



OPEN Uncovering genetic determinants of antioxidant properties in Thai landrace rice through genome-wide association analysis

Putut Rakhmad Purnama^{1,2}, Pipob Suwanchaikasem², Supaporn Junbuathong³, Somsong Chotechuen⁴, Peerapon Mounng-Ngam³, Waraluk Kasettranan⁵, Chanita Paliyavuth⁵, Monnat Pongpanich⁶, Sittiruk Roytrakul⁷, Luca Comai⁸, Teerapong Buaboocha^{9,10} & Supachitra Chadchawan^{2,10}✉

Pigmented rice (*Oryza sativa* L.) is recognized as a source of natural antioxidant compounds, such as flavonoids, oryzanol, tocopherol, and anthocyanin. Because of their nutritional benefits, anthocyanin-enriched or pigmented rice varieties are feasible alternatives for promoting human health. Pigment biosynthesis in rice is a complex process that involves multiple biosynthetic and regulatory genes. This study aimed to identify the genes associated with antioxidant traits in local Thai rice and develop molecular markers for breeding programs. Single nucleotide polymorphisms (SNPs) in 233 rice cultivars, 119,541 SNPs from the promoters, and the first 5 exons of all genes were associated with antioxidant phenotypes determined using 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid), 2,2-diphenyl-1-picrylhydrazyl, ferric reducing antioxidant power, total flavonoid content, and total phenolic content assays through a mixed linear model (MLM) genome-wide association study (GWAS). GWAS analysis detected 336 quantitative trait nucleotides located at 93 loci, 83% of which were consistently mapped to the reported quantitative trait loci of antioxidant traits. Ten loci, including the *OsRc* gene, were consistently co-associated with all phenotypic traits. Two genetic markers for polymerase chain reaction detection were developed for the *OsTT8* gene and the *OsRc* promoter, both of which were positively correlated with antioxidant properties. These markers can be applied for antioxidant trait selection in rice breeding programs.

Keywords Anthocyanins, Genetic marker, Loci, Proanthocyanidins, QTNs

Rice (*Oryza sativa* L.) is an important cereal crop consumed as a staple and processed form, along with wheat, maize, and potatoes. High rice consumption in more than 175 countries has led to many international programs aimed at boosting human nutrition through the development of high-quality rice¹. Consumption of a single type of grain may result in deficiency disorders due to decreased concentrations of vital amino acids, minerals, and vitamins². Pigmented rice is recognized as a source of natural antioxidant compounds, such as flavonoids, γ -oryzanol, tocopherol, and anthocyanin, among others^{3,4}. Epidemiological research has suggested that the low prevalence of some chronic diseases in rice-consuming regions may be due to the antioxidant content of rice⁵. Phenolic compounds in rice, such as flavonoids, protect the cells against oxidative stress⁶. γ -oryzanol in

¹Bioinformatics and Computational Biology Program, Graduate School, Chulalongkorn University, Bangkok 10330, Thailand. ²Center of Excellence in Environment and Plant Physiology, Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ³Pathum Thani Rice Research Center, Thanyaburi 12110, Pathum Thani, Thailand. ⁴Division of Rice Research and Development, Rice Department, Bangkok 10900, Thailand. ⁵Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ⁶Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ⁷Genome Technology Research Unit, Proteomics Research Laboratory National Science and Technology Development Agency, Pathum Thani 12120, Thailand. ⁸Department of Plant Biology and Genome Center, University of California Davis, Davis, CA 95616, USA. ⁹Center of Excellence in Molecular Crop, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ¹⁰Omic Sciences and Bioinformatics Center, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. ✉email: supachitra.c@chula.ac.th; s_chadchawan@hotmail.com

rice bran is widely acknowledged as a beneficial dietary supplement and medicinal material that maintains and promotes human health⁷. Tocopherol, also a primary bioactive compound, is a source of vitamin E present in high amounts in rice bran, whole rice, and milled rice⁸.

Genome-wide association studies (GWASs) are widely used to evaluate hundreds of thousands of variations across numerous genomes and to identify those that are statistically associated with complex traits, such as plant height, flowering time, grain production, and other important agronomic features^{9–12}. GWAS mapping has also identified significant loci linked to rice grain quality^{13–15} and stress resistance traits^{16–19}.

In a study by Xu et al.²⁰, GWAS analysis used single nucleotide polymorphisms (SNPs) from the whole-genome sequences of 120 rice accessions and identified at least four loci significantly associated with antioxidant capacity (2,2'-azino-bis-3-ethylbenzthiazoline-6-sulphonic acid [ABTS] and 2,2-diphenyl-1-picrylhydrazyl [DPPH]), total phenolic content, and total flavonoid content. These loci include *Rd* gene (*LOC_Os01g44260*), which encodes dihydroflavonol-4-reductase, and *LOC_Os07g11020*, the *Rc* gene encoding a bHLH transcription factor involved in seed proanthocyanidin biosynthesis.

In this research, 233 rice cultivars were evaluated for their antioxidant traits using various assays, namely ABTS²¹, DPPH²², ferric reducing antioxidant power (FRAP)²³, total flavonoid content (TFC)²⁴, and total phenolic content (TPC)²⁵ assays, and were associated with SNPs to determine the regions important for antioxidant traits. Appropriate primers were designed and validated in rice population which is the subset of the population used for GWAS, to demonstrate their potential for the selection of antioxidant traits in rice breeding programs.

Results

Phenotypic description of five antioxidant properties

The phenotypic variations in ABTS, DPPH, FRAP, TFC, and TPC are shown in Fig. 1A and **Supplementary Table S1**. The nonpigmented rice cultivars (white) were distinct from the pigmented cultivars (red and black) in all antioxidant phenotypes. The principal component analysis (PCA) plot displayed tight clustering of all white-pericarp cultivars indicating little variation. Colored rice cultivars (red and black), on the other hand, were scattered (Fig. 1B), although black and red types were fairly distinct along the PC2 axis.

The ABTS value of white rice ranged from 2.01–22.91 mg Trolox equivalent antioxidant capacity (TEAC)/100 g, which was substantially lower than the ranges of red and black rice at 38.22–109.63 and 31.08–107.02 mg TEAC/100 g, ($p = 1.4 \times 10^{-14}$ and 4.8×10^{-11}), respectively. The DPPH values of black, red, and white rice ranged from 19.19–70.23, 9.84–64.06, and 0.76–14.13 mg TEAC/100 g, respectively. The black and red rice varieties 'Khao Niaw Dam Maw' and 'Khao Niaw Daeng Grahm Raed', respectively, had the highest DPPH values (70.23 and 64.06 mg TEAC/100 g, respectively). The FRAP values of black, red, and white rice ranged from 10,187.5–33,931.2, 8,722.7–28,352, and 3,070.3–5,957.3 mg TEAC/100 g, respectively. The mean TFC values of black and red rice were comparable (1184.03 and 1111.03 mg gallic acid equivalent [GAE]/100 g, respectively; $p = 0.31$) and were significantly higher than that of white rice (322.43 mg GAE/100 g; $p = 2.2 \times 10^{-16}$ for black rice and $p = 1.13 \times 10^{-09}$ for red rice). Moreover, the TFC value of the red rice cultivar 'Hawm Mali Daeng' (69.96 GAE/100 g) was lower than that of all white rice cultivars. The highest TPC values were found in red rice cultivars (476.89–1160.80 mg GAE/100 g), followed by black (356.44–782.22 mg GAE/100 g) and white (60.89–165.33 mg GAE/100 g) rice cultivars. Thus, the red rice cultivar 'Hawm Mali Daeng' had the highest TPC value among all cultivars (Fig. 1A).

Heatmap and hierarchical clustering were performed based on the antioxidant phenotypes of all cultivars in this study. The cultivars were grouped into three clusters. The rice cultivars with colored seeds were separated into two clusters, A and B, whereas all white rice cultivars were grouped in cluster C (Fig. 1C). Cluster A typically comprised red and black cultivars with high DPPH and FRAP values, whereas cluster B comprised several red and black cultivars having low antioxidant values, mostly in DPPH and FRAP, with a few also having low ABTS, TFC, and TPC values. A list of the pigmented rice in clusters A and B is presented in Table 1. Cluster C was composed of all the members of the white rice cultivars and had antioxidant content that was typically lower than that of clusters A and B. However, some small groups in this cluster comprised six white rice cultivars, namely 'Chai Nat', 'RD43', 'CH4', 'CH1', 'Aew Mod Daeng', and 'Re Hawm', with slightly higher DPPH content than other white rice cultivars.

Genotypic description of SNP markers

In total, 119,541 high-quality SNP markers from the promoters and first 5 exons of all genes, excluding the minor alleles with the frequency less than 5% and the variants from the accessions with missing data, from 233 rice accessions were used in the GWAS. The distribution and density of the SNPs across the 12 chromosomes are shown in Fig. 2. The total length of the rice genome was ~380 Mbp, with one SNP per 3.18 Kbp on average. Chromosome 11 had the highest number of SNPs (15,809) and highest variation rate (one SNP per 1.8 Kbp), whereas chromosome 9 had the lowest number of SNPs (5318 SNPs) and lowest variation rate (one SNP per 4.3 Kbp). Details of the distribution and density of the SNPs on all chromosomes are provided in **Supplementary Table S2**.

The PCA plot of the SNP variants from red, black, and white pericarp cultivars demonstrated that the genotypes associated with pericarp color were combined (Fig. 3A), indicating recent gene flow between the different phenotypes, although a set of white varieties form a homogeneous cluster. This finding differs from the PCA graph of the antioxidant phenotypes (Fig. 1B), in which the white rice cultivars were distinctly separated from the colored rice cultivars. The percentage of explained components was lower, with only 34.39% in PC1 and 19.36% in PC2. Interestingly, the genotypes of 'Khao Niaw Dam Maw', 'Khao Niaw Dam Chaw Mai Pai', 'Saeng Lisaw' (black rice cultivar), and 'Ble La' (red rice cultivar) were separated from the others in the PCA plot. This finding is echoed in the phylogenetic tree based on unweighted pair group method with arithmetic

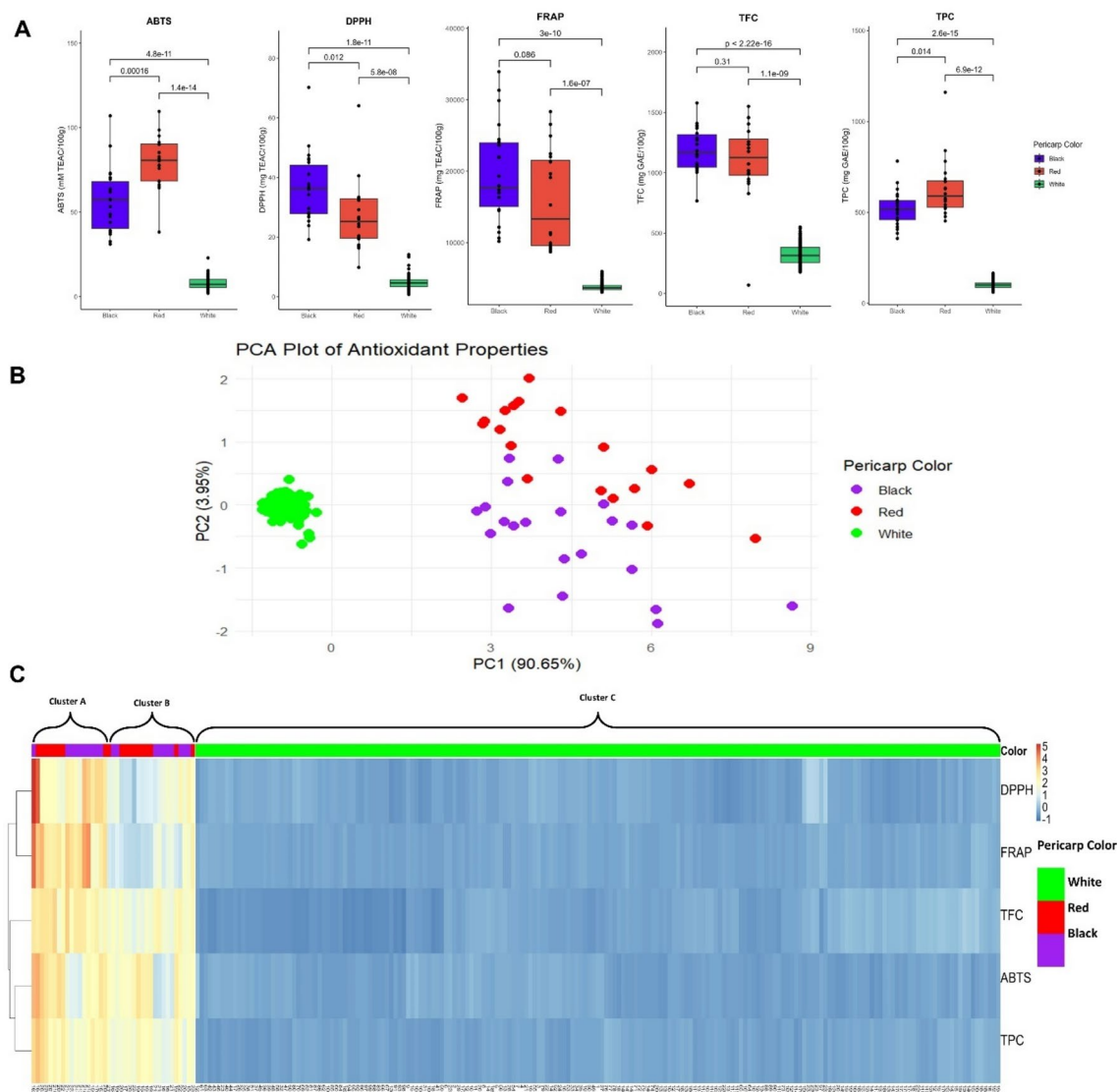


Fig. 1. Frequency distribution (A), principal component analysis (PCA) plot (B), and heatmap hierarchical clustering plot (C) of antioxidant properties in different pericarp colors. Three clusters are shown. The color of the top bar represents the pericarp color, and the number at the bottom indicates the accession number in this population (list is provided in **Supplementary Table S1**). Green, white rice cultivar; purple, black rice cultivar; and red, red rice cultivar. Number above box plot indicate *p*-value of t-test for each pairwise differences in antioxidants between pigmentation groups.

Cluster	Pericarp color	Cultivar name
Cluster A	Red (8)	'Khao' Niaw Daeng Graham Raed', 'Hawm Gra-dang-ngah', 'Rathu Heenati', 'Hawm Gulahb Daeng', 'Hawm Daeng (Photoperiod insensitive)', 'Nok Kum', 'Lai Mahk', and 'Loi Hah Ruang'
	Black (10)	'Mali Dam', 'Khao' Niaw Dam Maw', 'Khao' Niaw Dam Chaw Mai Pai', 'Niaw Dam Hawm Poo Khiaw', 'Mae Payah Tawng Dam', 'Gam Noi', 'Mali Dam', 'Dam Dahng', 'Gam Feuang', and 'Khao' Gam'
Cluster B	Red (12)	'Mai Daeng', 'Daeng Hawm Gulahb', 'Mali Daeng Bow', 'Mali Daeng San Nak', 'Mali Daeng Yao Nak', 'Hawm Mali Daeng', 'Niaw Sew Daeng', 'Sang Yod', 'Hawm Mali Daeng', 'Loi Hah Ruang', 'Tab Tim Chumpae (RD69)', and 'Ble La'
	Black (13)	'Khao' Glam', 'Khao' Niaw Dam 4', 'Gam Nai Prom', 'Meuang Nan', 'Gam Chiang Mai Ton Khiaw', 'Khao' Gam', 'Gam Nahng Payah', 'Hawm Nin', 'Riceberry', 'Gam Liaw', 'Niaw Dam Rai', and 'Saeng Lisaw'

Table 1. Pigmented rice cultivars were in clusters A and B. *Number in parentheses indicates number of cultivars.

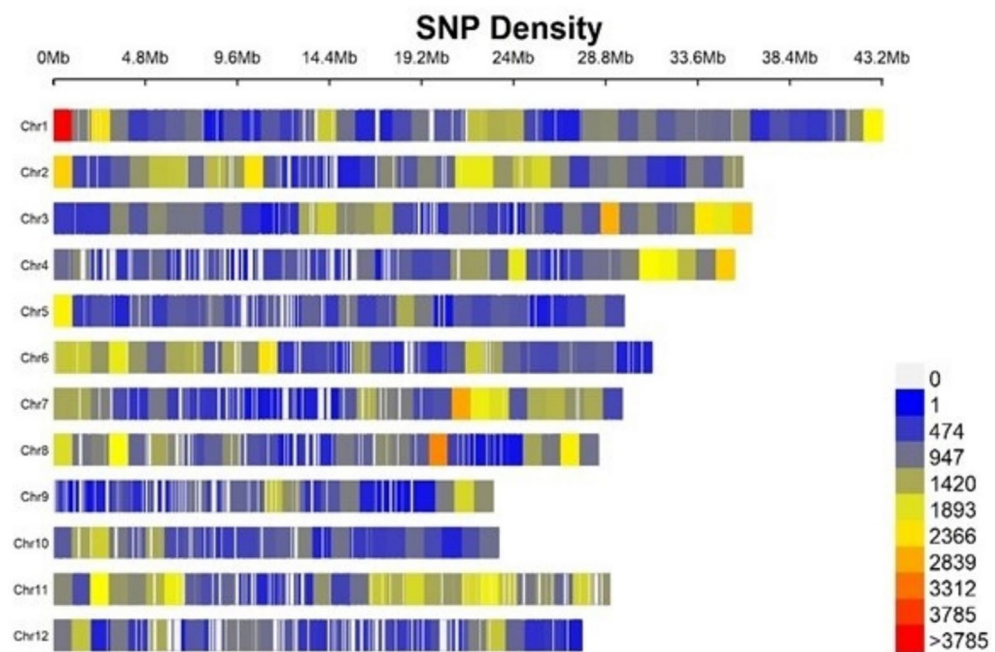


Fig. 2. SNP density plot of 119,541 variants in 233 rice accessions within 1 Mb window size. The x-axis represents the length of the chromosome in megabases (Mb), whereas the various colors indicate the densities.

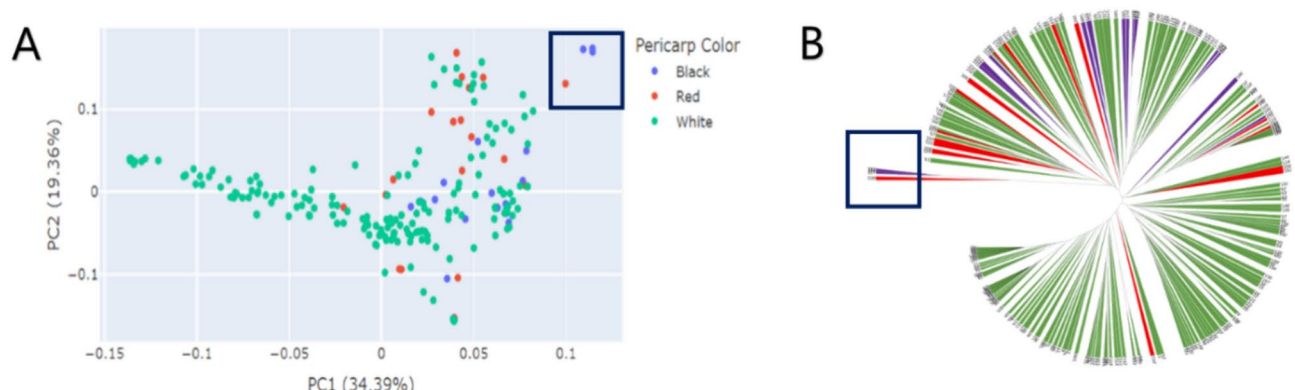


Fig. 3. Principal component analysis (PCA) plot of 119,541 SNP markers (A) and phylogenetic tree based on UPGMA (B). The square shows three black rice cultivars and one red rice cultivar, demonstrating the variance dissociation from other rice cultivars in this study.

mean (UPGMA) (Fig. 3B), where these four cultivars form a distinct clade, suggesting divergence from the other cultivars in this study.

Detecting genes associated with antioxidant properties using GWAS

A total of 336 quantitative trait nucleotides (QTNs) were detected to be associated with five phenotypic traits linked to antioxidant properties. These QTNs were annotated at 93 loci across 12 chromosomes. The Manhattan plot indicated that many SNPs affected phenotype. Several signals passed the threshold level using the mixed linear model (MLM) method, which was $p < 0.05/119,541 \sim 4 \times 10^{-7}$ (Fig. 4; **Supplementary Table S3**).

As shown in Fig. 4A, most QTNs associated with ABTS were located on chromosome 7. The QTN at the peak was snp72189, overlapped with *LOC_Os07g11020*, which is *Rc* gene, encoding a bHLH transcription factor regulating proanthocyanidin production in seeds. A few other QTNs were found on chromosomes 2, 3, 6, 10, and 11.

More QTNs were associated with DPPH than with ABTS (Fig. 4B). Several QTNs were annotated as loci on chromosome 3, including *LOC_Os03g38330* (disease resistance RPP13-like protein 1) and *LOC_Os03g38390* (GDSL-like lipase/acyl hydrolase). Similar to its association with ABTS, *LOC_Os07g11020* was the most

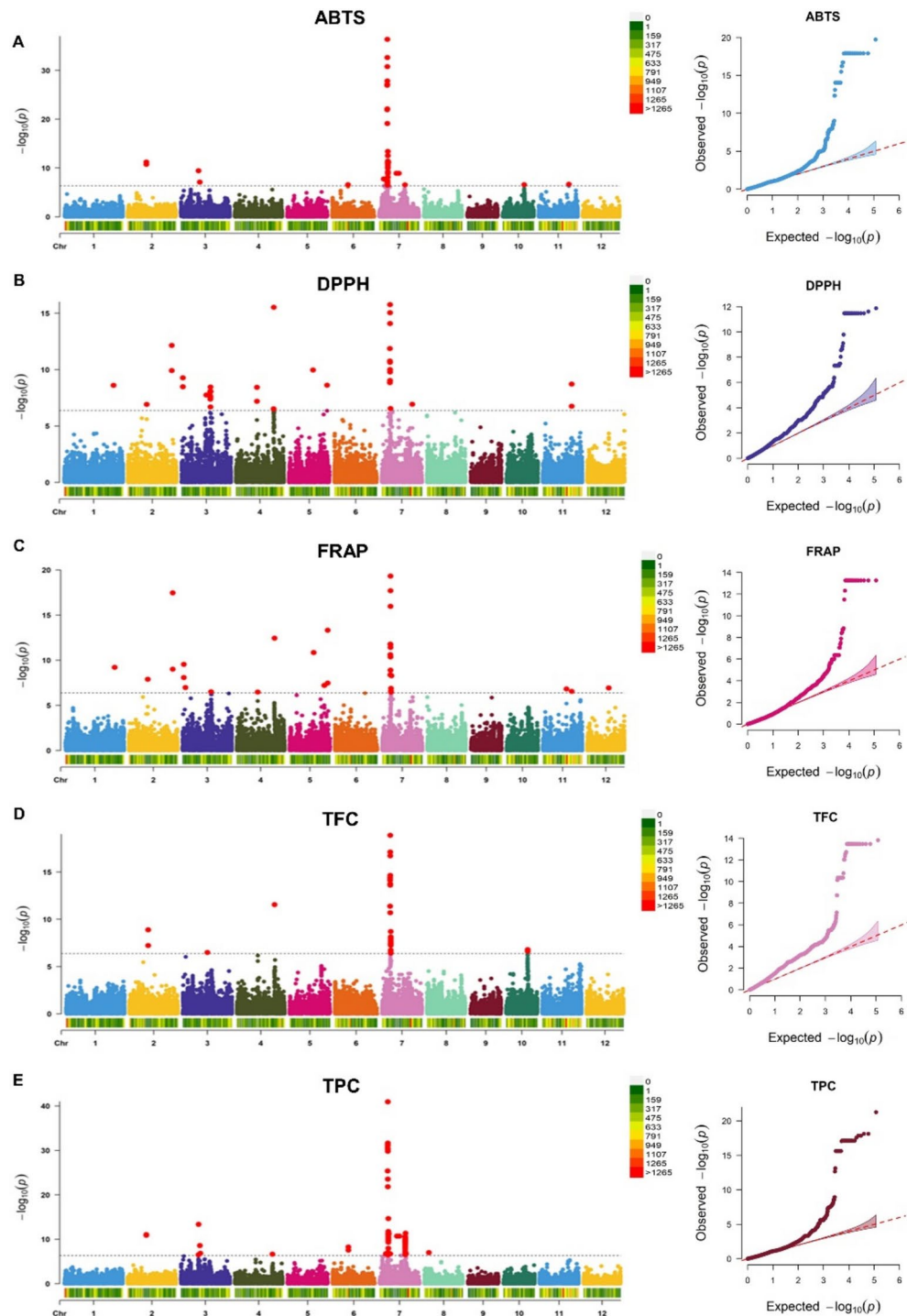


Fig. 4. The Manhattan plot (left) consists of five association antioxidants traits: (A) ABTS, (B) DPPH, (C) FRAP, (D) TFC and (E) TPC. The red dots represent QTNs that pass-through threshold value $< 4 \times 10^{-7} p$ -value ($\sim 6.39 \log_{10}(p)$). The bar below represents the density of QTNs in each chromosome. The chart (right) presents the quantile–quantile (Q–Q) plot comparing the observed and expected p-values from the association analyses.

significant locus. *LOC_Os04g47080* (anthocyanin regulatory Lc protein, putative, expressed) had the second highest significance value.

The associations with FRAP showed a trend similar to that of DPPH (Fig. 4C). QTNs were spread across all chromosomes, except chromosomes 6, 8, 9 and 10. The most significant SNPs overlapped with *LOC_Os07g11050* (IPP transferase), *LOC_Os07g11060.1* (expressed protein), and *LOC_Os07g11070* (DUF630/DUF632 domains containing protein) with the same p -values ($p = 4.73 \times 10^{-20}$). The SNPs in these loci were also most significantly

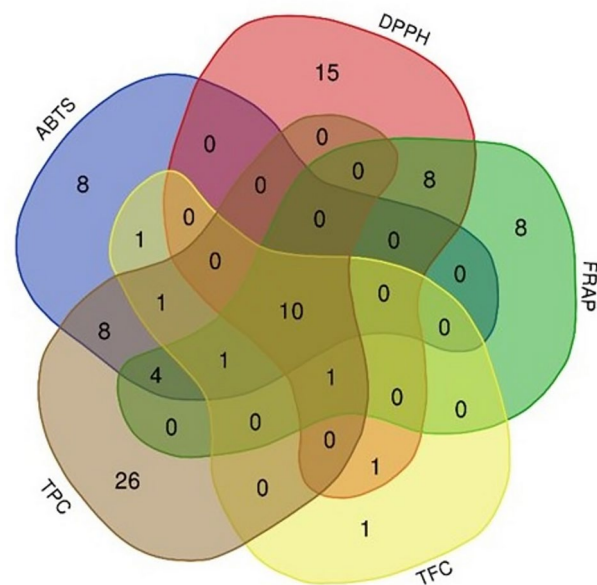


Fig. 5. The Venn diagram of co-associated loci in all five phenotypic traits.

No	Loci	Annotation
1	<i>LOC_Os07g10960</i>	expressed protein
2	<i>LOC_Os07g10970</i>	leucine zipper protein-like, putative, expressed
3	<i>LOC_Os07g11000</i>	structural constituent of ribosome, putative, expressed
4	<i>LOC_Os07g11020</i>	rc—bHLH transcription factor regulating proanthocyanidin production in seeds, expressed
5	<i>LOC_Os07g11050</i>	IPP transferase, putative, expressed
6	<i>LOC_Os07g11060</i>	expressed protein
7	<i>LOC_Os07g11070</i>	DUF630/DUF632 domains containing protein, putative, expressed
8	<i>LOC_Os07g11100</i>	hypothetical protein
9	<i>LOC_Os07g11110</i>	NAD-dependent epimerase/dehydratase family protein, putative, expressed
10	<i>LOC_Os07g11650</i>	LTPL164—protease inhibitor/seed storage/LTP family protein precursor, expressed

Table 2. Ten loci that co-associated with all phenotypic traits based on GWAS.

associated with TFC. (Fig. 4D; $p = 1.34 \times 10^{-19}$). This indicates that the same genetic architecture may be involved in the antioxidant traits detected using FRAP and TFC. The SNPs in *LOC_Os07g11020* were detected as the peak of TPC association (Fig. 4E) with the lowest p -value across all associations ($p = 1.15 \times 10^{-41}$). To identify the loci that overlapped with the SNPs consistently found in multiple associations, a Venn diagram was generated (Fig. 5).

Among the 93 loci associated with antioxidant properties, 10 were associated with all phenotypic traits (Table 2, Fig. 5). All 10 co-associated loci are located on chromosome 7. Nine of these were clustered together, whereas *LOC_Os07g11650* was located 0.3 Mb away from the others. *LOC_Os07g11020* with the lowest p -value (1.16×10^{-41}) was detected by GWAS of TPC trait.

These 10 loci were considered as common loci involved in antioxidant properties, as they were found in all different traits. Some other loci were considered specific markers for particular antioxidant traits, such as *LOC_Os10g30750* encoding Appr-1-p processing enzyme family protein that was only associated with TFC (Fig. 5, Supplementary Table S4). TPC had the most unique loci (26 loci), followed by DPPH (15 loci), and ABTS and FRAP (8 loci in each trait).

Of the 93 genes, 77 (83%) were mapped to previously reported quantitative trait loci (QTLs) on antioxidant traits, which were not only the direct antioxidant capacity but also the chemicals contributed to antioxidant capacity, such as the phenolic and flavonoid contents²⁸, phenyl propanoid glycosides³⁰, vitamin E²⁹, steryl ferulates⁵¹, anthocyanin, proanthocyanidin^{32,27}, folate⁵⁰ and tocotrienol⁴⁹. However, some SNPs on chromosomes 3, 5, and 8 were not consistent with previously reported QTLs. There are 32 SNPs located between 21.1–21.4 Mb of chromosome 3, covering 15 loci that have not been reported to be involved with antioxidant traits. The loci containing the highest number of SNPs are *LOC_Os03g38330*, encoding disease resistance RPP13-like protein 1, *LOC_Os03g38350*, encoding expressed protein, and *LOC_Os03g38390*, encoding GDLS-like lipase. None of them have been reported to be associated with antioxidant traits in rice seeds. Three SNPs in two loci at the far

end of the long arm of chromosome 5 (position 28 Mb) were detected via association with DPPH and FRAP, whereas only one SNP on the short arm of chromosome 8 at position 3.2 Mb has not been reported to be related to antioxidant traits. In the two latter cases, the signal is less convincing than the one on chromosome 3 and could be artefacts. In future studies, the loci between 21.1–21.4 Mb in the long arm of chromosome 3 should be considered for their role and functions (Fig. 6).

Linkage disequilibrium and haplotype analysis

The LD analysis was conducted to evaluate the region's degree of linkage. During meiosis, QTNs in the same block tend to be inherited together. GWAS mapping may identify false positive QTNs close to a real QTN. The average LD for SNPs at 1000 kb distance from each other was 0.76 (Supplementary Figure S1). In this study, we analyzed LD in chromosome 4 and 7, especially neighboring *snpc40579* and *snpc72189* (Fig. 7 and 8), since those two QTNs located within significant loci, *LOC_Os04g47080/Ra* gene and *LOC_Os07g11020/Rc* gene, respectively.

An LD analysis on chromosome 4 shows that QTNs surrounding *snpc40579* along the distance between 27.8 k to 28.0 k (Fig. 7A, Block 2) are inherited together more frequently than would be expected by chance. Moreover, as shown in Figs. 7B and 7C, there are eight haplotypes with frequencies higher than 0.01 within block 2. Black rice cultivars were typically found in haplotype 8. Red rice cultivars were found across haplotypes 1, 3, 5, and 6. Meanwhile, white rice cultivars are found across all haplotypes except haplotype 8. Various remaining haplotypes with a frequency less than 0.01 actually showed a high correlation with antioxidant levels.

On the other hand, LD analysis on chromosome 7 focused on the region between 6.0 Kb and 6.1 Kb surrounding *snpc72189* (Fig. 8A). At this distance, there were two blocks, Block 2 and 3. Block 2 consisted of three haplotypes (2 homozygous and 1 heterozygous) with 12 SNP each (Fig. 8B and C). Haplotype 1 consisted of all group cultivars. Haplotype 2 was typically dominated by red and white. Meanwhile, one black and one red rice were two cultivars with heterozygote genotypes (Fig. 8B). A black rice cultivar, 'Khao Niaw Dam Maw', was typically a member of haplotype 1 with the highest antioxidant levels. At the same time, a red rice cultivar 'Khao Niaw Daeng Graham Raed', was the highest antioxidant level in haplotype 2. Block 3 contained five haplotypes with frequencies larger than 0.01 (Fig. 8B and C). Black rice cultivars were found across haplotypes 1, 2, followed by red rice cultivars in haplotypes 3, 4, and 5. In comparison, white rice cultivars were spread across haplotypes 1, 2, and 3 (Fig. 8B).

Antioxidant markers were designed based on *OsRa* and *OsRc*

Snpc40579 is located on *LOC_Os04g47080* and overlapped with *OsRa*³⁴. Four associations (DPPH, FRAP, TFC, and TPC) showed that this QTN had a significantly low *p*-value ($p = 2.31 \times 10^{-07} - 3.01 \times 10^{-16}$). Therefore, this gene was selected for the development of genetic markers. *snpc72189* is located on *LOC_Os07g11020* and overlapped with *OsRc*³⁵. This QTN had the most consistent association with all phenotypic traits, with a significantly low

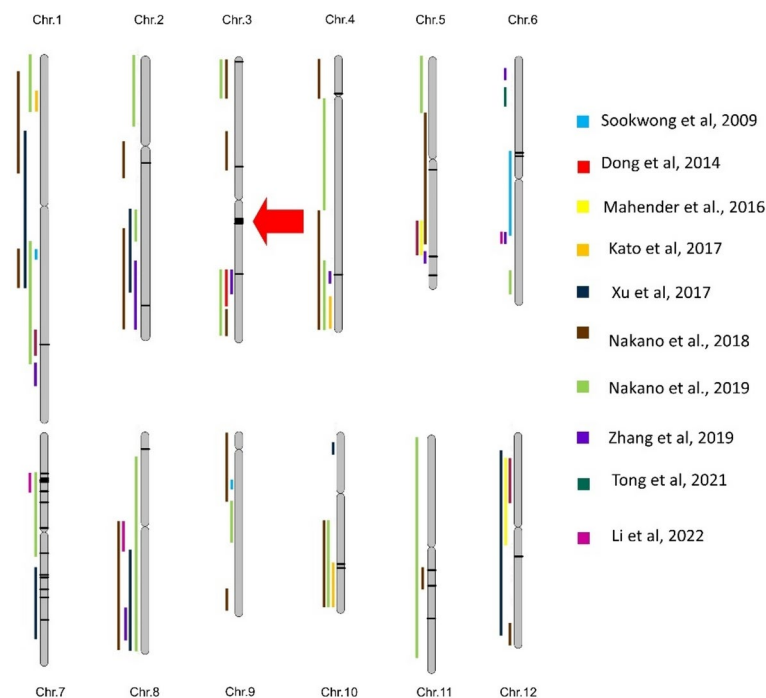


Fig. 6. The mapping of previously reported QTLs is shown with different color bars located beside the chromosomes in comparison with the detected loci from GWAS in this study, as shown by the horizontal black bars in the chromosomes. The red arrow indicates the potential QTN relating to antioxidants that have not been reported by QTLs marker before.

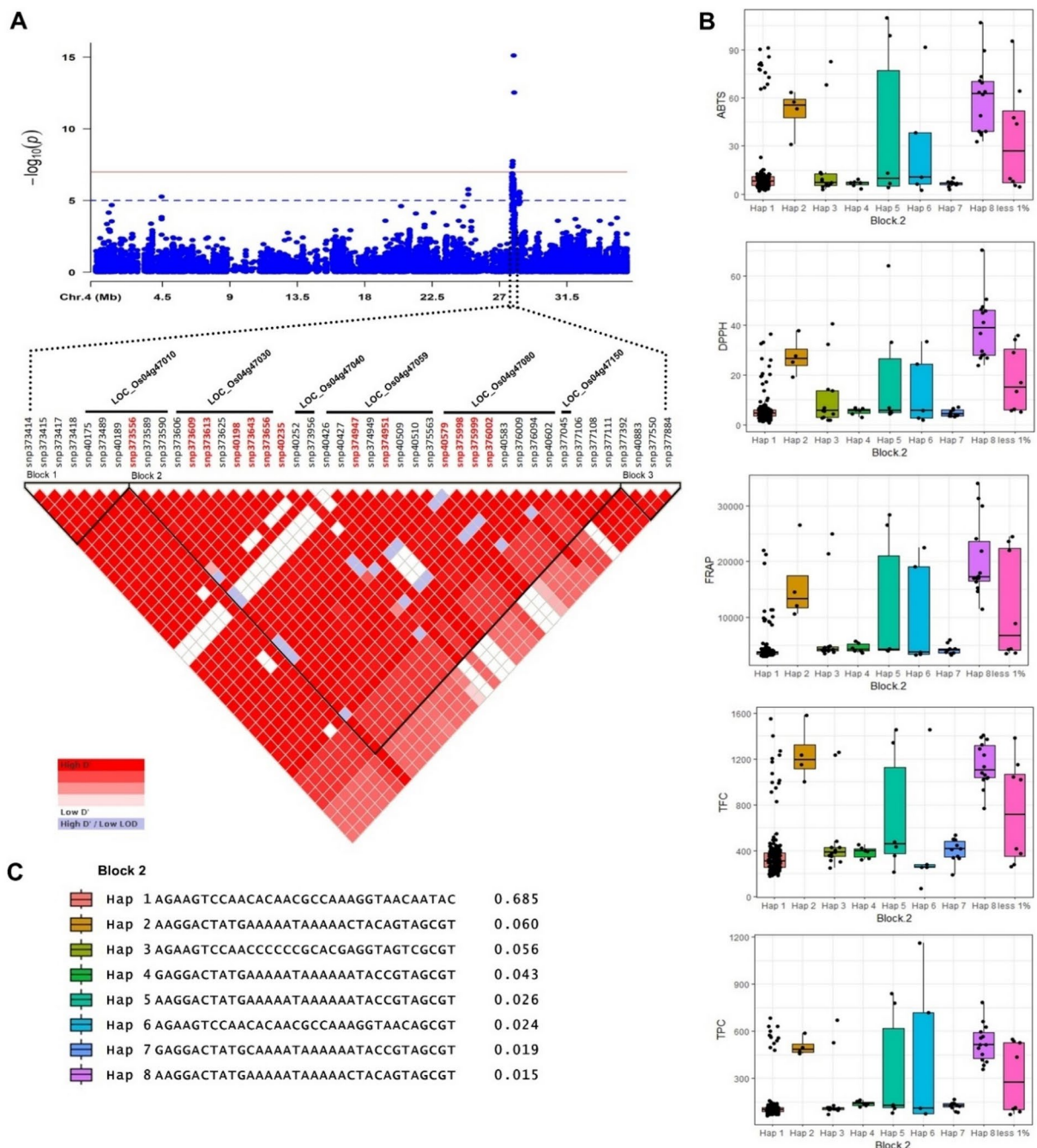


Fig. 7. Linkage disequilibrium near snp40579 and distribution of haplotype of each phenotype. **(A)** Local Manhattan plot (top) and LD heatmap (bottom). The significant SNPs within the block are represented in red font. **(B)** Boxplot of distribution of each haplotype in Block 2 regarding all antioxidant phenotypes. **(C)** The sequences of eight haplotypes in Block 2 have frequencies >0.01. Numbers on the right-hand side represent the frequency of each haplotype.

p -value ($p = 2.26 \times 10^{-11} - 1.16 \times 10^{-41}$). This gene was selected as the second region for the development of genetic markers that can discriminate between rice cultivars with low and high antioxidant levels.

The genetic markers targeted snp40579 (chr4) and snp72189 (chr7). For each SNP, approximately 2.5 Kb was expanded in both directions of the selected SNPs to determine regions with (insertion-deletions) INDELs showing an association with antioxidant traits. Marker efficiency was tested on 96 rice cultivars: 44 colored rice and 52 white rice cultivars. The cultivars are listed in Supplementary Table S5.

For snp40579 (chr4), the marker was designed in the region of *LOC_Os04g47059*, which was previously annotated as *OsTT8* (Transparent Testa 8)³⁶, as insufficient INDELs were found in *LOC_Os04g47080* (*OsRa*). *OsTT8* is located between *OsRa* and nearby QTNs in *LOC_Os04g47030*. The expected sizes of the PCR products

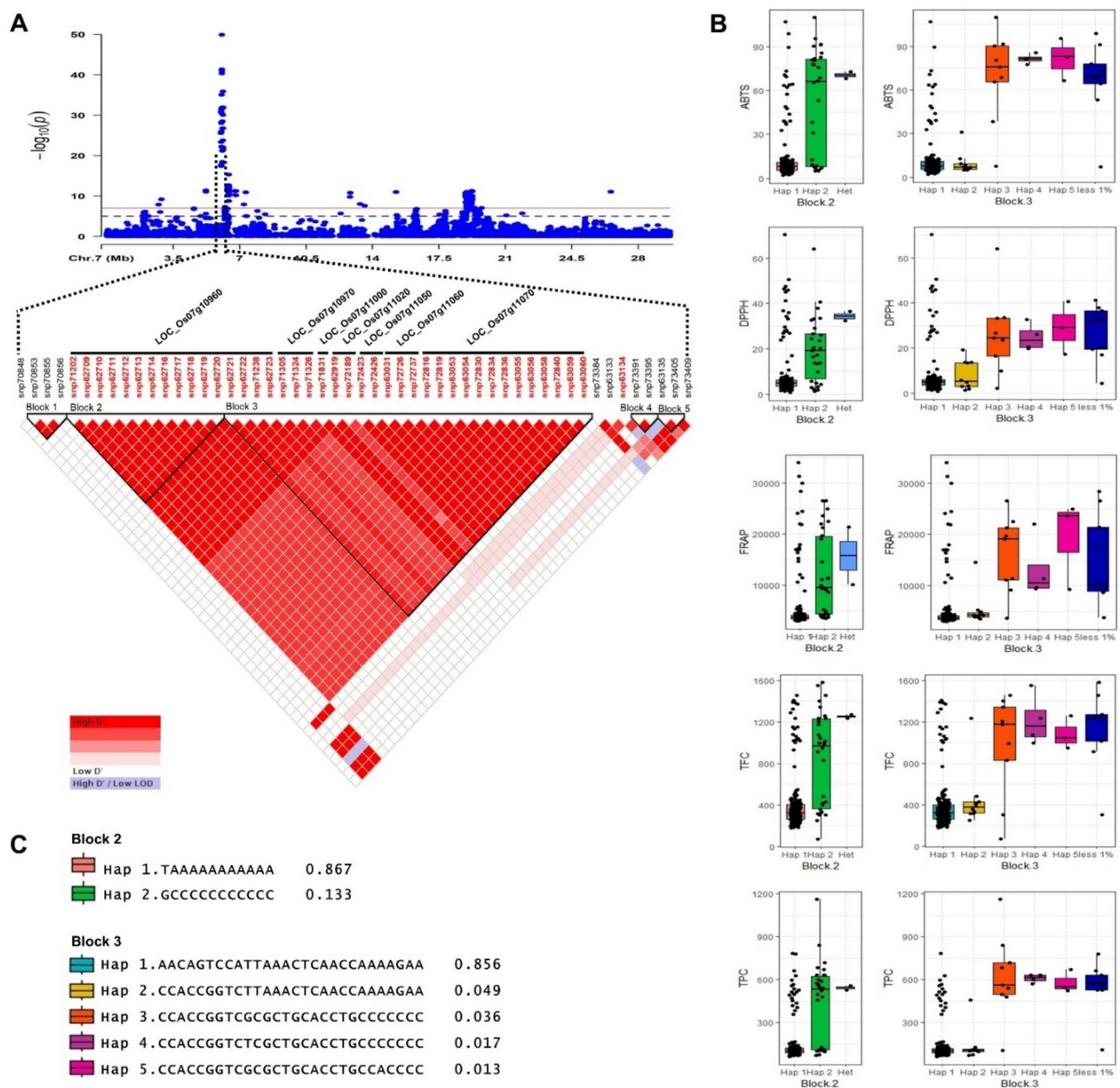


Fig. 8. Linkage disequilibrium near snp72189 and distribution of haplotype of each phenotype. **(A)** Local Manhattan plot (top) and LD heatmap (bottom). The significant SNPs within the block are represented in red font. **(B)** Boxplot of distribution of each haplotype in Block 2 and Block 3 regarding all antioxidant phenotypes. **(C)** The sequences of haplotypes in Block 2 and Block 3 have frequencies > 0.01. Numbers on the right-hand side represent the frequency of each haplotype.

were 510 bp for allele 1 (Hom1), in cultivars with low antioxidant levels, and 464 bp for allele 2 (Hom2), in cultivars with high antioxidant levels. The example of PCR banding patterns was shown in **Supplementary Figure S2**.

Poin-biserial correlation was used to determine the association of PCR product patterns amplified with the marker in *OsTT8* gene and the values of five antioxidant properties (ABTS, DPPH, FRAP, TFC, and TPC; Fig. 9; **Supplementary Table S5**). All correlations show a positive trend, with R-values ranging from 0.49–0.76. The lowest R-value was observed for ABTS, whereas the highest was detected for DPPH. All black rice cultivars possessed the Hom2 allele, whereas all white rice cultivars possessed the Hom1 allele. Nevertheless, several colored rice cultivars with high antioxidant levels behaved the same as white rice cultivars (possessing allele Hom1). We found that 15 red rice cultivars had the Hom1 allele (**Supplementary Table S5**). We assumed that the genetic background of Thai rice cultivars is unique and may not be discriminated using this marker alone. As indicated in the PCA and phylogenetic tree analyses (Fig. 3), the genetic makeup of the colored rice cultivars was mixed with that of the white rice cultivars.

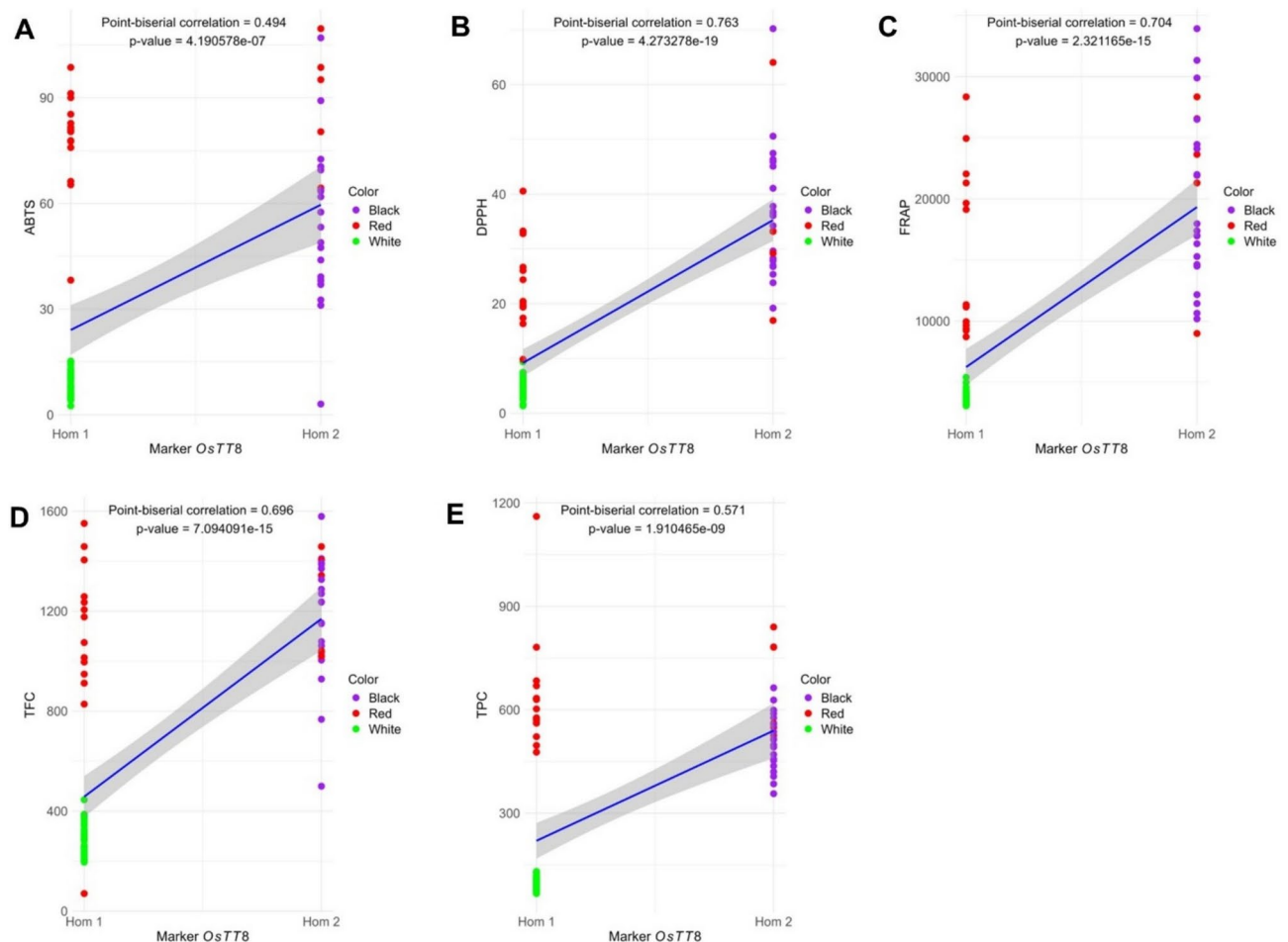


Fig. 9. Point-biserial correlation and linear regression of the first marker in the *OsTT8* gene with five antioxidant phenotypes: **(A)** ABTS, **(B)** DPPH, **(C)** FRAP, **(D)** TFC, and **(E)** TPC. Green, white rice cultivars; purple, black rice cultivars; red, red rice cultivars.

The second marker was focused on a position near snp72189 (*OsRc*). A primer pair was designed to cover the promoter region of *OsRc*. The expected fragment size amplified from the genomic DNA of white rice was 241 bp and was designated as the Hom3 allele, whereas that from the genomic DNA of colored rice was 216 bp and was designated as the Hom4 allele. The PCR product patterns showing Hom3 and Hom4 alleles were shown in **Supplementary Figure S3**.

The Point-biserial correlations between alleles and antioxidant traits in the 96 cultivars showed a positive trend, with an R-value ranging from 0.35–0.70 (Fig. 10A–E). DPPH was the least affected by this second marker, as indicated by an R-value of 0.35. This finding contrasts with that of the *OsTT8* marker (R-value = 0.76) (Fig. 9B). Correlation of *OsRc* with ABTS had the highest R-value (R-value = 0.70) (Fig. 10A). Surprisingly, the correlation between ABTS and the *OsTT8* marker was the lowest among all correlations with antioxidant properties (Fig. 9A). This finding suggests that the antioxidant properties detected by the different methods reflect the different genes involved in antioxidant regulation.

All the red rice cultivars possessed the Hom4 allele, whereas all the white rice cultivars possessed the Hom3 allele. Similar to the *OsTT8* marker, 22 black rice genotypes contained the Hom3 allele. Furthermore, six red rice cultivars possessed the Hom2 (*OsTT8* marker) and Hom4 (*OsRc* marker) alleles (**Supplementary Table S5**).

Discussion

Thailand is a center for rice diversity with more than 5,000 types of rice³⁷ and since the green revolution has been a leading global exporter of rice³⁸. Thailand has four rice production regions based on ecological resource conditions and topographic factors: (1) the northern area, which produces 23% of all paddy rice; (2) the northeastern area, which produces 44% of fragrant rice; (3) the central plain area, which produces 26% of deep-water rice; and (4) the southern area, where is the origins of multiple types of rice, white, red and black rice³⁹. The pigmented rice cultivars studied here were distributed in the northern, northeastern, and southern parts of Thailand whereas the white rice cultivars were mostly distributed in the central and eastern regions (Fig. 11). For instance, specific rice genotypes in the highlands of northern Thailand have higher monomeric anthocyanin and pigment content than those in the lowlands⁴⁰. Local high-quality rice varieties have gained market share owing

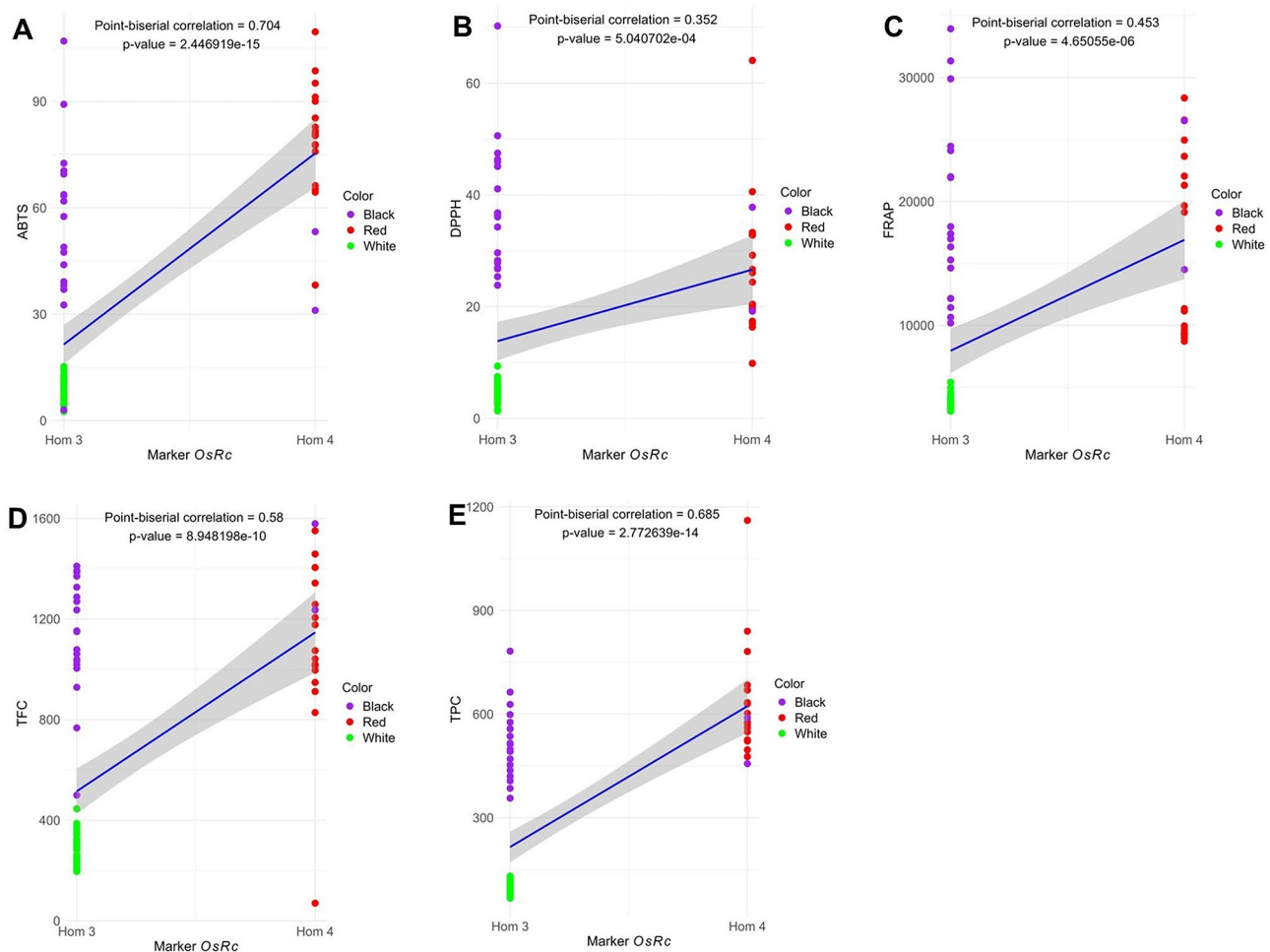


Fig. 10. Point-biserial correlation and linear regression of the second marker in the *Rc* gene promoter with five antioxidant phenotypes: (A) ABTS, (B) DPPH, (C) FRAP, (D) TFC, and (E) TPC. Green: White rice cultivars, Purple: Black rice cultivars, and Red: Red rice cultivars.

to growing interest both locally and internationally. Cultivars with pigmented pericarps are notable for their potential health benefits and commercialization⁴¹.

Pigmented rice is widely known to have antioxidant levels superior to those of white rice. However, color variations may affect the antioxidant levels. In this study, the red rice cultivars had significantly higher ABTS and TPC values than the black rice cultivars. In contrast, DPPH values were higher in the black rice cultivars. Meanwhile, FRAP and TFC values did not significantly differ between the red and black rice cultivars. According to Thaworn, et al.⁴², red Thai rice cultivars have much higher ABTS and TPC values, but the same TFC values as purple rice cultivars. ‘Lai Mak’, ‘Chaw Pli Khao’ (red rice cultivar), and ‘Gam Feueng’ (purple rice cultivar) exhibited the highest antioxidant properties. Higher levels of TPC in red rice than in purple rice were attributed to protocatechuic acid (11.08 and 5.77 mg/100 g) and syringic acid (0.42 and 0.07 mg/100 g)⁵. Proanthocyanidins may have potent antioxidant activity superior to that of anthocyanins due to the high content of ortho-dihydroxyl groups⁴³. Nevertheless, another study reported that black rice cultivars had a net tendency for the highest values based on DPPH, FRAP, TFC, and TPC, followed by red, purple, and brown rice cultivars⁵. Consistent with the results of previous studies, variations in antioxidant properties within the same pericarp type were found in this study such as the phenolic acid⁴⁴, polyphenol⁴⁵, and phenolic compound concentration⁴⁶.

Variation in metabolites depends on two possible causes: changes in regulatory factors and changes in biosynthetic enzymes. In the Thai rice varieties, *LOC_Os07g11020* was the locus most consistently associated with all antioxidant traits. It is also known as the *OsRc* gene^{35,47,48}, which encodes an MYB-helix-loop-helix (bHLH) motif domain regulatory protein and is involved in proanthocyanidin synthesis in rice pericarp. Previous studies have demonstrated the association of the *OsRc* gene with pericarp color. Yang, et al.⁴⁹ reported that a GWAS using the presence-absence and degree of pericarp color methods showed a high density of signals on chromosome 7 with the signal peak on the *OsRc* gene. Nitika, et al.⁵⁰ also mentioned similar GWAS results which *OsRc* was the main detected gene in pericarp color using 1,349,269 SNPs and 325 diverse rice accessions. Moreover, five out of 10 loci that co-associated with all phenotypic traits (*LOC_Os07g10970*, *LOC_Os07g11000*, *LOC_Os07g11050*, *LOC_Os07g11060*, *LOC_Os07g11070*) were located in same LD block with *LOC_Os07g11020*.

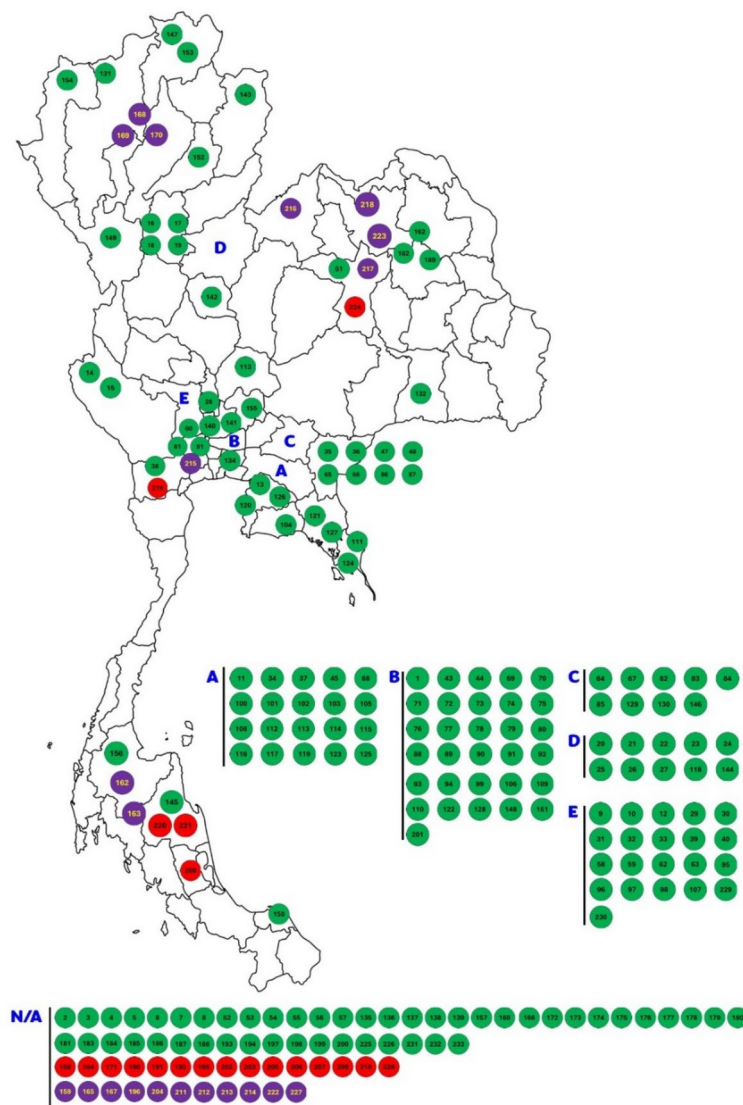


Fig. 11. Geographic distribution of Thai rice cultivars used in the current study. The cultivar numbers are listed in Supplementary Table S1. The colors of the circles represent the pericarp color: green, white rice; red, red rice; and purple, black rice. N/A indicates that there is no information available regarding the location of the cultivar.

/ *OsRc* (Fig. 7A). This finding indicates that nearby loci tend to be inherited along with *LOC_Os07g11020* instead of having their own effect.

As shown in Fig. 6, approximately 83% of the QTNs detected by GWAS were mapped with known QTL reported by several publications, including Sookwong, et al.³³, Dong Wei, et al.³², Mahender, et al.⁵¹, Kato, et al.⁷, Xu, et al.³⁰, Nakano, et al.²⁹, Nakano, et al.²⁷, Zhang, et al.²⁸, Tong, et al.³¹, and Li, et al.²⁶. The finding of our GWAS analysis, therefore, indicates that several of the same QTL identified in previous studies contributed to antioxidant properties in Thai rice varieties.

The *Ra* gene has been reported to regulate antioxidant traits in rice. *Ra* encodes the anthocyanin regulator Lc (*LOC_Os04g47080.1*), a bHLH transcription factor. GWAS detected this locus in association with all antioxidant traits, except the ABTS detection method. QTL mapping has indicated that this locus regulates steryl ferulate content²⁹, phenylpropanoid glycosides²⁷, and vitamin E²⁸. Therefore, both *OsRa* and *OsRc* detected in this GWAS were consistent with previous reports.

A gene associated with antioxidants in the present study is *OsTT8*³⁶ (*LOC_Os04g47059.1*), is closely linked to *OsRa* around 40 kb away and located in same LD block (Fig. 7A). The *OsTT8* is also known as *OsKala4*⁵² and *OsB2*³⁴. This locus is essential for the formation of black-purple pigmentation in rice pericarp. *OsTT8* encodes a bHLH transcription factor, which was well characterized in multiple species as anthocyanin regulator. Mutations that inactivate *OsTT8* are recessive and affect anthocyanin production resulting in transparent or chalky white grains, hence the name "Transparent Testa 8" (TT8)^{53–55}.

Based on Point-biserial's correlation, the *OsTT8* marker was strongly associated with black rice cultivars because black rice varieties always display the Hom2 allele, while red rice cultivars possess either the Hom1 or Hom2 alleles. Compared with the other phenotypic traits, DPPH was most positively correlated with this region (R -value = 0.763). Ghasemzadeh, et al.⁵⁶ reported that the DPPH radical scavenging activities of the free and bound fractions in black rice were higher than those in red and brown rice bran. Moreover, Laokuldilok, et al.⁵⁷ reported that a higher amount of anthocyanin was found in black bran (1135–2562 mg/g), followed by red bran (188 mg/g), and that it was absent in normal bran. The black rice cultivars also contain higher γ -oryzanol (2483–4057 mg/g) and α -tocopherol (24–27 μ g/g) levels compared with red rice cultivars (1859 mg/g, and 16 μ g/g, respectively).

In contrast, the marker on the *Rc* promoter was significantly associated with red rice cultivars because of the consistency of the red rice genotype possessing the Hom4 allele, as opposed to black rice cultivars (possessing the Hom3 or Hom4 allele). *OsRc* encodes an rc—bHLH transcription factor regulating proanthocyanidin production in seeds. The phenotypic traits ABTS and TPC significantly correlated with this region (R -value = 0.704 and 0.685, respectively). This was confirmed by the phenotypic distribution shown in Fig. 1, which demonstrated that red rice cultivars had significantly higher ABTS and TPC values than black rice. Our results are similar to those of Thaworn, et al.⁴² reported that the TPC value in local Thai rice was higher in red rice than in black rice cultivars. However, contrasting findings reported by Wisetkomolmat, et al.³⁷ showed that the ABTS, DPPH, FRAP, TFC, and TPC levels of black rice cultivars ('Bien Koo 5 CMU', 'K 4 CMU', 'Kum Akha 1 CMU', 'Kum Doi Sake', 'Kum Chao Morchor 107', 'Sang 5 CMU', and 'Pieisu 1 CMU') were higher than those of a red rice cultivar ('Yamuechaebia 3 CMU'). Finally, FRAP and TFC appeared to positively correlate with both markers in the black and white cultivars.

In conclusion, our investigation of 223 Thai rice cultivars revealed those with pigmented seeds display higher antioxidant content. Furthermore, much of the variation in antioxidant content depends on few regulatory and biosynthetic loci. We did, however, identify a novel QTL on chromosome 3 that deserves further characterization. We also identified a small and distinct clade of four pigmented varieties that suggest selection for specific properties likely to include anthocyanin pigments.

Materials and methods

Plant materials

A collection of 233 accessions was obtained from the Center of Excellence in Environment and Plant Physiology (Department of Botany, Chulalongkorn University, Thailand) as previously described by¹⁴. The accessions included 191 white, 20 red, and 22 black pericarp cultivars (**Supplementary Table S1**). All rice varieties were planted in the field simultaneously in 2017, and the seeds were collected to determine antioxidant levels using ABTS, DPPH, FRAP, TFC, and TPC assays.

One rice population consisting of 52 white rice cultivars and 44 pigmented rice cultivars was chosen for PCR validation. DNA samples were obtained from the shoot tissue of seedlings at the 2-week-old stage. DNA samples were used to validate the designed genetic markers via PCR.

Determination of total antioxidant capacity, TFC, and TPC

The antioxidant capacities of rice extracts were determined using the ABTS, DPPH, and FRAP assays^{21–23}. Results are expressed as the TEAC per 100 g of dry weight whole-grain rice. TFC was calculated according to Shen et al.⁴⁶, and TPC was calculated according to Zhang et al.²⁵. The results are expressed as mg GAE per 100 g of dry weight whole-grain rice. The measured values for these antioxidant assays are listed in **Supplementary Table S1**.

The box plots were used to illustrate the distribution of antioxidant phenotypes among white, red, and black rice cultivars. The student's *t*-test was used to analyze the significant difference of antioxidant level among rice pigmentation group. The first two principal components of the PCA were calculate using *prcomp()* function in R Program and visualized using *ggplot2* package⁵⁸ to represent the phenotypic variation. A heatmap with a hierarchical clustering plot was constructed using R package "pheatmap"⁵⁹ based on Euclidean distances to identify clustering patterns in the rice populations based on their antioxidant properties.

Thai rice genome sequencing

Genotype data consisting of 345,448 SNPs were obtained from the whole-genome sequences of 233 local Thai rice accessions. The rice genome of these 233 accessions were sequenced using Illumina NovaSeq system (Illumina, Inc., San Diego, CA, USA). The whole-genome sequencing data exhibited high quality, with a mean coverage of 92.5%, an average sequencing depth of 37.46x, a mean base quality score of 36.35, and a mean mapping quality score of 41.61, ensuring robust and accurate variant calling across the genome.

Variant calling and preparation for GWAS

Burrow-Wheeler Aligner (BWA) version 0.5.7–1⁶⁰ was used to align raw reads against genome assembly IRGSP-1.0. Variants were called following genome analysis toolkit (GATK) best practice pipeline using GATK version 3.3–0⁶¹. Variants with minor allele frequencies below 5% were excluded, and all sites with missing data were filtered out, resulting in a missing site proportion of 0. Variants were filtered based on the following criteria: for a heterozygous call, the minimum coverage was set to 5, with each of the two major alleles comprising at least 20% of the reads, and a minimum total coverage of 10 was required. Multiallelic variants were discarded. For a homozygous call, a minimum coverage of 6 was used, or a minimum of 3 if at least 10 accessions had positions with coverage of 6. Variants from the first five exons of each gene were selected for GWAS analysis to enhance the detection of causative associations within genes and increase the likelihood of detecting significant association which optimize the power of study in the given the sample size. The density of SNPs in each chromosome then

was visualized using R package 'CMplot'⁶². A PCA was also generated to illustrate the population stratification within rice population using R function `prcomp()` and package 'ggplot'⁵⁸. A phylogenetic tree was calculated based on UPGMA using VCF2PopTree⁶³ to show the genotypic relationships among cultivars. Some cultivars were emphasized as potential parental lines based on their antioxidant levels and genetic backgrounds for further rice breeding programs.

Identification of genes associated with antioxidant properties using GWAS

GWAS was performed to determine the association between SNP markers and the five antioxidant traits. GWAS was conducted on GEMMA v0.96 using MLM⁶⁴, which treats the SNP markers as fixed effects and phenotypic traits, kinship matrix (K), and PCA for population structure as random effects. Bonferroni correction ($0.05/m$; m = number of SNP markers) was used for the single-locus method; the threshold for association was corrected to 4.18×10^{-7} for the GWAS with 119,541 SNPs. Manhattan plots were generated using the R package 'CMplot'⁶². SNPs with a p -value below the cutoff threshold were selected as significant SNPs and used for annotation analysis against the Rice Genome Annotation Project database (<http://rice.uga.edu>)⁶⁵. The candidate loci were then obtained based on the QTN positions that either hit gene positions or are located 2.5 Kb near the gene positions based on the LD-plot decay.

The potential genes responsible for antioxidant traits were determined via functional annotation and comparison with previously reported QTLs underlying antioxidant traits. The Conserved Domain Database tool (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) of National Center for Biotechnology Information was used to predict the conserved domain of each candidate gene⁶⁶. The reported QTLs associated with antioxidant properties were collected from publications over the last 10 years, and the physical locations were retrieved from the Gramene database (<https://archive.gramene.org/markers/>). They were subsequently mapped together along the candidate gene using the Chromosome Map Tool (<http://viewer.shigen.oryzavw/maptool/MapTool.do>)^{67,68}. A Venn diagram was generated to identify the loci that were co-associated with all the traits by Drew Venn Diagram (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

Linkage disequilibrium analysis

The linkage disequilibrium (LD) decay was plotted to visualize how the association between alleles at different loci decreases with increasing distance between those loci. Pairwise LD values between markers were calculated as the squared correlation coefficient (r^2) using PLINK 1.9⁶⁹. Then the LD decay conducted by division of marker pairs within the 1000-kb region into bins 125 kbp and then was visualized by ggplot⁵⁸.

The LD-plot associations were done for the region with high association signal in the Manhattan Plot. The SNPs with lowest p -value were selected from corresponding chromosome and extended to the up and down stream regions until the region moved to another LD block. The LD-plot association was then visualized by Haploview⁷⁰. The sequences of each haplotype from the selected block were then plotted as a boxplot based on the phenotype distribution.

Development of genetic markers in local Thai rice cultivars

Genetic markers were developed to emphasize the presence of INDELS in the candidate genes of rice cultivars with different antioxidant levels and pigmentation. Primer pairs were designed to amplify fragments of candidate genes containing all INDELS, leading to a size difference of a minimum of 50 bp, must hit QTNs from any associations and should not have heterozygous allele, using our in-house bash script run in linux environment. These genetic markers were validated using 52 white and 44 pigmented rice cultivars (23 black and 21 red cultivars), which is the subset in the population used for GWAS with high and low antioxidant levels, respectively (Supplementary Table S5). The varying PCR amplicon sizes represent the presence of INDELS in each marker. Finally, Point-biserial's correlation was used to evaluate the correlation between antioxidant levels in rice and genotypic differences in the markers. The primers used were as follows: for the first marker (*OsTT8*), Fwd, CG ATCATCACTAGGCGATTG and Rev, CATGGCATCGTCCCTTATG; and for the second marker (promoter region of *OsRc*), Fwd, AAGAATAAAAACCGTGCTGACG and Rev, GGATGATCGAGATCGATAAGTAAAG.

Data availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. The genome data sets in this study are available and open upon request to corresponding author.

Received: 16 August 2024; Accepted: 18 December 2024

Published online: 09 January 2025

References

1. Yamuangmorn, S. & Prom-u-Thai, C. The potential of high-anthocyanin purple rice as a functional ingredient in human health. *Antioxidants* **10**, 833. <https://doi.org/10.3390/antiox10060833> (2021).
2. Bouis, H. E., Chassy, B. M. & Ochanda, J. O. Genetically modified food crops and their contribution to human nutrition and food quality. *Trends Food Sci. Technol.* **14**, 191–209. [https://doi.org/10.1016/S0924-2244\(03\)00073-6](https://doi.org/10.1016/S0924-2244(03)00073-6) (2003).
3. Min, B., McClung, A. M. & Chen, M. H. Phytochemicals and antioxidant capacities in rice brans of different color. *J. Food Sci.* **76**, C117–C126. <https://doi.org/10.1111/j.1750-3841.2010.01929.x> (2011).
4. Sudan, J. et al. Explicating genetic architecture governing nutritional quality in pigmented rice. *PeerJ* **11**, e15901. <https://doi.org/10.7717/peerj.15901> (2023).
5. Goufo, P. & Trindade, H. Rice antioxidants: phenolic acids, flavonoids, anthocyanins, proanthocyanidins, tocopherols, tocotrienols, γ -oryzanol, and phytic acid. *Food Sci. Nutr.* **2**, 75–104. <https://doi.org/10.1002/fsn3.86> (2014).

6. Saikia, S., Dutta, H., Saikia, D. & Mahanta, C. L. Quality characterisation and estimation of phytochemicals content and antioxidant capacity of aromatic pigmented and non-pigmented rice varieties. *Food Res. Int.* **46**, 334–340. <https://doi.org/10.1016/j.foodres.2011.12.021> (2012).
7. Kato, T., Matsukawa, T. & Horibata, A. Quantitative trait loci responsible for the difference in γ -oryzanol content in brown rice between japonica-type and indica-type rice cultivars. *Plant Prod. Sci.* **20**, 459–466. <https://doi.org/10.1080/1343943X.2017.1372109> (2017).
8. Shammugasamy, B., Ramakrishnan, Y., Ghazali, H. M. & Muhammad, K. Tocopherol and tocotrienol contents of different varieties of rice in Malaysia. *J. Sci. Food Agric.* **95**, 672–678. <https://doi.org/10.1002/jsfa.6742> (2015).
9. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967. <https://doi.org/10.1038/ng.695> (2010).
10. Wang, A. et al. Genome-wide association study-based identification genes influencing agronomic traits in rice (*Oryza sativa* L.). *Genomics* **113**, 1396–1406. <https://doi.org/10.1016/j.ygeno.2021.03.016> (2021).
11. Zhang, P., Zhong, K., Zhong, Z. & Tong, H. Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.). *BMC Plant Biol.* **19**, 1–12. <https://doi.org/10.1186/s12870-019-1842-7> (2019).
12. Zhao, K. et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**, 467. <https://doi.org/10.1038/ncomms1467> (2011).
13. Abhijith, K. P. et al. Genome-wide association study reveals novel genomic regions governing agronomic and grain quality traits and superior allelic combinations for Basmati rice improvement. *Front. Plant Sci.* **13**, 994447. <https://doi.org/10.3389/fpls.2022.994447> (2022).
14. Praphasanobol, P. et al. Genome-wide association study of starch properties in local thai rice. *Plants* **12**, 3290. <https://doi.org/10.3390/plants12183290> (2023).
15. Sachdeva, S. et al. Multi-model genome-wide association studies for appearance quality in rice. *Front. Plant Sci.* **14**, 1304388. <https://doi.org/10.3389/fpls.2023.1304388> (2024).
16. Guo, Z. et al. Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. *Mol. Plant* **11**, 789–805. <https://doi.org/10.1016/j.molp.2018.03.018> (2018).
17. Pantaliao, G. F. et al. Genome wide association study (GWAS) for grain yield in rice cultivated under water deficit. *Genetica* **144**, 651–664. <https://doi.org/10.1007/s10709-016-9932-z> (2016).
18. Lekklar, C. et al. Genome-wide association study for salinity tolerance at the flowering stage in a panel of rice accessions from Thailand. *BMC Genom.* **20**, 1–18. <https://doi.org/10.1186/s12864-018-5317-2> (2019).
19. Patishtan, J., Hartley, T. N., Fonseca de Carvalho, R. & Maathuis, F. J. Genome-wide association studies to identify rice salt-tolerance markers. *Plant Cell Environ.* **41**, 970–982. <https://doi.org/10.1111/pce.12975> (2018).
20. Xu, F., Bao, J., Kim, T.-S. & Park, Y.-J. Genome-wide association mapping of polyphenol contents and antioxidant capacity in whole-grain rice. *J. Agric. Food Chem.* **64**, 4695–4703. <https://doi.org/10.1021/acs.jafc.6b01289> (2016).
21. Re, R. et al. Antioxidant activity applying an improved ABTS radical cation decolorization assay. *Free Radic. Boil. Med.* **26**, 1231–1237. [https://doi.org/10.1016/s0891-5849\(98\)00315-3](https://doi.org/10.1016/s0891-5849(98)00315-3) (1999).
22. Mensor, L. L. et al. Screening of Brazilian plant extracts for antioxidant activity by the use of DPPH free radical method. *Phytother. Res.* **15**, 127–130. <https://doi.org/10.1002/ptr.687> (2001).
23. Benzie, I. F. & Strain, J. J. The ferric reducing ability of plasma (FRAP) as a measure of “antioxidant power”: the FRAP assay. *Anal. Biochem.* **239**, 70–76. <https://doi.org/10.1006/abio.1996.0292> (1996).
24. Chang, C.-C., Yang, M.-H., Wen, H.-M. & Chern, J.-C. Estimation of total flavonoid content in propolis by two complementary colorimetric methods. *J. Food Drug Anal.* <https://doi.org/10.38212/2224-6614.2748> (2002).
25. Zhang, Q. et al. A simple 96-well microplate method for estimation of total polyphenol content in seaweeds. *J. Appl. Phycol.* **18**, 445–450. <https://doi.org/10.1007/s10811-006-9048-4> (2006).
26. Li, K. et al. Genetic variation and association mapping of phenolic, flavonoid content and antioxidant capacity in USDA rice mini-core collection. *Genet. Resour. Crop Evol.* **69**, 1685–1694. <https://doi.org/10.1007/s10722-021-01333-y> (2022).
27. Nakano, H., Takai, T. & Kondo, M. Identification of quantitative trait loci for the concentrations of phenylpropanoid glycosides in brown rice. *ACS Omega* **4**, 17317–17325. <https://doi.org/10.1021/acsomega.9b02030> (2019).
28. Zhang, X. et al. Genetic dissection of vitamin E content in rice (*Oryza sativa*) grains using recombinant inbred lines derived from a cross between ‘Zhenshan97B’ and ‘Nanyangzhan’. *Plant Breed.* **138**, 820–829. <https://doi.org/10.1111/pbr.12739> (2019).
29. Nakano, H., Takai, T. & Kondo, M. Quantitative trait loci regulate the concentrations of sterol ferulates in brown rice. *Cereal Chem.* **95**, 800–810. <https://doi.org/10.1002/cche.10099> (2018).
30. Xu, T. et al. QTL mapping for anthocyanin and proanthocyanidin content in red rice. *Euphytica* **213**, 1–11. <https://doi.org/10.1007/s10681-017-2035-9> (2017).
31. Tong, J.-P., Han, Z.-S. & Han, A.-N. Mapping of quantitative trait loci for purple stigma and purple apiculus in rice by using a zhenshan 97B/minghui 63 RIL population. *Czech J. Genet. Plant Breed.* <https://doi.org/10.17221/20/2021-CJGPB> (2021).
32. Wei, D. et al. Identification of QTLs underlying folate content in milled rice. *J. Integr. Agric.* [https://doi.org/10.1016/S2095-3119\(13\)60537-7](https://doi.org/10.1016/S2095-3119(13)60537-7) (2014).
33. Sookwong, P. et al. Cross-fertilization for enhancing tocotrienol biosynthesis in rice plants and QTL analysis of their F2 progenies. *J. Agric. Food Chem.* **57**, 4620–4625. <https://doi.org/10.1021/jf900394t> (2009).
34. Xia, D. et al. How rice organs are colored: the genetic basis of anthocyanin biosynthesis in rice. *Crop J.* **9**, 598–608. <https://doi.org/10.1016/j.cj.2021.03.013> (2021).
35. Furukawa, T. et al. The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **49**, 91–102. <https://doi.org/10.1111/j.1365-3113.2006.02958.x> (2007).
36. Chen, M. et al. TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in Arabidopsis. *Plant Physiol.* **165**, 905–916. <https://doi.org/10.1104/pp.114.235507> (2014).
37. Wisetkomolmat, J. et al. Comparative analysis of nutritional components and phytochemical attributes of selected Thai rice bran. *Front. Nutr.* **9**, 833730. <https://doi.org/10.3389/fnut.2022.833730> (2022).
38. Maclean, J., Hardy, B. & Hettel, G. *Rice Almanac: Source book for one of the most important economic activities on earth* 4th edn. (International Rice Research Institute IRRI, 2013).
39. Xu, L. et al. Paddy rice mapping in thailand using time-series sentinel-1 data and deep learning model. *Remote Sens.* **13**, 3994. <https://doi.org/10.3390/rs13193994> (2021).
40. Rerkasem, B., Jumrus, S., Yimyan, N. & Prom-u-Thai, C. Variation of grain nutritional quality among Thai purple rice genotypes grown at two different altitudes. *ScienceAsia* <https://doi.org/10.2306/scienceasia1513-1874.2015.41.377> (2015).
41. Rerkasem, B. The rice value chain: A case study of Thai rice. *Chiang Mai Univ. J. Soc. Sci. Humanit.* <https://doi.org/10.1298/CMUJASR.2017.0001> (2017).
42. Thaworn, S., Chadchawan, S., Paliyavuth, C. & Kasettranon, W. Clustering of white, red and purple rice cultivars according to their total phenolic content, total flavonoid content and antioxidant capacity in their grains. *Agric. Nat. Resour.* <https://doi.org/10.3404/4/j.anres.2021.55.1.12> (2021).
43. Cai, Y.-Z. et al. Structure–radical scavenging activity relationships of phenolic compounds from traditional Chinese medicinal plants. *Life Sci.* **78**, 2872–2888. <https://doi.org/10.1016/j.lfs.2005.11.004> (2006).
44. Walter, M. et al. Antioxidant properties of rice grains with light brown, red and black pericarp colors and the effect of processing. *Food Res. Int.* **50**, 698–703. <https://doi.org/10.1016/j.foodres.2011.09.002> (2013).

45. Goffman, F. & Bergman, C. Rice kernel phenolic content and its relationship with antiradical efficiency. *J. Sci. Food Agric.* **84**, 1235–1240. <https://doi.org/10.1002/jsfa.1780> (2004).
46. Shen, Y. et al. Total phenolics, flavonoids, antioxidant capacity in rice grain and their relations to grain color, size and weight. *J. Cereal Sci.* **49**, 106–111. <https://doi.org/10.1016/j.jcs.2008.07.010> (2009).
47. Sweeney, M. T. et al. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* **3**, e133. <https://doi.org/10.1371/journal.pgen.0030133> (2007).
48. Sweeney, M. T., Thomson, M. J., Pfeil, B. E. & McCouch, S. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**, 283–294. <https://doi.org/10.1105/tpc.105.038430> (2006).
49. Yang, W. et al. Genome-wide association study of pericarp color in rice using different germplasm and phenotyping methods reveals different genetic architectures. *Front. Plant Sci.* **13**, 841191. <https://doi.org/10.3389/fpls.2022.841191> (2022).
50. Nitika, R. et al. Identification of genomic loci governing pericarp colour through GWAS in rice (*Oryza sativa* L.). *Ind. J. Genet. Plant Breed.* <https://doi.org/10.31742/ijgpb.82.1.1> (2021).
51. Mahender, A., Anandan, A., Pradhan, S. K. & Pandit, E. Rice grain nutritional traits and their enhancement using relevant genes and QTLs through advanced approaches. *Springerplus* **5**, 1–18. <https://doi.org/10.1186/s40064-016-3744-6> (2016).
52. Kim, D.-H. et al. An OsKala3, R2R3 MYB TF, is a common key player for black rice pericarp as main partner of an OsKala4, bHLH TF. *Front. Plant Sci.* **12**, 765049. <https://doi.org/10.3389/fpls.2021.765049> (2021).
53. Chen, M. H., Pinson, S. R., Jackson, A. K. & Edwards, J. D. Genetic loci regulating the concentrations of anthocyanins and proanthocyanidins in the pericarps of purple and red rice. *Plant Genome* <https://doi.org/10.1002/tpg2.20338> (2023).
54. Oikawa, T. et al. The birth of a black rice gene and its local spread by introgression. *Plant Cell* **27**, 2401–2414. <https://doi.org/10.1105/tpc.15.00310> (2015).
55. Wang, C. & Shu, Q. Fine mapping and candidate gene analysis of purple pericarp gene Pb in rice (*Oryza sativa* L.). *Chin. Sci. Bull.* <https://doi.org/10.1007/s11434-007-0472-x> (2007).
56. Ghasemzadeh, A., Karbalaii, M. T., Jaafar, H. Z. & Rahmat, A. Phytochemical constituents, antioxidant activity, and antiproliferative properties of black, red, and brown rice bran. *Chem. Cent. J.* **12**, 1–13. <https://doi.org/10.1186/s13065-018-0382-9> (2018).
57. Laokuldilok, T., Shoemaker, C. F., Jongkaewwattana, S. & Tulyathan, V. Antioxidants and antioxidant activity of several pigmented rice brans. *J. Agric. Food Chem.* **59**, 193–199. <https://doi.org/10.1021/jf103649q> (2011).
58. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 2nd edn, Vol. 2 (Springer Cham, 2016).
59. Kolde, R. & Kolde, M. R. Package 'pheatmap'. *R package* **1**, 790. <https://doi.org/10.32614/CRAN.package.pheatmap> (2015).
60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
61. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the cloud: using Docker, GATK, and WDL in Terra* (O'Reilly Media, 2020).
62. Yin, L. et al. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteom. Bioinform.* **19**, 619–628. <https://doi.org/10.1016/j.gpb.2020.10.007> (2021).
63. Subramanian, S., Ramasamy, U. & Chen, D. VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. *PeerJ* **7**, e8213. <https://doi.org/10.7717/peerj.8213> (2019).
64. Zhang, Z. et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360. <https://doi.org/10.1038/ng.546> (2010).
65. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 1–10. <https://doi.org/10.1186/1939-8433-6-4> (2013).
66. Lu, S. et al. CDD/SPARCLE: The conserved domain database in 2020. *Nucl. Acids Res.* **48**, D265–D268. <https://doi.org/10.1093/nar/gkz991> (2020).
67. Youens-Clark, K. et al. Gramene database in 2010: updates and extensions. *Nucl. Acids Res.* **39**, D1085–D1094. <https://doi.org/10.1093/nar/gkq1148> (2010).
68. Kurata, N. & Yamazaki, Y. An integrated biological and genome information database for rice. *Plant Physiol.* **140**, 12–17. <https://doi.org/10.1104/pp.105.063008> (2006).
69. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**, 559–575. <https://doi.org/10.1086/519795> (2007).
70. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265. <https://doi.org/10.1093/bioinformatics/bth457> (2005).

Acknowledgements

We would like to thank the Comai and CEEPP laboratory members for their technical support.

Author contributions

S.C., T.B. and L.C. contributed to the concept of the project. S.C., T.B., L.C., M.P., P.R.P. provided the methods. S.J., S.Cho, P.M., C.P. and W.K. provided plant materials. P.P. and P.R.P. investigated and P.P., M.P., P.R.P. and W.K. analyzed the data. S.C. and T.B. performed data curation. P.R.P. wrote the original draft. S.C., L.C., T.B., P.P. and M.P. reviewed and edited the second draft. L.C. and S.C. finalized the submitted manuscript. S.C. was project administration. This project was supervised by M.P., S.C., S.R. and T.B. All authors reviewed the manuscript.

Funding

This study was supported by the Agricultural Research Development Agency grant (grant number PRP6105020640) and the 90th Anniversary of Chulalongkorn University Scholarship under the Ratchadapisek Somphot Endowment Fund (GCUGR1125652065D no.1–65). PRP was supported by a CU-NSTDA doctoral scholarship.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83926-y>.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024