PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol 12(3): 1-9
https://doi.org/10.14719/pst.6822

HORIZON
e-Publishing Group

RESEARCH ARTICLE

Genomic insights into phenolic content: Multi-environment based marker-trait association mapping in rice (*Oryza sativa* L.)

Nallaiyan Karunya[1], Geetha Seshadri[1]*, Rajeswari Sivakami[1*], Raveendran Muthurajan[2], Jayakanthan Mannu[3], Manonmani Swaminathan[1], Uma Doraiswamy[3], Anandhi Krishnan[1], Ananda Lekshmi[1] & Almas Fathima[1]

[1]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India
[2]Director of Research, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India
[3]Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

OPEN ACCESS

ARTICLE HISTORY

Received: 20 December 2024
Accepted: 01 January 2025

Available online
Version 1.0 : 13 February 2025
Version 2.0 : 29 July 2025

Check for updates

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE

Karunya N, Seshadri G, Sivakami R, Muthurajan R, Mannu J, Swaminathan M, Doraiswamy U, Krishnan A, Lekshmi A, Fathima A. Genomic insights into phenolic content: Multi-environment based marker-trait association mapping in rice (*Oryza sativa* L.). Plant Science Today. 2025; 12(3): 1-9. https://doi.org/10.14719/pst.6822

Abstract

Phenolic acids are crucial for human health due to their potent antioxidant, anti-inflammatory, and antimicrobial properties, which help protect against chronic diseases and support overall well-being. In this study, 44 rice accessions were evaluated for total phenolic content in three different locations of Tamil Nadu and the marker trait association was done using 208 SSR markers. Among the association panel, *Mappillai Samba* was identified as having the highest total phenolic content of 1049.936 mg GAE/100 g. The phylogenetic analysis grouped the panel of entries into five genetic structure groups which nearly matched the geographical distance among the entries. Marker-trait association studies using GLM and MLM revealed that SSR markers RM287 and RM19358 were significantly associated with total phenolic content, explaining 23.4 % and 19.7 % of the observed variability, respectively. These markers were located in genomic regions linked to candidate genes involved in the biosynthesis of trans-cinnamate 4-monooxygenase (C4H) and phenylalanine ammonia-lyase (PAL), key enzymes in the phenolic acid pathway. Identifying these markers provides valuable tools for marker-assisted selection, enabling the development of biofortified rice varieties with enhanced phenolic content. Such advancements promise to improve rice nutritional quality and promoting public health through dietary interventions.

Keywords

C4H; GLM; MLM; PAL; SSR markers; total phenolic content

Abbreviation

C4H: Trans-cinnamate 4-monooxygenase

G × E: Genotype x Environment

GLM: General Linear Model

MAS: Marker-Assisted Selection

mg GAE/100 g: Milligrams of Gallic Acid Equivalent per 100 grams

MLM: Mixed Linear Model

PAL: Phenylalanine ammonia-lyase

PIC: Polymorphism Information Content

TPC: Total Phenolic Content

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple crops, providing sustenance to more than half of the global population. As a primary source of calories and nutrients in many regions, its quality plays a pivotal role in consumer acceptance and market value. Therefore, improving rice quality is a key objective in rice breeding programs worldwide (1). While rice quality has traditionally been assessed based on its physical characteristics and cooking properties, recent research has increasingly emphasized its therapeutic potential. This is particularly relevant in developing countries, where rice serves as a primary dietary staple (2). The growing challenge of global food security is exemplified by concerning statistics, with approximately 733 million people experiencing hunger in 2023 and 2.33 billion facing moderate to severe food insecurity, according to the United Nations and the World Food Programme (3,4). In parallel, there is an increasing demand for healthier food options, with functional foods enriched with phenolic compounds emerging as a promising strategy to enhance nutritional quality and mitigate the impact of chronic diseases (5). By incorporating these bioactive compounds into food systems, it is possible to address both public health needs and the global issue of food insecurity.

Whole rice grains, especially pigmented rice have emerged as a significant focus due to their rich phytochemical profiles, particularly in phenolic acids. The consumption of pigmented rice can enhance overall health by mitigating oxidative stress and reducing inflammation, which are key factors in the development of chronic conditions such as diabetes, cardiovascular diseases, and certain cancers (6,7). Phenolic acids play a crucial role in modulating metabolic pathways and improving glucose homeostasis (8). Furthermore, the regular intake of whole rice grain has been associated with improved gut health due to its high fiber content and prebiotic properties. As interest in functional foods continues to grow, biofortified rice with high phenolic content stands out as a promising dietary option that not only meets nutritional needs but also offers substantial health benefits (9). Therefore, breeding high-yield rice varieties with high phenolic content can address both food security and health improvement.

In this study, 416 rice accessions (361 white, 50 red, and 6 black) were genotyped using 100 SSR markers, identifying the SSR marker RM346 as significantly associated with phenolic content, thereby enhancing the understanding of the genetic factors influencing phenolic variation in rice (10). A study of 164 rice accessions identified 23 significant marker-trait associations for antioxidant traits using 155 SSR markers, with the Rc gene markers showing the strongest links, particularly Rid12, RM484, RM162, and RM5371 associated with phenolic content, flavonoid content, and antioxidant capacity (11). This study addresses the gap in the marker-trait association mapping of phenolic content by linking SSR markers to functional genes in phenolic acid biosynthesis, which were previously underexplored. This contributes valuable molecular tools for biofortification and improving the nutritional quality of rice, advancing current knowledge in rice breeding and public health. Considering these points, the present study was carried out to identify the best donor for improving phenolic content and to identify the SSR markers to accelerate the breeding programme.

Materials and Methods

The 44 rice accessions, obtained from the Ramiah Gene Bank, Department of Plant Genetic Resources, Tamil Nadu Agricultural University, were selected to represent a broad spectrum of genetic diversity. This association mapping panel included landraces, cultivars, improved varieties, and exotic collections from diverse geographical regions such as India, Colombia, Philippines, and Indonesia. The selection captured variability in TPC, spanning low, moderate, and high TPC, ensuring a comprehensive representation of genetic and phenotypic diversity in global rice germplasm.

The panel was evaluated across three distinct agro-climatic zones of Tamil Nadu i.e., Western, Cauvery Delta, and North Western zones in the *Kharif* 2021 to identify the interaction effect of environment on TPC. The seeds were dehusked using Lab Mini Rice Mill (LTJM 2099, GOYOJO, China) and then powdered using a cyclone sample mill equipped with a 0.5 mm screen (Udy Corporation, Fort Collins, Colorado- USA).

TPC was determined by the Folin-Ciocalteu colorimetric method (12). The pulverized samples (250 mg) were extracted using 80 % ethanol (10 ml) and centrifuged (3000 rpm for 20 min) (5430 R, Eppendorf, Germany). The extracts (200 µl) were then mixed with freshly prepared 1:1 Folin-Ciocalteu reagent (500 µl) and 20 % sodium carbonate (1000 µl), followed by incubation for 30 min. The absorbance was recorded at 660 nm using a spectrophotometer (LMSPUV1900, LABMAN, India). TPC was expressed as mg of Gallic acid (mg GAE/100 g) of dry weight.

Genomic DNA was extracted from 15-day-old seedlings using the CTAB extraction method (13). Polymorphic SSR markers (208) were used in this study. A Bayesian model-based clustering analysis was conducted using STRUCTURE v2.3.4, following the admixture model (14). Ten independent runs were performed, each with a burn-in period of 10,000 iterations followed by 100,000 Monte Carlo iterations, considering a range of subpopulations (K = 1–10). The optimal number of subpopulations was determined using an ad hoc statistic (DeltaK), which evaluates the rate of change in log probability between successive K-values using STRUCTURE SELECTOR (15). Additionally, a neighbor-joining phylogenetic tree was constructed using TASSEL software (version 5.0).

Statistical analysis

Statistical analysis was performed using the META-R tool (13), which utilizes the 'lme4' package (14). BLUEs were calculated from three environments using the following equation:

$$y_{ijk} = \mu + E_i + R_{j(i)} + G_k + GE_{ik} + e_{ijk}$$

Where y_{ijk} = trait of interest; μ = overall mean; E_i = random effect of the i^{th} environment; $R_{j(i)}$ = random effect of the j^{th} replicate nested in the ith environment; G_k = fixed effect of the k^{th} genotype, GE_{ik} = genotype by environment ($G \times E$) interaction; e_{ijk} = residual error. The calculated BLUEs were used for association mapping. Statistical analysis of phenotypic data was performed using R software (version 4.1.3).

Association analysis was performed using two models, the General Linear Model (GLM) (16) and Mixed Linear Model (MLM) (17). The association mapping analysis was performed using TASSEL software (version 5.0). The threshold for significant associations was calculated using a *p*-value < 0.05. To identify candidate genes linked with TPC, gene loci within 1 Mb up and downstream were extracted from the Rice Genome Annotation Project website (https://rice.uga.edu/).

Results and Discussion

Total phenolic content

TPC ranged from 105.087 to 1049.936 mg GAE/100 g with an average of 368.165 mg GAE/100 g. Among 44 genotypes, 12, 21 and 11 genotypes showed low (< 200 mg GAE/100 g), moderate (200-500 mg GAE/100 g) and high (> 500 mg GAE/100 g) TPC respectively (Table 1). These findings indicated the presence of a broad genetic diversity for TPC (18). The highest TPC was recorded in *Mappillai Samba* (1049.936 mg GAE/100 g) and the lowest in RPHP 134 (105.087 mg GAE/100 g). The grains of *Mappillai Samba* had numerous bioactive compounds, including β-sitosterol, campesterol, stigmasterol, squalene, trans-4-coumaric acid, p-coumaric acid, chorismic acid, 7-hydroxyflavone, genistein, gamma-tocotrienol, alpha-tocopherol, spermine, and putrescine, which contribute to its health benefits. The pigmented rice accessions exhibited higher TPC compared to the non-pigmented rice accessions (19–21).

Table 1. Details of rice accessions and BLUEs values of total phenolic content

Sl.No	Code	Genotypes	Origin/source	Parentage	TPC (mg GAE/ 100g)
1	RG1	*Mappilai Samba*	Tamil Nadu, India	Landrace	1049.936
2	RG3	*Senkar*	Tamil Nadu, India	Landrace	256.291
3	RG4	*Murugankar*	Tamil Nadu, India	Landrace	735.927
4	RG7	*Kudaivazhai*	Tamil Nadu, India	Landrace	898.515
5	RG9	*Kuruvai kalanjiyam*	Tamil Nadu, India	Landrace	876.468
6	RG12	*Vellai chithiraikar*	Tamil Nadu, India	Landrace	154.754
7	RG15	*Palakachaka*	Tamil Nadu, India	Landrace	263.730
8	RG18	CHIR 11	West Bengal, India	Improved chinsurah	299.042
9	RG20	*Kalvazhai*	Tamil Nadu, India	Landrace	222.496
10	RG22	IR 36	IRRI, Philippines	IR1561-228 -l2/IR1737//CR94-13	294.478
11	RG32	*Thogai samba*	Tamil Nadu, India	Landrace	647.016
12	RG33	*Malayathan Samba*	Tamil Nadu, India	Landrace	236.314
13	RG39	*Kattu Ponni*	Tamil Nadu, India	Landrace	353.108
14	RG41	*Godavari samba*	Tamil Nadu, India	Landrace	296.744
15	RG42	*Erapalli Samba*	Tamil Nadu, India	Landrace	161.890
16	RG44	*Mangan samba*	Tamil Nadu, India	Landrace	173.727
17	RG48	*Kalarkar*	Tamil Nadu, India	Landrace	158.287
18	RG50	*Sornavari*	Tamil Nadu, India	Landrace	179.782
19	RG51	RPHP 134	Kerala, India	Njavara	105.087
20	RG57	RPHP 103	Uttarkhand, India	Pant Sugandh Dhan 17	504.860
21	RG60	*Ramakuruvaikar*	Tamil Nadu, India	Landrace	316.691
22	RG66	*Seevanasamba*	Tamil Nadu, India	Landrace	168.130
23	RG68	IG63 (EC 728711-117674)	IRRI, Philippines	Caawa/Fortuna	258.256
24	RG69	RPHP 48	Uttarakhand, India	Bindii	163.693
25	RG72	*Aarkadu kichili*	Tamil Nadu, India	Landrace	289.313
26	RG76	*Mattakuruvai*	Tamil Nadu, India	Landrace	732.133
27	RG77	*Karuthakar*	Tamil Nadu, India	Landrace	795.392
28	RG83	RPHP 93	Uttarkhand, India	Type-3 Dehraduni basmati	202.954
29	RG92	IG 49(EC 729102-121052)	IRRI, Philippines	Menakely: IRGC51021-1	228.325
30	RG99	IG 31 (EC728844-117829)	Colombia	Oryzica Ilanos 5	250.286
31	RG102	*Varakkal*	Tamil Nadu, India	Landrace	602.660
32	RG103	*Mattaikar*	Tamil Nadu, India	Landrace	516.400

33	RG105	IG 6 (EC 729592-121642)	IRRI, Philippines	SOM CAU 70::IRGC8227-1	236.076
34	RG106	*Katta samba*	Tamil Nadu, India	Landrace	176.314
35	RG110	*Norungan*	Tamil Nadu, India	Landrace	786.729
36	RG118	*Ponmani Samba*	Tamil Nadu, India	Landrace	231.870
37	RG161	*Panamarasamba*	Tamil Nadu, India	Landrace	148.530
38	RG163	*Mikuruvai*	Tamil Nadu, India	Landrace	268.571
39	RG176	*Kodai*	Tamil Nadu, India	Landrace	229.791
40	RG182	ARB 59	Karnataka, India	Variety	207.078
41	RG184	IG 28(EC 728892-117880)	Indonesia	Seratoes hari	198.201
42	RG188	RPHP 8	Andhra Pradesh, India	24(k)	202.688
43	RG190	IG 26 (IC0590943-121899)	IRRI, Philippines	Basmati 370: IRGC 3750-1	219.132
44	RG192	*Nootri pathu*	Tamil Nadu, India	Landrace	901.610

The results of stability analysis indicated that there is no significant influence of environment on TPC. Based on the AMMI biplot, *Mappilai Samba, Murugankar, Kudaivazhai, Palakachaka, Mattaikar* and *Mikuruvai* recorded higher mean values and were found nearer to the origin (Fig. 1). These genotypes exhibited relatively uniform TPC across the three locations, resulting in low genotype × environment (G × E) interactions. This uniformity indicates that their phenotypic expression is largely stable and minimally influenced by environmental variability. Such stability underscores their potential for consistent performance across diverse agroecological conditions, making them valuable for breeding and cultivation. Further, these genotypes were found to be stable across the three environments and hence these genotypes can be used as donors to enhance the TPC (Table 2).

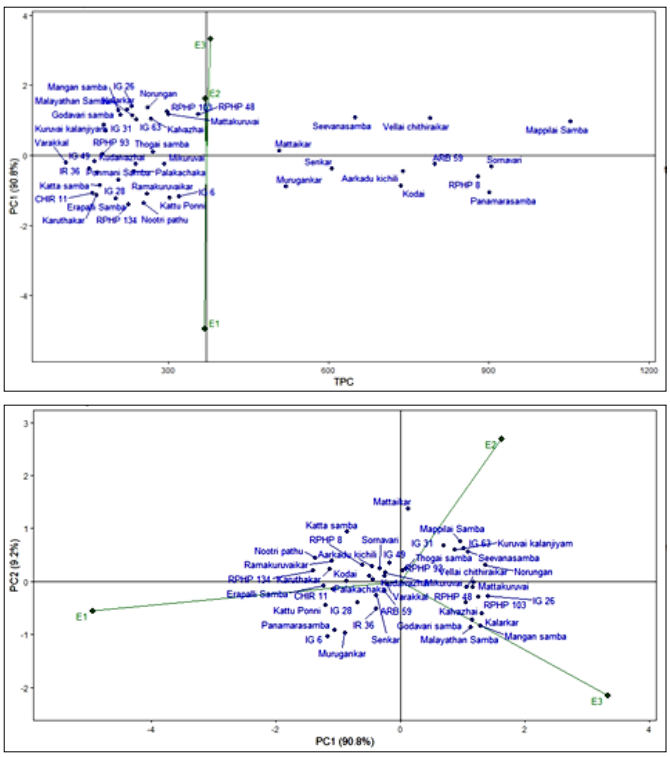


Fig. 1. AMMI biplots for total phenolic content.

Population structure analysis

Polymorphism Information Content (PIC) is a measure of the informativeness of a genetic marker, reflecting both the number of alleles at a locus and their frequency distribution. The PIC values for a set of genetic markers are a useful tool for evaluating the informativeness of these markers in population diversity studies. Markers with PIC values greater than 0.5 are highly informative, those with PIC values between 0.5 and 0.25 are considered moderately informative, while markers with PIC values less than 0.25 remain less informative. In the current study, 42 %, 56 % and 2 % of SSRs were less, moderately and highly informative, which demonstrates the significance of these markers for genetic diversity analysis in rice. The average SSR PIC value of 0.49 obtained in this study was slightly higher than that reported in previous studies on rice (Supplementary Table 1)(22–24).

Table 2. ANOVA for Stability (AMMI) for interactive traits – Total Phenolic Content

	df	SS	MSS
ENV	2	43.0466	21.5233
REP(ENV)	6	537.9322	89.65537
GEN	43	25723390	598218.4**
GEN: ENV	86	4827.321	56.13164
PC1	44	4382.826	99.60968
PC2	42	444.4951	10.58322
Residuals	258	143281.5	555.3545
Total	481	25876910	53798.15

In the present study, the population structure gave five distinct subpopulations across the rice accessions (SP1, SP2, SP3, SP4 and SP5). The subpopulations were identified based on the maximum Delta K value (ΔK = 5) (Fig. 2). SP1, SP2 and SP5 predominantly consist of genotypes with medium and high TPC, while SP3 and SP4 primarily contain low TPC accessions, respectively. Additionally, these sub-populations were categorized into distinct groups based on their geographical origin. SP1 comprised eight genotypes, which were composed of landraces from South Asia (India). SP2 also comprised eight genotypes, among which six were from South Asia (India), and two from South East Asia (Indonesia and Philippines). SP3 comprised four genotypes, among which two were from South Asia (India) and two from Southeast Asia (Philippines). SP4 comprised five genotypes, among which four were from South Asia (India) and one from South

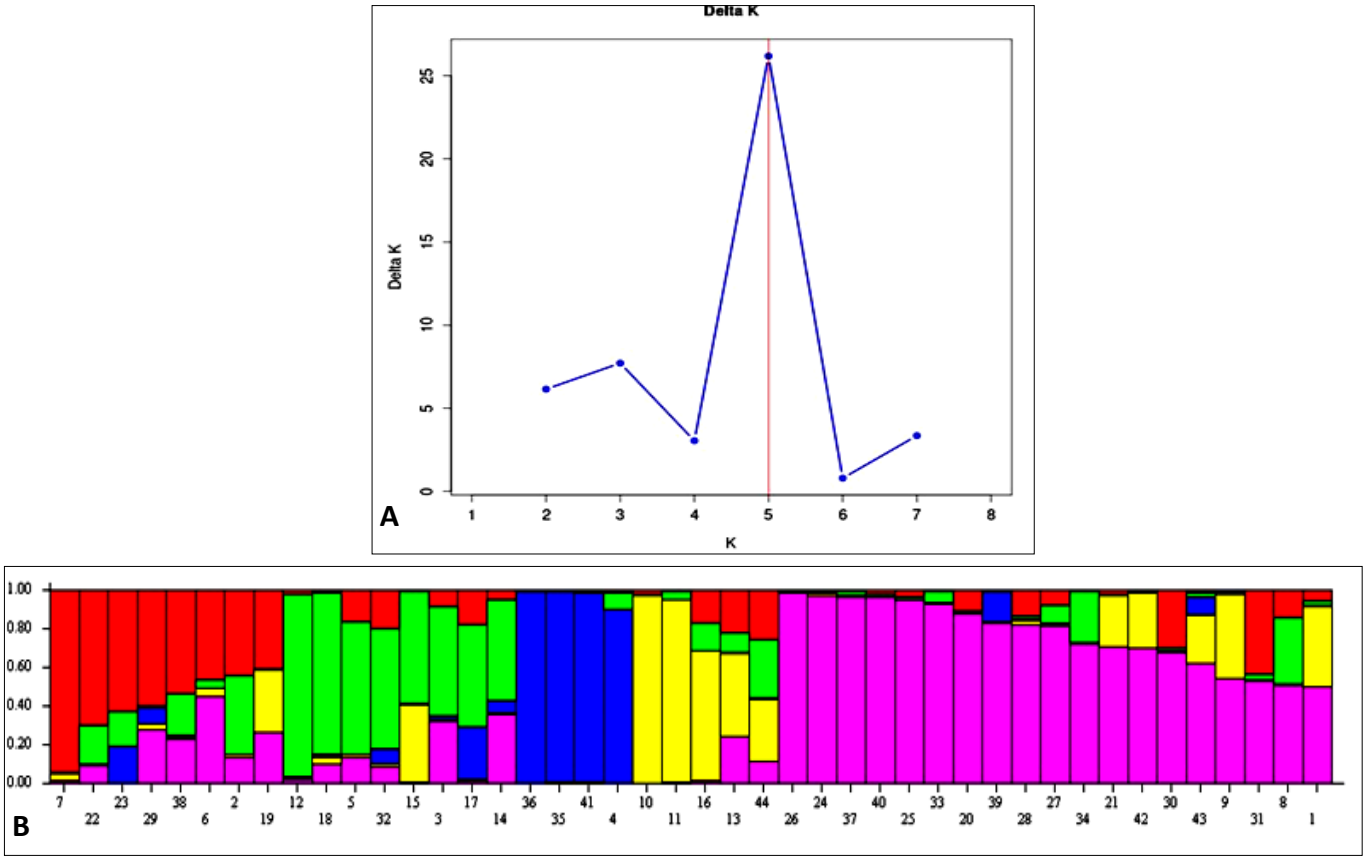


Fig. 2. (A) Magnitude of delta K from STRUCTURE analysis of 44 germplasm; (B) Population structure of 44 rice accessions. Population structure of all genotypes was divided based on genetic diversity detected by 208 SSR markers with K=5.

America (Colombia). SP5 comprised nineteen genotypes, among which seventeen were from South Asia (India), and two from South East Asia (Philippines). It clearly shows that each subpopulation is diverse from the others (25)

Genetic diversity analysis

The analysis of molecular variance (AMOVA) conducted on the 44 rice genotypes revealed highly significant genetic differences ($p < 0.001$) between sub-populations and within genotypes. However, no significant difference ($p > 0.001$) was observed between genotypes within sub-populations. Among the total genetic variations observed in the 44 rice genotypes, 9 % was attributed to genetic differentiation between the sub-populations, 53 % to genetic differentiation within the genotypes, and the remaining proportion (38 %) was due to genetic differences between genotypes within the sub-populations (Table 3). Additionally, the overall fixation index (Fst) and number of migrants (Nm) among the sub-populations were 0.176 and 2.529, respectively. This indicates the moderate level of gene flow and ensures the maintenance of genetic connectivity between sub-populations, while simultaneously permitting the development of genetic differentiation. Such results suggest a well-defined genetic architecture within the rice genotypes, where both intra-population processes, such as gene flow and mutation, and inter-population dynamics, like selection and genetic drift, contribute to the overall shaping of genetic diversity. This balance of forces underscores the complex evolutionary history of these genotypes and highlights the importance of both local adaptation and broader population-level interactions in maintaining their genetic variability.

The alleles (Na) ranged from 2.278 in SP3 to 3.014 in SP5, with an overall mean of 2.6366. The effective alleles (Ne) were highest in SP2 (2.131) and lowest in SP5 (1.982), with an average of 2.0398. Observed heterozygosity (Ho) ranged from 0.189 in SP4 to 0.218 in SP2, with an average of 0.2038. Expected heterozygosity (He) values ranged from 0.421 in SP5 to 0.476 in SP2, with an average of 0.4456. The lower Ho compared to He was expected due to rice's self-pollinating nature (26), resulting in a relatively higher degree of inbreeding within the population (27). The unbiased expected heterozygosity (uHe) ranged from 0.432 in SP5 to 0.507 in SP2, with a mean of 0.482. The fixation index (F), which measured inbreeding structure, varied from 0.553 in SP3 to 0.611 in SP5, with an average of 0.5802. The percentage of polymorphic loci (PPL) was

Table 3. Analysis of molecular variance among 44 genotypes based on the 208 SSR markers

Source	df	MSS	Estimated variance	Proportion of variation (%)	Nm	Fst
Between sub-populations	4	162.413	5.084	9 (<0.001)	2.529	0.176
Between genotypes within sub-population	39	81.191	29.772	53 (>0.001)		
Within genotypes	44	21.648	21.648	38 (<0.001)		
Total	87		56.503	100		

highest in SP5 (97.17 %) and lowest in SP4 (86.32 %), with an overall average of 92.264 %. The high percentage of polymorphic loci further emphasizes the genetic variability present, which is crucial for breeding programs aimed at enhancing resilience and adaptability (Table 4).

Table 4. Genetic diversity indices for the five rice sub-populations based on 208 SSR markers

Pop	Na	Ne	Ho	He	uHe	F	PPL (%)
SP1	2.731	2.099	0.209	0.469	0.500	0.596	96.23
SP2	2.811	2.131	0.218	0.476	0.507	0.564	94.81
SP3	2.278	2.000	0.202	0.436	0.498	0.553	86.79
SP4	2.349	1.987	0.189	0.426	0.473	0.577	86.32
SP5	3.014	1.982	0.201	0.421	0.432	0.611	97.17
Average	2.6366	2.0398	0.2038	0.4456	0.482	0.5802	92.264

Na - No. Alleles; Ne - No. Effective Alleles; Ho - Observed Heterozygosity; He - Expected Heterozygosity; uHe - Unbiased Expected Heterozygosity; F - Fixation Index; PPL – percentage of polymorphic loci

The population pairwise fixation indices, estimate genetic differentiation among populations due to genetic structure. An Fst value of 0.25 or higher suggests significant differentiation, values between 0.15 and 0.25 indicate moderate differentiation and values of 0.05 or less signify insignificant differentiation. This framework allows researchers to assess the extent of genetic structure and inform conservation and breeding strategies based on the level of genetic exchange between sub-populations. SP1 and SP2 had a low Fst value of 0.066, indicating minimal genetic differentiation between these populations. SP3 showed higher differentiation compared to the other populations, with Fst values of 0.148 (with SP1), 0.151 (with SP2), 0.183 (with SP4), and 0.165 (with SP5), suggesting significant genetic divergence from the rest. SP4 exhibited moderate differentiation with SP1 (0.080) and SP2 (0.087), while showing a higher Fst with SP3 (0.183), indicating greater divergence from SP3. SP5 had relatively low Fst values with SP1 (0.061), SP2 (0.072), and SP4 (0.094), suggesting these populations are genetically closer to SP5. However, SP5 showed higher differentiation from SP3 (0.165). These Fst values suggest that SP3 is the most genetically distinct population, while SP1, SP2, SP4 and SP5 are more closely related (Table 5).

Table 5. Population's pairwise genetic differentiation index (Fst)

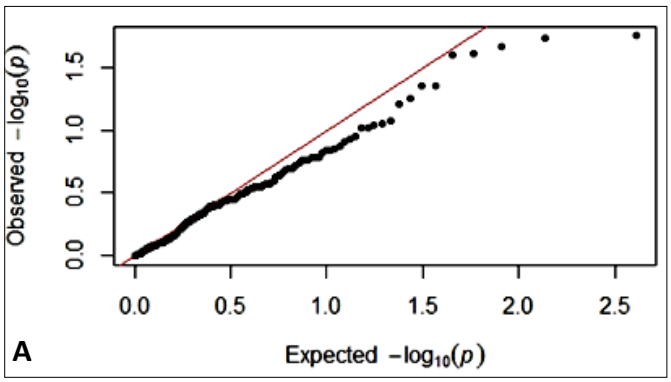
	SP1	SP2	SP3	SP4	SP5
SP1	0.000				
SP2	0.066	0.000			
SP3	0.148	0.151	0.000		
SP4	0.080	0.087	0.183	0.000	
SP5	0.061	0.072	0.165	0.094	0.000

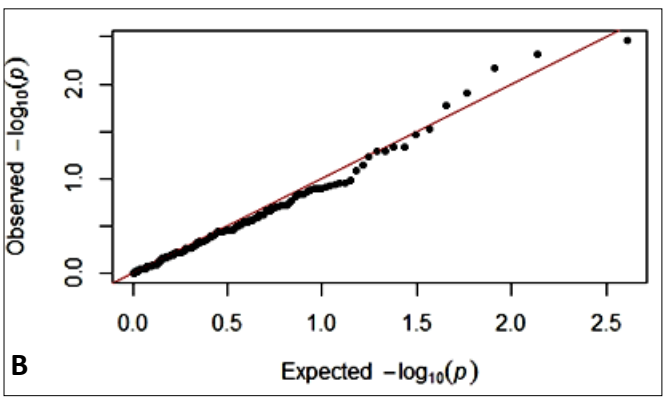
Marker trait association

The association between SSR markers and TPC was examined by two different models viz., GLM and MLM. The GLM with PCA model is expected to reduce the false positives that arise due to only population structure. The MLM with PCA and K model includes the kinship matrix in the model and is expected to reduce the false positives that arise from family relatedness. Both GLM and MLM are reported to control false positives better than ANOVA. The MLM model is reported to perform better than the GLM model alone by controlling false positives.

Based on the Q-Q plots, the MLM model showed deviations from the straight line, with upward inflation indicating false positives and downward deflation suggesting false negatives. In contrast, the GLM model exhibits Q-Q plots that closely follow the 1:1 line, with a slight upward deviation at the tail. This pattern suggests that both false positives and false negatives are well-controlled, indicating the presence of true associations and causal polymorphisms.

The association analysis using both the General Linear Model (GLM) and Mixed Linear Model (MLM) approaches identified several significant marker-trait associations across chromosomes 1, 2, 3, 4, 5, 8, and 12. In the GLM analysis, the R^2 value ranged from 9.40 % (RM230 on chromosome 8) to 28.52 % (RM8136 on chromosome 1). Similarly, the MLM analysis showed R^2 values ranging from 10.11 % (RM230 on chromosome 8) to 42.93 % (RM8136 on chromosome 1) (Fig. 3, Table 6).





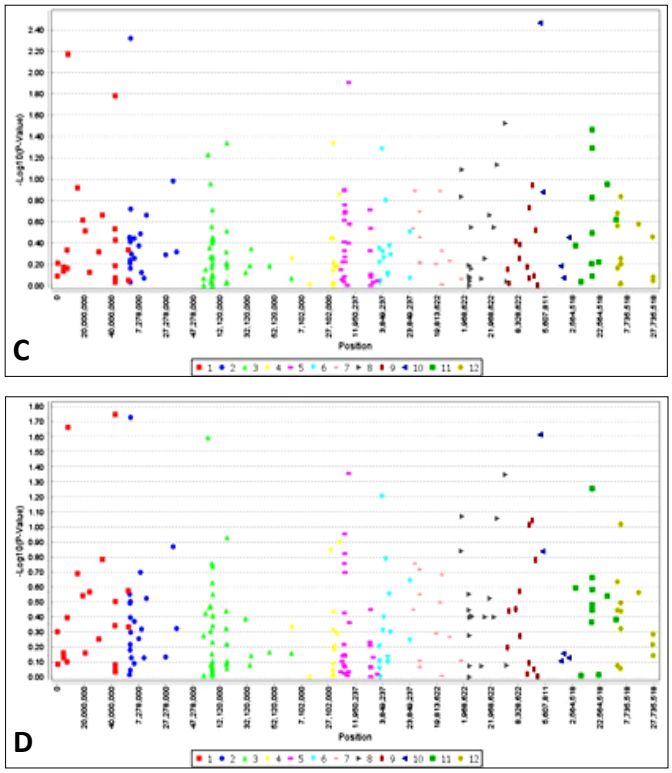


Fig. 3. (A-B) Quantile-quantile (QQ) plot of GLM and MLM; (C-D) The Manhattan plot of GLM and MLM for total phenolic content.

Table 6. Marker-trait associations and marker effect derived from 208 SSR markers and 44 rice accessions

Marker	Chromosome	Position (bp)	*p-value*	R^2
GLM				
RM243	1	7970836	0.00676	0.22569
RM8136	1	42926406	0.01659	0.28521
RM12381	**2**	**1505683**	**0.00479**	**0.23708**
RM1334	3	17421657	0.04576	0.15689
RM5511	4	30927988	0.04582	0.18805
RM6034	**5**	**7034442**	**0.01243**	**0.2049**
RM230	8	32538000	0.02995	0.09399
RM474	10	1819051	0.00344	0.24777
RM26762	11	17174930	0.03440	0.13253
MLM				
RM8136	1	42926406	0.01788	0.42927
RM243	1	7970836	0.02174	0.25675
RM12381	**2**	**1505683**	**0.01868**	**0.26689**
RM14432	3	3434709	0.02548	0.24625
RM6034	**5**	**7034442**	**0.04410**	**0.21038**
RM230	8	32538000	0.04494	0.10106
RM474	10	1819051	0.02434	0.24927

RM12381, and RM 6034 were found nearer to the candidate genes related to the synthesis of TPC.

Out of all these markers, two SSR markers viz., RM12381 (Chr. 2) and RM6034 (chr 5) were found to be associated with TPC. These markers were found nearer to the candidate genes related to the synthesis of TPC. When the sequential positions of these markers were hit in the genomic regions of the rice sequence available in the NCBI database, the marker RM12381 was found to be associated with the gene at the position Chr02:15717050 – 15718956 kb (Os02g0467000) which encodes the enzyme trans-cinnamate 4- monooxygenase. It was also observed, that, the marker RM6034 was found to be linked with the gene at the position Chr05:20953643 – 20955793 (Os02g0626100) encoding phenylalanine ammonia-lyse. Trans-cinnamate 4 -monooxygenase (C4H) and phenylalanine ammonia-lyase (PAL) are key enzymes in the phenylpropanoid pathway, which plays a crucial role in the biosynthesis of TPC. PAL catalyzes the first step in this pathway, converting phenylalanine into trans-cinnamic acid by deaminating phenylalanine. Trans-cinnamic acid is then hydroxylated by C4H to produce *p*-coumaric acid, a precursor for numerous phenolic compounds. In wheat, the expression levels of the phenylalanine ammonia-lyase (PAL6) and cinnamate 4-hydroxylase (C4H) genes were significantly elevated in plants cultured in a medium enriched with phenylalanine, indicating a regulatory response in the phenylpropanoid biosynthetic pathway (28) (Fig. 4).

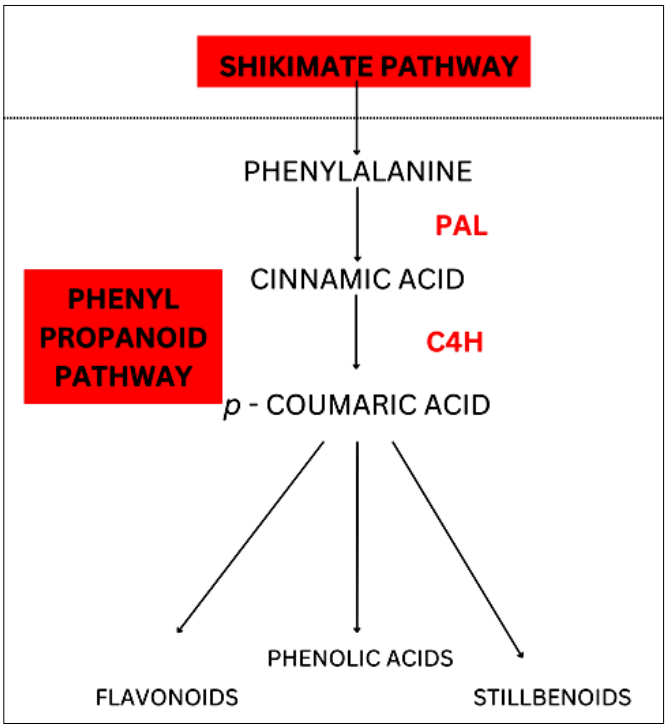


Fig. 4. General phenylpropanoid pathway. PAL - phenylalanine ammonia-lyase; C4H - cinnamate 4-hydroxylase.

The interplay between C4H and PAL is critical for plant adaptation to stress. By regulating the biosynthesis of phenolic compounds, these enzymes contribute to structural defences and antioxidant systems that mitigate oxidative damage caused by biotic and abiotic stress. Enhanced expression of these enzymes leads to increased accumulation of protective metabolites, improving overall plant health and survival under adverse conditions. Therefore, targeting these pathways could be a valuable strategy for developing stress-resistant crop varieties.

Earlier studies identified markers such as Rid12, RM346, RM484, RM162, and RM5371, which showed strong associations with phenolic content, flavonoid content, and antioxidant capacity (7,8). However, these markers were not directly linked to the genes involved in the phenylpro-

panoid pathway, which is responsible for phenolic content. In contrast, the present study identified two key markers, RM12381 (Chr 2) and RM6034 (Chr 5), associated with genes encoding trans-cinnamate 4-monooxygenase (Os02g0467000) and phenylalanine ammonia-lyase (Os02g0626100). These enzymes play vital roles in the phenylpropanoid pathway. By linking these markers to specific biosynthetic genes, this study makes a significant contribution to understanding phenolic biosynthesis and provides valuable tools for improving rice nutritional quality.

Conclusion

The enhancement of phenolic content in rice holds considerable promise for improving both food security and public health. The findings highlight the significance of pigmented rice varieties, particularly for their phytochemical profiles that offer potential benefits in managing oxidative stress and supporting metabolic health. The genotypes *Mappilai Samba, Murugankar, Kudaivazhai, Palakachaka, Mattaikar,* and *Mikuruvai* exhibited consistently high TPC with minimal G × E interactions, indicating stable phenotypic expression across environments. Their stability and high mean TPC values make them ideal donors for marker-assisted selection (MAS) programs targeting phenolic enrichment. As the two significant markers, RM12381 and RM6034 have been confirmed to be linked to the candidate genes responsible for the biosynthesis of TPC, these markers could be used for improving therapeutic traits in high-yielding rice varieties through Marker Assisted Breeding methods. Incorporating these genotypes and markers into breeding pipelines can accelerate the breeding programs focused on biofortified rice with functional health benefits. Continued exploration of these genetic markers can advance the development of nutritionally superior rice varieties, supporting both sustainable agriculture and health improvement goals in rice-dependent regions worldwide.

Acknowledgements

We are very grateful to DST-SERB program (SPG/2022/000058), New Delhi, India for funding this project.

Authors' contributions

KN and GS conceptualized the study, developed the methodology, conducted formal analysis, performed investigation and data curation and prepared the original draft; RS, RM, JM, MS, UD, AK, AL, AF supervised, validated, acquired funding and reviewed and edited the manuscript, with all authors approving the final version.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

Supplementary data

Supplementary Table 1 Details of genetic analysis of 208 SSR loci across the 44 rice accessions

References

1. Tiozon RJN, Sartagoda KJD, Fernie AR, Sreenivasulu N. The nutritional profile and human health benefit of pigmented rice and the impact of post-harvest processes and product development on the nutritional components: A review. Crit Rev Food Sci Nutr. 2023;63:3867–94. https://doi.org/10.1080/10408398.2021.1995697
2. Tieri M, Ghelfi F, Vitale M, Vetrani C, Marventano S, Lafranconi A, et al. Whole grain consumption and human health: an umbrella review of observational studies. Int J Food Sci Nutr. 2020;71:668–77. https://doi.org/10.1080/09637486.2020.1715354
3. World Food Programme. Global report on food crises (GRFC) 2024. https://www.wfp.org/publications/global-report-food-crises-grfc
4. World Health Organization. Hunger numbers stubbornly high for three consecutive years as global crises deepen: UN report [Internet]. https://www.who.int/news/item/24-07-2024-hunger-numbers-stubbornly-high-for-three-consecutive-years-as-global-crises-deepen--un-report
5. World Health Organization. The state of food security and nutrition in the world 2024 [Internet]. https://www.who.int/publications/m/item/the-state-of-food-security-and-nutrition-in-the-world-2024
6. Mbanjo EGN, Kretzschmar T, Jones H, Ereful N, Blanchard C, Boyd LA, Sreenivasulu N. The genetic basis and nutritional benefits of pigmented rice grain. Front Genet. 2020;11(229):1–18. https://doi.org/10.3389/fgene.2020.00229
7. Chattopadhyay K, Bagchi TB, Sanghamitra P, Sarkar S, Anilkumar C, Marndi BC, et al. Mapping genetic determinants for grain physicochemical and nutritional traits in brown and pigmented rice using genome-wide association analysis. Euphytica. 2023;219(57):1–17. https://doi.org/10.1007/s10681-023-03184-3
8. Kruk J, Aboul-Enein BH, Duchnik E, Marchlewicz M. Antioxidative properties of phenolic compounds and their effect on oxidative stress induced by severe physical exercise. J Physiol Sci. 2022;72(19)1–24. https://doi.org/10.1186/s12576-022-00845-1
9. Hu H, Zhao Y, Feng Y, Yang X, Li Y, Wu Y, et al. Consumption of whole grains and refined grains and associated risk of cardiovascular disease events and all-cause mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Am J Clin Nutr. 2023;117:149–59. https://doi.org/10.1016/j.ajcnut.2022.10.010
10. Shao Y, Jin L, Zhang G, Lu Y, Shen Y, Bao J. Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice. Theor Appl Genet. 2011;122:1005–16. https://doi.org/10.1007/s00122-010-1505-4
11. Li K, Li Q, Wang LY, Ren H, Ge Y. Genetic variation and association mapping of phenolic, flavonoid content and antioxidant capacity in USDA rice mini-core collection. Genet Resour Crop Evol. 2022;69:1685–94. https://doi.org/10.1007/s10722-021-01333-y
12. Singleton VL, Rossi JA. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Am J Enol Vitic. 1965;16:144–58. https://doi.org/10.5344/ajev.1965.16.3.144
13. Doyle JJ. Isolation of plant DNA from fresh tissue. Focus (Madison). 1990;12:13–15.
14. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59. https://doi.org/10.1093/genetics/155.2.945

15. Li Y, Liu J. Structure Selector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 2018;18:176–77. https://doi.org/10.1111/1755-0998.12719

16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904–09. https://doi.org/10.1038/ng1847

17. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38:203–08. https://doi.org/10.1038/ng1702

18. Rajagopalan VR, Manickam S, Muthurajan R. A comparative metabolomic analysis reveals the nutritional and therapeutic potential of grains of the traditional rice variety Mappillai Samba. Plants. 2022;11:543. https://doi.org/10.3390/plants11040543

19. de Mira NVM, Massaretto IL, Pascual C de SCI, Marquez UML. Comparative study of phenolic compounds in different Brazilian rice (*Oryza sativa* L.) genotypes. J Food Compost Anal. 2009;22:405–09. https://doi.org/10.1016/j.jfca.2008.06.012

20. Nath S, Bhattacharjee P, Bhattacharjee S, Datta J, Dolai AK. Grain characteristics, proximate composition, phytochemical capacity and mineral content of selected aromatic and non-aromatic rice accessions commonly cultivated in the North-East Indian plain belt. Appl Food Res. 2022;2(100067):1–8. https://doi.org/10.1016/j.afres.2022.100067

21. Sahu R, Mandal S, Das P, Ashraf GJ, Dua TK, Paul P, et al. The bioavailability, health advantages, extraction method and distribution of free and bound phenolics of rice, wheat and maize: A review. Food Chem Adv. 2023;3(100484):1–19. https://doi.org/10.1016/j.focha.2023.100484

22. Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, et al. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One. 2013;8:e84136. https://doi.org/10.1371/journal.pone.0084136

23. Adeboye KA, Oyedeji OE, Alqudah AM, Börner A, Oduwaye O, Adebambo O, Daniel IO. Genetic structure and diversity of upland rice germplasm using diversity array technology (DArT)-based single nucleotide polymorphism (SNP) markers. Plant Genet Resour. 2020;18:343–50. https://doi.org/10.1017/S1479262120000362

24. Thant AA, Zaw H, Kalousova M, Singh RK, Lojka B. Genetic diversity and population structure of Myanmar rice (*Oryza sativa* L.) varieties using DArTseq-based SNP and SilicoDArT markers. Plants. 2021;10:2564. https://doi.org/10.3390/plants10122564

25. Kimwemwe PK, Bukomarhe CB, Mamati EG, Githiri SM, Civava RM, Mignouna J, et al. Population structure and genetic diversity of rice (*Oryza sativa* L.) germplasm from the Democratic Republic of Congo (DRC) using DArTseq-Derived single nucleotide polymorphism (SNP). Agron. 2023;13:1906. https://doi.org/10.3390/agronomy13071906

26. Sleper DA, Poehlman JM. Breeding field crops: Gene Recombination in Plant Breeding. 5th ed. Wiley-Blackwell; 2006.

27. Ndjiondjop MN, Semagn K, Zhang J, Gouda AC, Kpeki SB, Goungoulou A, et al. Development of species diagnostic SNP markers for quality control genotyping in four rice (*Oryza sativa* L.) species. Mol Breed. 2018;38(131):1–13. https://doi.org/10.1007/s11032-018-0885-z

28. Feduraev P, Skrypnik L, Riabova A, Pungin A, Tokupova E, Maslennikov P, Chupakhina G. Phenylalanine and tyrosine as exogenous precursors of wheat (*Triticum aestivum* L.) secondary metabolism through PAL-associated pathways. Plants. 2020;9(476):1–19. https://doi.org/10.3390/plants9040476