Identification of anthocyanin biosynthesis genes in rice pericarp using PCAMP

Xinghai Yang1,†, Xiuzhong Xia1,†, Zongqiong Zhang1,†, Baoxuan Nong1, Yu Zeng1, Yanyan Wu2, Faqian Xiong3, Yuexiong Zhang3, Haifu Liang3, Yinghua Pan1, Gaoxing Dai1, Guofu Deng1,* and Danting Li1,*†

1Rice Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China
2Biotechnology Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China
3Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China

Received 8 January 2019; accepted 12 April 2019.
*Correspondence (Tel 86 771 3243019; fax 86 771 3246373) (ricegi@163.com) and (dengguofu@gxaas.net)
†These authors contributed equally to this paper.

Keywords: multiple pools, whole-genome resequencing, SNP-index, anthocyanin, candidate genes.

Anthocyanins are a kind of biologically active flavonoids, which have strong anti-oxidation and anti-mutation functions as phytonutrients and have important effects on human health. The anthocyanin metabolic pathway has been extensively studied in Arabidopsis thaliana, Petunia hybrida and Zea mays, which involves many structural genes and regulatory genes. However, only a few anthocyanin biosynthesis-related genes have been identified in rice, such as Rd (Furukawa et al., 2007), OsCHI (Hong et al., 2012) and Kala4 (Oikawa et al., 2015). The traditional method of mapping quantitative trait loci (QTLs) is only for two corresponding alleles and is time-consuming and labour-intensive. High-throughput sequencing technologies have become the new strategies for mapping the important traits of crops, such as simultaneous mapping and mutation identification by deep sequencing (SHOREmap) (Schneeberger et al., 2009), next-generation mapping (NGM) (Austin et al., 2011), mutation mapping (MutMap) (Abe et al., 2012), QTL-seq (Takagi et al., 2013) and genome-wide association study (GWAS) (Liu and Yan, 2019) can rapidly identify the genes for plant traits. However, SHOREmap requires a much larger sample size; the NGM studies the genes belongs to the recessive homozygous mutant phenotype; MutMap mainly identifies the single gene-controlled quality traits; QTL-seq constructs only two pools showing extreme opposite trait values for a given phenotype in a segregating progeny and maps 1–2 major genes for target trait; GWAS is applicable to natural population with a large sample size and thus its cost is high, and it is also difficult to detect the rare mutations and minor effective genes.

Here, we introduced Pair-wise Comparison Analysis for Multiple Pool-seq (PCAMP), an optimized method of QTL-seq to identify the genomic candidate regions involved in anthocyanin biosynthesis in rice pericarp. In this protocol, the second filial generation (F2) progeny generated by crossing two parents with different target traits were divided into n (n ≥ 3) subpopulations according to their phenotypes. Thirty phenotypically identical individuals were selected from each subpopulation, and their DNA samples were extracted to form a pool for sequencing. Finally, we compared the SNP-index between every two Pool-seq to map the genomic candidate regions.

Donglanmomi (DLMM) is a rice variety with high anthocyanin content (1797.82 μg/g DW). It was crossed to Huanghuazhan (HHZ) with very low anthocyanin content (3.68 μg/g DW) to generate F1 progeny, and F2 progeny were derived from self-pollination of the F1 progeny. After the rice seeds were fully matured, the progeny segregated in a 601:195 ratio for coloured pericarp and white pericarp phenotypes, respectively, conforming to a 3:1 segregation ratio (chi-squared test: χ² = 0.11, nonsignificant) and indicating that a gene plays an important role in anthocyanin biosynthesis in rice pericarp. Previous research showed that this gene was Kala4 (Oikawa et al., 2015). Subsequently, the F2 progeny were divided into four subpopulations according to the anthocyanin content of 796 individuals, and the DNA samples of 30 individuals in each subpopulation were mixed in equal amounts to form four pools: B1, B2, B3 and W, respectively (Figure 1a).

The DNA of DLMM, HHZ, B1, B2, B3 and W was sequenced using Illumina HiSeq X Ten high-throughput sequencing technology. After data filtration, the total base of six samples together was 161.48 Gb; of which, DLMM, HHZ, B1, B2, B3 and W accounted for 36.55 Gb, 39.63 Gb, 22.11 Gb, 21.64 Gb, 21.89 Gb and 19.66 Gb, respectively. Single nucleotide polymorphisms (SNPs) of six samples were detected through GATK software.

To identify the genomic candidate regions responsible for anthocyanin biosynthesis in rice pericarp, we compared the SNP-index between any two different pools. Distance method was used to fit the ΔSNP-index, and the distribution of ΔSNP-index is...
shown in Figure 1b1–b6. For the genomic candidate regions with overlapping physical positions on the same chromosome, the intersection regions were selected as the final genomic candidate regions. Therefore, the regions showing a significant association with anthocyanin biosynthesis-related genes in rice pericarp are shown in Figure 1c.

Three genomic candidate regions were adjacent to or contained the cloned genes of anthocyanin biosynthesis (Figure 1c). Rd was found to be involved in the proanthocyanidin biosynthesis of rice pericarp (Furukawa et al., 2007). The expression levels of Rd between DLMM and HHZ were significantly different (Figure 1d1). The sequences of DLMM and HHZ were amplified...
with PCR primer (F: ccatcacaagtgcaaggt, R: agtcgtcggctggtagg), and the products were sequenced. The 43rd base of the second exon of the Rd of HHZ was changed from C to A causing premature termination of translation of mRNA (Figure 1d2). Why is Rd located at the upstream of the genomic candidate region (1.19 Mb)? The number of SNPs in the genomic region nearby Rd was greatly reduced (Figure 1e). Thus, the false-positive result may be resulted from a decrease in nucleotide polymorphism within this genomic region.

OsCHI is a key gene involved in flavonoid metabolic pathway (Hong et al., 2012). The expression levels of OsCHI between DLMM and HHZ were significantly different (Figure 1d3). Ra is located in the candidate region on chromosome 4, which encodes the basic helix-loop-helix (bHLH) transcription factor, which plays a regulatory role in the anthocyanin biosynthesis (Hu et al., 1996). Subsequently, Hu et al. (2000) indicated that Ra consisted of Ra1 and Ra2. Recently, Okawa et al. (2015) successfully cloned Kala4, a key gene responsible for anthocyanin accumulation in rice pericarp, which was found to be the same gene as Ra2. The expression levels of Kala4 between DLMM and HHZ were significantly different (Figure 1d4). The DNA of DLMM and HHZ was amplified by functional primers (F: agggagtctctgtcatagtttc, R1: gcgtggttggcctatcttac, R2: gcgtgctgtaactacaagct). The results showed that the promoter region of Kala4 in DLMM had a genomic fragment inserted (Figure 1d5), and this change was the causes of generation of the black rice traits (Okawa et al., 2015).

There are 61 SNPs with ΔSNP-index ≥ 0.67 in 26.59–30.92 Mb on chromosome 2. They included a homozygous variant site of ΔSNP-index = 1. The expression levels of LOC_Os02g91410 between DLMM and HHZ were significantly different (Figure 1d6), and this gene encodes glycosyltransferase. In the anthocyanin biosynthetic pathway, glycosylation modification affects its stability in cells.

Within the 8.76- to 10.07-Mb region on chromosome 3, there are 24 SNPs with ΔSNP-index ≥ 0.67 and two homozygous variant loci with ΔSNP-index = 1. The expression levels of LOC_Os03g18020 between DLMM and HHZ were significantly different (Figure 1d7). This gene encodes leucoanthocyanidin dioxygenase, a key enzyme involved in anthocyanin biosynthesis in rice pericarp; among them, the genes Rd, OsCH, and Kala4 have been cloned. The results showed that the PCAMP method may be a powerful tool for identifying multiple gene-controlled traits in rice.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31860371), National Key Research and Development Program of China (2016YFD0100101-03), China National Rice Research Institute (170102), Guangxi Natural Science Foundation of China (2015GXNSFA139054, 2018GXNSFA138124). The accession number of resequencing data is from SRR7903633 to SRR7903638. The authors declare no competing interests.

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H. et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. 30, 174–178.

Austin, R.S., Vidaurre, D., Stamatiou, G., Brett, R., Provart, N.J., Bonetta, D., Zhang, J. et al. (2011) Next-generation mapping of Arabidopsis genes. Plant J. 67, 715–725.

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Takamure, I. et al. (2007) The Rd and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. Plant J. 49, 91–102.

Hong, L., Qian, Q., Tang, D., Wang, K., Li, M. and Cheng, Z. (2012) A mutation in the rice chalcone isomerase gene causes the golden hull and internode 1 phenotype. Planta 236, 141–151.

Hu, J., Anderson, B. and Wessler, S.R. (1996) Isolation and characterization of Arabidopsis genes. Plant J. 725, 141–151.

Hu, J., Reddy, V.S. and Wessler, S.R. (2000) The OsCHI gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. Plant Mol. Biol. 42, 667–678.

Liu, H.J. and Yan, J. (2019) Crop genome-wide association study: a harvest of biological relevance. Plant J. 97, 8–18.

Okawa, T., Maeda, H., Oguchi, T., Yamaguchi, T., Tanabe, N., Ebana, K., Yano, M. et al. (2015) The birth of a black rice gene and its local spread by introgression. Plant Cell 27, 2401–2414.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.E. et al. (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods. 6, 550–551.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A. et al. (2013) OSHL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74, 174–183.