African Journal of Biotechnology Vol. 11(12), pp. 2957-2970, 9 February, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.1912
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

Association analysis of important agronomic traits in japonica rice germplasm

Jiangbo Zhou[1], Aiqing You[2], Zhujun Ma[1], Lili Zhu[1] and Guangcun He[1]*

[1]State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, Wuhan 430072, China.
[2]Hybrid Rice Research Center, Hubei Academy of Agriculture Sciences, Wuhan 430064, China.

Accepted 6 October, 2011

For unraveling complex genetic traits and for marker-assisted selection in plant breeding, using elite germplasm for quantitative trait loci (QTL) mapping provides advantages over traditional linkage analysis. Association or linkage disequilibrium (LD) mapping revolutionized genetic mapping in humans, and is increasingly being used to examine plant genetics; it has been suggested that it is an efficient way of determining the genetic basis of complex traits. In this study, a natural population comprising 128 japonica rice varieties was investigated during two years, for eleven important agronomic traits. The population was genotyped using 152 microsatellite markers across the whole genome. A unified mixed linear model was used to identify marker-trait associations, taking into account population structure and kinship. A total of 16 significant marker-trait associations were identified. It is noteworthy that the loci exhibiting significant marker-trait associations were in genomic regions also identified as such in previous reports, and some new QTL regions for important agronomic traits were also detected in this study. Our results demonstrate that using association analysis to examine different varieties is an efficient approach and provides indicators that can be used in rice breeding.

Key words: Association mapping, linkage disequilibrium, agronomic traits, rice.

INTRODUCTION

Association or linkage disequilibrium (LD) mapping, which revolutionized genetic mapping in humans (Donnelly, 2008) and is increasingly being applied to plants (Nordborg and Weigel, 2008), and is considered an efficient way of determining the genetic basis of complex traits. Comparing traditional linkage mapping which depends on restricted allelic variation with a small number of recombination events, using the association mapping there is no need to develop segregating populations such as F2, double haploid (DH) or back-cross populations, instead a natural collection of inbred lines or varieties can be used. There is also the potential for discovering the QTL responsible for multiple traits, and examination of a higher proportion of polymorphic molecular markers could provide better genome coverage than any bi-parental population (Flint-Garcia et al., 2005; Jannink et al., 2001; Nordborg and Weigel, 2008). Essentially, the method could exploit the historical recombination events that have occurred in natural populations, thus reducing the cost and time taken for analysis (Flint-Garcia et al., 2005; Wen et al., 2009; Yao et al., 2009). Moreover, QTL and molecular markers detected in the inbred lines and elite germplasm can be used directly in breeding programs.

Rice (*Oryza sativa* L.), an important agricultural crop and a model species, has been cultivated for more than 7,000 years (YunFei et al., 2007; Zong et al., 2007). The differentiation and diversification of annuals in South Asia have been accelerated by marked climatic changes following the last glaciation, dispersal of plants over wide latitudinal or altitudinal ranges, human selection and

*Corresponding author. E-mail: gche@whu.edu.cn. Tel: +86-27-87641314. Fax: +86-27-68752327.

Abbreviations: QTL, Quantitative trait loci; **LD,** linkage disequilibrium; **GWA,** genome-wide association; **MLM,** mixed linear model; **PIC,** polymorphism information content; **UPGMA,** unweighted pair group method of the arithmetic average.

manipulation of the cultural environment. Widespread dispersal of the Asian rice cultigen led to the formation of three eco-geographic races: indica, japonica and javanica (Chang, 1976). With its high level of genetic diversity, historical recombination events must have occurred in natural populations of rice. Association mapping is therefore feasible, and potentially very useful for rice. In particular, since rice has been completely sequenced, it has a relatively small genome and is well suited to Genome-wide association (GWA). Recently, the association between 123 SSR loci and the complex traits associated with yield and its components, such as kernel width, kernel length, kernel width/length ratio, and 1000-kernel weight were estimated by Agrama et al. (2007) in 103 rice germplasm accessions. Subsequently, Wen et al. (2009) used the mixed linear model (MLM) method to identify SSR marker loci associated with three agronomic traits - heading date, plant height and panicle length - on chromosome 7 from a diverse sample of Chinese rice germplasm. Relatively little research on genome-wide associations in rice has been reported, and much remains to be done to exploit fully the rice germplasm resources associated with complex agronomically useful genetic traits.

In this study, we determined the population structure in a natural population that consisted of 128 japonica rice varieties. Furthermore, we used the mixed linear model method in the TASSEL software to conduct association analysis for eleven important agronomic traits through genome-wide scanning. Finally, the potential application of association analysis results for parental selection or marker-assisted selection for the development of superior rice varieties is discussed.

MATERIALS AND METHODS

Sample collection

A total of 128 japonica rice varieties were used in this study. The varieties originated from different regions of China and have been widely used as parents in plant breeding during past decades. Only a few of the varieties had been introduced from other countries, including Italy, Japan and North Korea; one of these was the genome sequenced variety Nipponbare (Supplementary information Table S1).

Field experiments and phenotypic evaluation

For field studies, each rice variety was planted out in a field at the experimental farm of the Hubei Academy of Agricultural Sciences (Wuhan, China) using a randomized block design (RBD). There were two replicates for each variety in each phenotypic evaluation experiment during the summers of 2008 and 2009. Each plot consisted of three rows, each containing 10 mounds. For all varieties, seedlings aged about 30 days were transplanted into the field at a spacing of 16.7 cm between plants within each row and 26.7 cm between the rows. The middle five plants in the central row of each plot were sampled in order to examine agronomic traits. Eleven quantitative traits of agronomic and economic importance, particularly relating to grain yield, were evaluated: heading date (HD; in days), plant height (PH; in centimeters), panicles per plant (PPP), panicle length (PL; in centimeters), spikelets per panicle (SPP), filled grains per panicle (FGPP), percentage seed set (SS), grain density (GD, which is the number of grains per 10 cm of panicle length), 1000-grain weight (KGW; in grams), flag leaf length of the main panicle (FL; in centimeters), and grain yield per plant (GYPP; in grams). All quantitative traits were obtained from two replicates in the experiment during each year. For SPP, FGPP and GYPP, all panicles from a plant were evaluated. For each trait, means of the replicates were used in the data analyses.

SSR genotyping

Total genomic DNA was extracted from the leaf tissue of six seedlings per variety following the methods described by Murray and Thompson (1980). Due to preparatory work for the screening of molecular markers, minor heterozygous loci were eliminated and the homozygous loci over all varieties were used preferentially in the subsequent analysis. Heterozygous loci and rare alleles were excluded. Thus, a total of 152 polymorphic microsatellite markers, approximately evenly distributed on the 12 chromosomes of rice, were used for genotyping (supplementary information, Table S2). Microsatellite polymorphism was analyzed by PCR. Amplification of DNA was performed in a 10 µL reaction mix consisting of 20 to 50 ng template DNA, 10 × PCR buffer, 0.2 mmol dNTPs, 0.2 µM forward and reverse primers, and 0.5 unit of *Taq*DNA polymerase. The PCR products were detected using the optimized silver staining method in denaturing polyacrylamide gels (Han et al., 2008). Primer sequences and PCR amplification conditions for each set of primers and the physical positions of marker loci were based on published Nipponbare sequence databases (http://www.gramene.org; http://www.ncbi.nlm.nih.gov/).

Data analysis

Allelic diversity and population structure

Summary statistics for the allele diversity of 152 SSR markers in 128 varieties were calculated using the PowerMarker V3.25 program (Liu and Muse, 2005); these included allele number, gene diversity, and polymorphism information content (PIC). The population structure was estimated with the model-based (Bayesian) cluster software STRUCTURE 2.3.1 (Falush et al., 2003). All marker loci which were evenly distributed across all 12 chromosomes were used to estimate the population structure. The software was set to have a burn-in of 10,000 and a run length of 100,000 with an admixture model and correlated allele frequencies for inferring the number (K) of subpopulations, ranging from K = 1 to K = 10 for ten parallel calculations. Inferred ancestry of individuals was indicated by estimated *Ln* probabilities for the data. In the model-based clustering algorithm, each line was estimated to have memberships in multiple subgroups, indicated by membership coefficients.

Cluster analysis

The proportion of shared alleles distance was used to calculate estimates of genetic distance for all pairs of individuals with all 152 markers using PowerMarker v3.25 software (Liu and Muse, 2005). The calculation was performed according to the method of Chakraborty and Jin (1993). The genetic distance matrix was used in cluster analysis based on the unweighted pair group method of the arithmetic average (UPGMA), performed over 1000 bootstrap replicates.

Table S1. Details of the rice germplasm used in this study.

S/N	Variety	Origin	Year bred / introduced
1	Nonglin 1	Japan	1931
2	Zaoguixuanxi	Shandong, China	1970s
3	Huangjinguang	Japan	1982
4	Dachejing	Jiangsu, China	1956
5	Nongken58	Japan	1957
6	Zhongjing315	Jiangsu, China	/
7	Guihuahuang	Jiangsu, China	1960
8	Balilla	Italy	1958
9	02428	Jiangsu, China	1989
10	Wanxiangnuo	Jiangsu, China	/
11	Jingyue 1	Beijing, China	1967
12	Zhongzuo321	Beijing, China	1980s
13	Nipponbare	Japan	1971
14	Baigedao	Jiangsu, China	/
15	Guinuo80	Jiangsu, China	1972
16	Ying8433	Liaoling, China	1984
17	Guihuanuo	Jiangsu, China	/
18	Banteng 5	Japan	/
19	Zhenxiang24	Jiangsu, China	/
20	Aihuangzhong	Zhejiang, China	1970
21	Aipinuo	Zhejiang, China	/
22	Aili 3	Guangdong, China	/
23	Lindao 2	Shandong, China	1972
24	Baijin	Japan	1966
25	Nonglin22	Japan	/
26	Liming	Japan	1966
27	Luaojing	Japan	/
28	Tieludao	North Korea	/
29	Jijing13	Hebei, China	/
30	Fengrunyedihuangjin	Japan	1957
31	Lindao 1	Shandong, China	1972
32	Yanjing 2	Jiangsu, China	1979
33	Lisuibo	Japan	1981
34	Zhunuo	Japan	/
35	Laolaiqing	Anhui , China	/
36	Pijing 2	Jiangsu, China	1974
37	Tongyujing	Jiangsu, China	/
38	Tianjin218	Tianjin, China	/
39	Chuangui 4	Jiangsu, China	1968
40	Xiushui115	Zhejiang, China	1983
41	Liaojing326	Liaoning, China	/
42	Lujing 1	Shandong, China	1978
43	Huajing45	Liaoning, China	/
44	Jindao1311	Tianjin, China	/
45	Liaojing207	Liaoning, China	1987
46	Chengfu24	Sichuan, China	/
47	Liaojing294	Liaoning, China	1987

Table S1. countd

48	Shuangnuo 4	Zhejiang, China	1975
49	Liaoyannuo	Liaoning, China	1990s
50	Xijing 2	Shanxi, China	1976
51	Qiannuo203	Guizhou, China	2000s
52	Bajing 1	Xinjiang, China	/
53	Chengnuo68	Sichuan, China	/
54	Liaoyou 7	Liaoning, China	1987
55	Wuyunjing 8	Jiangsu, China	1995
56	Xiangjing49	Jiangsu, China	/
57	Jia0433	Jiangsu, China	/
58	Huajing 3	Jiangsu, China	/
59	Tongjing109	Jiangsu, China	1990s
60	Wuyujing15	Jiangsu, China	1999
61	Zhendao631	Jiangsu, China	/
62	Wuyunjing 7	Jiangsu, China	1995
63	Suxiangjing 1	Jiangsu, China	/
64	Nanjing41	Jiangsu, China	/
65	Wuxiangjing14	Jiangsu, China	1999
66	Xiushui46	Zhejiang, China	/
67	Xudao 2	Jiangsu, China	1974
68	Shennong91	Liaoning, China	/
69	Xudao444	Jiangsu, China	/
70	Nanjing31	Jiangsu, China	1972
71	Tongjing 5	Jiangsu, China	/
72	Aijing23	Liaoning, China	1974
73	Hainuo 3	Jiangsu, China	/
74	Nanjing33	Jiangsu, China	1973
75	Liaonong938	Liaoning, China	1993
76	Jijing 10	Hebei, China	/
77	Shennong8718	Liaoning, China	1987
78	Tongjing 3	Jiangsu, China	/
79	Aifeng 2	Liaoning, China	1965
80	Liaojing454	Liaoning, China	/
81	Shennong606	Liaoning, China	1995
82	Xudao5027	Jiangsu, China	1974
83	Shennong8801	Liaoning, China	1988
84	Shafeng75	Xinjiang , China	1982
85	Kenyu16	Hebei, China	1990s
86	Hainuo 2	Jiangsu, China	/
87	Tongjing 1	Jiangsu, China	/
88	Huayu 1	Tianjin, China	/
89	Nanjing35	Jiangsu, China	1980
90	Nanjing32	Jiangsu, China	1972
91	Huai68 (Huaidao 9)	Jiangsu, China	2000
92	Xudao 3	Jiangsu, China	1991
93	Zhendao431	Jiangsu, China	/
94	Wuxiangjing 9	Jiangsu, China	/
95	Xudao 4	Jiangsu, China	1991
96	Bing03123	Jiangsu, China	/
97	Huajing 2	Jiangsu, China	/
98	Guanglingxiangjing	Jiangsu, China	/

Table S1. contd

99	Wuyujing 3	Jiangsu, China	1990s
100	Huaidao 8	Jiangsu, China	2004
101	Xiangnuo 4	Beijing, China	/
102	Zhongjing79-12-2	Beijing, China	1979
103	Chengte235	Zhejiang, China	/
104	Hongqi16	Tianjin, China	1971
105	Shuangchengnuo	Jiangsu, China	1978
106	Zhejing66	Zhejiang, China	1979
107	Hongqi12	Tianjin, China	1970
108	Jijing 6	Hebei, China	1984
109	Jiangfeng 3	Jiangsu, China	1972
110	Tudao 8	Jiangsu, China	/
111	Supin20324	Jiangsu, China	/
112	Lianjing8410	Jiangsu, China	1984
113	Ningjing 6	Ningxia, China	1978
114	Kundao 2	Jiangsu, China	1976
115	Xiangjing 3	Zhejiang, China	/
116	Liaojing 9	Liaoning, China	/
117	Shannong13	Shandong, China	1978
118	Shibei43	Jiangsu, China	/
119	Shajiao 5	Xinjiang , China	1975
120	Jijing 8	Hebei, China	1985
121	Yuhong 3	Jiangsu, China	1972
122	Nanjing43	Jiangsu, China	/
123	Shennong265	Liaoning, China	1992
124	Ningjing 7	Ningxia, China	1971
125	Tudao 2	Jiangsu, China	1981
126	Aishuang 2	Zhejiang, China	1980
127	Honggui 1	Sichuan, China	/
128	Zhendao99	Jiangsu, China	/

Association mapping

Associations between traits and markers were calculated using TASSEL 2.1 with the MLM program in the (Q +K) model (Yu et al., 2006). The unified mixed linear model simultaneously takes into account population structure and kinship and can effectively control false associations. The significant marker-trait associations were indicated by a P-value with corresponding R^2 for each marker as the percentage of the total variation explained.

RESULTS

Phenotypic variations

The summary statistics for the phenotypic data relating to the eleven traits of the 128 japonica varieties investigated in this study are shown in Table 1. Marked variations were recorded among the rice varieties for the traits in the two study years. For example, the value of HD among these varieties ranged from 63 to 118 days in 2008 and 79 to 126 days in 2009. In addition, for PH, PPP, PL, SPP and FGPP the maximum values were approximately three times larger than the minimum values for the varieties studied. These relatively large levels of phenotypic variability were measured for all traits among rice varieties, indicating that this collection of elite japonica varieties was appropriate to use in association studies of agronomic traits.

Genome-wide allelic diversity at SSR loci

A total of 152 polymorphic SSR markers were used to genotype the 128 varieties, generating a total of 568 alleles (Supplementary information, Table S2). The average number of alleles per locus was 3.7368, ranging from 2 (RM315 on chromosome 1, etc) to 14 (RM591 on chromosome 10). The average gene diversity over all

Table S2. Marker information used for genotyping in the study.

Marker	Chromosome	Physical position (Mb)	Allele number	Gene diversity	PIC
RM462	1	0.082	3.0	0.4449	0.3589
RM6672	1	0.2797	11.0	0.7471	0.7096
RM499	1	0.3867	3.0	0.0309	0.0306
RM7278	1	1.792	3.0	0.1460	0.1389
RM6324	1	2.375	1.0	0.0000	0.0000
RM428	1	2.6065	1.0	0.0000	0.0000
RM522	1	5.2427	3.0	0.3439	0.3168
RM151	1	5.5564	5.0	0.4865	0.4222
RM259	1	7.4434	5.0	0.1760	0.1716
RM3234	1	7.5682	5.0	0.2158	0.2083
RM6685	1	7.8695	4.0	0.2719	0.2487
RM575	1	8.0719	3.0	0.1046	0.1015
RM292	1	9.5639	4.0	0.2416	0.2319
RM580	1	9.6034	10.0	0.8077	0.7855
RM600	1	10.4090	2.0	0.0308	0.0303
RM312	1	10.9855	2.0	0.0458	0.0447
RM8139	1	19.4045	1.0	0.0000	0.0000
RM9	1	23.6529	5.0	0.5818	0.5017
RM5	1	24.2992	7.0	0.5817	0.5246
RM7202	1	26.5304	2.0	0.1307	0.1222
RM2420	1	28.3638	1.0	0.0000	0.0000
RM1117	1	29.6216	1.0	0.0000	0.0000
RM3615	1	30.1335	3.0	0.2638	0.2477
RM128	1	31.0656	1.0	0.0000	0.0000
RM3709	1	32.2753	3.0	0.1317	0.1249
RM302	1	33.3151	3.0	0.1893	0.1741
RM212	1	33.3814	2.0	0.1235	0.1159
RM315	1	37.0620	2.0	0.3065	0.2595
RM472	1	38.2170	3.0	0.1180	0.1127
RM6292	1	39.5494	5.0	0.1621	0.1581
RM165	1	40.4339	2.0	0.0155	0.0154
RM8236	1	40.4699	5.0	0.3252	0.2898
RM529	1	40.9983	4.0	0.6184	0.5545
RM154	2	1.0838	4.0	0.4005	0.3680
RM7033	2	1.6608	2.0	0.1172	0.1103
RM236	2	2.1055	2.0	0.3827	0.3095
RM279	2	2.8821	4.0	0.5972	0.5519
RM423	2	3.8368	2.0	0.0605	0.0587
RM322	2	7.4334	9.0	0.5857	0.5062
RM561	2	18.7640	3.0	0.1876	0.1792
RM341	2	19.3360	6.0	0.5408	0.4390
RM166	2	34.3467	2.0	0.1065	0.1009
RM213	2	34.6523	4.0	0.5901	0.5429
RM535	2	35.7782	2.0	0.0155	0.0154
RM517	3	0.0061	5.0	0.4015	0.3734
RM523	3	1.2988	2.0	0.0605	0.0587
RM545	3	4.9280	4.0	0.4662	0.4265
RM487	3	21.9756	1.0	0.0000	0.0000
RM426	3	27.5447	5.0	0.3568	0.3167
RM504	3	28.0855	3.0	0.3788	0.3127
RM143	3	33.1264	5.0	0.2415	0.2316

Table S2. contd

RM227	3	34.8713	1.0	0.0000	0.0000
RM442	3	35.7813	1.0	0.0000	0.0000
RM200	3	38.5734	2.0	0.0155	0.0154
RM537	4	0.1819	5.0	0.3429	0.3304
RM335	4	0.6799	11.0	0.4822	0.4695
RM518	4	2.0217	4.0	0.3947	0.3587
RM1305	4	5.6396	2.0	0.3488	0.2879
RM5900	4	13.7484	2.0	0.2637	0.2289
RM7113	4	18.1426	7.0	0.5175	0.4801
RM185	4	18.5643	1.0	0.0000	0.0000
RM119	4	21.2271	1.0	0.0000	0.0000
RM252	4	24.0227	3.0	0.0309	0.0306
RM5030	4	30.4992	10.0	0.5869	0.5486
RM127	4	33.8813	2.0	0.0308	0.0303
RM131	4	34.3896	8.0	0.2445	0.2393
RM124	4	34.7057	1.0	0.0000	0.0000
RM153	5	0.1678	3.0	0.2751	0.2577
RM548	5	2.7965	1.0	0.0000	0.0000
RM267	5	2.8594	3.0	0.3199	0.2778
RM289	5	7.7871	1.0	0.0000	0.0000
RM1127	5	15.9373	3.0	0.3377	0.3122
RM440	5	19.8919	1.0	0.0000	0.0000
RM188	5	22.6506	3.0	0.1997	0.1885
RM534	5	23.9869	1.0	0.0000	0.0000
RM133	6	0.2270	1.0	0.0000	0.0000
RM435	6	0.5373	2.0	0.0155	0.0154
RM588	6	1.6114	1.0	0.0000	0.0000
RM190	6	1.7646	2.0	0.4347	0.3402
RM584	6	3.4166	3.0	0.4291	0.3726
RM136	6	8.7513	1.0	0.0000	0.0000
RM527	6	9.8623	6.0	0.3984	0.3544
RM7311	6	11.0457	5.0	0.3369	0.3166
RM3183	6	12.4470	2.0	0.4995	0.3747
RM2229	6	15.5786	4.0	0.4058	0.3813
RM2859	6	19.4688	4.0	0.5203	0.4142
RM4447	6	22.6796	6.0	0.6859	0.6332
RM1340	6	23.3432	4.0	0.5679	0.5269
RM162	6	24.0355	5.0	0.3413	0.3112
RM5314	6	24.8428	3.0	0.2123	0.1992
RM528	6	26.5548	6.0	0.4198	0.3997
RM340	6	28.5992	2.0	0.0605	0.0587
RM461	6	30.1135	4.0	0.2390	0.2255
RM6965	7	2.8092	13.0	0.6608	0.6359
RM8262	7	5.2977	4.0	0.4408	0.3964
RM180	7	5.7345	7.0	0.6415	0.5748
RM1253	7	6.9670	4.0	0.7250	0.6766
RM6018	7	9.9785	3.0	0.4784	0.3712
RM542	7	12.7113	4.0	0.6017	0.5374
RM3670	7	13.4382	3.0	0.4263	0.3478
RM418	7	18.1315	2.0	0.0155	0.0154
RM3186	7	18.6293	4.0	0.5441	0.4400
RM3826	7	20.8067	7.0	0.7920	0.7625

Table S2. Contd

RM6152	7	22.1268	2.0	0.0458	0.0447
RM10	7	22.1871	4.0	0.4007	0.3771
RM70	7	22.7751	4.0	0.2504	0.2339
RM6432	7	23.1590	4.0	0.4357	0.3861
RM6098	7	24.4656	1.0	0.0000	0.0000
RM505	7	24.5257	3.0	0.4803	0.4104
RM234	7	25.4720	4.0	0.1326	0.1272
RM18	7	25.6518	2.0	0.4928	0.3714
RM429	7	26.8053	4.0	0.3953	0.3601
RM1306	7	28.9462	13.0	0.4617	0.4506
RM152	8	0.6776	2.0	0.4479	0.3476
RM1295	8	0.9816	2.0	0.4999	0.3749
RM126	8	5.2159	2.0	0.1172	0.1103
RM3181	8	7.5474	3.0	0.5770	0.5026
RM339	8	17.8120	3.0	0.5052	0.3853
RM223	8	20.5199	4.0	0.5988	0.5345
RM6070	8	26.1895	4.0	0.1725	0.1623
RM447	8	26.4169	5.0	0.6606	0.5967
RM281	8	27.7654	3.0	0.3023	0.2799
RM264	8	27.7938	3.0	0.3491	0.3192
RM477	8	27.9450	2.0	0.0155	0.0154
RM316	9	1.0751	2.0	0.2847	0.2442
RM105	9	12.5447	3.0	0.6473	0.5716
RM3700	9	15.4272	3.0	0.3345	0.2957
RM3025	9	16.3858	5.0	0.6498	0.6045
RM328	9	19.7247	2.0	0.1034	0.0980
RM201	9	20.1743	2.0	0.1034	0.0980
RM474	10	1.8040	4.0	0.2020	0.1945
RM6179	10	2.5715	4.0	0.1337	0.1302
RM311	10	9.4872	6.0	0.4995	0.4635
RM304	10	18.3974	6.0	0.2014	0.1929
RM333	10	22.1138	8.0	0.7079	0.6616
RM591	10	22.6415	14.0	0.7308	0.7048
RM286	11	0.3837	8.0	0.5601	0.5278
RM3339	11	4.1361	4.0	0.2286	0.2198
RM202	11	8.9849	2.0	0.4369	0.3414
RM7391	11	9.9184	4.0	0.1476	0.1432
RM3428	11	13.4451	5.0	0.4662	0.3996
RM229	11	18.3727	9.0	0.7668	0.7300
RM2110	11	19.4232	4.0	0.6843	0.6300
RM6105	11	23.6121	5.0	0.2832	0.2580
RM144	11	28.2469	4.0	0.1747	0.1681
RM8215	12	1.5858	4.0	0.3729	0.3523
RM19	12	2.4324	1.0	0.0000	0.0000
RM6296	12	3.2006	1.0	0.0000	0.0000
RM309	12	8.8316	4.0	0.4413	0.3872
RM235	12	17.8492	4.0	0.4210	0.3897
RM270	12	23.2463	4.0	0.6393	0.5843
RM3813	12	26.1079	3.0	0.3624	0.3022

SSR loci was 0.3049 and the PIC value for the SSR loci was 0.2745. The japonica rice germplasms used in this

Table 1. Summary statistics for the eleven agronomical traits examined in the two years.

Trait	Year	Mean	SD†	Minimum	Maximum
HD	2008	86.33594	11.63235	63	118
	2009	98.71875	8.226075	79	126
PH	2008	92.92129	12.21326	67	148.8333
	2009	96.15406	14.15891	61.6	158.5
PPP	2008	7.958659	1.89982	3.916667	12.8
	2009	12.32396	2.978646	6	22.7
SPP	2008	122.3699	26.40253	62.77262	187.1306
	2009	107.6668	26.2581	36.03796	193.3479
FL	2008	25.75404	4.317971	15.5	47.75
	2009	28.5819	4.574088	20.4	43.9
FGPP	2008	104.2886	22.14251	48.43211	156.854
	2009	87.18871	21.04107	27.47377	154.5606
GYPP	2008	19.03782	4.799762	8.110417	32.334
	2009	23.03555	4.701164	12.003	33.90525
PL	2008	143.2693	34.95429	71	233.8333
	2009	207.1771	53.15833	116.5	384.7
GD	2008	68.02049	15.16148	38.52247	110.7867
	2009	63.93924	14.99763	31.15884	112.3373
SS	2008	0.856648	0.077054	0.376646	0.951014
	2009	0.813705	0.078781	0.569359	0.948462
KGW	2008	23.83515	2.478127	16.72466	31.25345
	2009	22.54035	2.364478	17.43051	29.45428

† Standard deviation. HD, Heading date; PH, plant height; PPP, panicles per plant; SPP, spikelets per panicle; FL, flag leaf length of the main panicle; FGPP, filled grains per panicle; GYPP, grain yield per plant; PL, panicle length; SS, percentage seed set; GD, grain density; KGW, 1000-grain weight.

study represent a wide sample of distinct varieties developed during past decades across China.

Evaluation of population structure and cluster analysis

The model-based clustering method was performed using 152 marker loci, with the population structure estimated for ten replicates (Supplementary information, Table S2). The number of subpopulations (K) was determined on the basis of values of posterior probability (*Ln*P (D)). For K = 1 to K = 4 the values of *Ln*P (D) tended to increase steadily and, for all ten replicates, the highest likelihood was reached when K = 4. For higher values of K, *Ln*P(D) either increased but not significantly or the runs did not produce consistent results. Thus, the 128 rice varieties were assigned to four sub-populations according to the results of the STRUCTURE analysis. On this basis, four sub-populations were assumed in the association mapping. Thus, the population structure matrix (Q) for K = 4 was used by MLM in the TASSEL software.

The phylogenetic tree generated from the SSR data suggested that the 128 varieties were grouped into four clusters (Figure 1). In general, all samples for each variety fell into a single clade, and the cluster analysis presented here was generally consistent with the empirically predefined groups based on their derivative lines. The rice varieties used in this study primarily originated from the Yangtze River valley and North China,

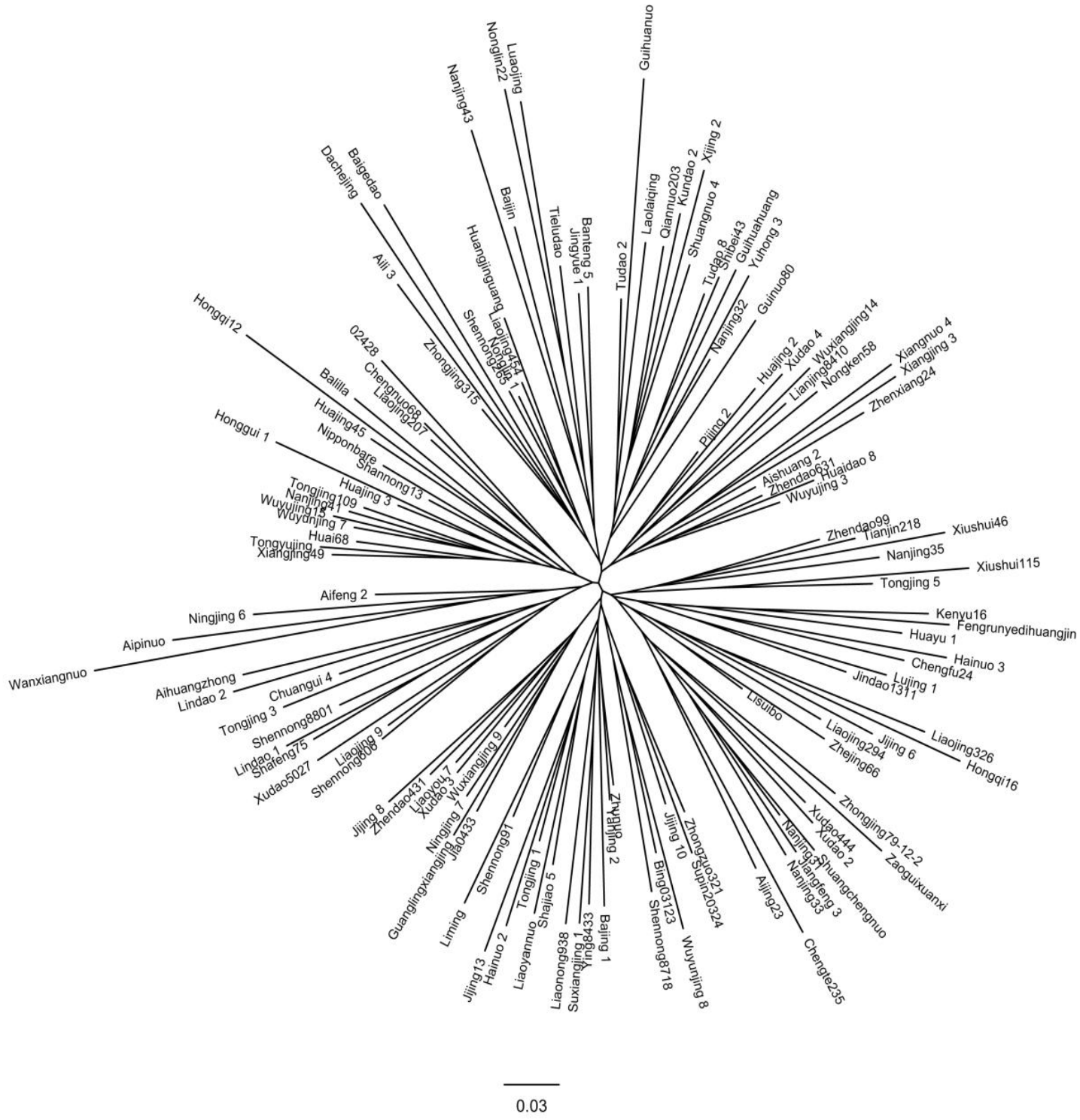

Figure 1. UPGMA dendrogram showing the relationships between the 128 rice varieties under consideration based on 152 SSR loci.

the main districts in which japonica rice is cultivated in China. The cluster analysis presented here provides another perspective on the relationships between the varieties and generally supports the population structure analysis.

Association analysis between agronomic traits and molecular markers

Using the MLM program in the TASSEL software, we identified marker-trait associations for the eleven

Table 2. Genome-wide associations between SSR markers and eleven agronomic traits with P<0.01.

Trait	Chromosome	Position (Mb)	Marker	P-value	R2†
HD	7	28.9462	RM1306	0.0000	0.0819
HD	1	40.4699	RM8236	0.0002	0.0723
HD	7	26.8053	RM429	0.0061	0.0367
PH	7	28.9462	RM1306	0.0000	0.1547
PH	8	26.1895	RM6070	0.0007	0.0396
PH	1	38.2170	RM472	0.0053	0.0199
PH	6	24.8428	RM5314	0.0078	0.0219
FL	7	12.7113	RM542	0.0016	0.0763
SPP	1	38.2170	RM472	0.0016	0.0710
FGPP	1	38.2170	RM472	0.0047	0.0692
PL	1	9.6034	RM580	0.0024	0.1620
SS	1	0.2797	RM6672	0.0000	0.2158
SS	8	26.1895	RM6070	0.0004	0.1285
KGW	11	4.1361	RM3339	0.0001	0.0941
KGW	7	28.9462	RM1306	0.0005	0.1353
KGW	1	40.4699	RM8236	0.0041	0.0556

†R^2 indicates the percentage of the total variation explained. HD, Heading date; PH, plant height; FL, flag leaf length of the main panicle; SPP, spikelets per panicle; FGPP, filled grains per panicle; PL, panicle length; SS, percentage seed set; KGW, 1000-grain weight.

agronomic traits evaluated. By carrying out genome-wide scanning, we detected a total of 16 significant marker-trait associations (P<0.01) (Table 2). All of the 12 significant SSR loci were indentified for the agronomic traits, with the R^2, percentage of the total variation explained ranging from 1.99 to 21.58%. Other parameters shown in Figure 2 were identified as follows:

Heading date (HD): We detected three locus with a significant association (P<0.01); RM1306 on chromosome 7 had the effect of explaining 8.19% of the total phenotypic variation.
Plant height (PH): We detected four loci with significant associations; RM1306 had the largest effect, explaining 15.47% of the total phenotypic variation.
Panicle length (PL): The marker locus significantly associated with PL was RM580 on chromosome 1, which explained 16.2% of the total phenotypic variation.
Spikelets per panicle (SPP): The marker loci significantly associated with SPP were RM472 on chromosome 1, which explained 7.1% of the total phenotypic variation.
Filled grains per panicle (FGPP): Only marker RM472 on chromosome 1 showed a genetic association with FGPP, explaining 6.92% of the total phenotypic variation.
Percentage seed set (SS): Two loci exhibited significant associations; of these, RM6672 on chromosome 1 had the largest effect, explaining 21.58% of the total phenotypic variation.
1000-grain weight (KGW): Three loci were identified as having significant associations; of these RM1306 on chromosome 7 had the largest effect, explaining 13.53% of the total phenotypic variation.
Flag leaf length of the main panicle (FL): We detected one locus RM542 on chromosome 7 with significant associations, explaining 7.63% of the total phenotypic variation.

We identified QTL had strong effects on multiple traits. For example, RM1306 on chromosome 7 had the greatest effect on the traits HD, PH and KGW, while RM472 on chromosome 1 and RM1306 on chromosome 7 were significantly associated with three traits simultaneously. Knowledge of these loci should make a valuable contribution to rice breeding programs.

DISCUSSION

Since rice is a species that has been completely sequenced, it is well suited to genome-wide association studies. The rice cultivars are, however, usually highly structured, owing to the species' ecological adaptations and population differentiation resulting from its complex breeding history involving both natural and artificial selection. Population structure is a strong confounding factor in association studies, especially with respect to traits that are important in local adaptation or diversifying selection and familial relatedness associated with recent co-ancestry (Nordborg and Weigel, 2008). Thus, spurious associations have been reported frequently, without giving consideration to population structure (Gupta et al., 2005). Many important crops like rice and wheat have an

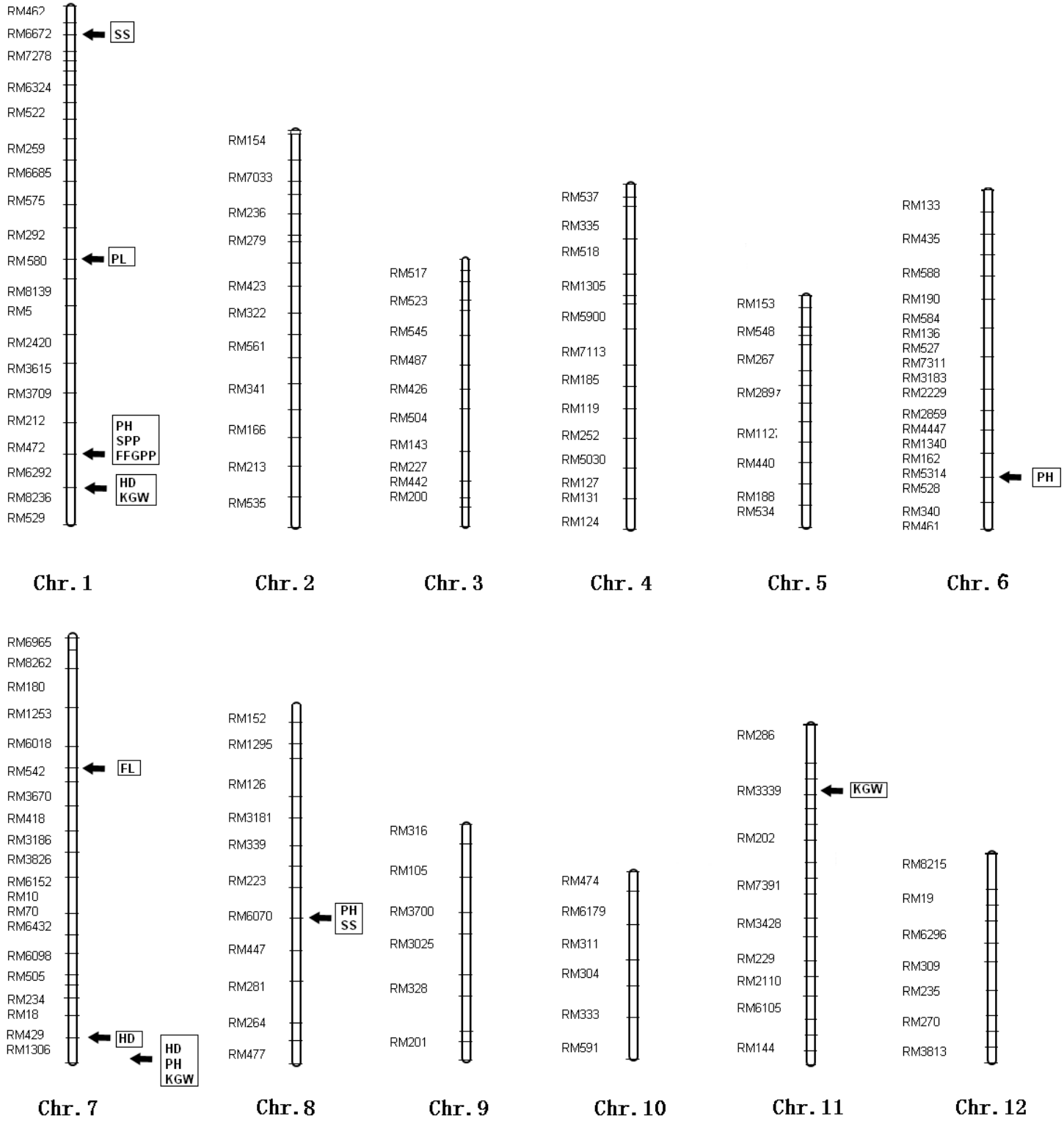

Figure 2. The marker loci significantly associated with the traits in this study (P<0.01).

intricate pattern of stratification associated with the germplasm as a result of their complex breeding history, leading to spurious associations being suggested and causing complications when undertaking association mapping (Ersoz et al., 2008). To address this problem, the unified mixed-model approach that accounts for

multiple levels of genetic relatedness simultaneously has recently been proposed and has been used for association mapping in various crop species. The markers used for genotyping the rice germplasm collections in this study were more than other rice association studies (Wen et al., 2009; Yan et al., 2009). One hundred and fifty-two polymorphic SSR markers used in our study were suitable for association mapping. Although, it still had some of the limitations in fine mapping, it probably was not subjected to ascertainment bias when looking at population structure. Here, we made use of the mixed linear model approach, which took into account population structure and familial relatedness, in order to reveal the association between SSR makers and eleven agronomic traits in rice. Such genome-wide association mapping should, therefore, be valid even in a selfing species such as rice with very high levels of population structure and much familial relatedness.

A total of 12 significantly associated loci ($P<0.01$) were detected in the experiments (Table 2). We found positional convergence between the significantly associated loci detected in our study and previously reported QTLs. The marker locus RM1306 significantly associated with HD and explaining the highest proportion of phenotypic variations in this study was located in the same genomic region on chromosome 7 as reported previously (Lin et al., 2000; Yamamoto et al., 1998; Yano et al., 1997). Marker locus RM429 on chromosome 7, significant with respect to HD, corresponded to a marker reported by Suh et al. (2005). In addition, we found many loci that explained a small amount of phenotypic variation as a result of the effectiveness and high resolution when examining polygenes or minor QTLs by association mapping. The high consistency with previous reports in the identification of candidate markers demonstrates the value of genome-wide association analysis of rice varieties.

Furthermore, we also identified multiple pleiotropic effects associated with the candidate markers. The marker locus RM1306 on chromosome 7 was significantly associated with HD, PH and KGW simultaneously, demonstrating its strong effect on phenotypic variations. These multiple pleiotropic effects associated with candidate markers could be used in marker-assisted selection to improve breeding efficiency. The association mapping findings were verified by conventional linkage mapping within previous reports. Thus, association mapping provides a powerful tool in unraveling the complex traits in plants; association mapping using pedigree-based germplasm provides high resolution for QTL mapping. Moreover, association mapping is likely to pinpoint superior alleles that have been captured by breeding practices and facilitate marker-assisted selection in plant breeding (Zhu et al., 2008). Thus, it provides essential clues that can be used in rice breeding.

Our data therefore demonstrate that using association analysis to examine important agronomic traits in rice varieties is an efficient way of identifying significant loci associated with these traits. Simultaneous associations can be identified and QTL values for the traits determined. This is can be potentially useful for plant breeding programs.

ACKNOWLEDGEMENT

This study was supported by the National Key Basic Research and Development Program of China (Grant No. 2006CB101702).

REFERENCES

Agrama HA, Eizenga GC, Yan W (2007). Association mapping of yield and its components in rice cultivars. Mol. Breed. 19: 341-356.

Chakraborty R, Jin L (1993). A unified approach to study hypervariable polymorphisms: Statistical considerations of determining relatedness and population distances. DNA fingerprinting: state of the science 67: 153-175.

Chang TT (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica, 25: 425-441.

Donnelly P (2008). Progress and challenges in genome-wide association studies in humans. Nature, 456: 728-731.

Ersoz ES, Yu J, Buckler ES (2008). Applications of linkage disequilibrium and association mapping in crop plants. Genomics-assisted crop improvement. pp. 97-119.

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.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, 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.

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.

Han Y, Teng C, Hu Z, Song Y (2008). An optimal method of DNA silver staining in polyacrylamide gels. Electrophoresis, 29: 1355-1358.

Jannink JL, Bink M, Jansen RC (2001). Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6: 337-342.

Lin H, Yamamoto T, Sasaki T, Yano M (2000). Characterization and detection of epistatic interactions of 3 QTLs, hd1, hd2, and hd3, controlling heading date in rice using nearly isogenic lines. Theor. Appl. Genet. 101: 1021-1028.

Liu K, Muse SV (2005). Powermarker: An integrated analysis environment for genetic marker analysis. Bioinformatics, 21: 2128-2129.

Murray MG, Thompson WF (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res. 8: 4321-4325.

Nordborg M, Weigel D (2008). Next-generation genetics in plants. Nature, 456: 720-723.

Suh J, Ahn SN, Cho YC, Kang KH, Choi IS, Kim YG, Suh HS, Hwang HG (2005). Mapping of QTLs for yield traits using an advanced backcross population from a cross between Oryza sativa and O. glaberrima. Korean J. Breed. 37: 214-220.

Wen W, Mei H, Feng F, Yu S, Huang Z, Wu J, Chen L, Xu X, Luo L (2009). Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). Theor. Appl. Gene. 119: 459-470.

Yamamoto T, Kuboki Y, Lin S, Sasaki T, Yano M (1998). Fine mapping of quantitative trait loci hd-1, hd-2 and hd-3, controlling heading date of rice, as single Mendelian factors. Theor. Appl. Gene. 97: 37-44.

Yan W, Li Y, Agrama HA, Luo D, Gao F, Lu X, Ren G (2009). Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). Mol. Breed. 24: 277–292.

Yano M, Harushima Y, Nagamura Y, Kurata N, Minobe Y, Sasaki T (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. Theor. Appl. Genet. 95: 1025-1032.

Yao J, Wang L, Liu L, Zhao C, Zheng Y (2009). Association mapping of agronomic traits on chromosome 2A of wheat. Gene. 137: 67-75.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, 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.

YunFei Z, GuoPing S, XuGao C (2007). Characteristics of the short rachillae of rice from archaeological sites dating to 7000 years ago. Chin. Sci. Bullet. 52: 1654-1660.

Zhu C, Gore M, Buckler ES, Yu J (2008). Status and prospects of association mapping in plants. Plant Genome, 1: 5-20.

Zong Y, Chen Z, Innes J, Chen C, Wang Z, Wang H (2007). Fire and flood management of coastal swamp enabled first rice paddy cultivation in east china. Nature, 449: 459-462.