RESEARCH ARTICLE

Association studies of dormancy and cooking quality traits in direct-seeded *indica* rice

SUNAYANA RATHI[1]*, K. PATHAK[2], R. N. S. YADAV[1], B. KUMAR[3] and R. N. SARMA[2]

[1]*Centre for Studies in Biotechnology, Dibrugarh University, Dibrugarh 786 004, India*
[2]*Department of Plant Breeding and Genetics, Assam Agricultural University, Jorhat 785 013, India*
[3]*Directorate of Maize Research, Indian Agricultural Research Institute, New Delhi 110 012, India*

Abstract

Association analysis was applied to a panel of accessions of Assam rice (*indica*) using 98 SSR markers for dormancy-related traits and cooking quality. Analysis of population structure revealed 10 subgroups in the population. The mean r^2 and D' value for all intrachromosomal loci pairs was 0.24 and 0.51, respectively. Linkage disequilibrium between linked markers decreased with distance. Marker-trait associations were investigated using the unified mixed-model approach, considering both population structure (Q) and kinship (K). Genome-wide scanning, detected a total of seven significant marker-trait associations ($P < 0.01$), with the R^2 values ranging from 12.0 to 18.0%. The significant marker associations were for grain dormancy (RM27 on chromosome 2), α-amylase activity (RM27 and RM234 on chromosomes 2 and 7, respectively), germination (RM27 and RM106 on chromosome 2), amylose (RM282 on chromosome 3) and grain length elongation ratio (RM142 on chromosome 4). The present study revealed the association of marker RM27 with traits like dormancy, α-amylase activity and germination. Simple correlation analysis of these traits revealed that these traits were positively correlated with each other and this marker may be useful for simultaneous improvement of these traits. The study indicates the presence of novel QTLs for a few traits under consideration. The study reveals association of traits like dormancy, α-amylase activity, germination, amylose content, grain length elongation ratio with SSR markers indicating the feasibility of undertaking association analysis in conjunction with germplasm characterization.

[Rathi S., Pathak K., Yadav R. N. S., Kumar B. and Sarma R. N. 2014 Association studies of dormancy and cooking quality traits in direct-seeded *indica* rice. *J. Genet.* **93**, 3–12]

Introduction

Association or linkage disequilibrium (LD) mapping which revolutionized genetic mapping in humans (Donnelly 2008) and plants (Nordborg and Weigel 2008), is considered as an efficient way of probing the genetic basis of complex traits. Association mapping is increasingly being utilized to detect marker-QTL linkage associations using plant materials routinely developed in breeding programmes (Shi *et al.* 2011). According to Shi *et al.* (2011), association mapping using breeding populations may be a more practical approach for cultivar development, considering that markers linked to major QTL can immediately be utilized in marker-assisted selection (MAS), once new QTLs are identified. For instance, in soybean (*Glycine max* L. Merr.) two markers, Satt114 and Satt239 were found to be associated with iron deficiency chlorosis loci using advanced breeding lines (Wang *et al.* 2008). Moreover, association mapping has validated the markers previously identified to be linked to various traits, which suggest practicability of this approach to use MAS in breeding programmes. In rice (*Oryza sativa* L.), microsatellite markers associated with yield and its components were identified in a variety trial, and many of them were located in regions where QTL had previously been identified (Agrama *et al.* 2007). One of the great advantages of association mapping lies in the fact that no mapping population needs to be developed, as the sampling of non-related individuals represent a series of advantage towards developing and validating MAS in breeding programmes (Jannink *et al.* 2001) as well as an opportunity for increasing the exploitation of germplasm accessions in the search for advantageous allele combinations.

Rice (*Oryza sativa* L.), an important agricultural crop, and a model species, has been cultivated for more than 7000 years (Agrama *et al.* 2007). Use of SSR markers to interpret population structure results in much greater resolution than use

*For correspondence. E-mail: rathisunayana@yahoo.co.in.

Keywords. Assam rice; α-amylase; dormancy; cooking quality; association mapping.

of other types of markers, because of the high level of polymorphism at SSR loci (Akkaya *et al.* 1992; Cho *et al.* 2000). In rice, the highly polymorphic nature of SSR motifs is coupled with a low level of homoplasy observed in *O. sativa* cultivars (Chen *et al.* 2002), providing an appropriate tool for population genetic studies.

Due to wide diversity in rice, several previous researches on rice reported a varied population structure (Zhang *et al.* 2011). Most studies widely accepted five major groups (*indica, aus*, aromatic, temperate *japonica* and tropical *japonica*) and they were detected in a sample of 234 rice varieties using association mapping analysis (Garris *et al.* 2005). Agrama *et al.* (2007) observed eight subpopulations corresponding to major geographic regions among 103 rice accessions. Likewise, seven subpopulations were detected within rice landraces in Guizhou province, China (Zhang *et al.* 2007). Two subgroups including *indica* and *japonica* as well as six sub-subgroups were found within a primary rice core collection of China (Wen *et al.* 2009). Seven subgroups were found within a 416 rice population from China (Jin *et al.* 2010). The varied numbers of subgroups might be due to different methods, different numbers of markers, different rice populations applied in population structure examination, which should be further studied. However, as far as we know, no information on the population structure of a rice collection of Assam (*indica* rice) assessed with SSR marker set is available.

LD defined as the non-random association of alleles at separate loci located on the same chromosome (Mackay and Powell 2007), is a prerequisite for association mapping. The distance at which LD declines with genetic or physical distance determines the marker density needed for achieving a reasonable mapping resolution. The extent of LD may vary among different genomic regions (Mather *et al.* 2007). Numerous studies on global germplasm collections indicate 25 cM as a reasonable resolution for association mapping in rice (Agrama *et al.* 2007; Agrama and Eizenga 2008; Li *et al.* 2011). In these populations, LD may decline over relatively short or long distances in the genome, making fine mapping possible. In sugar beet (*Beta vulgaris* L.), LD extended to 3 cM (Kraft *et al.* 2000) and LD in some *Arabidopsis* populations exceeded 50 cM (Nordborg *et al.* 2005). LD as a function of genetic distance is very common for distances <10 cM (Kraakman *et al.* 2004) in barley (*Hordeum vulgare* L.), in contrast with maize (*Zea mays* L.), for which the LD diminished after 2000 bp (Remington *et al.* 2001). LD decay between all pairs of SNP (single nucleotide polymorphism) loci in the region around the rice *xa*5 locus approached 0.1 after only 100 kb (Garris *et al.* 2003). The pattern of polymorphism LD decayed rapidly within 50 kb in *Arabidopsis* (Nordborg *et al.* 2005), while the LD extended over 3 cM in one human population (Eaves *et al.* 1998). These studies suggest that the extent of LD varies among different genomic regions and among different rice populations examined. However, to our knowledge, no earlier research is available on the LD of a rice collection of Assam, India (*indica* rice) with extensive genome-wide distributed SSR markers.

The cultivation of upland rice in Assam and Northeastern India under direct-seeded rainfed condition (March–July) is considered as playing gamble in the context of uncertainties in return. The upland rice encounters plenty of rainfall at the maturity stage along with rise in temperature and humidity, resulting in sprouting of rice grain within the panicle in the standing crop. Sprouted grains are imperfect in appearance, generally not accepted for milling and not suitable for human consumption, because of the presence of α-amylase (Bewley and Black 1985). Grain α-amylase activity is used as the major indicator of preharvest sprouting, because of the effect it has on flour properties and of its abundance in germinating grains. Grain dormancy is responsible for uniform germination, a trait which is particularly important for the grain quality, but is also relevant in the context of controlling preharvest sprouting (Nakamura *et al.* 2011). Moreover, development of high yield varieties of upland rice with a dormancy period of 20–30 days is of utmost importance to improve the productivity and quality of upland direct-seeded rice (Rathi *et al.* 2011). Improving grain quality is important challenge in rice breeding, thus priorities for the international market (Fan *et al.* 2005). Although quality assumes many aspects and is highly related to preference in diverse cultures, its characteristics are mainly defined by milling properties, grain size and shape, cooking and eating characteristics, and nutritional qualities (He *et al.* 1999). Among these, the most relevant are the appearance and cooking quality, reported to be directly related to amylose content, gel consistency and gelatinization temperature (Fan *et al.* 2005).

In this study, we determined the population structure in a natural population that consisted of 100 *indica* rice varieties. Further, we used the mixed linear model method in the TASSEL software to conduct association analysis for cooking quality traits through genomewide scanning. Finally, the potential application of association analysis results for parental selection or MAS for the development of superior rice varieties is discussed. To our knowledge, no studies have been reported to detect marker-trait association using association mapping method in Assam rice.

Materials and methods

Plant materials, biochemical and cooking quality traits evaluation

In this study a total of 100 *indica* rice genotypes originated from different regions of Assam, Northeastern India were used (table 1). All accessions were grown in net house during summer season for genetic purification, future seed stock and maintenance. Panicles from each plant were collected on the 35th day after heading when the grains were fully filled and were immediately stored at 4°C to maintain dormancy. The grains were oven dried at 100°C (±2°C) and converted into fine powder and kept in a desiccator for analytical works. The dormancy was assessed following Ikehashi (1972) and

Table 1. Description about the association mapping population use in study.

Genotypes	Name	Local/HYV/ improved	Place of collection	Characteristics	Genotypes	Name	Local/HYV/ improved	Place of collection	Characteristics
G1	Cheniahu	Local	Titabar	Pigmented	G51	Betguti Ahu	Local	Nagaon	White
G2	Kolong	Improved	Titabar	White	G52	IET-6148	Advanced breeding line	Titabar	White
G3	Kola Meghi Ahu	Local	Titabar	Pigmented	G53	IET-10898	Advanced breeding Line	Titabar	White
G4	ARC-10372	Local	Titabar	White	G54	Banglami	Local	Goalpapra	White
G5	DKH-36	Improved	Titabar	White	G55	Boga Ahu	Local	Titabar	White
G6	H-1120	Improved	Titabar	White	G56	Changa Ahu	Local	Goalpapra	Pigmented
G7	Sonamukhi	Local	Karimganj	Pigmented	G57	Kajoli Ahu	Local	Sonitpur	Pigmented
G8	Garem	Improved	Titabar	Pigmented	G58	AS-55	Local	Titabar	Pigmented
G9	Betguli Ahu	Local	Titabar	White	G59	Koi murali	Local	Goalpapra	Pigmented
G10	Ahu Joha-2	Local	Titabar	White	G60	Moren Chenkol	Local	Titabar	Pigmented
G11	Guni Ahu	Local	Dibrugarh	Pigmented	G61	Chidon	Local	Diphu	White
G12	Rikhoijoi	Local	Titabar	Pigmented	G62	Bizor-3	Local	Diphu	White
G13	Bongal Ahu	Local	Dibrugarh	White	G63	Bizor-1	Local	Diphu	White
G14	New Sali	Local	Titabar	White	G64	Bairing	Local	Diphu	Pigmented
G15	Rasi	HYV	Titabar	White	G65	Ahu I	Local	Nagaon	White
G16	IET-6223	Advanced breeding line	Titabar	White	G66	Basanta Bahar	HYV	Goalpapra	White
G17	Govind	HYV	Titabar	White	G67	Dubori Cinga	Local	Karimganj	Pigmented
G18	Jai Bangla	Local	Karimganj	White	G68	Garem I	Local	Titabar	Pigmented
G19	Kanua	Local	Karimganj	White	G69	As-156	Local	Titabar	Pigmented
G20	Kosamani	Local	Titabar	Pigmented	G70	As-330	Local	Titabar	Pigmented
G21	Baibi-2	Local	Diphu	Pigmented	G71	As-229	Local	Titabar	Pigmented
G22	Bongaloni	Local	Titabar	Pigmented	G72	As-195	Local	Titabar	White
G23	Jai Bangla1	Local	Karimganj	White	G73	As-39/13	Local	Titabar	Pigmented
G24	IET-10896	Advanced breeding line	Titabar	White	G74	As-177/2	Local	Titabar	Pigmented
G25	Ronga Gutia	Local	Titabar	Pigmented	G75	As-38/2	Local	Titabar	Pigmented
G26	Ch-63	HYV	Titabar	White	G76	As-305/2	Local	Titabar	Pigmented
G27	Nilaji-II	Local	Diphu	Pigmented	G77	As-180/2	Local	Titabar	Pigmented
G28	Rongadoria (local)	Local	Titabar	Pigmented	G78	As-317	Local	Titabar	Pigmented
G29	Koijanori III	Local	Titabar	Pigmented	G79	As-323	Local	Titabar	Pigmented
G30	Koijanori II	Local	Titabar	Pigmented	G80	As-324	Local	Titabar	White
G31	Koijanori I	Local	Titabar	Pigmented	G81	As-27	Local	Titabar	Pigmented
G32	Sarimohia Ahu	Local	Titabar	Pigmented	G82	As-327	Local	Titabar	Pigmented
G33	Tinimohia Ahu	Local	Titabar	White	G83	As-159/2	Local	Titabar	Pigmented
G34	Bahmori	Local	Titabar	Pigmented	G84	As-314	Local	Titabar	White
G35	Borkola Ahu	Local	Titabar	Pigmented	G85	As-312	Local	Titabar	Pigmented
G36	Daukola Maghi	Local	Titabar	Pigmented	G86	As-209	Local	Titabar	White
G37	Boga Bengena Gutia	Local	Titabar	Pigmented	G87	As-292	Local	Titabar	White
G38	Kola Bengena Gutia	Local	Sonitpur	White	G88	As-1224	Local	Titabar	White
G39	Kolamaghi	Local	Titabar	Pigmented	G89	As-204	Local	Titabar	Pigmented
G40	Sohalia Ahu	Local	Titabar	Pigmented	G90	As-178/3	Local	Titabar	White
G41	Kola Ahu	Local	Titabar	Pigmented	G91	As-206	Local	Titabar	White
G42	Panja Sali	Local	Titabar	White	G92	As-36/20	Local	Titabar	Pigmented
G43	Rongadoria-3	Local	Nagaon	Pigmented	G93	As-75	Local	Titabar	Pigmented
G44	Bongal doria	Local	Titabar	Pigmented	G94	As-313	Local	Titabar	Pigmented
G45	Iharsal	Local	Titabar	White	G95	As-310	Local	Titabar	Pigmented
G46	Begon Bisi	Local	Titabar	Pigmented	G96	As-105	Local	Titabar	Pigmented
G47	Goal Bhog	Local	Goalpapra	Pigmented	G97	As-47	Local	Titabar	White
G48	Poromi Ahu	Local	Titabar	Pigmented	G98	As-39	Local	Titabar	Pigmented
G49	Naga Ahu	Local	Titabar	Pigmented	G99	As-34	Local	Titabar	Pigmented
G50	Shesamthat	Local	Titabar	Pigmented	G100	As-90	Local	Titabar	Pigmented

Wan *et al.* (1997). Duration of dormancy was calculated for each genotype when 10% of the germination was achieved in each genotype (Rathi *et al.* 2011). Grain traits such as raw grain length, raw grain breadth, cooked grain length, cooked grain breadth, length elongation ratio and breadth elongation ratio were estimated as per standard protocol.

The α-amylase activity, starch content and total soluble sugars were estimated by the method described by Sadasivam and Manickam (1996). Amylose content was determined by the method described by Juliano (1971). Gel consistency was determined by the method of Cagampang *et al.* (1973). Gelatinization temperature (GT) was assessed indirectly as the alkali spreading value (ASV) of hulled kernels as per modified procedure of Little *et al.* (1958).

SSR analysis

The total genomic DNA from each of the genotypes included in the present study was extracted following the protocol of Plaschke *et al.* (1995) with slight modification. Hundred and twenty microsatellite primer pairs (Sigma Chemicals, St Louis, USA) were used for the amplification of DNA fragments containing simple sequence repeats from rice genotypes. SSR markers were screened at even distribution across the rice genome based on the published map information (Temnykh *et al.* 2001). Out of those, markers which showed clear-cut polymorphism with at least 10 bp differences among the amplicons in randomly chosen 20 genotypes included in the study were selected for the present research. Ninety-eight microsatellite markers were selected from original screening of 120 microsatellite primers for analysing the variability in hundred genotypes. The average genetic or physical distance between the selected SSR markers is 18 cM. Thermocycling was carried out in a GeneAmp PCR System 2400 (Applied Biosystems, Foster City, USA) and the 'touchdown' amplification conditions were: 94°C for 5 min; 2 × (94°C for 1 min, 65°C for 1 min and 72°C for 2 min); 2 × (94°C for 1 min, 62°C for 1 min and 72°C for 2 min); 4 × (94°C for 1 min, 59°C for 1 min and 72°C for 2 min); 25 × (94°C for 1 min, 55°C for 1 min and 72°C for 2 min); and a final extension of 72°C for 5 min. Touchdown PCR enhances the specificity of the initial primer–template duplex formation, avoid amplifying non-specific sequences and hence the specificity of the final PCR product. PCR products were analysed on 3% agarose gel and visualized by ethidium bromide. The photograph of the gel was digitally documented in Gel Documentation System (Ultra Violet Products, California, USA). The molecular weights of PCR products were calculated based on ladder of known molecular weight of 100 bp (Bangalore Genei, Bangalore, India).

Statistical analysis

Association mapping analyses were carried out using TASSEL 3.0 software (Bradbury *et al.* 2007). The mixed linear model (MLM) analyses were performed using a kinship K matrix and population structure Q matrix. The K matrix was generated based on 98 SSRs using kinship matrix function in TASSEL. Population structure consisted of a Q matrix that describes the per cent subpopulation parentage for each line in the analysis. These percentages were calculated by Structure 2.3.3 software (Pritchard *et al.* 2000). We set k (the number of subpopulations) from two to 20 and performed three runs for each k value. Length of burn-in period as well as numbers of iterations were set at 1,50,000. Since the likelihood for model parameter $k = 2$ was much higher than $k = 1$ and comparable with $k = 3$ or higher, we chose $k = 2$ and generated a Q matrix from 98 SSRs.

LD values for r^2 (Hill and Robertson 1968) and D' (Farnir *et al.* 2000) between SSR loci on the same linkage group (LG) were calculated using the software package TASSEL3 with the permutations test of 10,000. The LD among SSR markers were determined as per Cornell SSR 2001 based on the published map information (Temnykh *et al.* 2001). The significant marker-trait associations were indicated by P value with corresponding R^2 for each marker as the percentage of the total variation explained. The P value determines whether a QTL is associated with the marker and the R^2 marker evaluates the magnitude of the QTL effects. The pairs of loci were considered to be in significant LD if $P < 0.01$.

Results and discussion

The present investigation shows the presence of abundant genetic variation in the materials under investigation (table 2). Principal components analysis (PCA) was used to eliminate redundancy in the dataset. Altogether eight principal components (PC1 to PC8) were identified (table 2), accounting for more than 80% of the variation explained by the traits under study. PCA chooses independent or orthogonal axes, which are minimally correlated and represents linear combinations of the original characters (Clifford and Stephenson 1975). The relatively discriminating power of axes and their associated characters are measured by eigenvalues and factor scores, respectively (Ogunbodede 1997). In this study, we chose to follow the criterion used by Clifford and Stephenson (1975) and corroborated by Guei *et al.* (2005), which suggested that the first three PC are often the most important in reflecting the variation patterns among accessions, and the characters associated with these are more useful in differentiating accessions. According to this criterion, the first three components account for more than 59% of total variation (table 2), giving a clear idea of the structure underlying the variables analysed. However, Sanni *et al.* (2012) used the criterion of Raji (2002) that determine the cutoff limit for the coefficients of the proper vectors; this criterion treated coefficients greater than 0.3 as having a large enough effect to be considered important, while traits having a coefficient less than 0.3 were considered not to have important effects on the overall variation observed in the

Table 2. Description of the characters for 100 rice genotypes used for association mapping.

	*Percent of variance explained	Grain dormancy (%)	Duration of dormancy (Days)	Germination at three DAH (%)	α-Amylase activity at 3 DAG (U/mg maltose/min/ endosperm)	Germination at five DAH (%)	α-Amylase activity at 5 DAG (U/mg maltose/min/ endosperm)	Germination at seven DAH (%)	α-Amylase activity at seven DAG (U/mg maltose/ min/endosperm)
Minimum		0.57	1.00	0.57	0.03	0.57	0.19	0.57	0.43
Maximun		58.22	25.00	44.43	0.58	57.15	0.90	58.22	1.83
Mean		16.77	6.64	9.18	0.28	11.16	0.51	16.77	0.86
SE		0.28	0.45	0.23	0.01	0.31	0.01	0.28	0.04
PC1	30.82*	0.95	−0.72	0.91	0.49	0.93	0.35	0.95	0.53
PC2	10.34*	0.25	−0.14	0.19	−0.52	0.18	−0.61	0.25	−0.56
PC3	9.59*	0.04	−0.04	−0.02	0.10	−0.02	−0.04	0.04	−0.02
PC4	8.66*	0.02	0.00	0.08	0.14	0.08	0.28	0.02	0.23
PC5	8.09*	−0.05	−0.03	0.03	0.08	0.02	0.14	−0.05	0.26
PC6	6.19*	−0.06	0.14	−0.10	0.19	−0.09	0.27	−0.06	0.16
PC7	4.57*	−0.03	0.12	0.01	−0.36	0.00	−0.25	−0.03	−0.01
PC8	4.29*	−0.02	0.26	0.12	0.37	0.09	0.18	−0.02	−0.07

	*Per cent of variance explained	Germination at 10 DAH (%)	α-Amylase activity at 10 DAG (U/mg maltose/min/ endosperm)	Germination at 12 DAH (%)	α-Amylase activity at 12 DAG (U/mg maltose/min/ endosperm)	Grain length (mm)	Grain breadth (mm)	Cooked grain length (mm)	Cooked grain breadth (mm)
Minimum		0.57	0.73	1.00	0.46	5.00	2.00	6.50	3.00
Maximun		66.96	2.63	96.00	2.10	7.50	3.50	10.00	5.00
Mean		20.44	1.30	42.48	1.01	6.17	2.81	7.61	3.73
SE		0.72	0.02	0.34	0.02	0.07	0.06	0.05	0.05
PC1	30.82*	0.93	0.50	0.93	0.55	0.17	0.18	0.20	−0.24
PC2	10.34*	0.26	−0.58	0.23	−0.51	0.25	−0.05	0.34	0.16
PC3	9.59*	0.03	−0.08	0.00	0.13	−0.14	0.21	0.57	0.76
PC4	8.66*	0.01	0.02	−0.02	−0.05	−0.54	0.32	−0.02	0.14
PC5	8.09*	−0.05	0.30	−0.04	0.11	0.50	−0.65	0.33	0.13
PC6	6.19*	−0.06	−0.04	−0.05	−0.07	−0.04	0.02	0.39	0.04
PC7	4.57*	−0.04	0.18	−0.07	0.18	0.10	0.30	0.30	−0.15
PC8	4.29*	−0.06	−0.22	−0.11	−0.09	0.40	0.49	0.03	0.32

	*Per cent of variance explained	Length elongation ratio	Breadth elongation ratio	GT	GC	Starch	Amylose	Amylopectin	TSS
Minimum		1.00	1.00	2.00	55.00	66.00	14.00	47.10	0.35
Maximun		1.60	2.50	7.00	90.50	79.50	25.00	64.40	1.00
Mean		1.24	1.34	3.60	86.16	74.44	17.87	56.57	0.60
SE		0.02	0.04	0.00	1.43	0.11	0.06	0.12	0.01
PC1	30.82*	0.04	−0.32	0.19	0.17	−0.25	0.12	−0.28	0.07
PC2	10.34*	0.10	0.17	0.38	−0.18	0.28	0.10	0.18	0.26
PC3	9.59*	0.66	0.47	−0.30	−0.07	−0.41	−0.23	−0.21	−0.53
PC4	8.66*	0.44	−0.17	0.18	−0.30	0.68	−0.33	0.76	0.07
PC5	8.09*	−0.12	0.64	0.18	0.34	0.23	−0.35	0.39	0.19
PC6	6.19*	0.39	0.01	0.06	0.15	0.16	0.71	−0.27	0.60
PC7	4.57*	0.19	−0.38	−0.11	0.57	−0.03	−0.24	0.11	0.12
PC8	4.29*	−0.34	−0.10	0.19	0.12	−0.02	−0.09	0.04	0.03

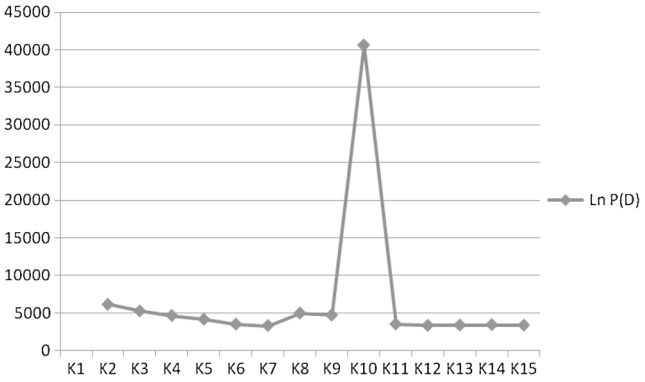

Figure 1. Population structure estimation in a set of *indica* rice. Estimated ln (probability of the data), which was calculated for K ranges from two to 15.

present study. The first PC accounted for more than 30% of total variance, whereby grain dormancy (%), germination (%) at 3, 7, 10 and 12 days after harvest (DAH); α-amylase activity at 3, 5, 7, 10 and 12 days after germination (DAG) were the variables that contributed most positively. In contrast, the variables contributing most negatively were breadth elongation ratio and duration of dormancy (days). The second PC accounted for more than 10% of total variance, in which gelatinization temperature and cooked grain length (mm) were identified as main contributor for classification with positive values. The third PC accounted for 9% of variation and was associated with cooked grain characteristics like length and breadth of cooked grain and their length-wise and breadth-wise elongation ratios.

A total of 194 alleles were detected with the set of 120 SSR markers on a panel of 100 genotypes. The average number of alleles per locus was 1.867, ranging from one (RM323, RM151, RM158, RM17, RM19, RM128, RM307, RM142, RM162, RM345, RM234, RM38, RM25, RM321, RM467, RM184) to three (RM592). Twelve loci presented 14 alleles with a frequency below 5% (rare alleles) and they were omitted from analysis to avoid an increase in variance errors in association analysis. The remaining 183 alleles from 98 loci, referred to as common alleles, were used to check the structure of 100 rice genotypes.

Population structure

Association mapping requires population structure to be taken into account to avoid false positive associations. The software Structure (Pritchard *et al.* 2000) was run for K (number of fixed subgroups or clusters) ranging from two to 20 on the entire set of genotypes using all SSR markers. The likelihood value of this analysis is shown in figure 1. Likelihood initially decreases and then increases abruptly at K10 and then again decreases (figure 1), after which it became almost constant. This could imply that the lines included in the analysis were not very diverse. However, the most significant change was observed when K was increased from nine to ten, which corresponds with the origin of the populations that can be divided as subgroups. Therefore, the structure results of K = 10 were considered the best possible partition as they showed a high consistency clustering based on genetic distance. Thus, 14, 4, 11, 26, 17, 28% of the lines were assigned to green, yellow, pink, cyan, brown, blue subgroups, respectively. A further study of the partitioning of lines can be seen in figure 2, which is the graphical representation of the placement of each line in the study into its corresponding cluster, for K = 10. Such a graph shows the number of lines in each cluster, and the per cent mixing of each line within each cluster, a useful visualization of admixture. Substructuring of the genotypes was also supported by the clustering done based on Neighbour joining (figure 3). We therefore used the respective Q matrix outputs of the four subpopulation runs (K = 10) for the structure-based association analysis.

LD of all polymorphic markers was studied in TASSEL with multiple permutations of 10,000 times. LD plot for r^2 and D' value between the markers has been found. The value for r^2 between the marker pairs was ranging from 0.0001 to

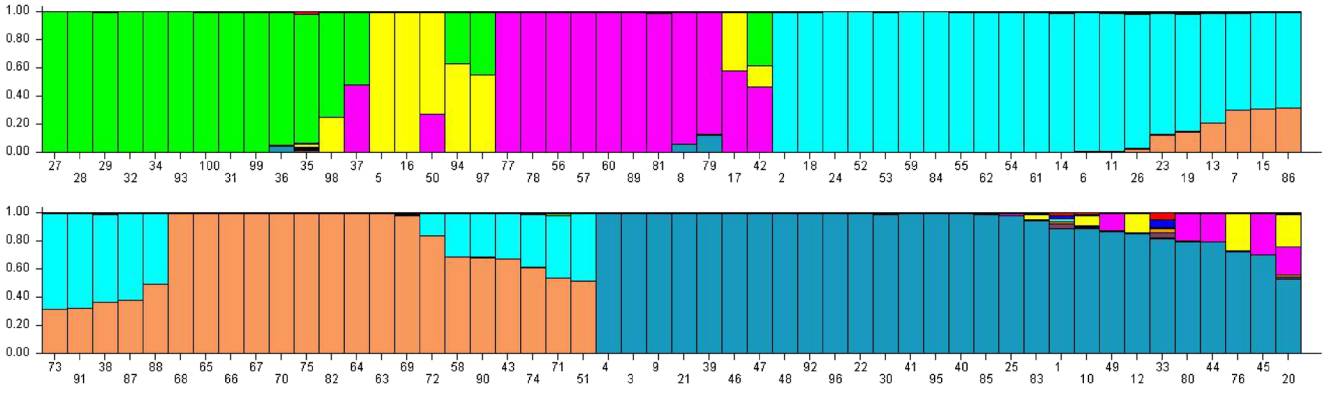

Figure 2. Bar plot showing genetic diversity of 100 rice genotypes using the program Structure. Each accession is divided into a number of hypothetical subpopulations based on the proportional membership (a vertical bar expressed as %) from K= 2 to K= 20, with the most divergent subpopulations were obtained at K= 10. Each group is represented by a different colour as listed: green, pink, yellow, cyan, brown, blue.

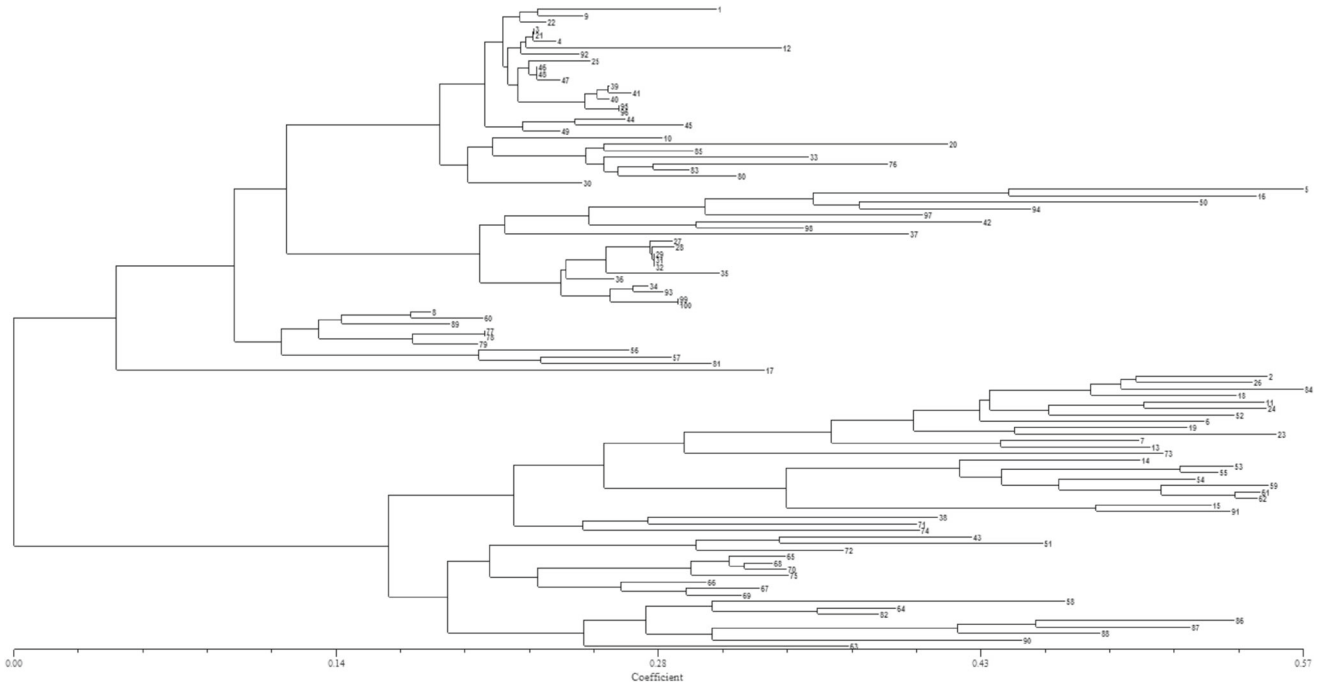

Figure 3. Neighbour-joining tree of 100 *indica* genotypes.

1.0 (table 3). However, for D' it was ranging from 0.003 to as high as 1.00. A statistical association between a neutral marker allele and the phenotype occurs when marker alleles are in gametic phase disequilibrium (GPD) (synonymously to term 'linkage disequilibrium') with alleles at a QTL. Over a series of generations, in an unstructured population (a randomly mating population with no complicating factors such as population subdivision, selection preference and immigration), only LD between QTL and markers closely linked to the QTL will sustain. However, most populations have some degree of structure or subdivision and the simple relationship between strength of LD and meiotic distance does not apply and therefore LD between unlinked loci often occur (Mackay and Powell 2007). Our association mapping population has also shown the partial structure which might have contributed for LD between unlinked loci. LD between unlinked markers may contribute to false positive which could have been taken care to some extent by structure and kinship studies. The LD between unlinked markers may also be due to the selection pressure put while improving the genotypes for desirable traits. High LD between markers may lead to requirement of less number of markers for population studies but have to compromise with the resolution. The high value of LD in rice may be due to its high self-pollinated nature. The LD between unlinked loci might be due to genetic drift, familial relatedness and population structure (Zhang *et al.* 2011). According to Farnir *et al.* (2000), when substantial LD between pairs of unlinked markers is detected, it is necessary to test simultaneously for linkage and LD to minimize the risk of type I error.

LD decay graph shows the decreasing order of the LD with respect to increase in chromosomal distance between the loci

Table 3. Association (R^2) of SSR markers with dormancy and cooking quality traits studied under different model.

Trait	Marker	Chromosome number	Position (cM)	*P*	Marker R^2*
α-Amylase activity at 12 DAG*	Satt14 (RM270)	2	66.0	8.90E-03	0.068
α-Amylase activity at 7 DAG	Satt60 (RM234)	7	88.2	5.78E-04	0.133
Amylose	Satt24 (RM282)	3	100.6	8.63E-03	0.076
Germination at 10 DAH* (%)	Satt14 (RM27)	2	66.0	1.19E-03	0.117
Germination at 12 DAH (%)	Satt14 (RM27)	2	66.0	2.38E-03	0.099
Germination at 3 DAH (%)	Satt14 (RM27)	2	66.0	3.75E-04	0.144
Germination at 5 DAH (%)	Satt14 (RM27)	2	66.0	4.32E-04	0.141
Germination at 5 DAH (%)	Satt17 (RM106)	2	123.2	9.59E-03	0.104
Germination at 7 DAH (%)	Satt14 (RM27)	2	66.0	7.32E-04	0.130
Grain dormancy (%)	Satt14 (RM27)	2	66.0	7.32E-04	0.130
Length elongation ratio	Satt33 (RM142)	4	99.0	3.84E-03	0.091

*R^2 percentage of the total variation explained. DAH, days after harvest; DAG, days after germination.

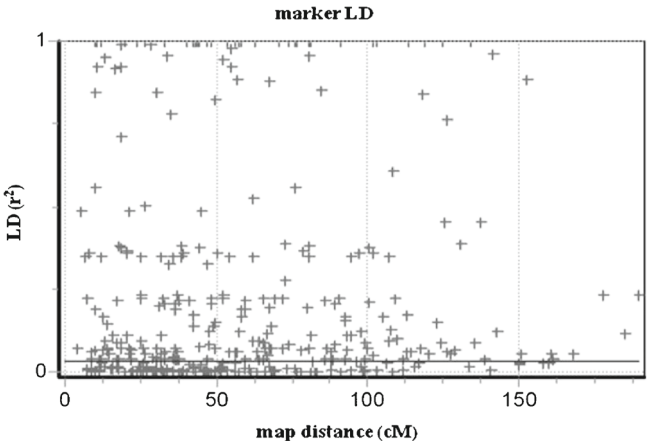

Figure 4. Scatterplot of the LD (r^2) of marker pairs as a function of the intermarker distance in cM. A genetic distance of 180 cM was chosen to represent unlinked loci on different chromosomes. LD analysis was performed on the entire population (100 genotypes).

(figure 4). Most of markers showing r^2 from 0 to 0.5 were below existing distance of 100 cM. However, the LD pertains up to 180 cM with value of 0.2. The high r^2 between the loci gives the effectiveness of the association mapping.

Association analysis between traits and molecular markers

Using the MLM program in the TASSEL software, we identified marker-trait associations for the 24 traits evaluated. By carrying out genomewide scanning, we detected a total of seven significant marker-trait associations ($P < 0.01$) (table 3). All the seven significant SSR loci were indentified for the traits studied, with the R^2, percentage of the total variation explained ranging from 12.0 to 18.0%.

Grain dormancy (GD): We detected one locus with a significant association ($P<0.01$); RM27 on chromosome 2 which had an effect of explaining 13.0% of the total variation.

Alpha amylase activity at seven days after germination: We detected one locus, RM234 on chromosome 7 with significant associations, which explained 13.0% of the total variation.

Alpha amylase activity at 12 days after germination: We detected one locus with significant associations; RM27 on chromosome 2 with an effect of explaining 6.0% of the total variation.

Germination at three days after harvest: The marker locus significantly associated with germination was RM27 on chromosome 2, which explained 14% of the total variation.

Germination at five days after harvest: Two marker loci with significant association were RM27 and RM106 on chromosome 2 with 14% and 10% variation, respectively.

Germination at seven days after harvest: The marker locus significantly associated with germination was RM27 on chromosome 2, which explained 13% of the total variation.

Germination at 10 days after harvest: The marker locus significantly associated with germination was RM27 on chromosome 2, which explained 11.7% of the total variation.

Germination at 12 days after harvest: The marker locus significantly associated with germination was RM27 on chromosome 2, which explained 9.9% of the total variation.

Amylose: The marker locus significantly associated with germination was RM27 on chromosome 2, which explained 13% of the total variation.

Grain length elongation ratio: An association of marker locus RM142 on chromosome 4 was detected with variation explained by it as 9%.

Estimation of population diversity and structure of germplasm accessions could provide pivotal information for resource management, association mapping and crop breeding. This study is based on using landraces of Assam, which is regarded as rich source of diversity. These landraces are old cultivated genotypes that have not been bred through strict breeding principles, therefore can be relatively closer to such primitive forms. Useful genetic variation in the landraces can therefore be used more judiciously in breeding programmes without narrowing down the existing genetic variation in the cultivated germplasms (Vanniarajan *et al.* 2012). The present study identified several markers with few traits. PCA also revealed the importance of those traits in which association was detected.

Seed dormancy, as suggested by germination percentage is an important parameter for a breeding programme. RM27 is associated with germination indicating dormancy is located on chromosome 2. Earlier studies with Assam rice as parent in mapping population failed to detect any QTL on chromosome 2. It suggests that there is possibility of detecting another gene associated with dormancy from chromosome 2. Cai and Morishima (2000) found QTLs for seed dormancy on all chromosomes except for 4 and 10 which supports the present findings. The markers identified by Cai and Morishima (2000) are near to markers identified to be associated with dormancy in the present study. Cai and Morishima (2000) detected a QTL for seed dormancy on chromosome 2 flanked by markers amp1-RZ476. This marker interval also encompasses the marker RM27 in the gramene map, with which the present study also identified a region associated with the same trait.

The present study revealed a significant association of markers RM27 and RM106 on chromosome 2 with germination. Jiang *et al.* (2011) also reported linkage of the marker S02054–S02057 on chromosome 2 with seed germination capability in rice. The linkage map comparison suggests that similarity of the chromosomal region identified with that of Jiang *et al.* (2011) for germination.

Germination at different dates and α-amylase activity is less studied. So no QTL information is available, warranting the needs for development of biparental population with these traits. The α-amylase activity is considered as a reliable indicator of dormancy as it is synthesized *de novo*

during the germination process (Vieira *et al.* 2002; Ullrich *et al.* 2009). However, information on genetic control of α-amylase activity in rice is limited. The present study revealed two markers for α-amylase activity, RM27 and RM234, on chromosome 2 and 7, respectively. Existing reports revealed the presence of two QTLs for α-amylase activity, (qAAA6-1 between RG653-G342, and qAAA6-2, between Waxy-C1496) on chromosome 6 (Cui *et al.* 2002). So the presence of new QTLs on chromosomes 2 and 7, as indicated by this study, cannot be ruled out, which needs further corroboration through biparental mapping.

Amylose content is known to be related with variation at the *waxy* locus on chromosome 6 with several modifiers (He *et al.* 1999; Lanceras *et al.* 2000; Septiningsih *et al.* 2003; Aluko *et al.* 2004; Takeuchi *et al.* 2007). According to the gramene database (http://www.gramene.org) for rice, so far 51 QTLs for amylose content have been identified (Sabouri *et al.* 2012). Apart from a major QTL on chromosome 6, another eight QTLs (*qAC-1, qAC-2, qAC-3, qAC-4, qAC-6a, qAC-6b, qAC-6c* and *qAC-6d*) were mapped for amylose (Sabouri *et al.* 2012). Though the present study could not detect any association between markers on chromosome 6 and with amylose content, the association of RM282 on chromosome 3 with amylose content supports the presence of minor QTL for amylose content on chromosome 3 as reported by Sabouri *et al.* (2012).

Length elongation ratio is an important quality parameter in rice. Ge *et al.* (2005) reported three QTLs for length elongation of cooked rice on chromosome 2, 6 and 11. However, identification of chromosomal regions for this trait on chromosome 4 by marker RM142 indicates novelty of genetic control, warranting its further investigation.

The present study revealed the association of RM27 with traits like dormancy, α-amylase activity and germination. Simple correlation analysis of these traits revealed that these traits were positively correlated with each other and this marker may be useful for simultaneous improvement of these traits.

Conclusion

This study demonstrated the feasibility of conducting association analysis together with germplasm characterization of a local rice collection using SSR markers. The expected high LD of rice inbred lines and cultivars, although facilitating the detection of marker-trait associations makes gene identification more difficult, since LD spans many thousands of base pairs. However, for breeding purposes the correlations detected by association analysis may be sufficient for marker-assisted selection and mining alleles related to important traits in germplasm collections. Knowledge on these loci should make a valuable contribution to rice breeding programmes. The information generated will be useful for selection of good donor lines for the traits under study and the marker-trait associations identified could be used for fine-resolution mapping of the genes/QTLs underlying a trait.

References

Agrama H. A. and Eizenga G. C. 2008 Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. *Euphytica* **160**, 339–355.

Agrama H. A., Eizenga G. C. and Yan W. 2007 Association mapping of yield and its components in rice cultivars. *Mol. Breed.* **19**, 341–356.

Akkaya M. S., Bhagwat A. A. and Cregan P. B. 1992 Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **132**, 1131–1139.

Aluko G., Martinez C., Tohme J., Castano C., Bergman C. and Oard J. H. 2004 QTL mapping of grain quality traits from the interspecific cross "*Oryza sativa* L. and *O. Glaberrima*". *Theor. Appl. Genet.* **108**, 1445–1452.

Bewley J. D. and Black M. 1985 *Seeds: physiology of development and germination,* pp. 329–347. Plenum Press, New York, USA.

Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y. and Buckler E. S. 2007 TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Cagampang G. B., Perez C. M. and Juliano B. O. 1973 A gel consistency test for eating quality of rice. *J. Sci. Food Agric.* **24**, 1589–1594.

Cai H. W. and Morishima H. 2000 Genomic regions affecting seed shattering and seed dormancy in rice. *Theor. Appl. Genet.* **100**, 840–846.

Chen X., Cho Y. G. and McCouch S. R. 2002 Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Genet. Genomics* **268**, 331–343.

Cho Y. G., Ishii T., Temnykh S., Chen X., Lipovich L., McCouch S. R. *et al.* 2000 Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 713–722.

Clifford H. T. and Stephenson W. 1975 *An introduction to numerical classification,* pp. 229. Academic Press, London, UK.

Cui K. H., Peng S. B., Xing Y. Z., Xu C. G., Yu S. B. and Zhang Q. 2002 Molecular dissection of seedling-vigor and associated physiological traits in rice. *Theor. Appl. Genet.* **105**, 745–753.

Donnelly P. 2008 Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731.

Eaves I. A., Barber R. A. and Merriman T. R. 1998 Comparison of linkage disequilibrium in populations from UK and Finland. *Am. J. Hum. Gen.* **63**, suppl. A1212.

Fan R. Z., Spinka C., Jin L. and Jung J. S. 2005 Pedigree linkage disequilibrium mapping of quantitative trait loci. *Eur. J. Hum. Genet.* **13**, 216–231.

Farnir F., Coppieters W., Arranz J. J., Berzi P., Cambisano N., Grisart B. *et al.* 2000 Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* **10**, 220–227.

Garris A. J., McCouch S. R. and Kresovich S. 2003 Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5*locus of rice (*Oryza sativa* L.). *Genetics* **165**, 759–769.

Garris A. J., Tai T. H., Coburn J., Kresovich S. and McCouch S. R. 2005 Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**, 1631–1638.

Ge X. J., Xing Y. Z., Xu C. G. and He Y. Q. 2005 QTL analysis of cooked rice grain elongation, volume expansion, and water absorption using a recombinant inbred population. *Plant Breed.* **124**, 121–126.

Guei R. G., Sanni K. A., Abamu F. J. and Fawole I. 2005 Genetic diversity of rice (*Oryza sativa* L.). *Agron. Afr.* **5**, 17–28.

He P., Li S. G., Qian Q., Ma Y. Q., Li J. Z., Wang W. M. *et al.* 1999 Genetic analysis of rice grain quality. *Theor. Appl. Genet.* **98**, 502–508.

Hill W. G. and Robertson A. 1968 Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231.

Ikehashi H. 1972 Induction and test of dormancy of rice seeds by temperature condition during maturation. *Jpn. J. Breed.* **22**, 209–216.

Jannink J. L., Bink M. C. A. M. and Jansen R. C. 2001 Using complex plant pedigrees to map valuable genes. *Trends Plant Sci.* **6**, 337–342.

Jiang W., Lee J., Jin Y. M., Qiao Y., Piao R., Jang S. M. *et al.* 2011 Identification of QTLs for seed germination capability after various storage periods using two RIL populations in rice. *Mol. Cell* **31**, 385–392.

Jin L., Lu Y., Xiao P., Sun M., Corke H. and Bao J. 2010 Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet.* **121**, 475–487.

Juliano B. O. 1971 A simplified assay for milled rice amylose. *Cereal Sci. Today* **16**, 334–338.

Kraakman A. T. W., Niks R. E., Van den Berg P. M. M. M., Stam P. and van Eeuwijk F. A. 2004 Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* **168**, 435–446.

Kraft T., Hansen M. and Nilsson N. O. 2000 Linkage disequilibrium and fingerprinting in sugar beet. *Theor. Appl. Genet.* **101**, 323–326.

Lanceras J. C., Huang Z. L., Naivikul O., Vanavichit A., Ruanjaichon V. and Tragoonrung S. 2000 Mapping of genes for cooking and eating qualities in Thai Jasmine rice (KDML105). *DNA Res.* **7**, 93–101.

Li X., Yan W., Agrama H., Jia L., Shen X., Jackson A. *et al.* 2011 Mapping QTLs for improving grain yield using the USDA rice mini-core collection. *Planta* **234**, 347–361.

Little R. R., Hilder G. B. and Dawson E. H. 1958 Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem.* **35**, 111–126.

Mackay I. and Powell W. 2007 Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* **12**, 57–63.

Mather K. A., Caicedo A. L., Polato N. R., Olsen K. M., McCouch S., Purugganan M. D. *et al.* 2007 The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* **177**, 2223–2232.

Nakamura S., Sameri M., Pourkheirandish M., Matsumoto T., Yano M., Sato K. *et al.* 2011 Identifying a candidate for the barley grain dormancy QTL SD2. In *Proceedings of Plant and Animal Genomes XIX Conference*, January 15–19, 2011, San Diego, USA.

Nordborg M. and Weigel D. 2008 Next-generation genetics in plants. *Nature* **456**, 720–723.

Nordborg M., Hu T. T., Ishino Y., Jhaveri J., Toomajian C., Zheng H. *et al.* 2005 The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196.

Ogunbodede B. A. 1997 Multivariate analysis of genetic diversity in kenaf. *Hibiscus cannabinus* (L.). *Afr. Crop Sci. J.* **5**, 127–133.

Plaschke J., Ganal M. W. and Röder M. S. 1995 Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* **91**, 1001–1007.

Pritchard J. K., Stephens M. and Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Raji A. A. 2002 Assessment of genetic diversity and heterotic relationships in African improved and local cassava (*Manihot esculenta* Crantz) germplasm. Ph.D. thesis. University of Ibadan, Nigeria.

Rathi S., Baruah A. R., Chowdhury R. K. and Sarma R. N. 2011 QTL Analysis of seed dormancy in indigenous rice of Assam, India. *Cereal Res. Commun.* **39**, 137–146.

Remington D. L., Thornsberry J. M., Matsuoka Y., Wilson L. M., Whitt S. R., Doebley J. *et al.* 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci.* **98**, 11479–11484.

Sabouri A., Rabiei B., Toorchi M., Aharizad S. and Moumeni A. 2012 Mapping quantitative trait loci (QTL) associated with cooking quality in rice (*Oryza sativa* L.). *Aust. J. Crop Sci.* **6**, 808–814.

Sadasivam S. and Manickam A. 1996 Biochemical methods. 2nd edition, pp. 179–186. New Age International, New Delhi, India.

Sanni K. A., Fawole I., Ogunbayo A., Tia D., Somado E. A., Futakuchi K. *et al.* 2012 Multivariate analysis of diversity of landrace rice germplasm. *Crop Sci.* **52**, 494–504.

Septiningsih E. M., Trijatmiko K. R., Moeljopawiro S. and McCouch S. R. 2003 Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor. Appl. Genet.* **107**, 1441–1433.

Shi C., Navabi A. and Yu K. 2011 Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. *BMC Plant Biol.* **11**, 52–62.

Takeuchi Y., Nonoue Y., Ebitani T., Suzuki K., Aoki N., Sato H. *et al.* 2007 QTL detection for eating quality including glossiness, stickness, taste and hardness of cooked rice. *Breed. Sci.* **57**, 231–242.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S. and McCouch S. 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* **11**, 1441–1452.

Ullrich S. E., Lee H., Clancy J. A., del Blanco I. A., Jitkov V. A., Kleinhofs A. *et al.* 2009 Genetic relationships between preharvest sprouting and dormancy in barley. *Euphytica* **168**, 331–345.

Vanniarajan C., Vinod K. K. and Pereira A. 2012 Molecular evaluation of genetic diversity and association studies in rice (*Oryza sativa* L.) *J. Genet.* **91**, 9–19.

Vieira A. R., Vieira M. D. G. G. C., Fraga A. C., Oliviera J. A. and Santos C. D. D. 2002 Action of gibberellic acid (GA_3) on dormancy and activity of α-amylase in rice seeds. *Rev. Bras. de Sementes* **24**, 43–48.

Wan Y., Cohen J. and Guerra R. 1997 A permutation test for the robust sub-pair linkage method. *Ann. Hum. Genet.* **61**, 79–87.

Wang J., McClean P. E., Lee R., Goos R. J. and Helms T. 2008 Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. *Theor. Appl. Genet.* **116**, 777–787.

Wen W., Mei H., Feng F., Yu S., Huang Z., Wu J. *et al.* 2009 Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.) *Theor. Appl. Genet.* **119**, 459–470.

Zhang D. L., Zhang H. L., Wei X. H., Qi Y. W., Wang M. X., Sun J. L. *et al.* 2007 Genetic structure and diversity of *Oryza sativa* L. in Guizhou, China. *Chin. Sci. Bull.* **52**, 343–351.

Zhang P., Li J., Li X., Liu X., Zhao X. and Lu Y. 2011 Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS One* **6**, e27565.

Received 4 January 2013, in final revised form 14 August 2013; accepted 26 August 2013
Published on the Web: 13 March 2014