Comparative genome-wide association studies for plant production traits under drought in diverse rice (*Oryza sativa* L.) lines using SNP and SSR markers

C. Muthukumar, T. Subathra, J. Aiswarya, V. Gayathri and R. Chandra Babu*

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

Rice is the major staple food crop for more than half of the world's population, but its productivity is often reduced by drought, especially when grown under rainfed conditions. Identification of molecular markers associated with plant production traits under drought, especially in the target populations of the environment (TPE) presents an opportunity to improve rainfed rice production using genomics tools. Marker–trait associations were studied using 1168 simple sequence repeat (SSR) markers and 911,153 single nucleotide polymorphisms (SNPs) with 17 diverse rice lines from different geographical regions and hydrological habitats. STRUCTURE analysis discriminated the rice accessions into three subpopulations. Significant genotypic linkage disequilibrium (LD) was found in the rice accessions using SSR markers. A total of 130 and 118 water–trait associations were obtained with SSR and SNP markers respectively, under stress. Comparison of SSR and SNP marker–trait associations revealed 23 consistent associations. Five marker–trait associations with genic SNPs were observed out of 23 associations. These genomic regions may be potential candidates for application in marker-assisted breeding of rice cultivars suitable for water-limited environments.

Keywords: Drought tolerance, linkage disequilibrium, marker–trait association, rice.

RICE is a staple food for nearly half of the world's population[1]. To meet the projected demands due to increasing global population, the world's rice production has to increase by 25% or more by 2030 (ref. 2). Drought is a major abiotic stress limiting rice production and yield stability in rainfed ecosystems[3] and it reduces rice productivity by 13–35% (ref. 4). Nearly 38% of the total rice area is under rainfed lowland and upland conditions, but it contributes only 21% of the total rice production and a large portion of projected increase has to come from this water-limited area[5]. Mapping genomic regions for drought tolerance and their use in marker-assisted breeding is considered to hasten development of high-yielding rice for water-scarce environments. Although traditional QTL mapping is an important tool in QTL tagging[6–9], it is resource-intensive and time-consuming. These limitations can be overcome with the use of association mapping[10]. This is a powerful tool used for high-resolution mapping of loci underlying quantitative traits such as drought tolerance. It takes advantages of accumulated historic recombination events in the natural population and is promising for identifying causative polymorphisms of complex traits[11]. Independent marker development which is distributed throughout the genome with available statistical methods was used to detect population structure[12,13]. The successful association analysis must be able to avoid spurious association from population structure or unequal relatedness within population[14]. Yu *et al.*[15] demonstrated population structure (*Q*) with relative kinship (*K*) matrices for mixed-model association mapping to correct the linkage disequilibrium (LD) caused by population structure and familial relatedness. Thus, marker-based kinship estimates might be appropriate for association mapping approaches and this method has been shown to perform better than other alternatives[15,16].

The extent of LD varies among the populations within the species and also across the genome of the species under study[17], and distribution pattern has the potential to enhance and accelerate genetic resource management activities[18]. Simple sequence repeat (SSR) and simple nucleotide polymorphism (SNP) are the most informative genetic markers useful for genetic diversity studies[19–21] and mapping[22,23]. Agrama *et al.*[24] used the mixed linear model (MLM) method to disclose the associations between 123 SSR markers and yield components in rice. Zhao *et al.*[25] studied 130 rice accessions with 170 SSR markers to identify marker–trait associations by MLM for grain quality. There are numerous reports on genome-wide association studies (GWAS) in rice using SNPs. Zhao *et al.*[26] reported GWAS using 44,100 SNPs with 413 rice accessions. Huang *et al.*[27] used ~3.6 million SNPs with 517 rice landraces. Lakew *et al.*[28] demonstrated association mapping of drought-related traits in barley using SSR and SNP markers. Yang *et al.*[29] reported that

*For correspondence. (e-mail: chandrarc2000@yahoo.com)

SSR markers provided more information on genetic diversity and performed better at clustering all lines into groups than SNPs. Since both the DNA markers are efficient at association mapping and population studies, we used both SSR and SNP markers for association mapping by MLM. Though there are numerous studies on the discovery of significant marker–trait associations through GWAS, only few of these have been validated. Identifying marker–trait associations in target populations of the environment (TPE) will help increase their application in breeding. Thus, the study was conducted with the specific objective of identifying consistent SSR and SNP markers for plant production traits under drought in target rainfed environment using 17 diverse rice accessions.

Materials and methods

Plant materials and field experiment

Seventeen rice accessions out of 20 lines included in the Oryza SNP panel were used in the present study and tested for plant phenology and production traits. N22, Azucena and Moroberekan are drought-tolerant[30], and FR13A is submergence-tolerant and moderately drought-tolerant[31]. Though the number of genotypes used in this study is small, they represent diverse variation in terms of hydrological habitat and response to drought (Table 1). Association analyses of diverse germplasm are perfectly suited for sampling a wide range of alleles with high resolution[32]. Seeds of the rice accessions were received from the International Rice Research Institute (IRRI), Philippines. The geographical origin, ecotype and hydrological habitat/drought response of the accessions are given in Table 1. The rice lines were tested for phenology and plant production traits under drought in rainfed environment. Phenotyping of the rice lines was done under drought during rainfed (northeast monsoon) season (September 2012–January 2013) in experimental fields of Agricultural Research Station (ARS), Tamil Nadu Agricultural University (TNAU), Paramakudi located in TPE. The rice lines were evaluated under irrigated (flooded) condition during May–October 2013 in the Paddy Breeding Station (PBS), TNAU, Coimbatore.

The rice lines were evaluated in 2.0×0.4 sq. m plots under drought in rainfed condition at ARS, Paramakudi in randomized block design with three replications. Seeds were hand-dibbled into dry soil (80 kg/ha) before monsoon. With initial rainfall of 52.5 mm, the seeds germinated and seedlings emerged. However, there was no rain after 40 days after emergence (DAE) and hence irrigation was given once at 60 DAE to save the plants from desiccation. After this, stress developed coinciding with reproductive stage. Stress continued and severe leaf rolling (LR) was observed in almost 90% of the lines. The reproductive stage drought stress was severe and some of the entries dried in this stress period. During this trial, a total of 354 mm rainfall was received during the crop growth. NPK fertilizers were applied at the rate of 50 : 25 : 25 kg/ha.

The rice lines were also evaluated in flooded (irrigated) condition in PBS, Coimbatore. Seeds were sown in raised nursery beds and 26-day-old seedlings were transplanted in the main field with a spacing of 20×10 cm in 1.2×0.2 sq. m plots. The plots were surface irrigated to field capacity at regular intervals. NPK fertilizers were applied at the rate of 100 : 50 : 50 kg/ha. Insect and weed control measures were applied periodically as required in both trials.

In both trials, observations were recorded for days to 50% flowering as the number of days from sowing to flowering in 50% of plants of each accession. At maturity, plant height, number of reproductive tillers and grain yield per plant were recorded from three plants selected random per replication. Spikelet fertility was calculated as the ratio of the number of filled grains to the total number of grains per panicle. All the matured panicles from the selected plants per accession were used for spikelet fertility measurement. All the plants in a plot were harvested to record grain and straw yields per plot and expressed in kg/ha. Above-ground biomass was arrived at by summing up grain and straw yields. Harvest index was estimated as a ratio of grain yield to above-ground biomass.

Statistical analysis

Analysis of variance for each trait in different experiments was done as mixed models using PROC MIXED procedure of SAS V.9.3 (SAS Institute Inc.)[33], where lines were kept as fixed effect and other variables were assigned as random. Broad sense heritability of different traits was estimated within a year for each location. The variance components for calculating broad sense heritability for each trait were calculated using SAS program PROC VARCOMP with REML method. The broad sense heritability was calculated as

$$H = \frac{\sigma_G^2}{\sigma_G^2 + (\sigma_e^2 / r)},$$

where σ_G^2 is the genotypic variance, σ_e^2 the residual variance and r is the number of replications. PROC SUMM was used for summary statistics calculations for the two trials in SAS.

Molecular markers

A set of 911,153 SNPs for the 17 accessions was obtained from OryzaSNP project, an oligomer array-based

Table 1. Geographical origin, ecotype and nature of 17 rice (*Oryza sativa*) accessions used in the study

IRGC accession no.	Designation	Origin	Variety class	Ecotype
117275	Pokkali	India	Landrace	*Indica*
117266	Dular	India	Landrace	*Aus*
117279	Tainung 67	Taiwan	AV	Temperate *Japonica*
117281	Aswina	Bangladesh	Landrace	Deep-water type 3
117273	N22	India	Landrace*	*Aus*
117276	Sadu-Cho	Korea	Landrace	*Indica*
117267	FR13A	India	Landrace**	*Aus*
117264	Azucena	Philippines	Landrace*	Tropical *Japonica*
117265	Dom-Sufid	Iran	Landrace	*Aromatic*
117277	Shan-Huang Zhan-2 (SHZ-2)	China	AV	*Indica*
117270	M202	United States of America	AV	Temperate *Japonica*
117280	Zhenshan 97B	China	AV	*Indica*
117271	Minghui 63	China	AV	*Indica*
117274	Nipponbare	Japan	AV	Temperate *Japonica*
117272	Moroberekan	Guinea	Landrace*	Tropical *Japonica*
117269	Li-jiang-Xin-Tuan-Hei-Gu (LTH)	China	AV	Temperate *Japonica*
117268	IR64-21	Philippines	AV	*Indica*

IRGC, International Rice Genebank Collection; AV, Advanced Variety; *Drought tolerant, **Submergence tolerant and moderately drought tolerant.

re-sequencing effort using Perlegen Sciences technology[34] and used for further studies. A total of 1168 SSR markers selected randomly covering the 12 rice chromosomes were obtained from http://www.gramene.org and used for genotyping the 17 rice accessions and for developing a separate SSR genotypic dataset. Thus, the average number of polymorphic SSR and SNP markers was 42 and 13,333 respectively, for each chromosome.

Genomic DNA was extracted using fresh seedling leaves following a CTAB procedure[35]. The quantity and quality of DNA was assessed in 0.8% agarose gel and concentration was adjusted to 50 ng/µl by comparing DNA standards. Polymerase chain reaction (PCR) amplification was performed in a volume of 20 µl containing each SSR primer of 1 µm (Sigma Aldrich, USA), 100 µM deoxy nucleotide, 1× *Taq* buffer, 0.02 U *Taq* polymerase (Bangalore Genei, India) and 50 ng of template DNA. PCR was performed in a thermal cycler (Master Cycler Gradient, Eppendorf, Germany). After 5 min at 94°C, the PCR involved 36 cycles of amplification, each cycle comprising 1 min at 94°C, 1 min at 57°C (depending on the annealing temperature of the markers), 1 min at 72°C and with a final extension step of 5 min at 72°C. The digested PCR products of the SSR marker were separated by electrophoresis in 3% agarose (BioWhittaker Molecular Applications, Vallensbaek Strand, Denmark)[36] in 0.5× tris-borate EDTA (TBE) buffer. Information on primer sequences and PCR amplification conditions for each set of primers is available at http://www.gramene.org/.

Genetic diversity and population cluster

The number of alleles per locus, major allele frequency, gene diversity, polymorphism information content (PIC) were calculated from the SSR genotypic data using PowerMarker version 3.25 (ref. 37). Nei's distance[38] was calculated and used for the unrooted phylogeny reconstruction using neighbour-joining method as implemented in PowerMarker, and MEGA 5.0 was used to visualize the tree[39]. The software package, STRUCTURE version 2.3.1 was applied to infer historical lineage which used to show clusters (Q) of genotypes[12]. The optimum number of populations (K) was selected with a burn-in period of 100,000 steps followed by 100,000 (Monte Carlo Markov chain replicates). The range of genetic clusters is set from $K = 1$ to $K = 10$. Each value of K was replicated 5 times[40]. On-line available tool 'structure harvester' was used (http://taylor0.biology.ucla.edu) to calculate the final population structure.

Linkage disequilibrium estimation

Using SSR markers, interallelic LD was computed using the software MIDAS (Multiallelic Interallelic Disequilibrium Analysis Software)[41] by accounting the multiallelic SSR markers in calculating r^2 in all the possible interallelic associations. Since they are phase-unknown genotypic data, haplotype frequency was estimated using EM algorithm and it stratifies them as two locus haplotypes into N/N, N/Y and Y/Y. N/Y and Y/Y are used for LD calculation, excluding the rare alleles N/N. Significant P-value was identified using Yates-corrected chi-square test and the LD decay plot was plotted using pairwise r^2 values against distance (kb). Map positions of the markers were based on Gramene marker database. The overall LD (r^2) decay plot with physical distance (bp) among the SSR loci was evaluated by nonlinear regression (NLR).

Hill and Weir[42] model was used for NLR fitting of the expectation of r^2.

$$E(r^2) = \left[\frac{(10+C)}{(2+C)(11+C)}\right]\left[1 + C\frac{(3+C)(12+12C+C^2)}{n(2+C)(11+C)}\right],$$

where C is the population recombination parameter ($C = 4Nc$, N being the effective population size and c the recombination fraction between the loci pair considered), and C was replaced with product of C and genetic distance in cM. McNally *et al.*[34] reported a detailed analysis and discussion on LD using 911,153 SNPs.

Mixed linear model for association analyses

The kinship matrix K was calculated on the basis of the 505 SSR marker loci using the software package SPAGeDi[43]. MLM was applied to study the links between marker loci and phenology while plant production traits under drought stress and non-stress conditions were analysed using the TASSEL (Trait Analysis by aSSociation Evolution and Linkage) version 3.1 software[44], taking into account $Q + K$ with P3D method implemented in TASSEL. The P-value (marker) was used for determining whether a marker (QTL) is associated with the trait. The 911,153 SNP dataset from rice lines was analysed with the GWAS tool GAPIT R package[45]. The PC matrix was generated automatically by setting GAPIT parameters pca.total to 3, and applied for kinship matrix estimations and compressed MLM[45].

Results and discussion

Variation in phenology and plant production traits

In the drought stress trial there was no rainfall after 40 DAE till maturity and the moisture in the field was insufficient to install PVC pipes for water-table measurements. Significant variation was observed among the rice lines for phenology and plant production traits both under drought stress and non-stress conditions. Trait mean, range, standard deviation and broad sense heritability are presented in Table 2. Since there was no rainfall after 40 DAE, plants under irrigated plots also suffered stress due to depletion of water in the borewell. In this study, heritability range for grain yield was 0.93 and ~0.80 for the remaining traits. Jin *et al.*[46] calculated the heritability for association mapping using 416 rice accessions and high heritability (approximately > 0.90 range) was reported. Li *et al.*[47] reported high heritability using 203 rice accessions for 14 traits in association mapping. High heritability for grain yield in drought stress indicates uniformity of drought phenotyping and high stability of the identified association. Moderate heritability (0.56) was observed in previous studies (unpublished) using the same set of rice accessions. This moderate to high heritability under drought stress indicates the suitability of grain yield as an important criterion[48,49].

Mean days to 50% flowering of the rice lines were 102 and 87 in stress and non-stress condition respectively. Mean spikelet fertility across rice lines was 83% and 85% in stress and non-stress condition respectively (Table 2). Grain yield ranged from 558 to 1667 and 44 to 6607 with a mean of 946 and 2874 kg/ha in stress and non-stress condition respectively. The heritability for grain yield was 0.93 and 0.98 in drought stress and non-stress condition respectively.

Genetic diversity using SSR markers

Assessment of the genetic variation and structure of diversity panels of lines represent important information for genetic analyses and identification of quantitative trait loci by means of association mapping[50]. A total of 1168 SSR markers distributed across the rice genome were used to genotype the 17 rice accessions in this study. Among them, 505 were polymorphic and used for further study. These markers detected a total of 1514 alleles among the rice accessions. The allelic richness was 3.12, ranging from 1 to 12 among 17 rice accessions as compared to the range 1 to 11 reported earlier in rice[51]. The number of alleles per marker ranged from 1 to 12 with an average of 3.12 (Table 3). PIC ranged from 0.0500 to 1.000 with an average of 0.4000. Out of the 505 polymorphic markers, 157 (31%) markers were highly informative (PIC ≥ 0.5), 231 markers (46%) were reasonably informative (PIC = 0.25–0.5) and 117 markers (23%) were less informative (PIC ≤ 0.25). The PIC value of 0.40 in this study is higher compared to earlier reports for rice[51,52]. The gene diversity ranged from 0.0571 to 1.0000 with an average of 0.4549 (Table 3). The gene diversity was arrived at using PowerMarker[37]. Thus, the 17 accessions used in this study have wide genetic diversity and are good candidates for GWAS of complex traits such as drought resistance in rice.

Population structure discrimination using SSR markers

Population structure is an important component in association mapping analyses because it can be a source of Type I error in an autogamous species such as rice[53]. A model-based approach STRUCTURE has been used frequently in association mapping studies[12,13]. STRUCTURE analysis discriminated the 17 accessions into three subpopulations, POP1, POP2 and POP3 (Figure 1). The subpopulation POP1 had nine accessions, while POP2 and POP3 had six and two accessions respectively. The accessions in each of the three subpopulations are clustered

Table 2. Trait mean, range, standard deviation (SD) and broad sense heritability (*H*) for phenology and plant production traits of rice lines under drought stress (WS) and non-stress (WW) conditions

Trait	Season	Mean	Range	SD	*H*
Days to 50% flowering	WS	102	91–114	5	0.78
	WW	87	54–122	15	0.92
Plant height (cm)	WS	56	27–71	10	0.81
	WW	96	55–131	18	0.81
Number of tillers	WS	5	4–7	1	0.42
	WW	9	2–14	3	0.75
Number of productive tillers	WS	4	2–6	1	0.82
	WW	9	2–14	3	0.92
Number of grains/panicle	WS	59	36–87	14	0.71
	WW	77	32–141	28	0.86
Number of chaffs/panicle	WS	11	7–18	3	0.91
	WW	15	0–45	12	0.97
Spikelet fertility (%)	WS	83	63–92	6	0.89
	WW	85	64–100	10	0.94
Grain yield (kg/ha)	WS	946	558–1667	277	0.93
	WW	2874	44–6607	1801	0.98
Straw yield (kg/ha)	WS	2686	1383–4767	827	0.79
	WW	6402	323–19532	2391	0.82
Total biomass (kg/ha)	WS	4245	2571–7009	827	0.76
	WW	9263	1347–21777	6264	0.87
Harvest index	WS	0.27	0.08–0.58	0.12	0.82
	WW	0.37	0.02–0.89	0.20	0.99

Table 3. Number of alleles per locus, gene diversity, polymorphism information content (PIC), major allele frequency (MAF) and heterozygosity of 17 rice accessions

	Alleles	Allele/locus	Gene diversity	PIC	MAF	Heterozygosity
Total	1514	3.1216	0.4549	0.4000	0.6454	0.3020
Minimum	1	1	0.0571	0.0500	0.1429	0.0588
Maximum	12	12	1.0000	1.0000	1.0000	1.0000

as follows: six *indica* accessions, two *japonica* and one *aus* type clustered in POP1; four *japonica* with two *aus* types in POP2, and one *aus* and *japonica* in POP3 (Table 3). Thus, this three-group model ($K = 3$) was found to sufficiently explain the genetic structure among the 17 accessions (Figure 1). In the SSR-dataset, pairwise relatedness coefficients between individuals have been measured using SPAGeDi software[43]. STRUCTURE clustered the accessions into three clusters, and both the STRUCTURE and PowerMarker analyses reached similar conclusions regarding population structure among these accessions. The results support each other in this study, with some exception. In the SNP dataset, default kinship and PCA = 3 were used as implemented in GAPIT R package to apply compressed mixed linear model (CMLM) approach for GWAS[45].

A neighbour-joining tree was also constructed based on Nei's genetic distance (Figure 1). It revealed genetic relationships fairly consistent with the STRUCTURE-based membership assignment for most accessions. However, a few rice accessions were displayed as admixtures in different clusters. For instance, cluster 1 with *aus* genotypes had Tainung 67, LTH and Moroberekan, cluster 2 with *indica* improved genotypes had Nipponbare and M202, and cluster 3 with *indica* landraces had Azucena as admixtures.

Linkage disequilibrium using SSR markers

The significant associations between molecular polymorphism and particular phenotypes, as well as the resolving power of LD mapping techniques, depend on knowledge of the LD extent and the rate of decay of LD with physical distance[12]. In this study, LD of haplotypic interalleles among all loci distributed in 12 linkage groups for the accessions was estimated by avoiding Hardy–Weinberg

equilibrium (HWE) assumption using reconstructed haplotypic data as implemented in MIDAS software. A total of 19,008 allelic pairs of the most frequent alleles represented as N/Y and Y/Y were used for LD estimation and 356 allelic pairs were significant with Yates-corrected χ^2 by the level of $P \leq 0.05$ and showed $r^2 \geq 0.1$. The minimum and maximum interallelic r^2 value was 0.10 and 0.81 respectively, with a mean of 0.26. D' and r^2 are most commonly used measures of LD[8,54], but r^2 has more reliable sampling properties than D' in cases with low allele frequencies, especially for self-pollinated species such as rice[55]. In this study, interallelic LD was estimated using MIDAS[41]. Based on the non-linear regression (NLR) curve, it is clear that LD in these rice accessions decays faster as in other studies (Figure 2). A set of rice accessions with diverse origins, LD as r^2, which is also an indication of marker–trait correlations, is the most appropriate LD quantification measure for association mapping[8]. LD decay using SNP-set was extensively discussed for this population by McNally *et al.*[34], who observed a strong population structure of the OryzaSNP set in these genotypes. Extent of LD varies among different genomic regions[56], rice accessions studied[57] and markers used.

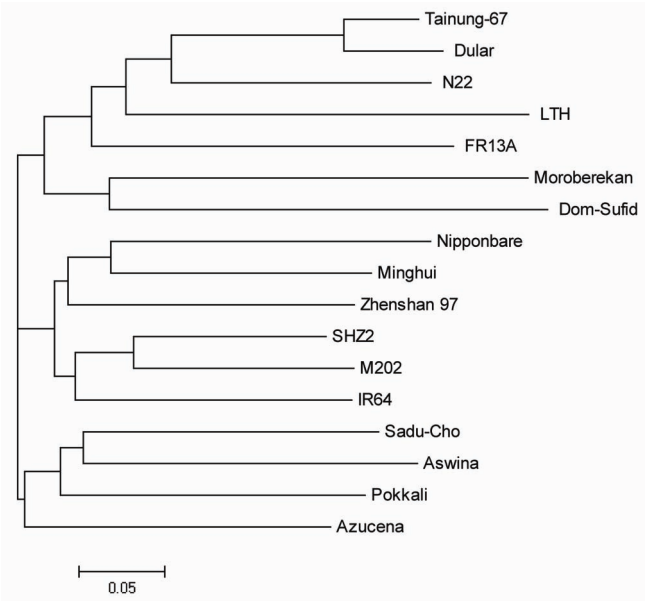

Figure 1. Neighbour-joining tree of 17 rice accessions.

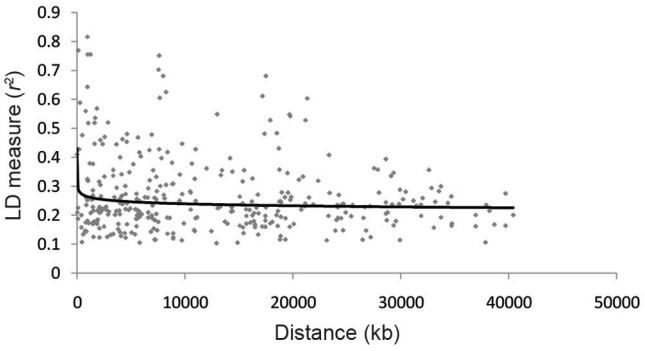

Figure 2. Linkage disequilibrium (LD) decay plot of SSR allele pairs as a function of genetic distance (kb) for rice accessions.

Association mapping by MLM using SSR markers

Association analysis was done using 505 polymorphic SSR markers. $Q + K$ model was used with MLM approach to obtain the marker–trait association. Based on MLM analysis, a total of 130 and 165 marker–trait associations were obtained for stress and non-stress conditions respectively, for 12 traits (data not shown). Association studies using SSR markers showed the significant marker–trait association of plant phenology and production traits in TPE.

GWAS using SNPs

A total of 12 traits with two environment combinations were analysed and population structure was controlled with default principle component analysis (PCA) matrix in GAPIT R package. A total of 911,153 SNPs were used for SNP–trait association analyses. The number of genome-wide significant SNP associations detected on chromosome 1 to 12 was 16, 8, 8, 7, 12, 5, 14, 10, 9, 5, 9, 15 and 8, 12, 10, 8, 1, 10, 6, 9, 10, 8, 15, 13 for stress and non-stress conditions respectively.

Comparison of association mapping using SSR versus SNP

The analysis revealed that 117 and 110 SNP–trait associations, and 130 and 165 SSR–trait associations were recorded respectively, under stress and non-stress conditions. There are 23 SNPs–trait associations for different traits and these genomic regions are close to the 19 SSR marker–traits which are associated for different traits under stress condition (Table 4).

SNP and SSR marker-trait association

The marker–trait associations detected using SNP were compared with those identified using SSR with their physical positions, and the regions detected in both SSR and SNP marker analysis were deduced. A total of 23 marker–trait associations were common across markers and these markers were previously reported by several workers using traditional QTL mapping studies for various traits. For example, on chromosome 4 the markers TBGU205562-RM3866 were associated with grain yield and total biomass in SNP and SSR analysis respectively. This region was reported earlier for a number of grains[58] and leaf chlorophyll content (SPAD)[59], drought responsive

Table 4. Comparative association between SSR and SNP markers for phenology and plant production traits under drought stress in rice diverse lines

Chr	SNP	Position	Traits	SSR	Position	Traits
Chr2	TBG111995	29580273	Spikelet fertility (%)	RM6933	29331976	No. of chaffs/panicle, grain yield (kg/ha)
Chr3	TBG171695	34232085	Plant height (cm)	RM7000	33589253	Days to 50% flowering
Chr4	TBGU205562	23912824	Grain yield (kg/ha)	RM3866	23333714	Total biomass (kg/ha), straw yield (kg/ha)
	TBG211863	28956555	No. of grains/panicle	RM470	28248941	Spikelet fertility (%)
	TBGU211863	28956555	No. of grains/panicle	RM470, RM317	28248941	Spikelet fertility (%)
Chr5	TBGU232483	5597188	Grain yield (kg/ha)	RM5997	6798666	Straw yield (kg/ha)
	TBGU232325	5476814	Spikelet fertility (%)	RM4915	3991545	No. of grains/panicle, no. of chaffs/panicle
Chr6	TBGU283441	9598156	No. of tillers	RM527	9874150	Grain yield (kg/ha)
	TBGU297986	22163663	No. of tillers	RM6298	23353565	Straw yield (kg/ha)
	TBG275192	4736178	Straw yield (kg/ha)	RM6176, RM8258	4623191	Straw yield (kg/ha)
Chr7	TBGU325164	18865445	Days to 50% flowering	RM5583	19307928	Spikelet fertility (%)
	TBG334303	27456467	No. of chaffs/panicle	RM5720	28616499	No. of chaffs/panicle, no. productive tillers
	TBGU334303	27456467	No. of chaffs/panicle	RM5720	28616499	No. of chaffs/panicle, no. productive tillers
	TBG325285	19020761	Plant height (cm)	RM5420	19304494	Single plant yield (g)
	TBGU325285	19020761	Plant height (cm)	RM5420	19304494	Single plant yield (g)
Chr8	TBG363179	26622302	Days to 50% flowering	RM6075	27764004	No. of productive tillers, no. of chaffs/panicle
	TBG356974	20279623	Spikelet fertility (%)	RM342	19953040	Grain yield (kg/ha), plant height (cm)
	TBG356980	20279808	Spikelet fertility (%)	RM342	19953040	Grain yield (kg/ha), plant height (cm)
	TBGI356980	20279808	Spikelet fertility (%)	RM342	19953040	Grain yield (kg/ha), plant height (cm)
Chr10	TBGU414867	21938337	No. of grains/panicle	RM5352	20672962	No. of chaffs/panicle, no. of productive tillers,
Chr11	TBG432787	17273065	Days to 50% flowering	RM6272	16400644	Total biomass (kg/ha), no. of grains/panicle
	TBGU424331	6090858	Grain yield (kg/ha)	RM3625	6591206	Days to 50% flowering
Chr12	TBGU464298	6691638	No. of productive tillers	RM7119	6694741	Spikelet fertility (%)

Chr, Chromosome.

AAP7 gene and *dwarf11* gene. The region on chromosome4_TBG211863_RM470 was associated with number of grains and spikelet fertility in SNP and SSR trait associations respectively, and DRO2 QTL[60] was located in this region. Chromosome5_TBGU232483_RM5997 was found to be associated for grain yield in SNP analysis, and straw yield, number of grains and number of chaffs in SSR analysis. The database qtaro.abr.affrc.go.jp/ also reported this region for grain yield, and transposon protein was found in this region. On chromosome 6, the SNP TBG275192 linked to straw yield was located in the same region as RM6176 and RM8258 linked to spikelet fertility. Importantly, Chromosome8_TBG363179_RM6075 was associated for days to 50% flowering in SNP–trait association and number of productive tillers and number of chaffs for SSR–trait association; the same region comprises DREB1G and heat shock factor class *B2b* genes. On chromosome10, TBGU414867 was linked to number of grains, and SSR RM5352 was liked with number of chaffs and number of productive tillers. This region was earlier reported for osmotic adjustment (OA)[61]. Chromosome11_TBG432787_RM6272 was associated with days to flowering in SNP–trait association and total biomass and number of chaffs in SSR–trait association. Moncada *et al.*[62] reported this region for the number of grains, grain weight and grain yield; the region contains *RALFL33* gene. Chromosome11_TBGU424331_RM3625 was linked to grain yield with SNP–trait association, and days to 50% flowering with SSR–trait association, this region was reported for yield per plant by Moncada *et al.*[62]. In SNP–marker trait association, 5 out of 23 associations were observed as polymorphic genic regions; the SNP_TBG111995 associated with spikelet fertility on chromosome 2 harbors the *ATPase* gene; the SNPs TBG356974 and TBGI356980 on chromosome 8 associated with spikelet fertility have MAP kinase gene; the SNP_TBGU414867 on chromosome 10 linked with number of grains has cytoplasmic peptidoglycan synthase gene and the SNP_TBG432787 on chromosome 11 associated with days to 50% flowering harbors rapid alkalinization factor gene. These genes, viz. ATPase[63], MAP kinase[64] and rapid alkalinization factor[65] were classified as drought responsive genes.

To sum up, 23 marker–trait associations were found consistent across SSR and SNP marker analysis. These regions were also earlier reported to be linked to drought resistance traits in rice using conventional QTL mapping studies. For instance, comparative GWAS analysis with SNP and SSR markers showed associations for grain yield and yield component traits under drought in TPE with SNP and SSR markers on chromosomes 4, 5, 6 and 11 within 2.27, 4.7, 0.44 and 3.7 cM regions respectively. The marker–trait association on chromosome 6 with SNP

and SSR marker RM6176 was 1.29 cM close to RM2434-RM6773, which was fine mapped in IR62266/Norungan RI lines for grain yield under drought stress in the same laboratory (unpublished). These regions are reported for *OsDREB1c* gene, and root dry weight by conventional QTL mapping studies, which confirms that these marker–trait associated regions are useful candidates for developing resilient rice cultivars suitable for water-limited environments using marker-assisted breeding strategies.

1. Kondo, M., Aguilar, A. and Abe J Morita, S., Anatomy of nodal roots in tropical upland and lowland rice varieties. *Plant Prod. Sci*., 2000, **3**, 437–445.
2. Seck, P. A., Diagne, A., Mohanty, S. and Wopereis, M. C. S., Crops that feed the world 7: Rice. *Food Security*, 2012, **4**, 7–24.
3. Suji, K. K. *et al*., Mapping QTLs for plant phenology and production traits using indica rice (*Oryza sativa* L.) lines adapted to rainfed environment. *Mol. Biotechnol*., 2012, **52**, 151–160.
4. Jongdee, B., The importance of leaf water potential and osmotic adjustment on growth and grain yield of rice (*Oryza sativa* L.) genotypes under water deficit conditions. Ph D thesis, The University of Queensland, Brisbane, Australia, 1998, p. 238.
5. Khush, G. S., Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol*., 1997, **35**, 25–34.
6. Hansen, M., Kraft, T., Ganestam, S., Sall, T. and Nilsson, N. O., Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genet. Res*., 2001, **77**, 61–66.
7. Stella, A. and Boettcher, P. J., Optimal designs for linkage disequilibrium mapping and candidate gene association tests in livestock populations. *Genetics*, 2004, **166**, 341–350.
8. Gupta, P. K., Rustgi, S. and Kulwal, P. L., Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol*., 2005, **57**, 461–485.
9. Stich, B. *et al.*, Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol. Breed*., 2006, **17**, 217–226.
10. Ross-Ibarra, J., Morrell, P. L. and Gaut, B. S., Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 8641–8648.
11. Stich, B. *et al*., Comparison of mixed-model approaches for association mapping. *Genetics*, 2008, **1754**, 1745–1754.
12. Pritchard, J. K., Stephens, M. and Donnelly, P., Inference of population structure using multilocus genotype data. *Genetics*, 2000, **155**, 945–959.
13. Pritchard, J. K., Stephens, M., Rosenberg, N. A. and Donnelly, P., Association mapping in structured populations. *Am. J. Hum. Genet*., 2000, **67**, 170–181.
14. Forabosco, P., Falchi, M. and Devoto, M., Statistical tools for linkage analysis and genetic association studies. *Expert Rev. Mol. Diagn*., 2005, **5**, 781–796.
15. Yu, J. M. *et al.*, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet*., 2006, **38**, 203–208.
16. Zhao, K. *et al.*, An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet*., 2007, **3**, e4.
17. Oraguzie, N., Wilcox, P., Rikkerink, H. and De Silva, H., Linkage disequilibrium. In *Association Mapping in Plants*, Springer, New York, NY, 2007, pp. 11–39.
18. Haussmann, B. I. G. *et al.*, Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theor. Appl. Genet*., 2004, **109**, 1005–1016.
19. Russell, J. R. *et al.*, A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. *Mol. Breed*., 2000, **6**, 553–568.
20. Sjakste, T., Rashal, I. and Roder, M., Inheritance of microsatellite alleles in pedigrees of Latvian barley varieties and related European ancestors. *Theor. Appl. Genet*., 2003, **106**, 539–549.
21. Hamza, S., Ben, H. W., Rebai, A. and Harrabi, M., SSR-based genetic diversity assessment among Tunisian winter barley and relationship with morphological traits. *Euphytica*, 2004, **135**, 107–118.
22. Varshney, R. K. *et al.*, A high density barley microsatellite consensus map with 775 SSR loci. *Theor. Appl. Genet*., 2007, **114**, 1091–1103.
23. Varshney, R. K. *et al.*, Identification and validation of a core set of informative genic SSR and SNP markers assaying functional diversity in barley. *Mol. Breed*., 2008, **22**, 1–13.
24. Agrama, H. A., Eizenga, G. C. and Yan, W., Association mapping of yield and its components in rice cultivars. *Mol. Breed.*, 2007, **19**, 341–356.
25. Zhao, W. G., Chung, J. W., Kwon, S. W., Lee, J. H., Ma, K. H. and Park, Y. J., Association analysis of physicochemical traits on eating quality in rice (*Oryza sativa* L.). *Euphytica*, 2013, **191**, 9–21.
26. Zhao, K. *et al.*, Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.*, 2011, **2**, 467; doi:10.1038/ncomns1467.
27. Huang, X. *et al*., Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genet*., 2010, **42**, 961–967.
28. Lakew, B., Henry, R. J., Ceccarelli, S., Grando, S., Eglinton, J. and Baum, M., Genetic analysis and phenotypic associations for drought tolerance in *Hordeum spontaneum* introgression lines using SSR and SNP markers. *Euphytica*, 2013, **189**, 9–29.
29. Yang, X. *et al*., Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica*, 2011, **139**, 1045–1054.
30. McNally, K. L. *et al*., Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol*., 2006, **141**, 26–31.
31. Godwin, I. D. *et al*., RAPD polymorphisms among variant and phenotypically normal rice (*Oryza sativa* var. *indica*) somaclonal progenies. *Plant Cell Rep*., 1997, **16**, 320–324.
32. Flint-Garcia, S. A., Thornsberry, J. M. and Buckler, E. S., Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol*., 2003, **54**, 357–374.
33. SAS Institute Inc., SAS® 9.3, SAS Institute Inc, Cary, NC, 2012.
34. McNally, K. L. *et al.*, Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 12273–12278.
35. Gawel, N. J. and Jarret, R. L., A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Mol. Biol. Rep*., 1991, **9**, 262–266.
36. Sambrook, J. and Russell, D. W., Detection of DNA in agarose gels. *Cold Spring Harbor Protocols*, 2006, doi: 10.1101/pdb.prot4022.
37. Liu, K. and Muse, S. V., PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 2005, **21**, 2128–2129.
38. Nei, M., Tajima, F. and Tateno, Y., Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol*., 1983, **19**, 153–170.
39. Tamura, K., Dudley, J., Nei, M. and Kumar, S., MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol*., 2007, **24**, 1596–1599.
40. Hubisz, M., Falush, D., Stephens, M. and Pritchard, J., Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour*., 2009, **9**, 1322–1332.
41. Gaunt, T. R., Rodriguez, S., Zapata, C. and Day, I. N. M., MIDAS: software for analysis and visualization of interallelic disequilibrium between multiallelic markers. *BMC Bioinform*., 2006, **7**, 227.
42. Hill, W. G. and Weir, B. S., Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol*., 1988, **33**, 54–78.

43. Hardy, O. J. and Vekemans, X., SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes*, 2002, **2**, 618–620.
44. Bradbury, P. J. *et al.*, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 2007, **23**, 2633–2635.
45. Lipka, A. E. *et al.*, GAPIT: genome association and prediction integrated tool. *Bioinformatics*, 2012, **28**, 2397–2399.
46. Jin, L. *et al*., Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet*., 2010, **121**, 475–487.
47. Li, X. *et al.*, Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). *PLoS ONE*, 2012, **7**(1), e29350.
48. Kumar, R., Venuprasad, R. and Atlin, G. N., Genetic analysis of rainfed lowland rice drought tolerance under naturally-occurring stress in eastern India: heritability and QTL effects. *Field Crops Res*., 2007, **103**, 42–52.
49. Venuprasad, R., Lafitte, H. R. and Atlin, G. N., Response to direct selection for grain yield under drought stress in rice. *Crop Sci*., 2007, **47**, 285–293.
50. Mezmouk, S., Dubreuil, P., Bosio, M., De´cousset, L., Charcosset, A., Praud, S. and Mangin, B., Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theor. Appl. Genet*., 2011, **122**, 1149–1160.
51. Cho, Y. G. *et al.*, Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet*., 2000, **100**, 713–722.
52. Jain, S., Jain, R. and McCouch, S., Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. *Theor. Appl. Genet*., 2004, **109**, 965–977.
53. Agrama, H. A. and Eizenga, G. C., Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. *Euphytica*, 2008, **160**, 339–355.
54. Gaut, B. S. and Long, A. D., The lowdown on linkage disequilibrium. *Plant Cell*, 2003, **15**, 1502.
55. Abdurakhmonov, I. Y. and Abdukarimov, A., Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int. J. Plant Genomics*, 2008, 574927.
56. Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S. and Purugganan, M. D., The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics*, 2007, **177**, 2223–2232.
57. Swamy, B. P. M. *et al*., Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. *BMC Genomics*, 2011, **12**, 319.
58. Xiao, J., Li, J., Yuan, L. and Tanksley, S. D., Identification of QTL affecting traits of agronomic importance in a recombinant inbred population derived from a sub-specific rice cross. *Theor. Appl. Genet*., 1996, **92**, 230–244.
59. Abdelkhalik, A. F., Shishido, R., Nomura, K. and Ikehashi, H., QTL-based analysis of leaf senescence in an indica/japonica hybrid in rice (*Oryza sativa* L.). *Theor. Appl. Genet*., 2005, **110**, 1226–1235.
60. Uga, Y. *et al*., Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. *Nature Genet*., 2013, **45**, 1097–1102.
61. Robin, S. *et al.*, Mapping osmotic adjustment in an advanced back-cross inbred population of rice. *Theor. Appl. Genet.*, 2003, **107**, 1288–1296.
62. Moncada, P. *et al*., Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. *Theor. Appl. Genet*., 2001, **102**, 41–52.
63. Huda, K. M. K. *et al*., OsACA6, a P-type IIB Ca2 + ATPase promotes salinity and drought stress tolerance in tobacco by ROS scavenging and enhancing the expression of stress-responsive genes. *Plant J*., 2013, **76**, 997–1015.
64. Shen, H. *et al*., OsWRKY30 is activated by MAP kinases to confer drought tolerance in rice. *Plant Mol. Biol*., 2012, **80**, 241–253.
65. Atkinson, N. J., Lilley, C. J. and Urwin, P. E., Identification of genes involved in the response of *Arabidopsis* to simultaneous biotic and abiotic stresses. *Plant Physiol*., 2013, **162**, 2028–2041.

ACKNOWLEDGEMENT. The research was supported by the Department of Biotechnology, Government of India.

Received 23 November 2014; revised accepted 11 March 2015