Indian J. Genet. Plant Breed., (2023); 83(4): 457-468
https://doi.org/10.31742/ISGPB.83.4.1

ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Unfolding genomic regions and identification of novel marker-trait associations for therapeutic traits flavonoid and anthocyanin content in rice (*Oryza sativa* L.)

L. Ananda Lekshmi, R. Muthuvijayaragavan, S. Geetha*, M. Hemalatha, V. Allan, M. Dhasarathan, N. Devasena, G. Ariharasutharsan, N. Vishnu Varthini, D. G. Manu, P. Jeyaprakash, S. Aiyswaraya and R. Saraswathi

Abstract

The genetic diversity, structure analysis, and marker-trait association for total flavonoid content (TFC) and anthocyanin content (AC) were studied in rice using 79 germplasm lines with 210 polymorphic microsatellite markers. Among the landraces studied, RG83 (RPHP 93) had the highest level of TFC (227.78 QE/100g) and RG126 (Kallimadayan) contained the highest level of AC (0.416 CE/100g). The STRUCTURE analysis grouped the studied landraces into three genetic structure groups which nearly matched the geographical distance. The results of the marker-trait associations study using MLM with PCA and kinship indicated that the SSR markers RM287 and RM19358 were closely associated with AC while another two markers RM452 and RM 2197 were closely linked with TFC and these two markers were found to be in the genomic regions related to the candidate genes for TFC and AC biosynthesis. In addition, the markers RM10961, RM5963, and RM559 were also found to have significant marker-trait associations with both AC and TFC.

Keywords: Rice, Anthocyanin content, total flavonoids, marker-trait association

Introduction

The bioactive compounds such as total flavonoid content (TFC) and anthocyanin content (AC) in rice grains reduce the risk of developing many chronic diseases by their antioxidant capacity (Shen et al. 2009). Improving the therapeutic traits along with yield and cooking quality of rice varieties would help to address most of the present-day health conflicts as rice is the most produced and consumed cereal. However, breeding efforts to improve the therapeutic value of rice are very much limited compared to other traits. Molecular Marker Assisted Breeding is a better choice for realizing rapid trait improvement and, therefore, the present study was taken up to find the SSR markers closely associated with TFC and AC through using MLM (Mixed Linear Model) with PCA (Principal Component Analysis) and kinship analysis.

Materials and methods

A panel comprised of 79 rice accessions of landraces and exotic collections from different regions conserved at the Ramaiah Gene Bank, Department of Plant Genetic Resources, Tamil Nadu Agricultural University (TNAU) was utilized for the present study (Table 1). The panel entries were grown in *kharif* 2020 at Paddy Breeding Station, TNAU, Coimbatore, by following recommended agronomic practices and the harvested seeds were used for the analysis of TFC and

---

Department of Genetics and Plant Breeding, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India.

[1]Centre for Plant Biotechnology and Molecular Biology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India.

***Corresponding Author:** S. Geetha, Department of Genetics and Plant Breeding, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India, E-Mail:geethagovind1@gmail.com

How to cite this article: Lekshmi L.A., Muthuvijayaragavan R., Geetha S., Hemalatha M., Allan V., Dhasarathan M., Devasena N., Ariharasutharsan G., Varthini N.V., Manu D.G., Jeyaprakash P., Aiyswaraya S. and Saraswathi R. 2023. Unfolding genomic regions and identification of novel marker-trait associations for therapeutic traits flavonoid and anthocyanin content in rice (*Oryza sativa* L.). Indian J. Genet. Plant Breed., **83**(4): 457-468.

Source of support: Nil

Conflict of interest: None.

Received: May 2023 **Revised:** Sept. 2023 **Accepted:** Oct. 2023

Table 1. Details of rice accessions and mean values of total flavonoid and anthocyanin contents

Genotype code	Genotypes	Origin/source	Parentage	Total flavonoid (QE/100g)	Anthocyanin (CE/ 100g)	Genotype code	Genotypes	Origin/source	Parentage	Total flavonoid (QE/100g)	Anthocyanin (CE/ 100g)
RG1	Mapillai Samba	Tamil Nadu	Landrace	100.79	0.368	RG107	RH2-SM-1-2-1	Tamil Nadu	Swarna/ Moroberekan	164.29	0.292
RG2	CK 275	Tamil Nadu	CO(R) 50 x Kavuni	69.05	0.292	RG109	Vadival	Tamil Nadu	Landrace	53.17	0.198
RG3	Senkar	Tamil Nadu	Landrace	100.79	0.232	RG110	Norungan	Tamil Nadu	Landrace	164.29	0.262
RG4	Murugankar	Tamil Nadu	Landrace	132.54	0.268	RG115	IG 43 (EC 728788-117759)	IRRI, Philippines	-	148.41	0.246
RG6	CHIR 5	West Bengal	Improved chinsurah	132.54	0.269	RG117	IG 65 (EC 72904-120958)	IRRI, Philippines	GODAHEENATI/ IRGC31393-1	148.41	0.23
RG7	Kudaivazhai	Tamil Nadu	Landrace	164.29	0.292	RG118	Ponmani Samba	Tamil Nadu	Landrace	53.17	0.337
RG9	Kuruvaikalan-jiyam	Tamil Nadu	Landrace	84.92	0.251	RG126	Kallimadayan	Tamil Nadu	Landrace	132.54	0.416
RG12	Vellaichithiraikar	Tamil Nadu	Landrace	53.17	0.288	RG132	IG 33 (EC 728938-117935)	Jamaica	WC3397	69.05	0.287
RG15	Palakachaka	Tamil Nadu	Landrace	132.54	0.261	RG160	IG 72(EC 728650-117587)	IRRI, Philippines	TD 25/IRGC 32351-1	53.17	0.225
RG17	Sivapuchithiraikar	Tamil Nadu	Landrace	69.05	0.212	RG161	Panamarasamba	Tamil Nadu	Landrace	53.17	0.305
RG18	CHIR 11	West Bengal		132.54	0.272	RG162	IR 64	IRRI, Philippines	IR 5857-33-2-1/ IR 2061-465-1-5-5	180.16	0.307
RG20	Kalvalai	Tamil Nadu	Landrace	69.05	0.308	RG163	Mikkuruvai	Tamil Nadu	Landrace	69.05	0.274
RG22	IR36	IRRI, Philippines	IR1561-228 -l2/ IR1737//CR94-13	100.79	0.184	RG176	Kodai	Tamil Nadu	Landrace	100.79	0.295
RG25	Sornakuruvai	Tamil Nadu	Landrace	100.79	0.318	RG182	ARB 59	Karnataka	Variety	53.17	0.226
RG32	Thogai samba	Tamil Nadu	Landrace	53.17	0.305	RG184	IG 28(EC 728892-117880)	Indonesia	Seratoeshari	100.79	0.187
RG34	RPHP 125	Uttar Pradesh		69.05	0.212	RG188	RPHP 8	Andhra Pradesh	24(K)	148.41	0.253
RG39	KattuPonni	Tamil Nadu	Landrace	180.16	0.313	RG191	IG 15(EC 728910-117901)	China	Sze guenzim	132.54	0.282
RG41	Godavari samba	Tamil Nadu	Landrace	100.79	0.259	RG192	Nootripathu	Tamil Nadu	Landrace	53.17	0.273
RG42	Earapalli samba	Tamil Nadu	Landrace	53.17	0.262	RG26	Rascadam	Tamil Nadu	Landrace	84.92	0.222
RG44	Mangan samba	Tamil Nadu	Landrace	84.92	0.272	RG33	Malayalathan samba	Tamil Nadu	Landrace	53.17	0.266
RG45	RPHP 105	Manipur	Moirangphou (Land races)	148.41	0.256	RG63	IG71 (EC728651-117588)	IRRI, Philippines	TEPI BORO/ IRGC27519-1	69.05	0.207
RG48	Kalarkar	Tamil Nadu	Landrace	53.17	0.248	RG69	RPHP 48	Uttarakhand	Bindii	69.05	0.256

Contd.....

RG50	Sornavari	Tamil Nadu	Landrace	69.05	0.332	RG83	RPHP 93	Uttarkhand	Type-3 Dehraduni basmati	227.78	0.212
RG51	RPHP 134	Kerala	Njavara (Landrace)	53.17	0.215	RG105	IG 6 (EC 729592-121642)	IRRI, Philippines	SOM CAU 70::IRGC8227-1	116.67	0.292
RG53	IR 68144-2B-2-2-3-1-127	IRRI, Philippines	IR72 x Zawa bonday	116.67	0.246	RG116	RPHP 27	IRRI, Philippines	Azucena	180.16	0.268
RG55	IG 67 (EC729050-120988)	IRRI, Philippines	IR77384-12-35-3-12-1B/ IRGC117299-1	116.67	0.307	RG127	IG 10 (EC 729686-121743)	IRRI, Philippines	Hasan seralir/ GC79564-C1	69.05	0.224
RG57	RPHP 103	Uttarkhand	Pant Sugandh Dhan 17 (Pusa Basmathi 1 from Pusa 167/Karnal local)	100.79	0.225	RG134	IG 9 (EC 729601-121651)	IRRI, Philippines	Gemjyajyanam/ IRGC32411-C1	211.90	0.235
RG58	Kodaikulathan	Tamil Nadu	Landrace	69.05	0.249	RG190	IG 26 (IC0590943-121899)	IRRI, Philippines	Basmati 370/ IRGC 3750-1	148.41	0.33
RG59	RPHP 68	Odhisa	Subadhra	84.92	0.247	RG68	IG63 (EC 728711-117674)	IRRI, Philippines	Caawa/Fortuna	100.79	0.307
RG60	Rama kuruvaikar	Tamil Nadu	Landrace	116.67	0.217	RG112	IG 35 (EC 728858-117843)	Cote D'ivoire	Pate blane MN1	132.54	0.273
RG66	Seevanasamba	Tamil Nadu	Landrace	53.17	0.274	RG113	IG 45 -(EC 728768117736)	Puerto Rico	Fortuna	164.29	0.262
RG67	RPHP 106	Manipur	Akutphou (Lang phou/IR 1364-37-3-1)	180.16	0.33	RG123	IG 2 (EC 729808-121874)	IRRI, Philippines	BLUEBONNET 50/IRGC1181	100.79	0.274
RG72	Aarkadukichili	Tamil Nadu	Landrace	84.92	0.345	RG124	IG 29 (EC 728925-117920)	IRRI, Philippines	TO/782-20-1	164.29	0.239
RG74	ARB 65	Karnataka	Variety	69.05	0.162	RG130	IG 39 (EC 728779-117750	Honduras	Honduras	84.92	0.313
RG76	Mattakuruvai	Tamil Nadu	Landrace	84.92	0.246	RG131	RPHP 90	Andhra Pradesh	24K	116.67	0.256
RG77	Karuthakar	Tamil Nadu	Landrace	84.92	0.292	RG141	IG44 (EC 728762-117729)	U. S.A.	Edith	164.29	0.218
RG92	IG 49(EC 729102-121052)	IRRI, Philippines	Menakely/ IRGC51021-1	100.79	0.179	RG181	IG 52(EC 728756-117723)	Brazil	Dourado agulha	69.05	0.253
RG99	IG 31 (EC728844-117829)	Colombia	Oryzica llanos 5	180.16	0.275	RG189	IG 41 (EC 728800-117776)	Indonesia	Kaniranga	69.05	0.225
RG102	Varakkal	Tamil Nadu	Landrace	69.05	0.225				Mean	105.01	0.26
RG103	Mattaikar	Tamil Nadu	Landrace	148.41	0.252				Range	53.17 -227.78	0.16 - 0.42
RG106	Katta samba	Tamil Nadu	Landrace	53.17	0.316						

Table 2. Details of genetic analysis of 210 SSR loci across the 79 rice accessions

Marker	Chr. No	Major Allele Frequency	Allele No	Gene Diversity	Hetero-zygosity	PIC	Marker	Chr. No	Major Allele Frequency	Allele No	Gene Diversity	Hetero-zygosity	PIC
RM6784	1	0.32	5	0.77	0.03	0.73	RM3777	5	0.72	5	0.45	0	0.42
RM1	1	0.35	4	0.72	0.1	0.67	RM5874	5	0.55	3	0.53	0.01	0.42
RM1067	1	0.38	5	0.71	0.1	0.66	RM5973	5	0.76	4	0.4	0	0.37
RM8136	1	0.52	6	0.67	0.08	0.63	RM5994	5	0.73	3	0.42	0	0.37
RM3694	1	0.54	5	0.65	0.18	0.61	RM5575	5	0.78	5	0.38	0.01	0.35
RM243	1	0.41	4	0.66	0.77	0.6	RM3322	5	0.77	3	0.38	0	0.35
RM5	1	0.44	4	0.65	0.61	0.58	RM17898	5	0.77	3	0.38	0	0.35
RM6321	1	0.57	4	0.57	0.03	0.49	RM6034	5	0.71	3	0.42	0	0.34
RM1247	1	0.47	3	0.56	0.53	0.46	RM17956	5	0.72	4	0.46	0.46	0.43
RM237	1	0.49	3	0.56	0.87	0.46	RM574	5	0.8	3	0.33	0	0.29
RM24	1	0.53	4	0.55	0.63	0.46	RM1024	5	0.83	4	0.3	0.04	0.28
RM1282	1	0.54	3	0.54	0.03	0.44	RM161	5	0.84	4	0.29	0.04	0.27
RM283	1	0.71	4	0.47	0.03	0.43	RM7434	6	0.42	6	0.72	0.23	0.67
RM259	1	0.68	4	0.48	0.01	0.43	RM19708	6	0.51	4	0.65	0.65	0.6
RM12269	1	0.61	3	0.5	0.03	0.41	RM19469	6	0.47	3	0.61	0.04	0.52
RM128	1	0.72	3	0.44	0	0.4	RM5963	6	0.54	4	0.59	0.09	0.51
RM10916	1	0.73	4	0.43	0.43	0.39	RM19358	6	0.54	4	0.58	0.87	0.5
RM3362	1	0.77	3	0.37	0	0.32	RM19642	6	0.61	3	0.55	0.03	0.49
RM12261	1	0.82	4	0.31	0.03	0.3	RM19320	6	0.66	4	0.51	0.01	0.47
RM312	1	0.84	3	0.28	0.04	0.26	RM19329	6	0.7	4	0.48	0.03	0.45
RM238	1	0.94	3	0.12	0.05	0.12	RM314	6	0.54	3	0.54	0.75	0.44
RM324	2	0.39	6	0.74	0.08	0.7	RM19319	6	0.51	3	0.53	0.81	0.43
RM12793	2	0.58	4	0.59	0.01	0.53	RM162	6	0.73	4	0.44	0.01	0.41
RM71	2	0.47	3	0.6	0.39	0.52	RM584	6	0.57	3	0.51	0.71	0.4
RM555	2	0.55	4	0.58	0.04	0.51	RM19346	6	0.72	3	0.42	0.13	0.37
RM12350	2	0.53	3	0.58	0.03	0.5	RM19520	6	0.79	3	0.35	0.01	0.32
RM341	2	0.47	3	0.58	0.59	0.5	RM19620	6	0.85	4	0.27	0.03	0.26
RM12369	2	0.57	4	0.57	0.03	0.49	RM11	7	0.34	5	0.75	0.67	0.7
RM211	2	0.66	4	0.52	0.06	0.47	RM21867	7	0.44	3	0.65	1	0.58
RM110	2	0.68	4	0.48	0.01	0.43	RM125	7	0.5	5	0.64	0.92	0.58
RM12363	2	0.55	3	0.53	0.01	0.42	RM21041	7	0.58	5	0.61	0.03	0.57
RM233	2	0.57	3	0.52	0.76	0.42	RM118	7	0.47	5	0.61	0.76	0.54
RM236	2	0.51	2	0.5	0.99	0.37	RM248	7	0.65	4	0.52	0.39	0.47
RM48	2	0.51	2	0.5	0.99	0.37	RM21866	7	0.68	5	0.5	0.03	0.47

Contd.....

RM452	2	0.3	6	0.77	0.18	0.73	RM21264	7	0.66	4	0.51	0.01	0.47
RM6367	2	0.71	3	0.43	0.03	0.37	RM21262	7	0.65	4	0.51	0.01	0.46
RM300	2	0.78	3	0.37	0.01	0.34	RM455	7	0.77	3	0.38	0.03	0.35
RM12374	2	0.77	3	0.37	0	0.32	RM21263	7	0.81	4	0.33	0.03	0.31
RM12386	2	0.8	3	0.33	0.01	0.29	RM2584	8	0.29	5	0.79	0	0.75
RM5513	3	0.41	4	0.69	0.03	0.63	RM25	8	0.34	5	0.77	0.29	0.73
RM3766	3	0.38	4	0.69	0.1	0.63	RM6838	8	0.34	4	0.73	0	0.68
RM14607	3	0.48	5	0.67	0	0.63	RM447	8	0.41	5	0.7	0.13	0.65
RM5477	3	0.48	5	0.67	0	0.63	RM22597	8	0.47	5	0.69	0.1	0.65
RM14241	3	0.53	5	0.65	0	0.61	RM152	8	0.39	4	0.69	0.05	0.63
RM442	3	0.51	4	0.65	0.04	0.6	RM3507	8	0.53	5	0.65	0	0.61
RM14623	3	0.56	5	0.63	0.03	0.59	RM44	8	0.45	5	0.66	0.73	0.61
RM14639	3	0.37	3	0.66	0.87	0.59	RM22590	8	0.47	5	0.66	0.56	0.6
RM14331	3	0.38	3	0.66	0.03	0.59	RM408	8	0.61	4	0.57	0.03	0.53
RM55	3	0.52	4	0.63	0.05	0.58	RM22600	8	0.65	4	0.52	0.32	0.47
RM1334	3	0.51	4	0.64	0	0.57	RM7057	8	0.69	4	0.49	0.01	0.46
RM6881	3	0.43	3	0.65	0	0.57	RM3181	8	0.65	3	0.51	0	0.45
OSR13	3	0.53	5	0.61	0.03	0.55	RM256	8	0.65	3	0.51	0	0.45
RM2326	3	0.59	4	0.58	0	0.54	RM433	8	0.5	3	0.54	0.9	0.43
RM14432	3	0.47	3	0.62	0.01	0.54	RM284	8	0.78	3	0.37	0.01	0.34
RM14549	3	0.59	4	0.58	0.27	0.54	RM230	8	0.78	2	0.34	0.44	0.29
RM231	3	0.49	4	0.6	0.11	0.52	RM316	9	0.37	8	0.77	0.72	0.73
RM14622	3	0.56	3	0.56	0.03	0.48	RM205	9	0.33	5	0.76	0	0.72
RM15214	3	0.67	4	0.51	0	0.46	RM215	9	0.46	4	0.65	0.01	0.58
RM14638	3	0.65	4	0.51	0.01	0.46	RM107	9	0.61	4	0.57	0	0.52
RM15221	3	0.7	4	0.48	0.03	0.45	RM3025	9	0.5	3	0.59	0.47	0.5
RM14612	3	0.7	4	0.48	0.03	0.45	RM105	9	0.67	4	0.49	0.48	0.44
RM5480	3	0.54	3	0.54	0.35	0.44	RM24325	9	0.5	3	0.54	1	0.43
RM6283	3	0.54	3	0.54	0.03	0.44	RM464	9	0.49	3	0.52	0.94	0.41
RM14626	3	0.65	4	0.5	0.03	0.44	RM23711	9	0.72	3	0.44	0	0.39
RM15226	3	0.71	4	0.46	0.03	0.43	RM7424	9	0.67	3	0.45	0.63	0.36
RM6291	3	0.51	3	0.53	0	0.43	RM23788	9	0.7	3	0.43	0.34	0.34
RM7403	3	0.72	3	0.44	0	0.39	RM242	9	0.81	2	0.31	0.38	0.26
RM3441	3	0.73	3	0.42	0	0.37	RM222	10	0.33	4	0.73	0.01	0.68
RM3872	3	0.76	4	0.4	0	0.37	RM271	10	0.39	4	0.71	0.19	0.66
RM416	3	0.61	2	0.48	0	0.36	RM474	10	0.48	4	0.67	0	0.62
RM514	3	0.81	3	0.32	0	0.3	RM590	10	0.67	5	0.5	0.05	0.46

Contd.....

RM14548	3	0.8	3	0.33	0	0.29	RM171	10	0.68	3	0.47	0.03	0.41
RM16131	3	0.84	4	0.29	0.03	0.28	RM536	11	0.35	6	0.78	0.1	0.75
RM14254	3	0.63	4	0.55	0.01	0.5	RM144	11	0.3	6	0.77	0.18	0.73
RM3400	3	0.93	4	0.13	0.06	0.13	RM287	11	0.37	5	0.73	0.27	0.69
RM3029	3	0.93	3	0.13	0.06	0.13	RM26767	11	0.37	3	0.66	0.94	0.59
RM16643	4	0.34	6	0.77	0.06	0.74	RM26762	11	0.42	3	0.65	1	0.58
RM5709	4	0.39	4	0.71	0.04	0.66	RM26758	11	0.42	3	0.64	1	0.56
RM17377	4	0.46	5	0.69	0.04	0.64	RM26761	11	0.44	3	0.64	1	0.56
RM17470	4	0.63	5	0.57	0.04	0.54	RM224	11	0.63	6	0.57	0.09	0.54
RM3276	4	0.61	4	0.57	0.01	0.53	RM26756	11	0.69	4	0.49	0.01	0.46
RM5511	4	0.63	4	0.55	0.01	0.5	RM552	11	0.62	3	0.5	0.63	0.42
RM3687	4	0.65	4	0.54	0	0.5	RM206	11	0.8	5	0.35	0	0.33
RM3423	4	0.68	5	0.5	0.03	0.47	RM20A	12	0.42	4	0.68	0	0.62
RM7585	4	0.7	5	0.48	0.03	0.44	RM27589	12	0.35	3	0.67	0.87	0.59
RM255	4	0.74	4	0.43	0.03	0.4	RM27591	12	0.37	3	0.66	0.96	0.59
RM6992	4	0.72	3	0.44	0.01	0.39	RM5795	12	0.42	3	0.65	0.03	0.57
RM559	4	0.73	3	0.41	0.01	0.34	RM2197	12	0.44	3	0.64	0	0.57
RM17616	4	0.86	3	0.25	0	0.23	RM1227	12	0.66	3	0.51	0	0.46
RM5700	5	0.22	6	0.82	0.46	0.79	RM17	12	0.53	3	0.52	0.63	0.41
RM495	5	0.39	4	0.69	0.03	0.63	RM247	12	0.53	4	0.51	0.95	0.39
RM4915	5	0.42	4	0.69	0.01	0.63	RM27593	12	0.72	3	0.44	0	0.39
RM17962	5	0.48	5	0.67	0	0.63	RM277	12	0.67	3	0.46	0.08	0.37
RM437	5	0.4	4	0.68	0.39	0.62	RM20	12	0.73	3	0.4	0.52	0.33
RM413	5	0.53	5	0.64	0.22	0.6	Hv0740	-	0.35	5	0.75	0.19	0.71
RM7293	5	0.42	4	0.66	0.11	0.6	Hv0247	-	0.37	3	0.66	0.14	0.59
RM6841	5	0.56	5	0.63	0.03	0.59	RM14028	-	0.39	3	0.65	0.96	0.58
RM3870	5	0.58	5	0.62	0	0.58	RM14201	-	0.43	3	0.65	0	0.57
RM17959	5	0.58	5	0.61	0.03	0.57	RM24028	-	0.46	3	0.63	0.97	0.55
RM7118	5	0.59	4	0.58	0	0.54	RM20952	-	0.65	4	0.53	0.01	0.5
RM178	5	0.65	4	0.53	0.05	0.5	RM2770	-	0.65	4	0.51	0.01	0.46
RM17960	5	0.64	5	0.54	0.03	0.5	RM12381	-	0.68	3	0.48	0.04	0.43
RM18858	5	0.65	4	0.53	0.01	0.5	RM12334	-	0.72	3	0.44	0.01	0.39
RM3853	5	0.69	4	0.49	0.01	0.45	RM14606	-	0.76	3	0.39	0	0.36
RM1187	5	0.65	4	0.5	0.03	0.44	RM27592	-	0.65	2	0.46	0.05	0.35
							Mean		0.58	3.82	0.54	0.2	0.48

Chr. = Chromosome, PIC = Principal information content

AC. The TFC and TAC were estimated as per the methods suggested by Woisky and Salatino (1998) and Fuleki and Francis (1968), respectively. The DNA from the leaf tissues was isolated by CTAB method (Doyle and Doyle 1987). The details of the genetic analysis of 210 polymorphic SSR loci across the 79 rice accessions are given in Table 2.

The basic statistics of the marker allele was executed using GenAlEx 6.5 (https://biology-assets.anu.edu.au/GenAlEx/Welcome.html). The SSR scores were used to create a data matrix to analyze genetic relationships using the DARwin program version 2.11a (http://darwin.cirad.fr/.). The dendrogram was constructed based on a neighbor-joining algorithm using the marker data for all the rice genotypes following unweighted pair group method analysis (UPGMA) (Gawel and Jarret 1991). The population structure was executed using STRUCTURE HARVESTER (2.3.4) (Earl and VonHoldt 2012) to identify the subpopulation. The number of subpopulations is represented by K, was identified by a model-based approach. The analysis was executed using CLUMPAK for K 1-10 and five independent runs were assessed for each K value. The best K value was determined by ΔK given by Evanno et al. (2005) based on second-order change of likelihood function with respect to K found by Structure. The graph plotted for K *vs* ΔK showed best K. We used a burn-in period of 100,000 steps followed by 100,000 Monte Carlo Markov Chain replicates, as suggested by Pritchard et al. (2003).

The marker-trait association was done using the software program TASSEL 4.0.1. (http://www.maizegenetics.net). GLMs with Q, GLM with PCA (Principal Component Analysis) (Price et al. 2006), MLM with Q + K, and MLM with PCA+ K matrix (Kinship matrix for family relatedness estimates) (Yu et al. 2006) were used to test marker-trait associations (MTAs). The physical distance of the genomic region adjacent (1 Mb up and down) to the significant markers from the MTA was determined by integrating the genetic distance of the markers and the physical map of the Nipponbare reference genome (Assembly: IRGSP-1.0 (GCF_001433935.1)) in the *Oryza* genome browser of NCBI (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_001433935). The annotation information was also derived.

Results and discussion

Variations in therapeutic traits

The traits, TFC ranged from 53.17 to 227.78 mg QE /100 g with an average of 105.01 mg QE/100 g and AC ranged from 0.162 – 0.416 mg CE/100g with an average of 0.265 mg CE/100g (Table 1) and revealed the presence of wide range of genetic variation for the studied traits (Fig. 1.). In general, the pigmented genotypes had higher level of both TFC and AC. A similar estimation of genetic diversity using molecular markers and their genome-wide association was studied in *Vaccinium* spp. for anthocyanin content (Debnath et al. 2023).

Population structure analysis

The maximum Delta K value was observed for K = 2 (Fig. 2a). However, with structure analysis at the peak value of K at K = 2 the panel genotypes were grouped into two subclusters. Though these two sub-groups were differentiated based on geographic origin they did not get clearly differentiate for high and low TFC and AC. Therefore, the next peak K at K=3 was considered for analysis as suggested by Pradhan et al. (2019), which divided the panel into three sub-populations, OS1(high), OS2 (low) and OS3 (medium) based on the level of TFC and AC (Fig. 2b and 2c). Ohet al. (2018) observed a similar trend for anthocyanin content in rice at K = 2 and observed a similar pattern in the anthocyanin content of rice when considering a grouping at K=2 also reported that it was only at K = 4 that high anthocyanin genotypes were clearly distinguished from those with low anthocyanin levels. These results imply that, despite some variations in genetic traits, there is no significant distinction in the genetic

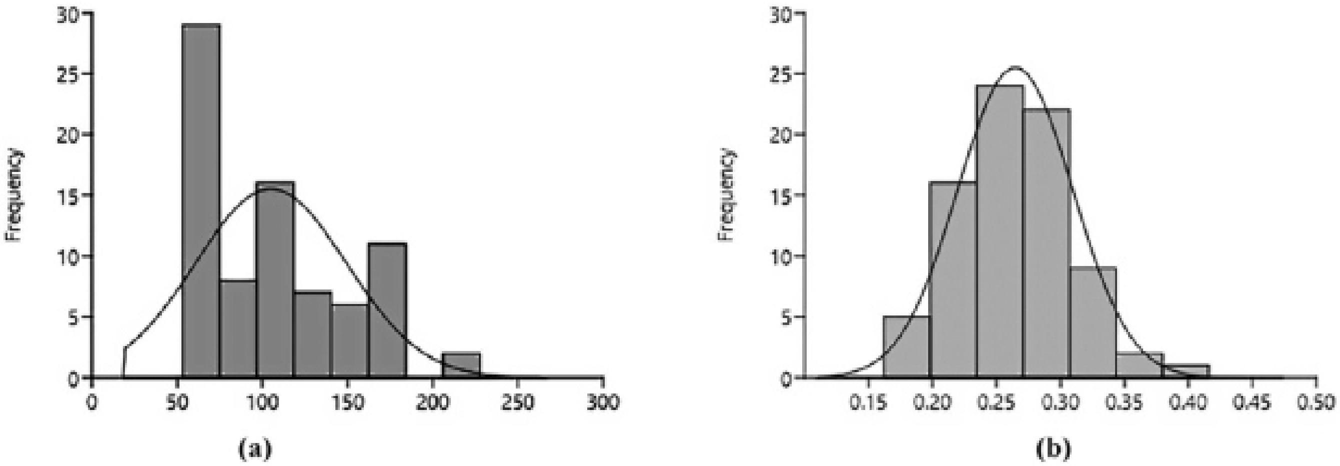

Fig. 1. (a) Frequency distribution of total flavonoids in rice germplasm, (b) Frequency distribution of anthocyanin in rice germplasm

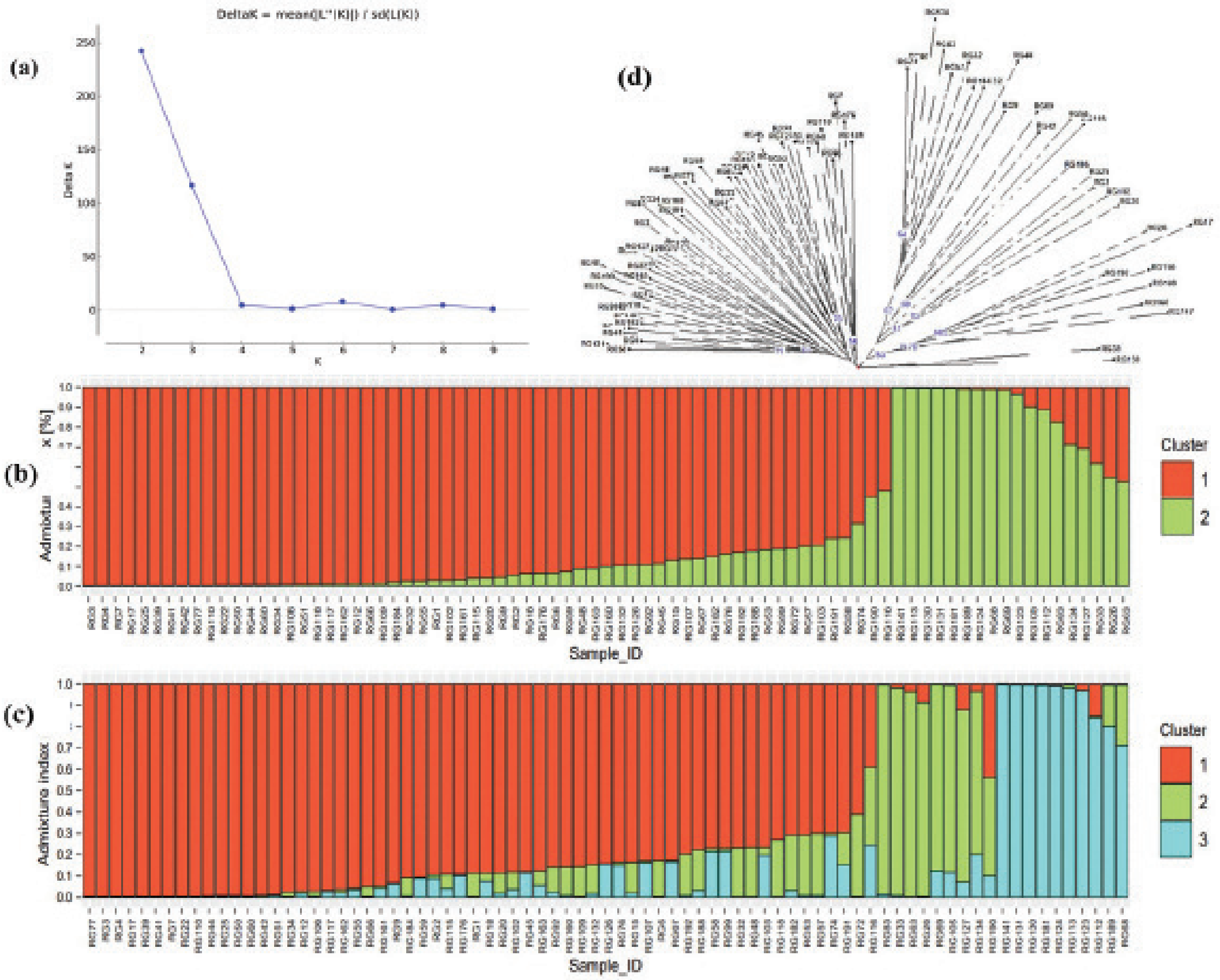

Fig. 2. (a) Magnitude of delta K from STRUCTURE analysis of 79 germplasm, where delta Kmean (|L"(K)|)/sd (L(K).(b) and (c) Population structure of 79 rice accessions. Population structure of all genotypes was divided based on genetic diversity detected by 214 SSR markers with K=2 (b) and k=3 (c). Three subpopulations were represented in (c) by different colours viz., subpopulation 1(59 genotypes), subpopulation 2 (10 genotypes), subpopulation 3 (10 genotypes). (d) A phylogenetic tree representing genotypic diversity computed by neighbour -joining algorithm (DARwin UP)

makeup between accessions with high anthocyanin and those without it. Similar results were also reported by Bastia et al. (2022) while grouping of rice germplasm based on six antioxidant traits including flavonoid and anthocyanin content, clear distinguishing was obtained only at K = 4.

In the present study, the three sub-populations, named OS1, OS2 and OS3, comprised 59, 10 and 10 accessions respectively. OS1 predominantly consists of genotypes with high anthocyanin content, while OS3 and OS2 primarily contain medium and low anthocyanin genotypes, respectively. Additionally, these sub-populations were clearly categorized into three distinct groups based on their geographical origin also. It clearly shows that each subpopulation is diverse from the others. The red bar inferred ancestry genotypes of OS1 were mainly associated to Indian origin genotypes and a majority of the genotypes, encompassing genotypes with high anthocyanin content and total flavonoid content clustered in OS1 were landraces of Tamil Nadu, India. Most of the genotypes within the light green bar category (OS2) were IRRI lines from the Philippines, with a minority originating from India. Approximately half of the genotypes within OS2 exhibited low grain total flavonoid content (<100 QE/100 g) and low anthocyanin content (<0.262 CE/100 g). The entire group represented by the blue bar (OS3) comprised exotic lines. The genotypes within OS3 fell into the medium grain total flavonoid content range (100-200 QE/100 g) and moderate anthocyanin content (0.33–.26 CE/100g).When the NJ (neighbor-joining) tree was constructed, three branches (I, II and III), as well as sub-branches in each main branch, were evident in the tree (Fig. 2d). The population structure analysis of barley genotypes revealed that the most evident level of

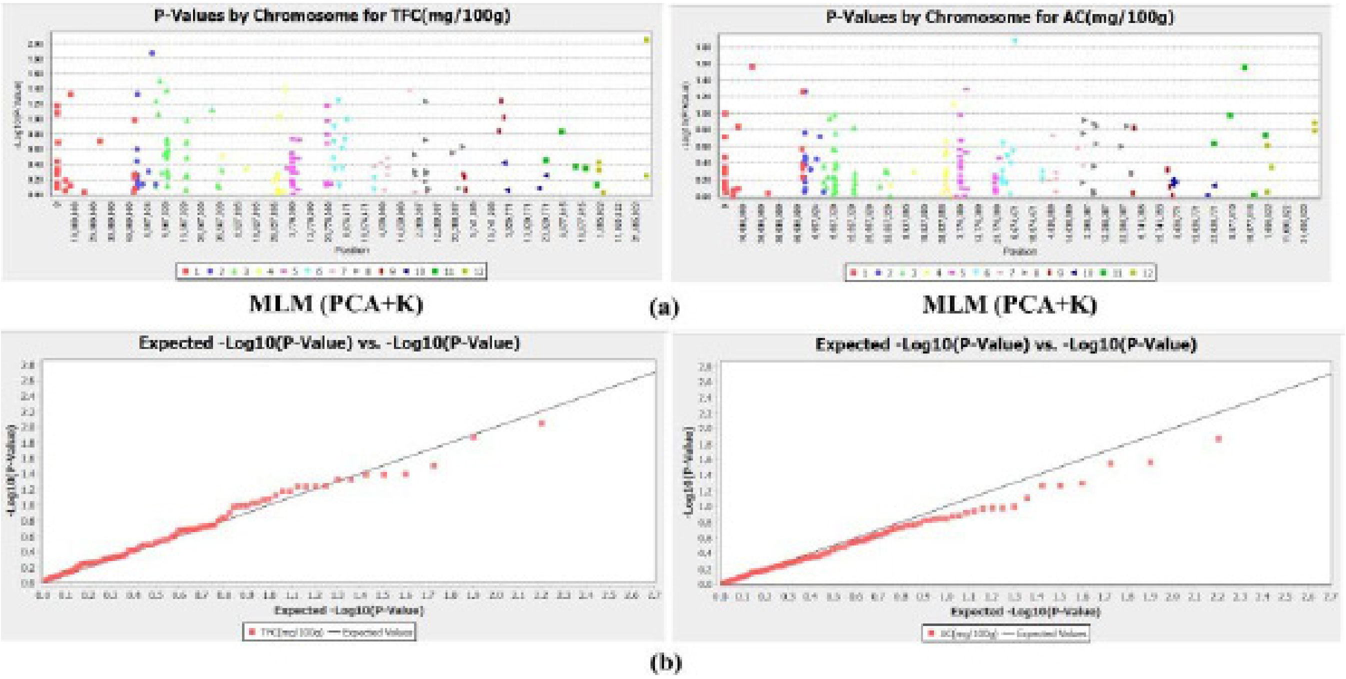

Fig. 3. (a) The Manhattan plot of the AC and TFC -SSR marker association. (b) Quantile-quantile (QQ) plot of the data shown in the Manhattan plot of MLM (PCA+K)

differentiation for TFC was obtained at k=7 (Han et al. 2018). On the other hand, the dendrogram constructed using the UPGMA clustering algorithm grouped the rice accessions into three clusters which did not show relevance to their centre of development.

Marker trait associations (MTA)

The association between SSR markers and the phenotypic data of the two therapeutic traits (TFC and AC) was examined by four different models *viz*., GLMs with Q, GLM with PCA, MLM with Q + K and MLM with PCA+K. Further, these four models were compared for the best fit and the MLM with PCA+K model was the better model fitted for the reducing false positive association results. Table 3 shows an overview and details of MTAs for MLM with PCA+K model with cut-off p<0.01 and p <0.05 (Fig. 3a). Quantile-quantile (QQ) plot of the data shown in the Manhattan plot of MLM (PCA + K) depicted in Fig. 3b shows the normal distribution of the traits among the germplasm. In total, 14 significant MTAs (TFC-8; AC-6) distributed on Chromosomes (Chr.), 1, 2, 3, 4, 5, 6, 7, 11, and 12 were detected at P<0.05, with marker R^2 ranging from 9.5 to 22.7% for AC and 6.2 to 24.8% for TFC, respectively. Out of all these markers, two SSR markers *viz*., RM287 (Chr. 11) and RM19358 (chr 6) were more associated with AC while another two SSR markers, RM452 (Chr. 2) and RM2197 (Chr. 12) for TFC. Bastia et al. (2022) have reported three SRR markers associated with flavonoid content through both GLM (General Linear Model) and MLM approaches in rice.

The significantly associated markers, RM287 and RM19358, identified through this study were found nearer to the candidate genes related to the synthesis of AC. When the sequential positions of these markers were hit in the genomic regions of rice sequence available in the NCBI database, the marker RM287 (Chr. 11: 16.77 Mb) identified for high AC was found to be associated with the gene at the position LOC9266684:14.28 – 14.29 Mb which encodes the enzyme probable 2-oxoglutarate-dependent dioxygenase ANS. This enzyme plays a key role in the biosynthetic pathway of flavonoids, especially anthocyanins. It is understood that another gene *CHS,* encoding Chalcone synthase enzyme is involved in the biosynthesis of phenyl propanoid compounds which form the precursors of both flavonoids and anthocyanins. It was also observed, that, the marker RM287 was found intermediate to the two genes encoding ANS and CHS enzymes which are essential for the biosynthesis of flavonoid and anthocyanin. It was observed that another significantly associated marker RM19358 (Chr. 6: 2.2 Mb) was found in proximity to the gene at the position LOC43401082: 2.394 – 2.397 Mb encoding for the enzyme, Anthocyanin 5 aromatic acyl transferase, crucial for the biosynthesis of flavonoids (Jiang et al. 2013). As the two significant markers, RM287 and RM19358 have been confirmed to be linked to the candidate genes responsible for the biosynthesis of TFC and AC, these markers could be used for improving therapeutic traits in high-yielding rice varieties through Marker Assisted Breeding methods. Oh et al. (2018) identified four potential genes implicated in encoding proteins within

Table 3. Marker-trait associations and marker effect derived from 210 SSR markers and 79 rice accessions

	GLM (PCA)						MLM (PCA+K)					
Trait	Marker	Chr	Pos	marker_F	P	r^2	Marker	Chr	Pos	marker_F	p	r^2
AC	RM5963	6	8,814,621	4.17491	0.0093	0.163	RM5963	6	8814621	3.85402	0.0136	0.177
	RM287	11	16,767,319	2.73394	0.0150	0.218	RM10916	1	15,069,064	3.87464	0.0274	0.128
	RM10916	1	15,069,064	3.32189	0.0444	0.114	RM287	11	16,767,319	2.43301	0.0281	0.227
	RM19358	6	2,200,330	3.2473	0.0449	0.082	RM6034	5	7,034,255	3.13595	0.0508	0.095
	RM6321	1	42,923,127	2.72215	0.0517	0.110	RM6321	1	42,923,127	2.68354	0.0542	0.109
	RM12369	2	1,306,615	2.72215	0.0517	0.110	RM12369	2	1,306,615	2.68354	0.0542	0.109
TFC	RM2197	12	27,350,686	7.66234	0.0082	0.140	RM2197	12	27,350,686	7.46486	0.0090	0.148
	RM452	2	9,563,257	3.27763	0.0105	0.188	RM452	2	9,563,257	3.13007	0.0134	0.217
	RM11	7	19,256,914	2.69834	0.0132	0.254	RM231	3	2,432,425	2.62976	0.0313	0.180
	RM231	3	2,432,425	2.99336	0.0169	0.177	RM559	4	35,151,595	3.35837	0.0406	0.093
	RM48	2	36,318,532	5.51783	0.0216	0.069	RM14606	3	6,351,926	4.3284	0.0414	0.062
	RM416	3	31,248,603	4.56104	0.0361	0.057	RM11	7	19,256,914	2.18081	0.0416	0.248
	RM215	9	21,189,110	2.9996	0.0365	0.111	RM259	1	7,445,627	2.79552	0.0471	0.120
	RM259	1	7,445,627	2.97316	0.0381	0.116	RM110	2	1,326,947	2.79552	0.0471	0.120
	RM110	2	1,326,947	2.97316	0.0381	0.116						
	RM205	9	22,720,624	3.0541	0.0382	0.163						
	RM14241	3	124,844	2.93748	0.0401	0.120						
	RM3507	8	6,158,070	2.93748	0.0401	0.120						
	RM5963	6	8,814,621	2.67811	0.0547	0.110						

	GLM (Q)						MLM (Q + K)					
	Marker	Chr	Pos	marker_F	P	r^2	Marker	Chr	Pos	marker_F	p	r^2
AC	RM5963	6	8814621	4.44024	0.0066	0.166	RM5963	6	8814621	3.75694	0.0148	0.166
	RM287	11	16767319	2.73962	0.0143	0.215	RM10916	1	15069064	3.77506	0.0292	0.120
	RM10916	1	15069064	3.67003	0.0320	0.120	RM287	11	16767319	2.39616	0.0295	0.215
TFC	RM11	7	19256914	2.8666	0.0086	0.261	RM2197	12	27350686	6.816	0.0119	0.130
	RM2197	12	27350686	7.29266	0.0094	0.127	RM559	4	35151595	4.10084	0.0205	0.109
	RM559	4	35151595	4.76882	0.0113	0.114	RM231	3	2432425	2.74774	0.0250	0.179
	RM231	3	2432425	3.14109	0.0128	0.179	RM452	2	9563257	2.69286	0.0275	0.177
	RM452	2	9563257	3.08831	0.0140	0.177	RM11	7	19256914	2.32054	0.0295	0.250
	RM20952	1	100	3.45032	0.0213	0.132	RM20952	1	100	2.90989	0.0407	0.119
	RM18858	5	23176482	3.45032	0.0213	0.132	RM18858	5	23176482	2.90989	0.0407	0.119
	RM536	11	8985132	2.91863	0.0286	0.165	RM259	1	7445627	2.76005	0.0485	0.112
	RM19346	6	1908016	3.61378	0.0320	0.091	RM110	2	1326947	2.76005	0.0485	0.112
	RM259	1	7445627	3.04142	0.0345	0.115						
	RM110	2	1326947	3.04142	0.0345	0.115						
	RM312	1	100	2.6206	0.0417	0.126						
	RM14612	3	6534336	2.83827	0.0448	0.116						
	RM15221	3	17398962	2.83827	0.0448	0.116						
	RM19329	6	1697615	2.83827	0.0448	0.116						
	RM14241	3	124844	2.78772	0.0473	0.111						
	RM3507	8	6158070	2.78772	0.0473	0.111						

the anthocyanin biosynthesis pathway. These genes include *Os01t0372500* (ANS1) on chromosome 1, responsible for encoding Leucoanthocyanidin dioxygenase 1. Remarkably, during our gene annotation study, we found that marker RM10916 (Chr.1) also codes for the same protein, aligning with their findings. Furthermore, another marker, RM559 (Chr.4), exhibited marker-trait associations with anthocyanin content (AC) and total flavonoid content (TFC) and encodes Flavanone-3-dioxygenase1. Interestingly, this coincides with the previously reported candidate gene *Os04t0662600* on chromosome 4, which encodes the same protein as identified by Oh et al. (2018) in their gene expression study pertaining to anthocyanin and TFC in rice. Moreover, the marker RM5963, which displayed a significant marker-trait association with therapeutic traits in the present study, also encodes a similar protein found in the candidate gene *Os06t0192100* located on chromosome 6, as reported by both Oh et al. (2018) and Oikawa et al. (2015). This particular protein plays a vital role in anthocyanin biosynthesis. Given that our findings align with the transcriptome study results of Oh et al. (2018), it underscores the validity of the identified marker-trait associations through this study. Shao et al. (2011) conducted research in which they employed association mapping to discover SSR markers associated with the total phenolic content in rice through association mapping of pigmented and non-pigmented rice genotypes. Furthermore, Fathima et al. (2022) have also reported the identification of molecular markers that are linked to flavonoids and anthocyanins in rice. Similarly, Oh1 et al. (2021) found molecular markers in citrus that can be used to detect the genes responsible for anthocyanin pigment production. A similar study conducted by Gupta et al. (2023) identified the SRR markers associated with total flavonoid content in lentils using association mapping. Thus, through this present study the SSR markers, RM287 and RM19358 have been identified to be linked with the anthocyanin and RM452 and RM2197 flavonoid content and hence could be validated in different genetic backgrounds. In addition, the markers RM10961, RM5963 and RM559 were also found to have significant marker-trait associations with both AC and TFC.

Authors' contribution

Conceptualization of research (SG); Designing of the experiments (SG); Contribution of experimental materials (LA, RM, NV, DGM, PJ, SA, RS, ND); Execution of field/lab experiments and data collection (SG, LA, RM, NV, DGM, SA); Analysis of data and interpretation (VA, SG, MD, MH); Preparation of manuscript (SG, MD, MH, LA, GA).

Reference

Bastia R., Pandit E., Sanghamitra P., Barik S.R., Nayak D.K., Sahoo A., Moharana A., Meher J., Dash P.K., Raj R. and Jena B.K. 2022. Association Mapping for Quantitative Trait Loci Controlling Superoxide Dismutase, Flavonoids, Anthocyanins, Carotenoids, γ-Oryzanol and Antioxidant Activity in Rice. Agronomy, **12**: 3036. https://doi.org/10.3390/agronomy12123036

Debnath S.C., Bhatt D. and Goyali J.C. 2023. DNA-Based Molecular Markers and Antioxidant Properties to Study Genetic Diversity and Relationship Assessment in Blueberries. Agronomy, **13**: 1518. https://doi.org/10.3390/agronomy13061518

Doyle J. J. and Doyle J. L.1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull., **19**: 11-15. http://worldveg.tind.io/record/33886

Earl D. A. and VonHoldt B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour., **4**: 359-361.

Evanno G., Regnaut S. and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol., **14**: 2611-2620.

Fathima M.A., Geetha S., Hemalatha M. and Amudha K. 2022. Identification and validation of genetic locus linked to flavonoid and anthocyanin content in rice using Bulk Segregant Analysis. Indian J. Genet. Plant Breed., **82**: 299-303. https://doi.org/10.31742/ISGPB.82.3.4

Fuleki T. and Francis F. 1968. Quantitative methods for anthocyanins. 1. Extraction and determination of total anthocyanin in cranberries. J. Food Sci., **33**: 72-77.https://doi.org/10.1111/j.1365-2621.1968.tb03658.x

Gawel N. and Jarret R. 1991. Chloroplast DNA restriction fragment length polymorphisms (RFLPs) in *Musa species*. Theor. Appl. Genet., **81**: 783-786. https://doi.org/ 10.1007/BF00224990

Gupta S., Dikshit H.K., Mishra G.P. and Aski M. 2023. Association mapping of total flavonoid content and antioxidant capacities in lentil (*Lens culinaris* Medik.). Indian J. Genet. Plant Breed., **83**: 326-334.https://doi.org/10.31742/ISGPB.83.3.4

Han Z., Zhang J., Cai S., Chen X., Quan X. and Zhang G. 2018. Association mapping for total polyphenol content, total flavonoid content and antioxidant activity in barley. BMC Genom., **19**: 1-10.https://doi.org/10.1186/s12864-018-4483-6

Jiang X.., Liu Y., Li W., Zhao L., Meng F., Yunsheng Wang, Huarong Tan, Hua Yang, Chaoling Wei, Xiaochun Wan, Liping Gao and Tao Xia. 2013. Tissue-Specific, Development-Dependent Phenolic Compounds Accumulation Profile and Gene Expression Pattern in Tea Plant (*Camellia sinensis*). PLOS ONE, **8**(4): https://doi.org/10.1371/journal.pone.0062315.https://doi.org/10.1371/journal.pone.0062315

Oh J.H., Lee Y.J., Byeon E.J., Kang B.C., Kyeoung D.S. and Kim C.K. 2018. Whole-genome resequencing and transcriptomic analysis of genes regulating anthocyanin biosynthesis in black rice plants. Biotech, **8**: 1-13.https://doi.org/10.1007/s13205-018-1140-3

Oh C.J., Woo J.K., Yi K.U., Park Y.C., Lee H.Y., Kim M., Park S., Yun S.H., Lee Y., Kim H.J. and Song K.J. 2021. Development of molecular markers for genotyping of Ruby, a locus controlling anthocyanin pigment production in Citrus with its functional analysis. Sci. Hortic., **289**: 110457.https://doi.org/10.1016/j.scienta.2021.110457

Oikawa T., Maeda H., Oguchi T., Yamaguchi T., Tanabe N., Ebana K., Yano M., Ebitani T. and Izawa T. 2015. 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

Pradhan S., Pandit E., Pawar S., Bharati B., Chatopadhyay K., Singh S., Dash P. and Reddy J. 2019. Association mapping reveals multiple QTLs for grain protein content in rice useful for biofortification. Mol. Genet. Genom., **294**(4): 963-983.https://doi.org/10.1007/s00438-019-01556-w

Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A. and Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies.Nat. Genet., **38**(8): 904-909.https://doi.org/10.1038/ng1847

Pritchard J. K., Wen W. and Falush D. 2003. Documentation for STRUCTURE software: Version 2.3. http://web.stanford.edu/group/pritchardlab/structure.html (accessed on 4 October 2023)

Shao Y., Jin L., Zhang G., Lu Y., Shen Y. and Bao J. 2011. Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice. Theor. Appl. Genet., **122**: 1005-1016. https://doi.org/10.1007/s00122-010-1505-4

Shen Y., Jin L., Xiao P., Lu Y. and Bao J. 2009. 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

Woisky R. G. and Salatino A. 1998. Analysis of propolis: some parameters and procedures for chemical quality control. J. Apic. Res., **37**: 99-105.https://doi.org/10.1080/00218839.1998.11100961

Yu S., Yu K., Tresp V., Kriegel H. P. and Wu M. Supervised probabilistic principal component analysis. In Proceedings of the 12th ACM SIGKDD International conference on Knowledge discovery and data mining, pp. 464-47. https://doi.org/10.1145/1150402.1150454