Mol Breeding (2007) 19:341–356
DOI 10.1007/s11032-006-9066-6

Association mapping of yield and its components in rice cultivars

H. A. Agrama · G. C. Eizenga · W. Yan

Received: 23 April 2006 / Accepted: 29 October 2006 / Published online: 11 February 2007

Abstract To make advances in rice breeding it is important to understand the relatedness and ancestry of introduced rice accessions, and identify SSR markers associated with agronomically important phenotypic traits, for example yield. Ninety-two rice germplasm accessions recently introduced from seven geographic regions of Africa, Asia, and Latin America, and eleven US cultivars, included as checks, were evaluated for yield and kernel characteristics, and genotyped with 123 SSR markers. The SSR markers were highly polymorphic across all accessions. Population structure analysis identified eight main clusters for the accessions which corresponded to the major geographic regions, indicating agreement between genetic and predefined populations. Linkage disequilibrium (LD) patterns and distributions are of fundamental importance for genome-wide mapping association. LD between linked markers decreased with distance and with a substantial drop in LD decay values between 20 and 30 cM, suggesting it should be possible to achieve resolution down to the 25 cM level. For the 103 cultivars, the complex traits yield, kernel width, kernel length, kernel width/length ratio, and 1000-kernel weight, were estimated by analysis of variety trial data. The mixed linear model method was used to disclose marker-trait associations. Many of the associated markers were located in regions where QTL had previously been identified. In conclusion, association mapping in rice is a viable alternative to QTL mapping based on crosses between different lines.

Keywords Linkage disequilibrium · Unified mixed-model method · Population structure · Kinship coefficient · Relatedness

Abbreviations

SSR	Simple sequence repeat
QTL	Quantitative trait loci
cM	CentiMorgan

H. A. Agrama (✉)
University of Arkansas Rice Research and Extension Center, Stuttgart, AR 72160, USA
e-mail: hagrama@uark.edu

G. C. Eizenga · W. Yan
United States Department of Agriculture – Agricultural Research Service, Dale Bumpers National Rice Research Center, Stuttgart, AR 72160-1090, USA

Introduction

Rice (*Oryza sativa* L.) genetic mapping often involves the development, genotyping, and phenotyping of doubled haploid (Li et al. 2003;

Hittalmani et al. 2002), recombinant inbred (Ge et al. 2005; Guo et al. 2005), or advanced backcross (Li et al. 2004; Jing et al. 2005) lines derived from an F_1 cross between different cultivars. For these mapping populations there is extreme disequilibrium between linked loci (Mather et al. 2004) and there should be no population structure, because of the random sampling, or disequilibrium between non-linked loci favoring the detection and approximate mapping of quantitative trait loci (QTL). In contrast, genetic mapping in humans relies on analysis of genotypic and phenotypic data sampled non-randomly from existing populations of complex but often unknown structure. In these populations, linkage disequilibrium (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. 2002). 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 *xa5* locus approached 0.1 after only 100 kb (Garris et al. 2003) and in *Arabidopsis* the pattern of polymorphism LD decayed rapidly, within 50 kb (Nordborg et al. 2005), in contrast with one human population for which the LD extended over 3 cM (Eaves et al. 1998). Most studies of species variation begin by sampling from populations defined on the basis of culture or geography and might not reflect underlying genetic relationships (Rosenberg et al. 2002; Foster and Sharp 2002). If a whole-genome scan is to be undertaken, trait mapping by allele association requires high marker density (Lander and Schork 1994; Jorde 1995; Jorde 2000; Risch 2000). SSRs are particularly useful for studying the population structure and demographic history of domesticated species such as rice (Garris et al. 2005), and are being extensively used to genotype rice germplasm collections (Yang et al. 1994; McCouch et al. 1997; Ishii and McCouch 2000; Ishii et al. 2001; Ni et al. 2002; Lu et al. 2005). Use of SSR markers to interpret population structure results in much greater resolution than use of other types of marker, 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. Although polymorphic SSRs are excellent molecular markers, because of their multi-allelism and the resulting high informativeness (Weber and May 1989), they may not be frequent enough for association studies (Ching et al. 2002). Size homoplasy of SSR alleles and allele reversion could also be a problem in some populations (Estoup et al. 1995; Viard et al. 1998). Rice has the smallest genome size (estimated as approx. 430 Mb) among cereal crops, which makes it most manageable at the whole-genome level.

The potential of LD and regression methods to identify and characterize loci/genes associated with different complex traits in true breeding lines has been demonstrated in barley and maize (Kraakman et al. 2004; Wilson et al. 2004). In contrast with calculating associations between pairs of loci or genes by the traditional LD technique, or using a single marker-trait regression scheme, a multivariate approach, known as discriminant analysis, has been used for whole-genome scans of microsatellite markers associated with economic traits in unrelated inbred lines of rice (Zhang et al. 2005). Of particular interest to rice breeders is the possibility of using existing germplasm resources for gene and allele discovery on the basis of association mapping strategies (Kruglyak 1999; Jorde 2000; Farnir et al. 2000). Understanding population structure is important to avoid identifying spurious associations between phenotype and genotype in association mapping (Pritchard and Rosenberg 1999; Pritchard et al. 2000; Pritchard and Donnelly 2001).

Detection of marker-trait associations in breeding germplasm has potential advantages over classical linkage analysis and QTL mapping (Jannink and Walsh 2002). For example, broader genetic variation in a more representative genetic background can be included in the analysis, LD mapping may achieve higher resolution, and

multi-trait phenotypic data stored in databases can be linked to marker characterizations of the cultivars involved. The last advantage is especially important when evaluation of the trait is time-consuming and expensive, as it is for yield, adaptability, and stability (Kraakman et al. 2004). This approach may, however, be limited by the structure of the population, eventually leading to spurious, inappropriate levels of LD for QTL mapping, and insufficient phenotypic or genetic diversity available within the gene pool (Condon and Smith 2005). The number of markers needed for genome-wide LD scanning depends on the level of LD, however; to be effective, LD-mapping requires a marker density compatible with the distances across which LD extends in the population of interest. Association mapping could complement and enhance previous QTL information for marker-assisted selection in wheat (Breseghello and Sorrells, 2006), barley (Kraakman et al. 2006), and maize (Yu and Buckler 2006). Within different sets of barley germplasm there was frequent disequilibrium among non-linked SSR markers, suggesting that association mapping without consideration of population structure would result in a high rate of false positive Type I errors (Mather et al. 2004). Use of aunified mixed-model approach to account for multiple levels of relatedness simultaneously, as detected by use of genetic markers, has improved control of both type I and type II error rates (Yu et al. 2006).

Multi-allelic markers, for example SSRs, have been used to characterize population structure in maize (Flint-Garcia et al. 2005; Remington et al. 2001) and rice (Garris et al. 2003); and LD-based associations in wheat, (Kruger et al. 2004), barley (Maccaferri et al. 2005), and *Lolium* (Skøt et al. 2005). The objective of this research was to use a large collection of blast (*Magnaporthe oryzae* B. Couch) resistant rice accessions to determine the utility of population structure analysis, linkage disequilibrium (LD), and association mapping of yield traits in evaluating rice germplasm accessions.

Materials and methods

Included in this study were 92 blast-resistant rice germplasm accessions from seven different regions of Africa, Asia, and Latin America (Table 1). The 91 accessions were recently added to the USDA-ARS National Plant Germplasm System. The eleven US cultivars were Bengal (PI561535), Cocodrie (PI606331), Drew (PI596758), Katy (PI527707), Kaybonnet (PI583278), LaGrue (PI568891), Lemont (PI475833), Newbonnet (PI474580), Wells (Moldenhauer et al. 2000), Saber (PI633624), and Zenith (CIor7787). TeQing (PI536047) is an older cultivar from Guangdong, China. Further information on the accessions and cultivars is available from the USDA-ARS National Plant Germplasm System (http://www.ars-grin.gov/npgs/).

The test was seeded in April 2000 and April 2001 in a split-block design (group as main block) with four replications at the University of Arkansas Rice Research and Extension Center in Stuttgart, AR, USA. The plots were 4.5 m long and six rows wide with 20 cm spacing between rows. At maturity, 3.6 m from the center, two rows in each plot were harvested for grain yield (GY). Other aspects of field management and measures of grain quality were as described by Dilday et al. (2001) and Yan et al. (2003). Seed (30 g) of each accession was dehulled to determine the kernel length (KL), kernel width (KW) and kernel length/width ratio (LWR) with the GrainCheck 312 video image system (FOSS Food Technology). Thousand kernel weight (TKW) was determined by weighing 1000 rice kernels.

Genomic DNA was extracted from leaf tissue by the methods described in Eizenga et al. (2006). One-hundred and twenty-three microsatellite markers located on the twelve chromosomes were selected from the core set developed and mapped by McCouch et al. (2002). The forward primers were labeled with FAM, TET, NED, or HEX fluorescent dyes and the reverse primers were unlabeled. DNA amplification was performed using an MJ Research PTC-100 96 Plus thermal cycler. PCRs were conducted in a 10-μL reaction mix containing 37.5 ng template DNA, 1× PCR buffer, 0.025 units *Taq* DNA polymerase (Qiagen, Valencia, CA, USA), 0.2 mmol dNTPs, and 0.8 pmol forward and reverse primers. Information on primer sequences and PCR amplification conditions for each set of primers are available at

Table 1 Rice germplasm accessions included in this study and their country of origin[a]

No.	Accession name	Country of origin[b]
1	Bhujon Kolpo	Bangladesh
2	Bogra	Bangladesh
3	Khoia	Bangladesh
4	Iac 47	Brazil
5	IRGA409	Brazil
6	Guang-6ai-4	China
7	02428	China-CD
8	Chunzhi No. 11	China-CD
9	Fu No83	China-CD
10	Kechengnuo No. 4	China-CD
11	Sheng 10	China-CD
12	Shufeng 117	China-CD
13	Shufeng 122	China-CD
14	Tie 90-1	China-CD
15	Tiejing No. 4	China-CD
16	Zhang 32	China-CD
17	Xiangzaoxian No. 1	China-HN
18	71198	China-HZ
19	Aijiaonante	China-HZ
20	Zanuo No. 1	China-HZ
21	Zhongyu No. 1	China-HZ
22	Zhongyu No. 6	China-HZ
23	Zhongzao No. 1	China-HZ
24	460	China-JT
25	2410	China-JT
26	4484	China-JT
27	4593	China-JT
28	4594	China-JT
29	4596	China-JT
30	4597	China-JT
31	4607	China-JT
32	4611	China-JT
33	4612	China-JT
34	4632	China-JT
35	4633	China-JT
36	4642	China-JT
37	4641(1)	China-JT
38	GP-2	China-JT
39	Gui 99	China-JT
40	R 147	China-JT
41	R 312	China-JT
42	Dian No. 1	China-KM
43	Egyptian Jasmine	Egypt
44	GZ-1368-5-4	Egypt
45	GZ-5578-2-1-2	Egypt
46	GZ-5594-23-1-2	Egypt
47	GZ-5830-48-2-2	Egypt
48	Ad 9246	Ivory Coast
49	Fkr 19 (Tox728-8)	Ivory Coast
50	Fkr 48	Ivory Coast
51	32 Xan Sc	Ivory Coast
52	Ita 406	Ivory Coast
53	Ita 416	Ivory Coast
54	Let 3137	Ivory Coast
55	S 992-F4-2-5-1-B	Ivory Coast
56	Tnau 7893	Ivory Coast
57	Tox 3093-35-2-3-3-1	Ivory Coast
58	Tox 3211-49-1-1-3-2	Ivory Coast
59	Tox 3241-21-2-2-3-2	Ivory Coast
60	Tox 3241-22-3-3-3	Ivory Coast
61	Tox 3241-31-2-1-3-1	Ivory Coast
62	Tox 3441-123-2-3-2-2-2	Ivory Coast
63	Tox 3553-34-3-2-3-2-2	Ivory Coast
64	Tox 3706-60-3-3-3	Ivory Coast
65	Tox 3706-6-3-3-2	Ivory Coast
66	Tox 3716-4-3-2-2-2-2	Ivory Coast
67	Tox 3717-25-3-1-3	Ivory Coast
68	Tox 3717-25-3-3-1	Ivory Coast
69	Tox 3717-25-3-3-2	Ivory Coast
70	Tox 3717-76-2-2-3	Ivory Coast
71	Tox 3717-81-1-1-3	Ivory Coast
72	Tox 3770-17-2-2-1	Ivory Coast
73	Tox 3771-144-2-1-1	Ivory Coast
74	Tox 3772-38-2-2-3	Ivory Coast
75	Tox 3772-40-3-2-2	Ivory Coast
76	Tox 3772-94-1-1-1	Ivory Coast
77	Tox 3779-51-2-2-2	Ivory Coast
78	Tox 3867-19-1-1-3-1-1-1	Ivory Coast
79	Tox 3869-34-1-3-1-1-3-3	Ivory Coast
80	Tox 3872-61-3-3-3-2-1	Ivory Coast
81	Tox 3894-41-2-3-1	Ivory Coast
82	Tox 4136-38-2	Ivory Coast
83	Tox 4251-313-3	Ivory Coast
84	Tox 3749-71-1-1-3-2-2	Ivory Coast
85	Wab450-24-3-2-P18-hb	Ivory Coast
86	Wab450-I-B-P-62-hb	Ivory Coast
87	Pyongyang 23	Korea, N.
88	IR56450-28-2-2 1	Philippines
89	Nj70507 17578	Philippines
90	RP2199-16-2-2-1	Philippines
91	S972B-22-1-3-1-1	Philippines
92	Bengal	USA
93	Cocodrie	USA
94	Drew	USA
95	Katy	USA
96	Kaybonnet	USA
97	LaGrue	USA
98	Lemont	USA
99	Newbonnet	USA
100	Saber	USA
101	Te Qing	China/USA
102	Wells	USA
103	Zenith	USA

[a] Adapted from Eizenga et al. (2006)

[b] Source of Chinese germplasm: CD - Chengdu, HN - Hunan, HZ - Hangzhou, KM - Kunming, JT - Joshua Tao

http://www.gramene.org/ (verified 3 July 2006). PCR products were separated by size using an ABI 3700 DNA analyzer (Applied Biosystems (ABI), Foster City, CA, USA). SSR fragment sizing was performed with Gene Scan software (ABI) using the local Southern method and default analysis settings after which alleles were identified with Genotyper software (ABI) and binned manually. For most markers fractional numbers were rounded to the nearest integer and alleles differing by 1 bp were declared different. SSR data, obtained from genotyping US cultivars with the same ABI 3700 DNA analyzer (Lu et al. 2005), was included for comparison. The map position of most of the SSR loci was inferred from McCouch et al. (2002).

The model-based software Structure (Pritchard et al. 2000) was used to infer population structure using a burn-in of 10,000, run length of 100,000, and a model allowing for admixture and correlated allele frequencies. Ten independent runs yielded consistent results. A model-based clustering algorithm was applied that identified subgroups with distinctive allele frequencies. This procedure, implemented in computer structure, places individuals into K clusters, where K is chosen in advance but can be varied for independent runs of the algorithm. The degree of admixture, alpha, was inferred from the data. When alpha is close to zero, most individuals are essentially from one population or another whereas alpha >1 means that most individuals are admixed (Falush et al. 2003). The range of possible tested Ks was from 2 to 10.

Distance-based analysis of the accessions using Euclidean inferred ancestry for each accession and the key for identifying the accessions was shown in the neighbor-joining tree using the unweighted pair group method using arithmetic averages (UPGMA). This is a hierarchical algorithm for clustering accessions into similar groups. The output of this clustering procedure is a dendrogram or tree with distance along the horizontal (top) axis and the accessions lines listed vertically down the side. Levels of genetic variation within and among populations identified by the cluster analysis were estimated from allelic frequencies using analysis of molecular variance, AMOVA (Weir and Cockreham 1984; Weir 1996). The software Arlequin 3.0 (Excoffier et al. 2005) performs the AMOVA procedure using SSR and standard multi-locus frequency data. Wright's F-statistics are a hierarchical series of fixation indices where F_{IS} represents the deviation from Hardy–Weinberg expectation within populations (approximately equal to the mean F across populations), F_{ST} measures the fixation of different alleles in different populations, and F_{IT} measures deviations from Hardy–Weinberg expectation across the population system as a whole.

Decay of LD with distance in cM between SSR loci within the genome was evaluated using PowerMarker 3.23 (Liu and Muse 2004). The LD decay was calculated using the statistical coefficient of determination (R^2) which is a measurement of correlation between a pair of variables (Hill and Robertson 1968). All association tests were run with the mixed linear model (MLM) method as described by Yu et al. (2006) in TASSEL 1.9.4 (http://www.maizegenetics.net/), a recently developed unified mixed-model method simultaneously taking into account multiple levels of both gross level population structure (Q) and finer scale relative kinship (K). The population structure matrix (Q) was identified by running Structure at $K = 7$. The relative kinship matrix (K matrix) was obtained by running SPAGeDi (Hardy and Vekemans 2002). Output from SPAGeDi was formatted to a text file readable by TASSEL. The P-value determines whether a QTL is associated with the marker and the R^2-marker evaluates the magnitude of the QTL effects (personal communication, Zhiwu Zhang, Cornell University).

Results

The 103 accessions had a mean grain yield (GY) of 7,301 kg ha^{-1}, a mean kernel length (KL) of 6.7 mm, a mean kernel width (KW) of 2.3 mm, a mean kernel length/width ratio (LWR) of 2.9, and a mean 1000-kernel weight (TKW) of 21.2 g (Table 2). Correlations of TKW with GY and with KW, and of KL with LWR were very significant ($P < 0.0001$). Correlation of KL with TKW was significant ($P < 0.001$) as was the

Table 2 Descriptive statistics for yield (GY), dehulled kernel length (KL), kernel width (KW), kernel length/width ratio (LWR), and 1000-kernel weight (TKW)

	GY (kg ha^{-1})	KL (mm)	KW (mm)	LWR	TKW (g)
Average	7301	6.7	2.3	2.9	21.2
Range	8822	2.4	1.1	3.0	11.2
Minimum	2897	5.4	1.8	1.0	16.2
Maximum	11719	7.8	2.9	4.0	27.4
Std. Dev.	1950	0.6	0.2	0.5	2.1
Correlations					
GY					
KL	0.019				
KW	0.186*	–0.481**			
LWR	–0.162*	0.846***	–0.748***		
TKW	0.348***	0.395**	0.496***	–0.064	

*$P < 0.01$, **$P < 0.001$, ***$P < 0.0001$

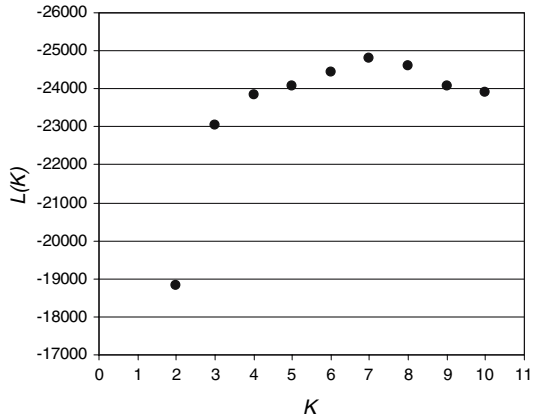


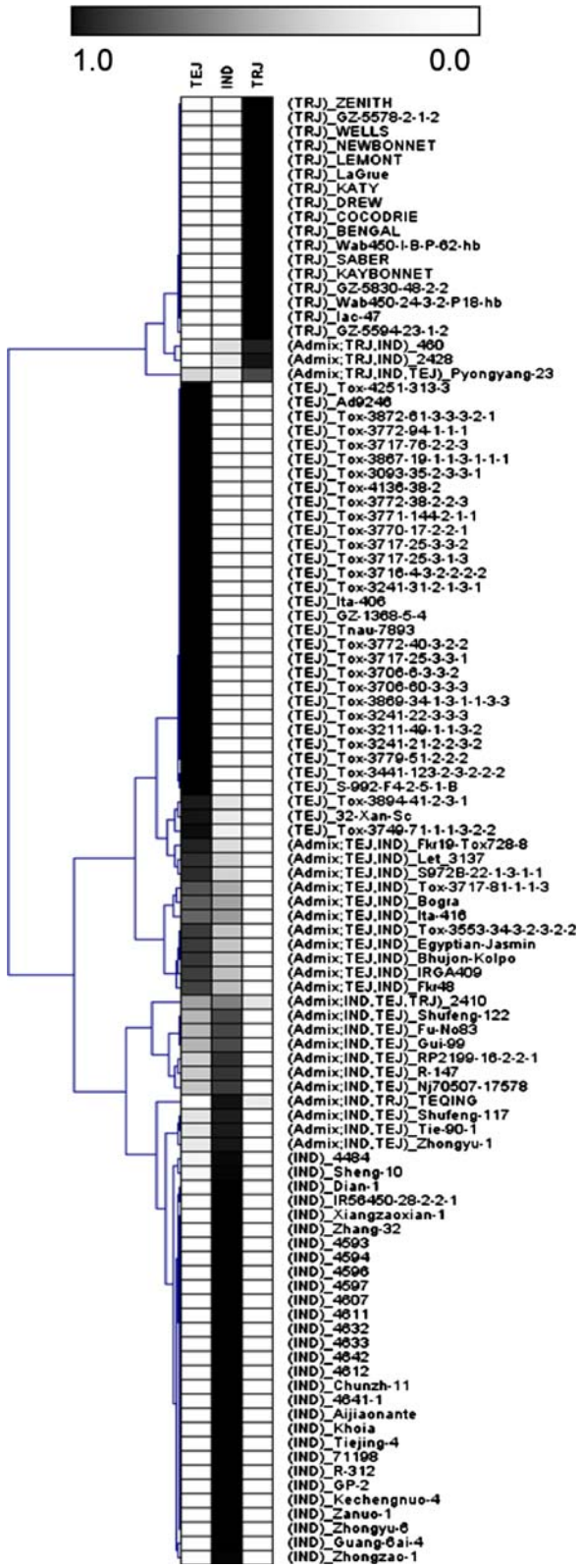
Fig. 1 Log probability of data, $L(K)$, averaged over the replicates

correlation of GY with KW ($P < 0.01$). There was a positive correlation between GY and KL but it was not significant. Significant negative correlations of LWR with KW ($P < 0.0001$) and GY ($P < 0.01$) were found. This indicates that increasing KW and KL results in increase kernel weight and, to a lesser extent, improves GY. The correlations also suggest that increased KW has more effect on GY than increased KL. The negative correlation between KW and KL ($P < 0.001$) indicates that as KW increases KL decreases.

All 123 SSR markers were polymorphic and produced a total of 1009 alleles among the 103 accessions assayed. The average number of alleles per locus was 8.2, ranging from 2 (RM338 on chromosome 3) to 21 (RM206 on chromosome 11). The average genetic diversity over all SSR loci was 0.635, ranging from 0.115 (RM512 on chromosome 12) to 0.881 (RM304 on chromosome 10).

Analysis of genetic distance and population structure provided evidence of significant population structure in these rice accessions and identified the highest likelihood value at $K = 7$ for all ten replicates (Fig. 1). Analysis of these data identified the major substructure groups when the number of populations was set at three, however, which was consistent with clustering based on genetic distance. In this worldwide sample, accessions from the same predefined population nearly always shared similar membership coefficients in inferred clusters (Fig. 2). At $K = 3$, most accessions were classified into one of the three groups, which corresponded to the traditional rice sub-species *indica* (29 accessions), *temperate japonica* (32), and *tropical japonica* (17) separated by relatively large genetic distance. In addition to these 78 accessions that were clearly assigned to a single population, where more than 85% of their inferred ancestry is derived from one of the model-based populations, 25 accessions in the sample were categorized as having a mixed ancestry, defined as an ''admix'' (Fig. 2). Although most of these were identified as admixture between *indica* and *temperate japonica* groups, other admixture combinations were also present.

Divergence among the accessions was found using all measures of population structure. The Structure model-based analysis for several theoretical population sizes (defined in this study as ancestral backgrounds) with the highest posterior probability was for a model with seven different backgrounds (Fig. 3, center). For these ancestral backgrounds, a burn of 10,000 runs followed by data collection on 100,000 runs seemed to be sufficient, giving reasonably consistent values of *ln prop data* $(X|K)$ across replicates. At $K = 7$, the backgrounds largely corresponded to major geographic regions with one background being US accessions. Several accessions, for example ''Tox3717-25-3-3-2'' and ''Shufeng-122'', had par-


1.0
0.0
TEJ
IND
TRJ
(TRJ)_ZENITH
(TRJ)_GZ-5578-2-1-2
(TRJ)_WELLS
(TRJ)_NEWBONNET
(TRJ)_LEMONT
(TRJ)_LaGrue
(TRJ)_KATY
(TRJ)_DREW
(TRJ)_COCODRIE
(TRJ)_BENGAL
(TRJ)_Wab450-I-B-P-62-hb
(TRJ)_SABER
(TRJ)_KAYBONNET
(TRJ)_GZ-5830-48-2-2
(TRJ)_Wab450-24-3-2-P18-hb
(TRJ)_Jac-47
(TRJ)_GZ-5594-23-1-2
(Admix;TRJ,IND)_460
(Admix;TRJ,IND)_2428
(Admix;TRJ,IND,TEJ)_Pyongyang-23
(TEJ)_Tox-4251-313-3
(TEJ)_Ad9246
(TEJ)_Tox-3872-61-3-3-3-2-1
(TEJ)_Tox-3772-94-1-1-1
(TEJ)_Tox-3717-76-2-2-3
(TEJ)_Tox-3867-19-1-1-3-1-1-1
(TEJ)_Tox-3093-35-2-3-3-1
(TEJ)_Tox-4136-38-2
(TEJ)_Tox-3772-38-2-2-3
(TEJ)_Tox-3771-144-2-1-1
(TEJ)_Tox-3770-17-2-2-1
(TEJ)_Tox-3717-25-3-3-2
(TEJ)_Tox-3717-25-3-1-3
(TEJ)_Tox-3716-4-3-2-2-2-2
(TEJ)_Tox-3241-31-2-1-3-1
(TEJ)_Ita-406
(TEJ)_GZ-1368-5-4
(TEJ)_Tnau-7893
(TEJ)_Tox-3772-40-3-2-2
(TEJ)_Tox-3717-25-3-3-1
(TEJ)_Tox-3706-6-3-3-2
(TEJ)_Tox-3706-60-3-3-3
(TEJ)_Tox-3869-34-1-3-1-1-3-3
(TEJ)_Tox-3241-22-3-3-3
(TEJ)_Tox-3211-49-1-1-3-2
(TEJ)_Tox-3241-21-2-2-3-2
(TEJ)_Tox-3779-51-2-2-2
(TEJ)_Tox-3441-123-2-3-2-2-2
(TEJ)_S-992-F4-2-5-1-B
(TEJ)_Tox-3894-41-2-3-1
(TEJ)_32-Xan-Sc
(TEJ)_Tox-3749-71-1-1-3-2-2
(Admix;TEJ,IND)_Fkr19-Tox728-8
(Admix;TEJ,IND)_Let_3137
(Admix;TEJ,IND)_S972B-22-1-3-1-1
(Admix;TEJ,IND)_Tox-3717-81-1-1-3
(Admix;TEJ,IND)_Bogra
(Admix;TEJ,IND)_Ita-416
(Admix;TEJ,IND)_Tox-3553-34-3-2-3-2-2
(Admix;TEJ,IND)_Egyptian-Jasmin
(Admix;TEJ,IND)_Bhujon-Kolpo
(Admix;TEJ,IND)_IRGA409
(Admix;TEJ,IND)_Fkr48
(Admix;IND,TEJ,TRJ)_2410
(Admix;IND,TEJ)_Shufeng-122
(Admix;IND,TEJ)_Fu-No83
(Admix;IND,TEJ)_Gui-99
(Admix;IND,TEJ)_RP2199-16-2-2-1
(Admix;IND,TEJ)_R-147
(Admix;IND,TEJ)_Nj70507-17578
(Admix;IND,TRJ)_TEQING
(Admix;IND,TEJ)_Shufeng-117
(Admix;IND,TEJ)_Tie-90-1
(Admix;IND,TEJ)_Zhongyu-1
(IND)_4484
(IND)_Sheng-10
(IND)_Dian-1
(IND)_IR56450-28-2-2-1
(IND)_Xiangzaoxian-1
(IND)_Zhang-32
(IND)_4593
(IND)_4594
(IND)_4596
(IND)_4597
(IND)_4607
(IND)_4611
(IND)_4632
(IND)_4633
(IND)_4642
(IND)_4612
(IND)_Chunzh-11
(IND)_4641-1
(IND)_Aijiaonante
(IND)_Khoia
(IND)_Tiejing-4
(IND)_71198
(IND)_R-312
(IND)_GP-2
(IND)_Kechengnuo-4
(IND)_Zanuo-1
(IND)_Zhongyu-6
(IND)_Guang-6ai-4
(IND)_Zhongzao-1

◀**Fig. 2** Estimated population structure (K = 3) for the 103 accessions as rice subgroups *tropical japonica* (TRJ), *temperate japonica* (TEJ), and *indica* (IND). Each accession is represented by row, which is partitioned into K colored segments according to the individual's estimated membership fractions in each of the K clusters

tial membership in multiple backgrounds. At K = 7, the cluster that included tropical japonica populations split into two clusters, of which one includes all US genotypes (Fig. 3).

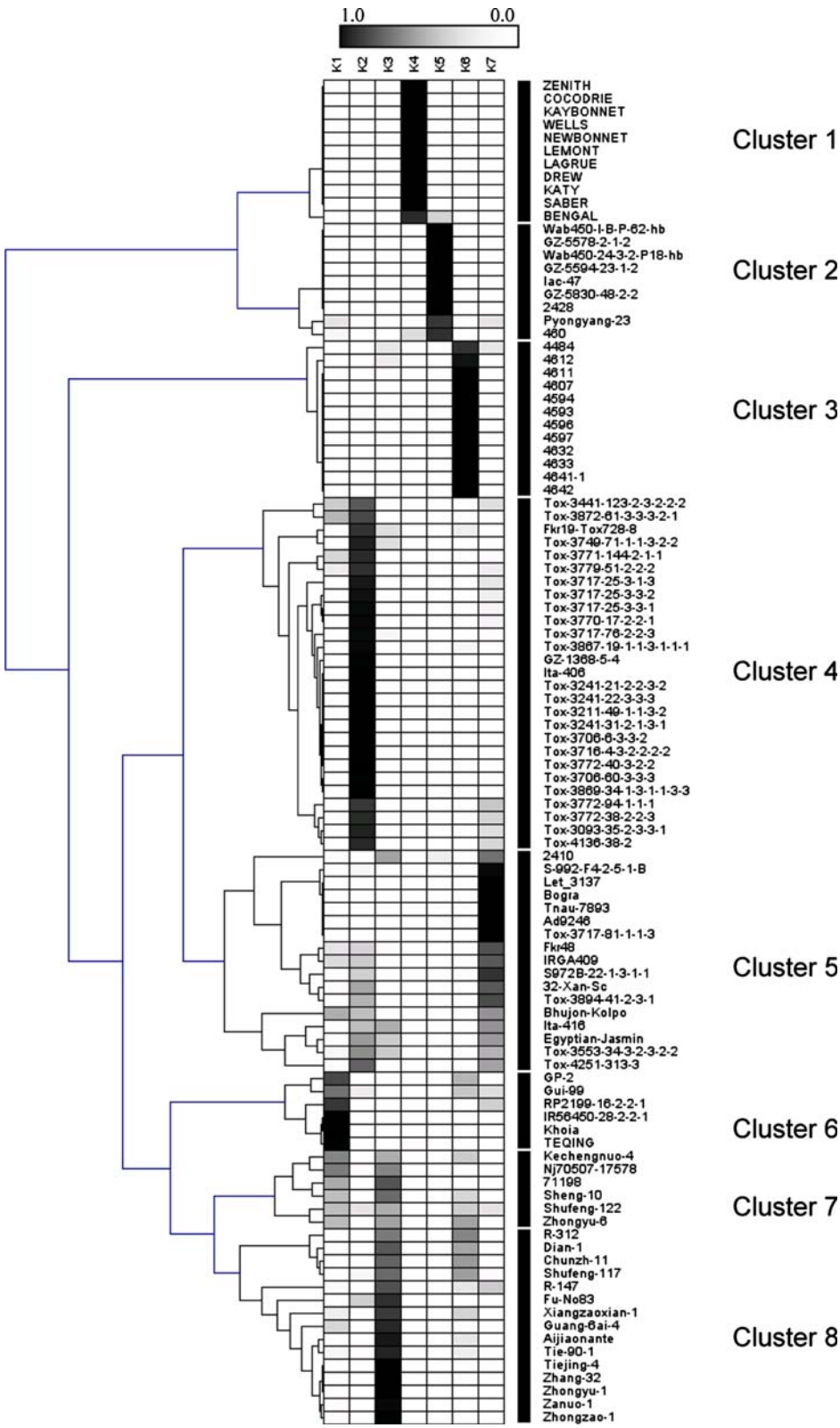
Distance-based analysis of the 103 accessions with Euclidean distance detected eight major clusters (Fig. 3, left and right sides). The accessions were classified into eight clusters by the UPGMA and algorithm neighbor joining tree based on the genetic similarity matrix. These clusters usually agreed with the origin of the accessions and the clustering previously defined by Eizenga et al. (2006). All US accessions were grouped in cluster 1; cluster 3 contained Chinese lines from J. Tao, cluster 4 contained the breeding lines obtained from the Africa Rice Center (WARDA), Bouake, Ivory Coast, cluster 7 mainly included Chinese lines from Chengdu, and cluster 8 contained the Chinese lines from Hangzhou (Fig. 3). Clusters 2, 5, and 6 had more than one ancestral background, defined as an admixture, and these clusters included accessions from Bangladesh, Brazil, Egypt, North Korea, the Philippines, and China.

The distribution of molecular genetic variation among and within the eight clusters of accessions was estimated by AMOVA, which revealed 41% of total variation was among the clusters whereas 51% of the variation was within the clusters (Table 3). Calculation of Wright's F statistics for all SSR loci revealed F_{IS} was 0.66, suggesting there was deviation from Hardy–Weinberg expectation for molecular variation within the clusters, F_{IT} was 0.71, signifying non-equilibrium conditions across clusters and a deficiency of heterozygous SSR loci, and F_{ST} was 0.312 indicating 31.2% of the total genetic variation was among the clusters.

Determination of F_{ST} for the polymorphic SSR loci across all accessions revealed F_{ST} ranged from 0.06 for RM120 on chromosome 11 to 0.94 for RM124 on chromosome 4 with an average of 0.381, indicating 38.1% of the total variation in allele frequency of the 103 accessions was because of genetic differences among clusters. Determination of the pair-wise F_{ST} values between the eight clusters (Table 4) indicated that genetic differentiation among clusters was highest for the combination of clusters 1 and 3 (F_{ST} = 0.719). The F_{ST} values for cluster 1 paired with clusters 4, 5, 6, 7, and 8 ranged from 0.649 to 0.533 and F_{ST} = 0.381 for clusters 1 and 2. These values confirm that the US accessions were quite different from most other accessions except those in cluster 2. The combination of clusters 7 and 8 had the lowest F_{ST} value (0.110), because of the shared ancestry of these clusters. The average F_{ST} value for all eight cluster combinations was 0.391.

The distribution of data points in the plot of LD (R^2) decay against distance (cM) within the twelve chromosomes (Fig. 4) showed that LD was not a simple monotonic function of the distance between markers. LD was very common for distances <30 cM and occasionally, LD occurred between SSR loci that were further apart. The R^2 between unlinked markers on different chromosomes was mainly <0.4. Some unlinked SSR markers associated with blast resistance were identified using these accessions by Eizenga et al. (2006) and had higher R^2. For example, R^2 values were 1.0 (RM206 on chromosome 11, OSM89 on chromosome 12), 0.96 (RM208 on chromosome 2, RM5963 on chromosome 6), and 0.91 (RM3431-ch 6, RM144-ch 11). In contrast with a-priori expectation, some marker pairs that were close together on the Gramene map were not correlated across the genotypes and so were in linkage equilibrium.

Association analysis (Table 5) identified marker-trait associations (P < 0.05) for all the traits evaluated, these included GY, KW, and LWR, associated five markers each, KL, with six markers, and TKW, with four markers. A total of 25 marker-trait associations were identified with 21 different SSR markers. Four markers were associated with two traits: RM85 on chromosome 3 was associated with KW and TKW, RM122 on chromosome 5 was associated with KL and TKW, RM459 also on chromosome 5 was associated with KL and TKW, and RM228 on chromosome 10 was associated with GY and LWR. Seventeen of the 25 associations were in regions where QTL associated with the given trait had previously


1.0
0.0
K1
K2
K3
K4
K5
K6
K7
ZENITH
COCODRIE
KAYBONNET
WELLS
NEWBONNET
LEMONT
LAGRUE
DREW
KATY
SABER
BENGAL
Wab450-I-B-P-62-hb
GZ-5578-2-1-2
Wab450-24-3-2-P18-hb
GZ-5594-23-1-2
Iac-47
GZ-5830-48-2-2
2428
Pyongyang-23
460
4484
4612
4611
4607
4594
4593
4596
4597
4632
4633
4641-1
4642
Tox-3441-123-2-3-2-2-2
Tox-3872-61-3-3-3-2-1
Fkr19-Tox728-8
Tox-3749-71-1-1-3-2-2
Tox-3771-144-2-1-1
Tox-3779-51-2-2-2
Tox-3717-25-3-1-3
Tox-3717-25-3-3-2
Tox-3717-25-3-3-1
Tox-3770-17-2-2-1
Tox-3717-76-2-2-3
Tox-3867-19-1-1-3-1-1-1
GZ-1368-5-4
Ita-406
Tox-3241-21-2-2-3-2
Tox-3241-22-3-3-3
Tox-3211-49-1-1-3-2
Tox-3241-31-2-1-3-1
Tox-3706-6-3-3-2
Tox-3716-4-3-2-2-2-2
Tox-3772-40-3-2-2
Tox-3706-60-3-3-3
Tox-3869-34-1-3-1-1-3-3
Tox-3772-94-1-1-1
Tox-3772-38-2-2-3
Tox-3093-35-2-3-3-1
Tox-4136-38-2
2410
S-992-F4-2-5-1-B
Let_3137
Bogra
Tnau-7893
Ad9246
Tox-3717-81-1-1-3
Fkr48
IRGA409
S972B-22-1-3-1-1
32-Xan-Sc
Tox-3894-41-2-3-1
Bhujon-Kolpo
Ita-416
Egyptian-Jasmin
Tox-3553-34-3-2-3-2-2
Tox-4251-313-3
GP-2
Gui-99
RP2199-16-2-2-1
IR56450-28-2-2-1
Khoia
TEQING
Kechengnuo-4
Nj70507-17578
71198
Sheng-10
Shufeng-122
Zhongyu-6
R-312
Dian-1
Chunzh-11
Shufeng-117
R-147
Fu-No83
Xiangzaoxian-1
Guang-6ai-4
Aijiaonante
Tie-90-1
Tiejing-4
Zhang-32
Zhongyu-1
Zanuo-1
Zhongzao-1
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8

◀**Fig. 3** Population structure and distance-based analysis of the 103 rice accessions using Euclidean-inferred ancestry for each accession. Initially, the accessions were divided into clusters based on UPGMA and neighbor joining (*left and right sides*). Next, the accessions were divided into seven ancestral backgrounds defined as *K* (*center*) based on analysis in Structure. Accessions were assigned to a single background or to two or more backgrounds if the genotype indicated the accession was admixed with membership in two or more different backgrounds and estimated on a scale from 1.0 (accession is from one *K* only) to 0.0 (accession is not from this *K*)

been identified (http://www.gramene.org/). Of the 25 marker-trait associations, seven were identified as explaining 20% or more of the total variation (R^2, Table 5) for GY (RM261, RM228), KL (RM284), LWR (RM7, RM228), and TKW (RM440, RM122). Only RM284 associated with KL was not in the region of a previously identified QTL for the associated trait.

Discussion

The genetic structure of rice has previously been documented (Glaszmann 1987; Parsons et al. 1999; Ni et al. 2002; Garris et al. 2005) but this analysis uses accessions recently introduced into the USA that have blast resistance (Eizenga et al. 2006). By use of Structure software with $K = 3$, the *O. sativa* rice accessions were significantly differentiated into three subgroups, *temperate japonica*, *tropical japonica*, and *indica*. The analysis revealed that several accessions had partial ancestry in more than one backgrounds. These accessions probably had a complex breeding history involving intercrossing and introgression between germplasm from diverse backgrounds, overlaid with strong selection pressure for agronomic and quality characteristics (Mather et al. 2004). Population structure analysis identified eight main clusters for the accessions that corresponded to major geographic regions. General agreement between the genetic and predefined clusters suggests that knowledge of the ancestral background can facilitate choices of parental lines in rice-breeding programs (Rosenberg et al. 2002). Although both rice-breeding efforts and domestication have had large effects on structuring the diversity of rice, the independent population histories of the groups have also shaped the gene pools (Garris et al. 2005).

Table 4 Pair-wise F_{ST} values between eight clusters as identified using Euclidean distance in Fig. 3

Population	1	2	3	4	5	6	7	8
1	–							
2	0.381	–						
3	0.719	0.601	–					
4	0.609	0.485	0.496	–				
5	0.561	0.418	0.392	0.169	–			
6	0.649	0.467	0.479	0.362	0.198	–		
7	0.589	0.412	0.360	0.291	0.168	0.142	–	
8	0.533	0.400	0.301	0.291	0.157	0.208	0.110	–

Values of F_{ST} were high when the eight clusters were considered, thus identifying large differences between the accessions. It has been demonstrated that markers with higher F_{ST} values have greater resolving power and produce more consistent genetic distance estimates (Watkins et al. 2003). The significant F_{ST} among the clusters suggests a real difference between these clusters, and heterosis might be observed for crosses between the accessions made to improve yield (N'Goran et al. 2000).

Linkage disequilibrium (LD) studies have now been conducted for more than a dozen plant systems, both at the individual gene level and at the level of whole genome. In individual species, these studies included:

Table 3 Analysis of molecular variance (AMOVA) for the eight clusters of rice accessions identified in Fig. 3

Fixation indices: $F_{IS} = 0.663$, $F_{ST} = 0.312$, $F_{IT} = 0.710$

Source of variation	d.f.	Sum of squares	Variance components	Percentage variation	*P*-value
Among clusters	7	1193.29	6.14	41.37	<0.001
Among accessions within clusters	95	1707.25	8.32	50.71	<0.001
Within accessions	103	136.50	1.32	7.92	<0.001
Total	205	3037.04	15.79		

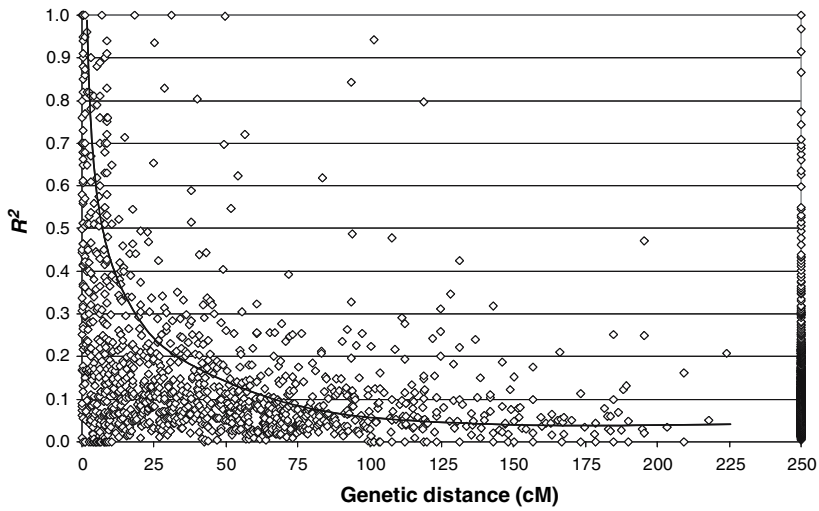


Fig. 4 Scatterplot of the LD (R^2) of marker pairs as a function of the intermarker distance in cM. A genetic distance of 250 cM was chosen to represent unlinked loci on different chromosomes. LD analysis was performed on the entire population (103 accessions)

(i) estimation of the extent of LD in different plant genomes or in different parts of the genome of an individual species,
(ii) measure of nucleotide diversity/haplotype structure,
(iii) assessment of the effect of selection/domestication,
(iv) identification of marker-trait associations, Gupta et al. (2005).

Table 5 Association (R^2) of SSR markers with yield (GY), dehulled kernel length (KL) and width (KW), length/width ratio (LWR), and 1000 dehulled kernels weight (TKW), as described in Table 2

Trait	SSR Marker[a]	Chromosome no.	Position (cM)	*P*	R^2 [b]
GY	RM416	3	191.6[c]	0.0515	0.1244
GY	RM261	4	35.4[c]	0.0016	0.2032
GY	RM447	8	124.6	0.0370	0.1950
GY	RM271	10	59.4[c]	0.0111	0.1909
GY	RM228	10	96.3[c]	0.0118	0.2542
KL	RM259	1	54.2[c]	0.0237	0.1016
KL	RM16	3	131.5[c]	0.0317	0.1155
KL	RM122	5	0.0[c]	0.0167	0.1226
KL	RM284	8	83.7	0.0197	0.3258
KL	RM202	11	54.0	0.0105	0.1012
KL	RM287	11	68.6	0.0491	0.1994
KW	RM468	3	202.3	0.0060	0.1351
KW	RM85	3	231.0	0.0103	0.1451
KW	RM459	5	93.6[c]	0.0041	0.1305
KW	RM248	7	116.6	0.0106	0.1867
KW	RM152	8	9.4[c]	0.0022	0.1318
LWR	RM109	2	0.0	0.0009	0.1287
LWR	RM7	3	64.0[c]	0.0171	0.2359
LWR	RM245	9	112.3	0.0006	0.1748
LWR	RM228	10	96.3[c]	0.0203	0.3213
LWR	RM147	10	99.8[c]	0.0139	0.1330
TKW	RM449	1	78.4[c]	0.0118	0.2199
TKW	RM85	3	231.0[c]	0.0429	0.1914
TKW	RM122	5	0.0[c]	0.0039	0.2661
TKW	RM459	5	93.6[c]	0.0509	0.1750

[a] Only SSR markers with a significant marker-trait association are reported ($P < 0.05$)

[b] R^2 indicates the percentage of the total variation explained

[c] QTL for the given trait previously reported in the same region (http://www.gramene.org/)

The LD between unlinked loci, for example RM208 and RM224 on chromosomes 2 and 11, and RM3431 and RM144 on chromosomes 6 and 11, was occasionally significant. These markers are linked with three blast-resistance (*Pi*-) genes located on these chromosomes (Fjellstrom et al. 2004). LD was high among unlinked SSR markers around *Pi* that were associated with blast resistance in these accessions. Our results suggest that mapping strategies exploiting LD around the region of *Pi*-loci may be particularly effective in rice, as it was in exploiting the haplotype diversity and LD surrounding the *xa5* locus (Garris et al. 2003). Further research will be conducted to describe the diversity and decay of linkage disequilibrium in this region of the rice genome, to characterize the extent of LD in resistant accessions, to determine if it is possible to reduce the number of candidate genes, and to analyze haplotype diversity in the context of population structure to determine the distribution of the resistance allele among ecotypes and to make predictions about the allelic diversity underlying the *Pi* phenotype. In contrast with the situation in the human, genome-wide LD mapping of which may require a marker density two times of magnitude higher than that required for conventional linkage mapping, the available 150 microsatellites could be sufficient for first-pass LD screening in rice genotypes (Kruglyak 1999). The corollary of this observation, however, is that the mapping resolution to be gained from LD is likely to be limited in these populations also. In this work we still observed a substantial drop in LD values between 20 and 30 cM, suggesting it should, nevertheless, be possible to achieve resolution down to the 25 cM level. The same observation on LD at larger distances was found in *Arabidopsis* (Nordborg et al. 2002) and barley (Kraakman et al. 2004). In sugar beet lines, however, LD was <3 cM (Kraft et al. 2000), in *Lolium perenne*, LD was <3.4 cM (Skøt et al. 2005), and in maize LD diminished over a distance of 2000 bp (Remington et al. 2001). Significant LD between pairs of unlinked markers was observed (McRae et al. 2002; Skøt et al. 2005), emphasizing the advantage of both linkage and LD testing. Many factors affect LD (Ardlie et al. 2002), but the most probable cause of the high level of LD in rice is that it is a self-pollinated species. Selection can also increase LD, for instance, by a hitchhiking effect in which the alleles at flanking loci of a locus under selection can be rapidly swept to high frequency or fixation (Kraakman et al. 2004). Further analyses will be required to evaluate the benefit of LD mapping at the sub-cM level in these populations.

For LD between two multi-allelic loci, r^2 (statistical coefficient of determination) and D′ (absolute ratio of deviation of haplotype frequencies from disequilibrium compared with its maximum value) are the most widely used measure of LD for each pair of alleles, or even for overall LD between all the alleles at two loci (Gupta et al. 2005). Whereas D′ measures only recombination differences, r^2 summarizes recombination and mutation history. Also, r^2 is indicative of how markers might be correlated with the QTL of interest, so r^2 is often preferred for association studies (Abdallah et al. 2003). In the sets of cultivated rice examined here, linkage disequilibrium as r^2 is present on a scale that could be useful for association mapping. Genome-scale association mapping should be possible, if adequate methods are implemented to control the effects of population structure. Model-based analyses of population structure, for example those conducted here, may be helpful for providing information that could be incorporated into association mapping analyses.

Fear of false-positive outcomes arising from population stratification has virtually dictated progress in human association study design and analysis methodology (Cardon and Bell 2001, Yu et al. 2006). Genome-scale association mapping should be possible if adequate methods are implemented to control for the effects of population structure. Association mapping without consideration of population structure would have a high rate of Type I error (false positive) because of spurious associations between non-linked loci. Model-based analysis of population structure similar to that conducted here has provided information that has been incorporated into association mapping analysis.

In the eight clusters identified in this study, F_{IS} was high, suggesting that most of the populations

deviated from the Hardy–Weinberg expectation within populations. The F_{IT} value, indicating nonequilibrium conditions across populations and deficiency of heterozygotes, was also high. Testing for association between the SSR multilocus genotype data associated with quantitative variation in the presence of population structure (Thornsberry et al. 2001; Remington et al. 2001) was applied using the MLM procedure in TASSEL (Yu et al. 2006). The positive results from these studies should encourage the further testing of these methods in different genetic systems, in the same way as we applied these procedures to rice. Association between markers and grain yield, dehulled kernel length, and width, length/width ratio, and 1000-kernel weight was examined in two ways—significance of marker-trait association (*P*-values) using TASSEL software and marker-trait associations found in other QTL studies reported in http://www.gramene.org/. This indicates that QTL detected in mapping populations from biparental mapping populations were widely presented in this set of accessions, and that they could be detected with LD mapping. Association between traits and marker regions that had not been implicated before to affect trait suggest new QTL. The trait markers association was indicative of a rapid decrease in correlation, suggesting LD across a short distance. Within different sets of barley germplasm there was frequent disequilibrium among non-linked SSR loci, suggesting that association mapping without consideration of population structure would have a high rate of false positive Type I error (Mather et al. 2004). A unified mixed-model approach to account for multiple levels of relatedness simultaneously, as detected by genetic markers, has resulted in improved control of both type-I and type-II error rates (Yu et al. 2006).

Plant genetics has an important and challenging goal of identifying the genetic variants that underlie complex traits. Two main approaches are available for mapping the relevant genes and identifying the variants that associate with the traits: linkage mapping in families and population-based genetic association studies. Linkage mapping has been very successful in finding genes for rare, Mendelian, monogenic disease resistance. For complex traits that involve variants at several loci, each of which contribute small amounts to the overall genetic contribution, linkage studies mainly identify only those loci with the strongest influence, however. The most significant finding of this paper is that the LD in this set of germplasm did not decay until 20–30 cM. These results could have important implications for association testing in rice. There are two previous studies of LD in rice. The first is that of Garris et al. (2003) who found LD to decay at 100 kb across one region on chromosome 5. In a second study Olsen et al. (2006) analyzed a 500-kb region on chromosome 6 and found a 250 kb selective sweep at the waxy locus that led to elevated LD in that region. This would indicate that LD decay at 250 kb was unusually high, because of selection on the waxy locus. Although the amount of LD will vary across the genome (because of recombination rates, selective pressures, etc.), these studies seem indicative of LD decaying in rice at 1 cM or less (assuming an average of 250 kb/cM across the genome). This is in contrast with the LD decaying at 20–30 cM in the current study. LD in some *Arabidopsis* populations exceeds 50 cM (Nordborg et al. 2002), however, and LD as a function of genetic distance is very common for distances <10 cM (Kraakman et al. 2004) in barley.

In theory, genetic association mapping has greater power than linkage studies to identify variants with weak effects that might contribute risk for common complex traits (Risch and Merikangas 1996). Whole-genome association studies have the advantage of enabling the entire genome to be assessed for trait-associated variants, rather than analyzing specific candidate genes. The disadvantage of such studies, however, is that a large amount of genotyping is required. This can be reduced by using a subset of markers to report on neighboring linked markers within the genome (Smith and O'Brien 2005). Application of association mapping to plant breeding seems to be a promising means of overcoming the limitations of conventional linkage mapping (Stich et al. 2005). Our results have shown that LD studies are an efficient means of indicating novel genes for important agronomic characters that subsequently can be validated in specific biparental crossing populations, and for confirm-

ing QTL that have been detected in mapping populations.

Acknowledgement The authors acknowledge the support of H. Raeann Refeld and Dr Hesham A. Agrama from the Arkansas Rice Research and Promotion Board. Technical contributions to this research were made by H. Raeann Refeld and Quynh P. Ho. Contributions of Melissa H. Jia, Gordon H. Miller and the late Mark A. Redus of the DB NRRC Genomics Core Facility also are acknowledged.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Abdallah JM, Goffinet B, Ayrolles CC, Pérez-Enciso M (2003) Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. Genet Sel Evol 35:513–532

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nature Rev Genet 2:91–98

Chen X, Cho YG, McCouch SR (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Mol Genet Genomics 268:331–343

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski A (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19. pp 1–14

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Condon F, Smith K (2005) Linkage disequilibrium and marker-trait associations within six-rowed barley breeding germplasm. Plant & Animal Genomes XIII Conference. January 15–19, 2005

Dilday RH, Yan W, Moldenhauer KAK, Gibbons JW, Lee FN, Bryant RJ (2001) Chinese and other foreign germplasm evaluation. In: Norman RJ, Meullenet J-F (eds) B.R. Wells Rice Research Studies 2000. University of Arkansas, Agricultural Experiment Station. Research Series 485:1–12

Eaves IA, Barber RA, Merriman TR (1998) Comparison of linkage disequilibrium in populations from UK and Finland. Am J Hum Gen A221

Eizenga GC, Agrama HA, Lee FN, Yan W, Jia Y (2006) Identifying novel resistance genes in newly introduced Blast resistant rice germplasm. Crop Sci 46:1870–1878

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50

Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol 12:1074–1084

Farnir F, Coppieters W, Arranz W, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fjellstrom RG, Conaway-Bormans CA, McClung AM, Marchetti MA, Shank AR, Park WD (2004) Development of DNA markers suitable for marker assisted selection of three *Pi* genes conferring resistance to multiple *Pyricularia grisea* pathotypes. Crop Sci 44:1790–1798

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell S, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Foster M, Sharp RR (2002) Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity. Genome Res 12:844–850

Gao LZ, Ge S, Hong DY (2000) Low levels of genetic diversity within populations and high genetic differentiation among populations of a wild rice, *Oryza granulate* Nees et Arn. Ex Watt. From China. Int J Plant Sci 161:691–697

Gao LZ, Cheng W, Ge S, Hong DY, Jiang WZ, Wang W (2001) Genetic erosion in northern marginal population of common wild rice *Oryza rufipogon* Griff. and its conservation, revealed by allozyme analysis. Hereditas 133:47–53

Gao LZ, Schaal BA, Zhang CH, Jia JZ (2002) Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. Using microsatellite and allozyme markers. Theor Appl Genet 106:173–180

Gao LZ, Zhang CH (2005) Comparisons of microsatellite variability and population genetic of two endangered wild rice species, *Oryza rufipogon* and *O. officinalis*, and their conservation implications. Biodivers Conserv 14:1663–1679

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

Ge XJ, Xing YZ, Xu CG, He YQ (2005) QTL analysis of cooked rice grain elongation, volume expansion, and water absorption using a recombinant inbred population. Plant Breed 124:121–126

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30

Guo LB, Xing YZ, Mei HW, Xu CG, Shi CH, Wu P, Luo LJ (2005) Dissection of component QTL expression in yield formation in rice. Plant Breed 124:127–132

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hittalmani S, Shashidhar HE, Bagali PG, Huang N, Sidhu JS, Singh VP, Khush GS (2002) Molecular mapping of quantitative trait loci for plant growth, yield and yield related traits across three diverse locations in a doubled haploid rice population. Euphytica 125:207–214

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

Ishii T, Xu Y, McCouch SR (2001) Nuclear- and chloroplast-microsatellite variation in A-genome species of rice. Genome 44:658–666

Jannink JL, Walsh B (2002) Association mapping in plant populations. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB International

Jing YH, Sun CQ, Tan LB, Fu YC, Zhang PJ, Xu ZJ, Chen WF, Wang XK (2005) Mapping QTLs controlling vascular bundle and panicle-related traits from Yuanjiang common wild rice (*Oryza rufipogon* Griff.). Acta-Genetica-Sinica 32:178–182

Jorde LR (1995) Linkage disequilibrium as a gene-mapping tool. Am J Hum Genet 56:11–14

Jorde LR (2000) Linkage disequilibrium and the search for complex disease genes. Genome Res 10:1435–1444

Kraakman ATW, Niks RE, van den Berg PM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Kraakman ATW, Martínez F, Mussiraliev B, van Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breed 17:41–58

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

Kruger SA, Able JA, Chalmers KJ, Langridge P (2004) Linkage disequilibrium analysis of hexaploid wheat. In: Plant & Animal Genomes XII Conference, 10–14 January, Town & Country Convention Center, San Diego, CA, p 321

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144

Lander ES, Schork NJ (1994) Genetic dissection of quantitative traits. Science 256:2037–2048

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch SR (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome 47:697–704

Li ZK, Yu SB, Lafitte HR, Huang N, Courtois B, Hittalmani S, Vijayakumar CHM, Liu GF, Wang GC, Shashidhar HE, Zhuang JY, Zheng KL, Singh VP, Sidhu JS, Srivantaneeyakul S, Khush GS (2003) QTL × environment interactions in rice. I. Heading date and plant height. Theor Appl Genet 108:141–153

Liu K, Muse S (2004) PowerMarker: new genetic data analysis software. Version 3.0. Free program distributed by the author over the internet from http://www.powermarker.net

Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai TH (2005) Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Sci 45:66–76

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–289

Mackill DJ (1995) Classifying japonica rice cultivars with RAPD markers. Crop Sci 35:889–894

Mather DE, Hyes PM, Chalmers KJ, Eglinton J, Matus I, Richardson K, VonZitzewitz J, Marquez-Cedillo L, Hearnden P, Pal N (2004) Use of SSR marker data to study linkage disequilibrium and population structure in *Hordeum vulgare*: Prospects for association mapping in barley. In: International barley genetics symposium, Brno, Czech Republic, 20–26 June 2004. pp 302–307

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair MW (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

McCouch SRM, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Carinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. Genetics 160:1113–1122

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (2000) The genetic structure of cocoa populations (*Theobroma cacao* L.) revealed by RFLP analysis. Euphytica 115:83–90

Ni J, Colowit PM, Mackill DJ (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42:601–607

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomanjian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfeisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol 3:1289–1299

Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan D (2006) Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region. Genetics 173:975–983

Parsons BJ, Newbury HJ, Jackcon MT, Ford-Lloyd BV (1999) The genetic structure and conservation of aus, aman and bro rices from Bangladesh. Genet Res Crop Evol 46:587–598

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228

Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor Popul Biol 60:227–237

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Remington DL, Thornsberry JM, Matsuola Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Risch NJ (2000) Searching for genetic determination for the new millennium. Nature 405:847–856

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol Breed 15:233–245

Smith MW, O'Brien SJ (2005) Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat Rev (online) doi:10.1038/nrgl1657

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Viard F, Franck P, Dubois MP, Estoup A, Jarne P (1998) Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. J Mol Evol 47:42–51

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington A-ME, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA, Jorde1 LB (2003) Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. Genome Res 13:1607–1618

Weber J, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Weir BS (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc., Inc., Sunderland, MA, USA

Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler IV ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Yan WG, Rutger JN, Bryant RJ, Lee FN, Gibbons JW (2003) Characteristics of newly-introduced accessions in the USDA-ARS rice quarantine program In: Norman RJ, Meullenet J-F (eds) B.R. Wells Rice Research Studies 2002. Univ. of Arkansas Agric Exp Stn Res Ser 504:112–124

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187–194

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17:155–160

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley J, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang N, Xu Y, Akash M, McCouch S, Oard JH (2005) Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. Theor Appl Genet 110:721–729