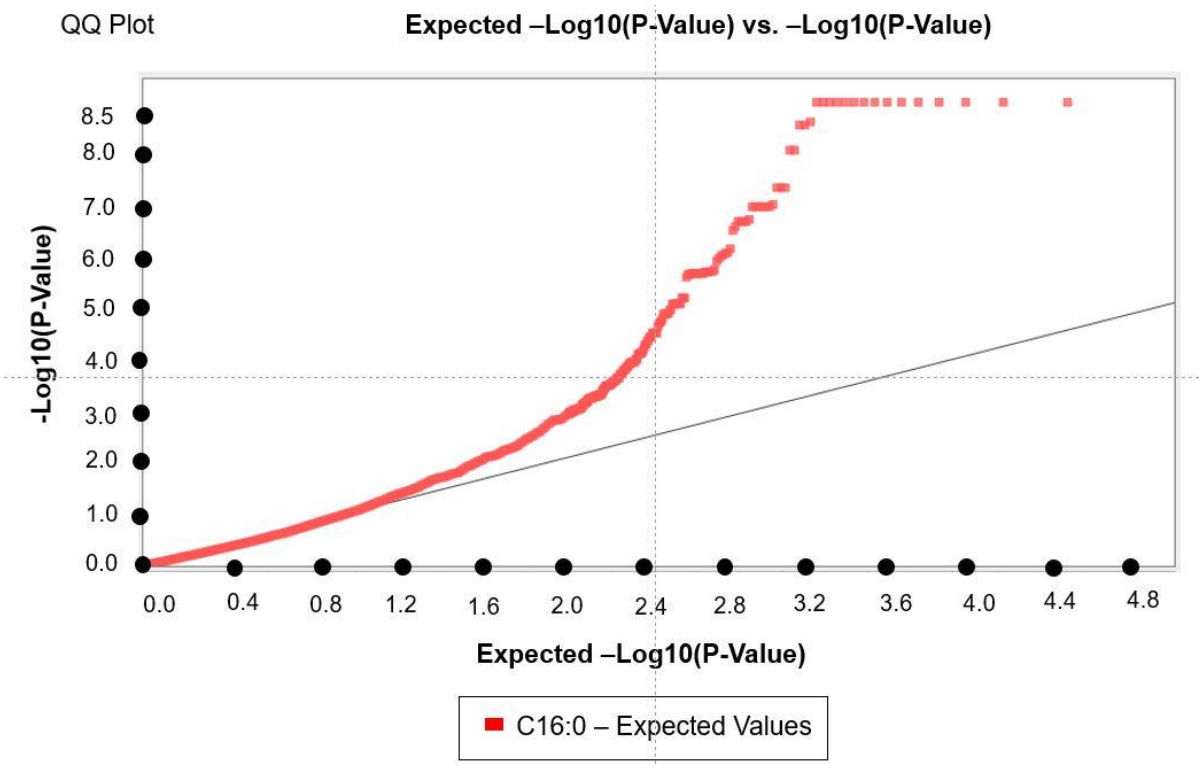

Figure 2. Quantile-quantile plot for observation of p-values and theoretical p-value.

Table 1. List of QTLs correlated with PA content.

No	QTL	Length	Marker
1	*qC16.0.1.1*	335.5kb	*Dj01_33384766F, Dj01_33477642R, Sj01_33484805F*
2	*qC16.0.1.2*	102.1kb	*Dj01_34335058R, Sj01_34412657F, Dj01_34417923R Sj01_34422873F*
3	*qC16.0.1.3*	445kb	*Dj01_35221487R, Sj01_35241374F, Sj01_35347280R Dj01_35347306F, Dj01_35400756R, Sj01_35433119R*
4	*qC16.0.11.1*	195.4kb	*Dj11_09579650F, Dj11_09595978F, Dj11_09628491R, Dj11_09502218F Dj11_09497568R, Dj11_09629685F*
5	*qC16.0.11.2*	35.6kb	*Sj11_16965449F, Sj11_16967657R*
6	*qC16.0.11.3*	45.6kb	*Sj11_17205202F, Sj11_17233812R*
7	*qC16.0.11.4*	51.8kb	*Dj11_17355923R*
8	*qC16.0.11.5*	30kb	*Dj11_15329900F*
9	*qC16.0.11.6*	56.6kb	*Dj11_15734723R*

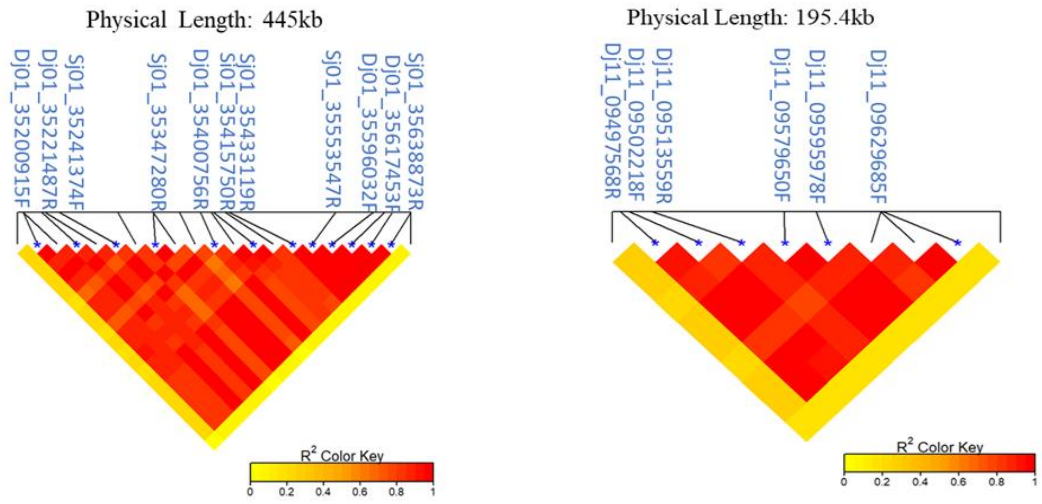

Figure 3. LD heatmap analysis revealed two haplotype blocks with outstanding R^2 in a condensed region.

Potential candidate genes associated to PA content in rice bran oil

A total of 187 candidate genes relating to PA content was found in chromosomes 1 and 11, where 141 genes are in chromosome 1 and 46 genes in chromosome 11. The highest number of candidate genes was on *qC16.0.1.3* with 69 genes, and the lowest locates on *qC16.0.11.2* with only 5 genes. Some promising candidate genes are reported in the Table 2.

Table 2. List of some promising candidate genes relating to PA content in rice.

QTL	Chr	Position	MSU ID	Gene description
qC16.0.1.1	1	33384766.. 33629815	*LOC_Os01g57770.1*	*OsPOP4* - Putative Prolyl Oligopeptidase homologue, expressed
			LOC_Os01g57940.1	Tyrosine protein kinase domain containing protein, putative, expressed
qC16.0.1.2	1	34335058.. 34422873	*LOC_Os01g59490.1*	FAD dependent oxidoreductase domain containing protein, expressed
			LOC_Os01g59530.1	*OsCML1* - Calmodulin-related calcium sensor protein, expressed
qC16.0.1.3	1	35200915.. 35638873	*LOC_Os01g60920.1*	*OsFBX30* - F-box domain containing protein, expressed
			LOC_Os01g61200.1	*GDSL-like lipase/acylhydrolase*, putative, expressed
			LOC_Os01g61570.1	*GDSL-like lipase/acylhydrolase*, putative, expressed
qC16.0.11.6	11	15712256.. 15768816	*LOC_Os11g27370.1*	UDP-glucoronosyl and UDP-glucosyl transferase domain containing protein, expressed
			LOC_Os11g27329.1	*OsSCP62* - Putative Serine Carboxypeptidase homologue, expressed
qC16.0.11.4	11	17336538.. 17388300	*LOC_Os11g29870.1*	*WRKY72*, expressed
			LOC_Os11g29910.1	Plastocyanin-like domain containing protein, putative, expressed
qC16.0.11.3	11	17204988.. 17250589	*LOC_Os11g29710.1*	*OsFBO4* - F-box and other domain containing protein, expressed
			LOC_Os11g29690.1	Oxidoreductase, 2OG-Fe oxygenase family protein, putative, expressed
qC16.0.11.2	11	16951754.. 16987343	*LOC_Os11g29290.1*	Cytochrome P450, putative, expressed
qC16.0.11.1	11	9497568..9 629685	*LOC_Os11g17110.1*	Disease resistance protein RPM1, putative

DISCUSSION

Genome-wide association studies on the PA content in Vietnamese rice revealed a number of candidate genes that are directly linked with the PA biosynthesis pathway. Some of them have already been demonstrated to be associated to PA content, but many others are novel genes and their function needs to be investigated in future experiments.

The *LOC_Os01g59490.1* gene located on QTL *qC16.0.1.2* is detected to encode FAD-dependent oxidoreductase domain containing protein. Eighteen *FAD* genes have been identified in rice. *OsFAD2-1* which belonging to the same family of *LOC_Os01g59490.1* has already been demonstrated to catalyze the conversion of C18:1 to C18:2 in rice grains. The functional analysis using CRISPR/Cas9 genome editing (Abe *et al*., 2018) or RNAi technology (Tiwari *et al*., 2016) has strongly confirmed the involvement of this gene in FA biosynthesis pathway. Therefore, *LOC_Os01g59490* is a very promising candidate gene in the synthesis of PA in rice.

The *LOC_Os01g61200.1* and *LOC_Os01g61570.1* genes located on *qC16.0.1.3* were also found to encode GDSL-like lipase/acylhydrolase. In rice, there is a large gene family encoding 114 GDSL lipase proteins (Chepyshko *et al*., 2012). The function of several genes encoded GDSL lipase proteins in fatty acid pathway has been clearly demonstrated in other studies (Huang *et al*., 2015; Ding *et al*., 2019; Eunhye Jo *et al*., 2021). In *Arabidopsis, SFARs (seed fatty acid reducers)* belonging the GDSL lipases family has been known to reduce the storage and composition of fatty acid in mature seeds (Huang *et al*., 2015). In *Geobacillus thermocatenulatus*, *lip29* and *est29* genes of GDSL lipase family showed strong lipolytic activity to C12-C16-containing-fatty acids and high lipolytic specificity toward C4-C8-containing-fatty acids (Eunhye Jo *et al*., 2021). Seed oil content was also increased in *Brassica napus* by regulating two closely related GDSL-motif-containing lipases (Ding *et al*., 2019). Therefore, in our study, these genes also encode the GDSL lipase protein, whose function has been illustrated to be involved in the the FA biosynthesis pathway in other studies.

The *LOC_Os11g27370.1* gene, which encodes UDP-glucuronosyltransferase domain containing protein, was detected on *qC16.0.11.6*, chromosome 11. The function of UDP-glucuronosyltransferase relating to FA has been demonstrated in animals. In rats, the level of FA docosahexaenoic acid (22:6) in the liver microsomal lipids during postnatal development was changed in correlation with levels of UDP-glucuronosyltransferase activity (Dannenberg *et al*., 1992). UDP-glucuronosyltransferase also involved in bile acid detoxification in mice (Staudinger *et al*., 2001). However, the enrolment of this enzyme in plant fatty acid biosynthesis has not been studied yet. Therefore, the particular function relating to fatty acid pathway, specifically PA of this gene should experimentally analyzed in rice as well as other plants.

CONCLUSION

In conclusion, our study has made significant advancements in understanding the diversity and genetic basis of PA content in rice. Through an analysis of a diverse collection of 161 Vietnam rice accessions, we have identified promising candidate

genes relating to PA content including *OsFAD, GDSL lipase* and *UDP-glucuronosyltransferase*. Under aligning with previous GWAS studies, some potential candidate genes contributing to PA content in rice are revealed for further functional studies. This investigation provides valuable insights into the genes underlying PA content in rice, paving the way for potential applications in breeding programs aimed at developing rice varieties with improved nutritional profiles.

Acknowlegement: *This research was supported by the Vietnam Academy of Science and Technology through grant number: CN4000.01/22-24.*

REFERENCES

Abe K, Araki E, Suzuki Y, Toki S, Saika H (2018) Production of high oleic/low linoleic rice by genome editing. *Plant Physiol Biochem* 131: 58–62.

Carta G, Murru E, Banni S, Banni S (2017) Palmitic acid: physiological role, metabolism and nutritional implications. *Front Physiol* 8: 902.

Chauhan I, and Singh L (2023) A comprehensive literature review of lipids used in the formulation of lipid nanoparticles. *Curr Nanomater* 8: 126–152.

Chepyshko H, Lai CP, Huang LM, Liu JH, Shaw JF (2012) Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (*Oryza sativa* L. *japonica*) genome: New insights from bioinformatics analysis. *BMC Genomics* 13: 1–19.

Dannenberg JA, Worman HJ, Scarlata S (1992) Developmental changes in the amount and functional state of UDP-glucuronosyltransferase. *Biochim Biophys Acta* 1116: 250–255.

Ding LN, Guo XJ, Li M, Fu ZL, Yan SZ, Zhu KM., Wang Z, Tan XL (2019) Improving seed germination and oil contents by regulating the GDSL transcriptional level in *Brassica napus*. *Plant Cell Rep* 38: 243–253.

Eunhye Jo, Kim J, Lee A, Moon K, Cha J (2021) Identification and characterization of a novel thermostable GDSL-type lipase from *Geobacillus thermocatenulatus*. *J Microbiol Biotechnol* 31: 483–491.

Ghazani SM, Marangoni AG (2016) *Healthy Fats and Oils*. In Wrigley C, Corke H, Seetharaman K, Faubion J, 2nd ed. Oxford Academic Press*:* 257-267.

Goffman FD, Pinson S, Bergman C (2003) Genetic diversity for lipid content and fatty acid profile in rice bran. *J Am Oil Chem Soc* 805: 485–490.

Hernández CMP, Toledo VL, Díaz CF, Ávalos Y, Burgos P, Narro C, Peña OD, Espinoza CJ, Cifuentes AF, Navarro AF, Riquelme C, Troncoso R, Criollo A, Morselli E (2019) Palmitic acid reduces the autophagic flux and insulin sensitivity through the activation of the free fatty acid receptor 1 (FFAR1) in the hypothalamic neuronal cell line N43/5. *Front Endocrinol* 10: 176.

Huang LM, Lai CP, Chen LFO, Chan MT, Shaw JF (2015) *Arabidopsis* SFAR4 is a novel GDSL-type esterase involved in fatty acid degradation and glucose tolerance. *Bot Stud* 56: 1–12.

Innis SM (2016) Palmitic acid in early human development. *Crit Rev Food Sci Nutr* 56: 1952–1959.

Kawahara Y, Bastide M, Hamilton JP, Kanamori H, Mccombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR (2013) Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice*

6: 3–10.

Kitta K, Ebihara M, Iizuka T, Yoshikawa R, Isshiki K, Kawamoto S (2005) Variations in lipid content and fatty acid composition of major non-glutinous rice cultivars in Japan. *J Food Compos Anal* 18: 269–278.

Limtrakul P, Semmarath W, Mapoung S (2020) Anthocyanins and proanthocyanidins in natural pigmented rice and their bioactivities. Phytochemicals in human health. *IntechOpen*.

Mai NTP, Nguyen LTT, Tran SG, To HTM (2023) Genome-wide association study reveals useful QTL and genes controlling the fatty acid composition in rice bran oil using Vietnamese rice landraces. *Funct Integr Genomics* 23: 150.

Morrison WR, and Lloyd M. Smith (1964) Preparation of fatty acid methyl esters and dimethylacetals from lipids with boron fluoride-methanol. *J Lipid Res* 5: 600–608.

Pearson TA, Manolio TA (2008) How to interpret a genome-wide association study. *JAMA* 299: 1335–1344.

Phung NTP, Mai CD, Mournet P, Frouin J, Droc G, Ta NK., Jouannic S, Le TL, Do NV, Gantet G, Courtois B (2014) Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *BMC Plant Biol* 14: 1–16.

Ren J, Mozurkewich EL, Sen A, Vahratian AM, Ferreri TG, Morse AN, Djuric Z (2013) Total serum fatty acid analysis by GC-MS: assay validation and serum sample stability. *Curr Pharm Anal* 9: 331-339.

Staudinger J, Liu Y, Madan A, Habeebu S, Klaassen CD (2001) Coordinate regulation of xenobiotic and bile acid homeostasis by pregnane X receptor. *Drug Metab Dispos* 29: 1467–1472.

Tiwari GJ, Liu Q, Shreshtha P, Li Z, Rahman S (2016) RNAi-mediated down-regulation of the expression of OsFAD2-1: Effect on lipid accumulation and expression of lipid biosynthetic genes in the rice grain. *BMC Plant Biol* 16: 189.

Ying JZ, Shan JX, Gao JP, Zhu MZ, Shi M, Lin HX. (2012) Identification of quantitative trait loci for lipid metabolism in rice seeds. *BMC Mol Plant* 5: 865–875.