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 the 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 $-\log_{10}(P) > 3.5$ 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 $p < 3.0E-4$. 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).

![Manhattan plot](image_url)

**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 $-\log_{10}(P \text{ value})$ 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](image_url)

**Figure 2.** Quantile-quantile plot for observation of p-values and theoretical p-value.
Table 1. List of QTLs correlated with PA content.

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

Figure 3. LD heatmap analysis revealed two haplotype blocks with outstanding R² 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.

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr</th>
<th>Position</th>
<th>MSU ID</th>
<th>Gene description</th>
</tr>
</thead>
<tbody>
<tr>
<td>qC16.0.1.1</td>
<td>1</td>
<td>33384766..33629815</td>
<td>LOC_Os01g57770.1</td>
<td>OsPOP4 - Putative Prolyl Oligopeptidase homologue, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os01g57940.1</td>
<td>Tyrosine protein kinase domain containing protein, putative, expressed</td>
</tr>
<tr>
<td>qC16.0.1.2</td>
<td>1</td>
<td>34335058..34422873</td>
<td>LOC_Os01g59490.1</td>
<td>FAD dependent oxidoreductase domain containing protein, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os01g59530.1</td>
<td>OsCML1 - Calmodulin-related calcium sensor protein, expressed</td>
</tr>
<tr>
<td>qC16.0.1.3</td>
<td>1</td>
<td>35200915..35638873</td>
<td>LOC_Os01g60920.1</td>
<td>OsFBX30 - F-box domain containing protein, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os01g61200.1</td>
<td>GDSL-like lipase/acylhydrolase, putative, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os01g61570.1</td>
<td>GDSL-like lipase/acylhydrolase, putative, expressed</td>
</tr>
<tr>
<td>qC16.0.11.6</td>
<td>11</td>
<td>15712256..15768816</td>
<td>LOC_Os11g27370.1</td>
<td>UDP-glucoronsyl and UDP-glucosyl transferase domain containing protein, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os11g27329.1</td>
<td>OsSCP62 - Putative Serine Carboxypeptidase homologue, expressed</td>
</tr>
<tr>
<td>qC16.0.11.4</td>
<td>11</td>
<td>17336538..17388300</td>
<td>LOC_Os11g29870.1</td>
<td>WRKY72, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os11g29910.1</td>
<td>Plastocyanin-like domain containing protein, putative, expressed</td>
</tr>
<tr>
<td>qC16.0.11.3</td>
<td>11</td>
<td>17204988..17250589</td>
<td>LOC_Os11g29710.1</td>
<td>OsFBO4 - F-box and other domain containing protein, expressed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOC_Os11g29690.1</td>
<td>Oxidoreductase, 2OG-Fe oxygenase family protein, putative, expressed</td>
</tr>
<tr>
<td>qC16.0.11.2</td>
<td>11</td>
<td>16951754..16987343</td>
<td>LOC_Os11g29290.1</td>
<td>Cytochrome P450, putative, expressed</td>
</tr>
<tr>
<td>qC16.0.11.1</td>
<td>11</td>
<td>9497568..9629685</td>
<td>LOC_Os11g17110.1</td>
<td>Disease resistance protein RPM1, putative</td>
</tr>
</tbody>
</table>
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 oxoreductase domain containing protein. Eighteen FAD genes have been identified in rice. OsFAD2-1 which belonged 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 enrollment 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.

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

REFERENCES


Kitta K, Ebihara M, Iizuka T, Yoshikawa R,


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.


