ORIGINAL PAPER

Genetic diversity and population structure of a diverse set of rice germplasm for association mapping

Liang Jin · Yan Lu · Peng Xiao · Mei Sun · Harold Corke · Jinsong Bao

Received: 4 June 2009 / Accepted: 12 March 2010 / Published online: 4 April 2010

Abstract Germplasm diversity is the mainstay for crop improvement and genetic dissection of complex traits. Understanding genetic diversity, population structure, and the level and distribution of linkage disequilibrium (LD) in target populations is of great importance and a prerequisite for association mapping. In this study, 100 genome-wide simple sequence repeat (SSR) markers were used to assess genetic diversity, population structure, and LD of 416 rice accessions including landraces, cultivars and breeding lines collected mostly in China. A model-based population structure analysis divided the rice materials into seven subpopulations. 63% of the SSR pairs in these accessions were in LD, which was mostly due to an overall population structure, since the number of locus pairs in LD was reduced sharply within each subpopulation, with the SSR pairs in LD ranging from 5.9 to 22.9%. Among those SSR pairs showing significant LD, the intrachromosomal LD had an average of 25–50 cM in different subpopulations. Analysis of the phenotypic diversity of 25 traits showed that the population structure accounted for an average of 22.4% of phenotypic variation. An example association mapping for starch quality traits using both the candidate gene mapping and genome-wide mapping strategies based on the estimated population structure was conducted. Candidate gene mapping confirmed that the *Wx* and starch synthase IIa (*SSIIa*) genes could be identified as strongly associated with apparent amylose content (AAC) and pasting temperature (PT), respectively. More importantly, we revealed that the *Wx* gene was also strongly associated with PT. In addition to the major genes, we found five and seven SSRs were associated with AAC and PT, respectively, some of which have not been detected in previous linkage mapping studies. The results suggested that the population may be useful for the genome-wide marker–trait association mapping. This new association population has the potential to identify quantitative trait loci (QTL) with small effects, which will aid in dissecting complex traits and in exploiting the rich diversity present in rice germplasm.

Introduction

The rapid human population growth in the world is boosting demand for a corresponding increase in crop grain yield. Understanding the molecular genetic control of phenotypic variation, such as yield and quality-related traits, remains a major task in the genetic study of important cereal crops. As the most important staple food crop for more than half of the world's population, rice (*Oryza sativa* L.) has been receiving increasing attention in genetic dissection of simple or complex traits using linkage mapping. Several quantitative trait loci (QTLs) have been successfully isolated and functionally characterized using map-based cloning techniques (Ren et al. 2005; Li et al.

Communicated by T. Sasaki.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1324-7) contains supplementary material, which is available to authorized users.

L. Jin · Y. Lu · P. Xiao · J. Bao (✉)
Institute of Nuclear Agricultural Sciences,
Key Laboratory of Zhejiang Province and Chinese Ministry
of Agriculture for Nuclear-Agricultural Sciences,
Zhejiang University, Hua Jiachi Campus,
Hangzhou 310029, People's Republic of China
e-mail: jsbao@zju.edu.cn

M. Sun · H. Corke (✉)
School of Biological Sciences, The University of Hong Kong,
Hong Kong, People's Republic of China
e-mail: hcorke@hku.hk

2006; Konishi et al. 2006; Song et al. 2007; Xue et al. 2008).

Association mapping is another effective approach to connecting structural genomics and phenomics in plants (Thornsberry et al. 2001), if information on population structure and linkage disequilibrium (LD) is available. Many important crops have complex population structure arising from a long domestication and breeding history (Flint-Garcia et al. 2003). If association analysis is performed on these populations without considering the effects of population structure, spurious association between genotypic and phenotypic variation may be detected because of the unequal allele frequency distribution between subgroups (Knowler et al. 1988). With the development of statistical methods, independent markers that are distributed throughout the genome can be successfully used to detect population structures (Pritchard et al. 2000a, b). In addition to population structure, the extent and distribution of LD across the genome also affects the resolution of association mapping (Remington et al. 2001). The pattern of LD is generally shaped by population history, but other factors such as population structure, selection, mutation, relatedness, and genetic drift also have an effect on LD. However, LD caused by linkage is the most important for association mapping (Stich et al. 2005).

In rice, population structure and its effect on diversity and LD have been reported before. Garris et al. (2005) detected five major groups from a diverse sample of 234 rice accessions including *indica*, *aus*, *tropical japonica*, *temperate japonica* and *aromatic* and suggested that a higher degree of resolution of population structure is needed to effectively utilize LD for association mapping. The rice population was highly structured and significant LD surrounding the *Xa5* locus was observed between sites up to 100 kb apart (Garris et al. 2003). 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. Although the level of LD may vary across the genome because of different recombination rates, selective pressures, etc., these studies seem that LD decays in rice at 1 cM or less (assuming an average of 250 kb/cM across the genome) (Agrama et al. 2007). Mather et al. (2007) used unlinked SNPs to determine the amount of background LD in five 500-kb regions of the rice genome in three major cultivated rice varieties (*indica*, *tropical japonica*, and *temperate japonica*) and in the wild ancestor of Asian rice, *Oryza rufipogon*, and found that the extent of LD is greatest in *temperate japonica* (approximately 500 kb or over), followed by *tropical japonica* (approximately 150 kb) and *indica* (approximately 75 kb), compared to LD in *O. rufipogon* which extends over a much short distance ($\ll$40 kb). However, other studies using different rice accessions indicated that LD decays at 20–30 cM (Agrama et al. 2007; Agrama and Eizenga 2008). These studies suggest that the extent of LD varies among different genomic regions (Mather et al. 2007), different rice accessions studied (Agrama and Eizenga 2008).

The objectives of this study were (1) to evaluate the population structure and genetic diversity of a set of rice materials mainly collected in China; (2) to detect the extent of LD between pairs of SSR markers on a genome-wide scale in rice; and (3) to demonstrate the utility of association mapping for starch quality as an example based on population structure.

Materials and methods

Plant materials

A total of 416 rice accessions were selected mostly from the germplasm centers and various rice breeding programs in China and several from USDA-ARS, Rice Research Unit, USA. The rice accessions with code numbers between BP201 and BP400 were landraces, while the others were cultivars or breeding lines (Supplementary Table 1).

Phenotypic data

All the accessions in this study were planted from late May to October in 2 years, 2006 and 2007, at the Zhejiang University farm, Hangzhou, China. Four rows with six plants per row were planted for each variety in the field. The leaf tissues of each accession were harvested for DNA extraction, and the rice grains were harvested for analysis of grain color and nutritional quality properties (Shen et al. 2009). Data for agronomic traits, such as heading date, plant height, flag leaf length, flag leaf width, and panicle length were averaged from five individuals. The grain length, grain width, grain length/width ratio were averaged from ten grains, and the 1,000-grain weight were measured in triplicate from 100 grains.

After being air-dried and stored at room temperature for 3 months, all rice samples were dehulled to brown rice using a Satake Rice Machine (Satake Corp. Japan). After measurement of the grain color (see below), the brown rice was then ground to pass through a 100-mesh sieve on a Cyclone Sample Mill (UDY Corporation, Fort Collins, Colorado, USA).

The color of rice grain sample (brown rice) was measured with a TC-PIIG automatic color difference meter (Beijing Optical Instrument Factory, Beijing, China). Color measurements were expressed as tristimulus parameters,

L^*, a^*, and b^*. L^* indicates lightness (100 white and 0 black). a^* indicates redness-greenness and b^* indicates yellowness-blueness. In addition, the chroma (C) value indicates color intensity or saturation, calculated as $C = (a^{*2} + b^{*2})^{1/2}$, and Hue angle was calculated as $H° = \tan^{-1}(b^*/a^*)$ (Bao et al. 2005; Shen et al. 2009).

Wholemeal brown flour (1 g) of each accession was extracted with 25 mL of methanol containing 1% HCl for 24 h at 24°C. The procedure was repeated twice. The methanolic extracts were centrifuged at 4,000*g* for 15 min, and the supernatants were pooled and stored at 4°C. Total phenolic content was assayed by a Folin–Ciocalteu colorimetric method and expressed as milligrams of gallic acid equivalent (mg GAE) per 100 g of dry weight (Bao et al. 2005; Shen et al. 2009). Total flavonoid content was determined by a colorimetric method, calculated using the standard rutin curve, and expressed as milligrams of rutin equivalent (mg RE) per 100 g of dry weight (Bao et al. 2005; Shen et al. 2009). Total antioxidant capacity of rice extracts was measured spectrophotometrically by the improved 2,2-azino-bis-(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS) radical cation method (Bao et al. 2005; Shen et al. 2009). Results were expressed in terms of Trolox equivalent antioxidant capacity (TEAC, mM Trolox equivalents per 100 g dry weight).

The apparent amylose content (AAC) and pasting properties of the milled rice flour were determined in a previous study (Bao et al. 2006a, c) and the data were used here for simulation testing of the marker–traits association.

Heritability

Analysis of variance (ANOVA) was carried out to determine genotypic and environmental variances among the traits measured in two environments using the general linear model procedure (Proc glm) with the SAS program version 8 (SAS Institute Inc., Cary, NC). Broad-sense heritability was calculated as $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2/n)$, where σ_g^2 is the genotypic variance, σ_e^2 is the environmental variance, n is the number of environments (Wang et al. 2007).

SSR marker genotyping

DNA was extracted following a CTAB procedure (Doyle 1991). Microsatellite markers (100) located on the 12 chromosomes were selected from the core set developed and mapped by McCouch et al. (2002). Each 20 µL PCR reaction consisted of 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% Triton X 100, 2 mM $MgCl_2$, 0.1 mM dNTPs, 200 nM primers, 1 unit of *Taq* polymerase, and 50 ng of genomic DNA. All amplifications were performed on a MG96G thermal cycler (LongGene Scientific Instruments, Co. Ltd, Hangzhou, China) under the following conditions: (1) predenature at 95°C for 5 min; (2) 35 cycles of run, each followed by denaturation at 95°C for 1 min, annealing at 55–60°C (dependent on primers) for 45 s and extension at 72°C for 1 min; (3) final extension at 72°C for 7 min. PCR products were run on 8% denaturing polyacrylamide gel at 100 V for 150 min and marker bands were visualized using silver staining. Information on primer sequences and PCR amplification conditions for each set of primers are available at http://www.gramene.org/.

Genetic diversity, phylogenetic analysis and population structure

The summary statistics including the number of alleles per locus, major allele frequency, gene diversity, polymorphism information content (PIC) values, and classical F_{st} values were determined using PowerMarker version 3.25 (Liu and Muse 2005, http://statgen.ncsu.edu/powermarker/).

Nei's distance (Nei et al. 1983) was calculated and used for the unrooted phylogeny reconstruction using neighbor-joining method as implemented in PowerMarker with the tree viewed using MEGA 4.0 (Tamura et al. 2007, http://www.megasoftware.net/).

Analysis of population structure among rice accessions was performed using the software package STRUCTURE (Pritchard and Wen 2004, http://pritch.bsd.uchicago.edu/software.html) in its revised version 2.2 (Falush et al. 2003, 2007). The optimum number of populations (K) was selected after five independent runs of a burn-in of 500,000 iterations followed by 500,000 iterations for each value of K (testing from $K = 2$ to $K = 10$). A model-based clustering algorithm was applied that identified subgroups with distinctive allele frequencies and placed individuals into K clusters, where K is chosen in advance but can be varied for independent runs of the algorithm. The most likely number of clusters (K) was selected by comparing the logarithmized probabilities of data [$\Pr(X|K)$] and α value for each value of K according to Pritchard et al. (2000a).

Linkage disequilibrium

Linkage disequilibrium was evaluated for each pair of SSR loci using TASSEL 2.0.1 (http://www2.maizegenetics.net/index.php?page=bioinformatics/tassel/index.html), both on all accessions and on the clusters as inferred by STRUCTURE. D' and r^2 LD measures modified for multiple loci were used (Hedrick 1987). Significance (P values) of D' for each SSR pair was determined by 100,000 permutations. For each SSR locus, the rare alleles (i.e., those present in less than 1% of the accessions) were combined into one allelic class described by Maccaferri et al. (2005).

Association analysis

To compare phenotypes of the seven groups identified by STRUCTURE, ANOVA was employed with the SAS program version 8 (SAS Institute Inc., Cary, NC), followed by multiple means comparisons among these groups. Association between marker alleles and different starch physicochemical property data was performed with TASSEL (trait analysis by association, evolution and linkage) Version 2.0.1 software, taking into account gross level population structure (*Q*) (Bradbury et al. 2007). The *P* value (marker) determining whether a marker (QTL) is associated with the trait and the R^2 (marker) indicating the fraction of the total variation explained by the marker were reported.

Results

Genetic diversity

A total of 100 SSR markers, randomly distributed across the genome, were used to evaluate the genetic diversity of the population. All of the 100 SSR markers were polymorphic across the 416 rice accessions and a total of 390 alleles were detected (Supplementary Table 1). The average number of alleles per locus was 3.9, ranging from 2 to 9. The average gene diversity was 0.4736, ranging from 0.0473 to 0.7618. The average PIC value was 0.4214, ranging from 0.0466 to 0.7216 (Supplementary Table 2).

Forty-seven rare alleles, defined as those alleles with a frequency less than 1% were identified at 43 loci. Among these rare alleles, 12 unique rare alleles were found in accession BP135.

Population structure

The model-based simulation of population structure using SSRs showed that likelihood was maximized and α minimized when the number of populations was set at seven, suggesting that these rice accessions can be grouped into seven subpopulations, as inferred from the model, here denoted as POP1, POP2, POP3, POP4, POP5, POP6 and POP7, respectively (Fig. 1; Supplementary Table 1). When

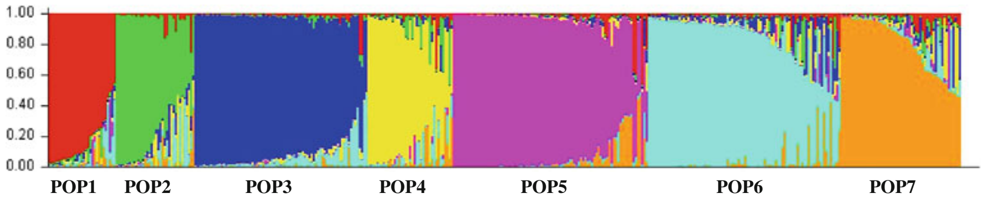

Fig. 1 Population structure of 416 rice accessions based on 100 SSRs ($K = 7$). Each accession is represented by a *thin vertical line*, which can be partitioned into *seven colored segments* that represent the estimated membership probabilities (*Q*) of the individual to the seven clusters, and then all accessions were sorted by *Q*

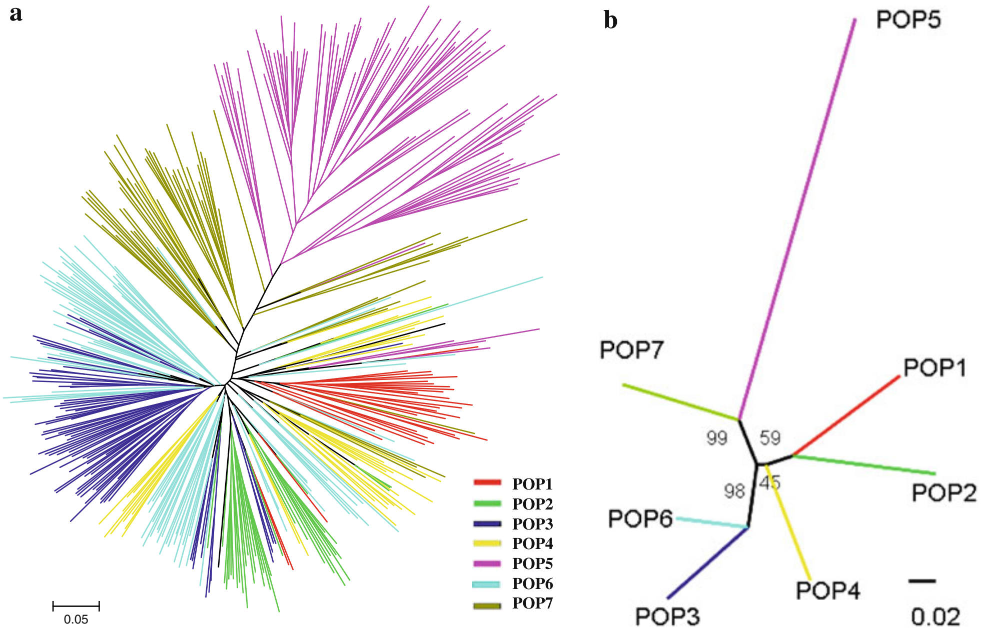

Fig. 2 Unrooted neighbor-joining trees of 416 accessions (**a**) and seven subpopulations (**b**) based on Nei's genetic distances

K was set to 2, most of rice accessions were unambiguously divided into *indica* and *japonica* groups. However, there were eight accessions (BP459, BP464, BP465, BP470, BP474, BP476, BP487, and BP532) showing a membership probability of less than 0.65 (Supplementary Fig. 1). These eight accessions were actually derived from *indica*/*japonica* hybridization. The typical *japonica* rice, such as BP017 (Xiushui 11), BP018 (Zheda 104), BP050 (Xiushui 110), BP518 (Ning 67B), BP579 (Jingxi 17), BP605 (Zhonghua 11) and tropical *japonica* rice Lemont (BP015) and Azucena (BP021) were all clustered into the *japonica* group. The typical *indica* rice such as BP001 (Zhefu 802), BP003 (Jiayu 293), BP004 (Zhe 733), BP005 (Zhefu 504) and BP009 (Zhenong 921) that were officially released in Zhejiang Province, China, during last three decades were all clustered into different *indica* groups. According to the membership pattern when $K = 7$, the rice accessions in POP1, POP2, POP3, POP4, POP6 and POP7 belonged to the *indica* group, whereas only those in POP5 belonged to the *japonica* group. The eight rice accessions (BP459, BP464, BP465, BP470, BP474, BP476, BP487, and BP532) mentioned above were assigned to POP5, but with a much smaller membership probability (Supplementary Table 1).

A neighbor-joining tree of 416 accessions was constructed based on Nei's genetic distance (Fig. 2a). It revealed genetic relationships fairly consistent with the STRUCTURE-based membership assignment for most of the accessions. However, a few rice accessions were displaying as admixtures in different clusters. For example, four rice accessions (BP459, BP463, BP465, BP474) were assigned to POP5 by STRUCTURE, but they were not clustered into the *japonica* group in the neighbor-joining tree (shown in pink color in Fig. 2a). These "misplaced" rice accessions were breeding lines derived from *indica*/*japonica* crosses, as explained above. With a few exceptions, the neighbor-joining tree showed that the rice accessions could be well differentiated according to their subspecies affiliation, i.e., *indica* or *japonica*.

Genetic relationships among populations

The overall F_{st} among the seven subpopulations was 0.2531 (95% confidence interval 0.2264–0.2810), with F_{st} for each locus ranging from 0.0089 to 0.7306. Pairwise comparison on the basis of the values of F_{st} could be interpreted as standardized population distances between two populations. The pairwise F_{st} value in this study ranged from 0.0734 between POP3 and POP6 to 0.4273 between POP2 and POP5, with an average pairwise F_{st} of 0.2220 (Table 1). The genetic distance data agreed with the F_{st} estimates. POP3 showed the smallest genetic distance with POP6 (0.04), whereas POP2 showed the greatest genetic distance with POP5 (0.2551). As expected, the greatest F_{st} or genetic distance was found between *japonica* (POP5) and *indica* (POP1, POP2, POP3, POP4, POP6 and POP7) groups. Within *indica* rice, POP1 showed the greatest genetic distance with POP3 (0.1126). A neighbor-joining tree of the seven subpopulations (Fig. 2b) was constructed based on pairwise Nei's genetic distances given in Table 1, showing that these subpopulations could be clustered into three groups. The groupings of POP5 with POP7, and POP3 with POP6 both had high bootstrap support (99 and 98%, respectively), whereas the grouping of POP1, POP2 and POP4 was not highly supported. The genetic relationships among the subpopulations are largely concordant with those revealed in Fig. 2a, where all 416 individual rice accessions were included in the neighbor-joining analysis.

Table 1 Pairwise estimates of F_{st} and Nei's genetic distance based on 100 SSR loci among the seven model-based subpopulations

Cluster	POP1	POP2	POP3	POP4	POP5	POP6	POP7
POP1		0.0831	0.1126	0.0799	0.2177	0.0823	0.0844
POP2	0.1591		0.0957	0.0720	0.2551	0.0890	0.0958
POP3	0.2238	0.2221		0.0854	0.2327	0.0400	0.1034
POP4	0.1477	0.1538	0.1460		0.2014	0.0614	0.0755
POP5	0.3935	0.4273	0.3865	0.3583		0.1980	0.1692
POP6	0.1804	0.2184	0.0734	0.1057	0.3424		0.0733
POP7	0.1709	0.1975	0.1819	0.1438	0.3018	0.1279	

Genetic distance estimates appear above the diagonal and pairwise F_{st} appears below the diagonal

Genetic diversity of subpopulations

The genetic diversity for each subpopulation was assessed (Table 2). POP7 had a highest gene diversity of 0.4114, with a total of 316 alleles or 3.16 alleles per locus, followed by POP5 with a gene diversity of 0.4077, a total of 330 alleles or 3.3 alleles per locus. Among the 390 alleles detected in the total populations, 36 (9.23%) were subpopulation-specific or private alleles. POP6 had more private alleles (20 or 5.13%) than others (Table 2).

Linkage disequilibrium

The extent of LD was assessed among all 4,951 pairs of SSRs loci for all accessions as well as for the seven subpopulations separately (Table 3). Across all accessions, as many as 62.8% of the total marker pairs were in LD (based

Table 2 Summary statistics for each subpopulation

Subpopulation	Sample size	Alleles	Alleles/locus	Gene diversity	PIC	No. of population-specific alleles
POP1	31	302	3.02	0.3678	0.3246	1
POP2	36	260	2.60	0.2944	0.2557	1
POP3	79	292	2.92	0.3327	0.2919	3
POP4	39	284	2.84	0.3611	0.3176	1
POP5	88	330	3.30	0.4077	0.3587	7
POP6	88	336	3.36	0.3762	0.3347	20
POP7	55	316	3.16	0.4114	0.3633	3
Overall	416	390	3.90	0.4736	0.4214	47

PIC polymorphism information content

Table 3 Percentage of SSR locus pairs in significant ($P < 0.05$) linkage disequilibrium (LD)

	Markers on the same chromosome		Markers from different chromosomes		Total	
	No. of locus pairs in LD[a]	Fraction of locus pairs	No. of locus pairs in LD[a]	Fraction of locus pairs	No. of locus pairs in LD	Fraction of locus pairs
All	210 (363)	0.579	2896 (4587)	0.627	3,106	0.628
POP1	18 (333)	0.054	240 (4,074)	0.059	258	0.059
POP2	32 (256)	0.125	229 (3,283)	0.070	261	0.074
POP3	31 (286)	0.108	365 (3,548)	0.103	396	0.103
POP4	33 (319)	0.103	275 (3,966)	0.069	308	0.072
POP5	88 (354)	0.249	1,004 (4,415)	0.227	1,092	0.229
POP6	44 (327)	0.135	536 (4,051)	0.132	580	0.133
POP7	43 (347)	0.124	505 (4,354)	0.116	548	0.117

[a] The values in parenthesis are total number of locus pairs

on D', $P < 0.05$) after Bonferroni correction. The percentage of LD for pairs of markers from the different chromosome was 62.7%, only slightly higher than the 57.9% for markers on the same chromosomes (Table 3).

The frequency of pairs of loci with significant LD ($P < 0.05$, based on D') was reduced by more than half when LD was calculated within each subpopulation (Table 3). The lowest percentage of locus pairs in LD (5.9%) was found in POP1, whereas POP5 had the highest percentage (22.9%).

At the whole population level, the r^2 values among all the SSR pairs ranged from 0 to 0.4556. Among the inter-chromosomal pairs, r^2 values ranged from 0.019 to 0.046 (Table 4). Figure 3 showed the distribution of r^2 values for the inter-chromosomal pairs for the whole population. Distributions were shifted toward the low r^2 values, which were expected as these SSRs were unlinked. The r^2 values reflect disequilibrium due to either chance or evolutionary forces that affect variation across the entire genome. The 75th percentile of the r^2 values was 0.0614, which was used as background LD to define elevated LD (Mather et al. 2007). As shown in Fig. 4, intrachromosomal LD was very common for distances <50 cM and occasionally LD occurred between SSR loci that were further apart (Fig. 3). Within the subpopulations, the 75th percentile background r^2 values of inter-chromosomal pairs ranged from 0.0212 (POP6) to 0.0515 (POP1) (Fig. 5). The intrachromosomal LD decaying pattern varied among different subpopulations. For POP1, POP3 and POP6, LD extended to ~50 cM in most cases, but some pairs extended to 150 cM, and the average LD decayed at 25–30 cM. For POP2 and POP4, most LD were found at <75 cM, some pairs extended to 100 cM, and the average LD decayed around 30–40 cM. For POP5 and POP7, the extent of LD went further to 100 cM and the average LD decayed around 50 cM.

Diversity of phenotypic traits

The agronomic and quality trait data on 416 rice accessions grown in 2006 and 2007 are summarized in Table 5. Substantial variation existed in nearly all the 25 traits measured in this diverse germplasm set. Heading date was the most striking example of phenotypic variation, ranging

Table 4 Linkage disequilibrium (LD) statistics D' and r^2 for each subpopulation

	Markers on the same chromosome		Markers from different chromosomes		Total	
	D'	r^2	D'	r^2	D'	r^2
All	0.267	0.028	0.267	0.027	0.267	0.027
POP1	0.617	0.044	0.627	0.046	0.627	0.049
POP2	0.607	0.052	0.588	0.038	0.589	0.039
POP3	0.449	0.019	0.450	0.019	0.450	0.019
POP4	0.532	0.036	0.533	0.035	0.533	0.036
POP5	0.370	0.025	0.373	0.024	0.373	0.024
POP6	0.467	0.015	0.477	0.019	0.476	0.019
POP7	0.412	0.029	0.409	0.028	0.409	0.028

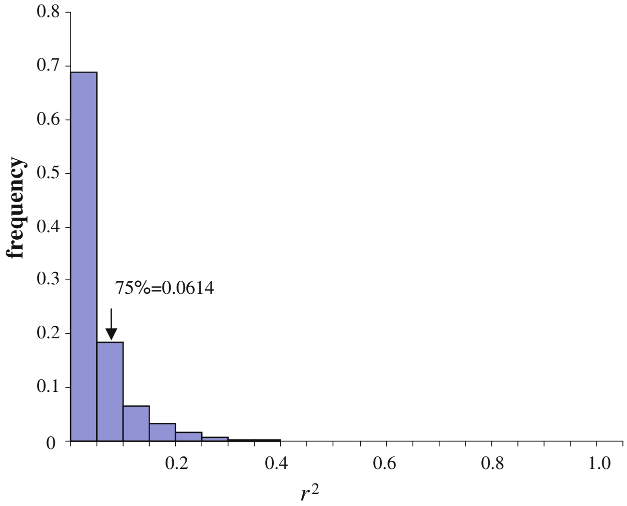

Fig. 3 Distribution of linkage disequilibrium r^2 values for inter-chromosomal SSR pairs

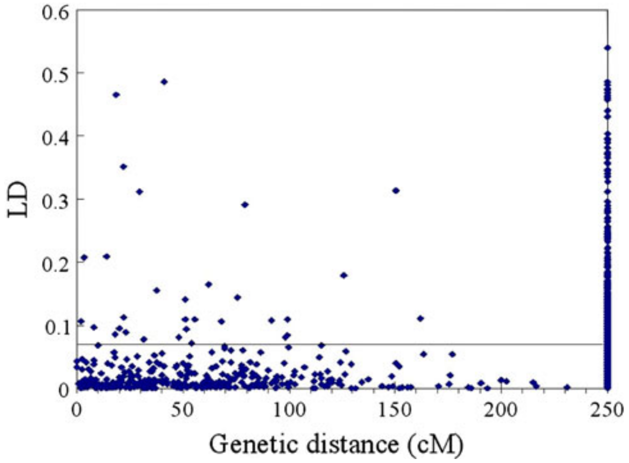

Fig. 4 Scatterplot of linkage disequilibrium (LD, r^2) between SSR pairs versus inter-marker genetic distance in cM for the whole population. The observed LD values for inter-chromosomal markers are compiled in a single file at 250 cM. The *horizontal line* indicates the 75th percentile from the distribution of inter-chromosomal SSR pairs (Fig. 3)

from 50 to 126 days after sowing. The 1,000-grain weight ranged from 12.6 to 34.8 g across 2 years. Genetic diversity also existed in starch quality (Bao et al. 2006c), grain color and nutritional quality (Shen et al. 2009).

The broad-sense heritability for agronomic traits was estimated (Table 5). Heading date, plant height, grain length/width ratio, panicle length and 1,000-grain weight were highly heritable, all with an estimated $H^2 > 0.90$. Heritability for grain length (0.686) and grain width (0.711) was a little lower (Table 5).

Population structure was the dominant factor in phenotypic variation, accounting for an average of 22.4% of the phenotypic variation across all traits in this study (Table 5). Flag leaf width and 1,000-grain weight were less affected by population structure, with less than 10% of the variation attributable to it. In contrast, plant height and grain length/width ratio were highly affected by population structure, with more than 40% of the variation attributable to it. The starch quality traits were affected by population structure to different degrees. More than 40% of variation in apparent amylose content (AAC) and gel hardness could be accounted for by population structure, whereas it might account for 11–28% of variation in other traits. The grain color traits and nutritional quality traits were less affected by population structure. Except for the $a*$ of grain color, population structure accounted for less than 10% of the phenotypic variation (Table 5). As the average effect of more than 20% is roughly equivalent to a major QTL detected in interval mapping or in association analysis, it is necessary to examine the effect of population structure before performing association test for any given trait.

An example of association test based on structured population: starch quality and starch synthesizing genes

Primary association mapping for starch quality using gene markers derived from major starch biosynthesizing genes was previously conducted without consideration of the

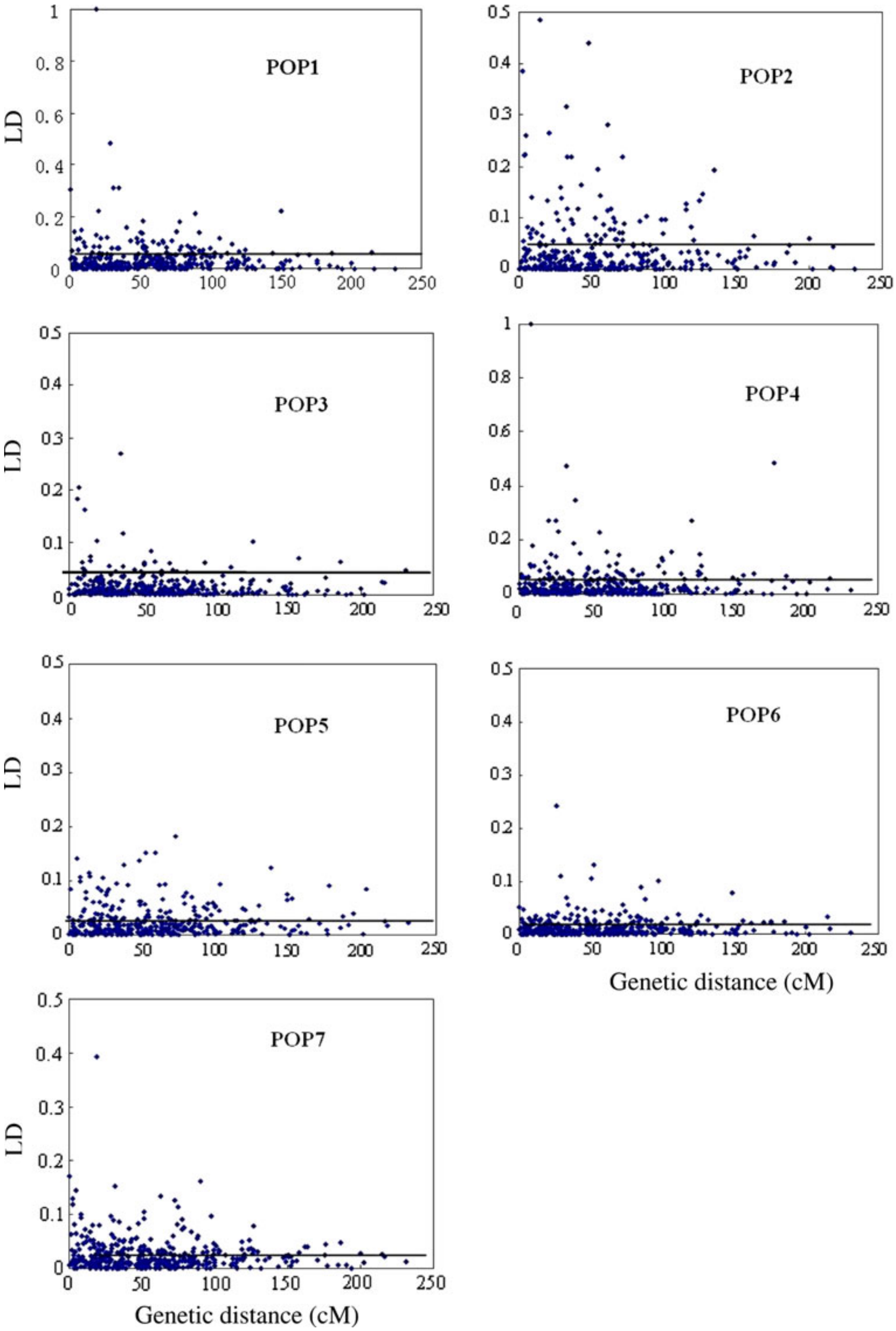

Fig. 5 Scatterplot of linkage disequilibrium (LD, r^2) between SSR pairs versus inter-marker genetic distance in cM for each of the seven subpopulations

effect of population structure (Bao et al. 2006a, b). Based on population structure identified in the present study, we reconducted association mapping for starch quality using seven gene markers and 100 SSRs. It is well known that the *Wx* gene is the major gene responsible for amylose synthesis, thus *Wx* was identified as the candidate gene for AAC ($P < 2.64 \times 10^{-95}$; Table 5). Other genes that were closely linked with *Wx*, such as *SS1* and *SSIIa*, were also associated with AAC (Table 6). In addition to the major genes, we found five SSRs that were also associated with AAC ($P < 0.001$), although each could explain only about 3% of the phenotypic variation (Table 7).

It is well known that the *SSIIa* gene is the major gene responsible for amylopectin synthesis, which determines the gelatinization temperature of starch. Thus, the *SSIIa* gene was identified as the major candidate gene for the

Table 5 Descriptive statistics and percentage of phenotypic variation explained by population structure for 25 traits

Trait	Year	Mean ± SD	Range	H^2	R^2[a]
Agronomic traits					
Heating date (days)	2006	73.2 ± 12.5	50.0–126.0	0.976	34.4
	2007	73.4 ± 12.4	51.0–119.0		32.7
Plant height (cm)	2006	105.3 ± 27.0	54.3–182.4	0.908	46.6
	2007	110.8 ± 30.2	53.0–188.1		42.9
Flag leaf length (cm)	2006	36.2 ± 9.5	15.4–77.3	0.782	23.1
	2007	36.8 ± 9.0	15.9–66.7		23.8
Flag leaf width (cm)	2006	1.73 ± 0.27	0.90–2.68	0.802	9.3
	2007	1.70 ± 0.26	1.00–2.63		7.9
Grain length (mm)	2006	8.44 ± 0.90	6.76–11.18	0.686	31.9
	2007	8.52 ± 0.91	6.77–11.21		30.7
Grain width (mm)	2006	2.99 ± 0.35	2.19–4.07	0.711	39.1
	2007	3.03 ± 0.36	2.26–4.86		42.1
Grain length/width ratio	2006	2.87 ± 0.56	1.87–4.64	0.991	40.1
	2007	2.86 ± 0.55	1.83–4.47		40.9
Panicle length (cm)	2006	24.2 ± 4.0	13.1–38.4	0.918	18.6
	2007	23.3 ± 3.6	13.1–23.3		15.5
1,000-grain weight (g)	2006	23.2 ± 3.3	12.9–33.4	0.906	6.2
	2007	22.8 ± 4.6	12.6–34.8		6.0
Starch quality traits[b]					
Apparent amylose content, AAC (%)		23.1 ± 5.7	1.6–32.6	NA	41.5
Peak viscosity (RVU)		241.7 ± 33.5	81.7–305.4	NA	11.5
Hot paste viscosity (RVU)		173.1 ± 30.9	43.0–248.1	NA	14.9
Cool paste viscosity (RVU)		307.5 ± 51.8	56.5–409.1	NA	21.3
Breakdown (RVU)		68.6 ± 19.7	19.6–141.3	NA	18.2
Setback viscosity (RVU)		65.9 ± 37.1	−82.6–193.2	NA	27.0
Pasting temperature, PT (°C)		74.9 ± 2.8	67.5–81.5	NA	18.3
Hardness (g)		29.2 ± 11.1	0.8–61.1	NA	43.6
Grain color traits					
*L**	2006	51.8 ± 9.5	14.1–67.4	NA	8.5
*a**	2006	3.46 ± 2.90	−12.9–13.8	NA	21.1
*b**	2006	14.3 ± 2.1	−1.8–18.7	NA	6.7
C	2006	15.1 ± 1.4	9.48–19.7	NA	9.5
H°	2006	72.5 ± 31.0	−193.3–101.4	NA	7.0
Nutritional quality traits					
Phenolics (mg GAE/100 g)	2006	208.1 ± 155.6	108.1–1,244.9	NA	8.0
Flavonoids (mg RE/100 g)	2006	134.6 ± 20.6	88.51–286.29	NA	5.0
Antioxidant capacity (mMTAEC)	2006	0.474 ± 0.752	0.0120–5.533	NA	8.4

NA not available

[a] Percentage of phenotypic variation explained by population structure

[b] These data were adapted from Bao et al. (2006c)

pasting temperature (PT) ($P = 4.43 \times 10^{-65}$; Table 6). The *Wx* is also associated with PT ($P < 0.0001$), whereas *SS1*, which is located between *SSIIa* and *Wx*, was not associated with PT. In addition, seven SSRs were found to be associated with PT ($P < 0.001$), of which RM253 and RM276 could explain 5–10% of the variation (Table 7).

Discussion

Identifying genetic variants underlying quantitative trait variation in crops is a challenging task facing plant breeders. Linkage mapping, and more recently, association or linkage disequilibrium (LD) mapping have been applied

Table 6 Association of starch synthesizing gene markers with apparent amylose content (AAC) and pasting temperature (PT)

Gene markers	AAC		PT	
	P	R^2	*P*	R^2
Wx SNP	1.02×10^{-105}	0.3979	1.04×10^{-4}	0.0313
Wx SSR	2.64×10^{-95}	0.3968	0.008	0.0464
SS1 SSR	6.51×10^{-11}	0.0689	0.1584	0.0109
SSIIa SNP	3.73×10^{-4}	0.0188	4.43×10^{-65}	0.4312
Sbe1 SSR	0.0217	0.0141	0.0162	0.0217
Sbe1 STS	0.0016	0.0145	0.0034	0.0179
Sbe3 SNP	0.8005	9.39×10^{-5}	0.9824	1.04×10^{-6}

Table 7 Association of other SSR markers with apparent amylose content (AAC) and pasting temperature (PT) ($P < 0.001$)

Markers	Chromosome	Position	*P*	R^2
AAC				
RM122	5	0	5.01×10^{-4}	0.0217
RM161	5	96.9	4.00×10^{-4}	0.0225
RM346	7	47	2.70×10^{-6}	0.0444
RM336	7	61	0.001	0.0328
RM126	8	57	4.13×10^{-4}	0.0187
PT				
RM253	6	37	3.88×10^{-5}	0.0520
RM276	6	40.3	1.13×10^{-10}	0.1042
RM346	7	47	2.48×10^{-4}	0.0449
RM278	9	77.5	1.38×10^{-5}	0.0523
RM222	10	11.3	7.50×10^{-4}	0.0395
RM484	10	97.3	9.76×10^{-4}	0.0295
RM206	11	102.9	5.52×10^{-4}	0.0460

to elucidate the genetic basis of natural variation in important quantitative traits. In the conventional linkage mapping studies, the LD required for mapping quantitative trait loci is generated as a result of the mating design, whereas association studies exploit the LD already present in the population of interest (Maurer et al. 2006). The basic components required for conducting association mapping have been proposed by Whitt and Buckler (2003) and Flint-Garcia et al. (2005), i.e., germplasm choice, estimation of population structure, trait evaluation, identification of candidate polymorphisms, and statistical analysis.

Genetic and phenotypic diversity in rice germplasm

Choice of germplasm is critical to the success of association mapping. Good examples of populations for use in association mapping in plant genetics are breeding and gene bank collections of cultivars, breeding lines, germplasm, etc. (Malosetti et al. 2007). The rice materials used in this study encompass landrace, cultivars, and breeding lines, mostly collected from rice germplasm centers and breeding programs. No wild rice accessions (*Oryza rufipogon*) were used in this study, because of their easy shattering of the grains. Also no rice material from other South Asian countries was used because some materials could not flower at the test location. In this study, the PIC values of our rice collections ranged from 0.0466 to 0.7216 with a mean of 0.4214 (Supplementary Table 1). Xu et al. (2005) reported an average PIC value of 0.74, ranging from 0.17 to 0.92, in the world collections of rice materials, and an average PIC value of 0.50, ranging from 0.02 to 0.88, in the US collections. Agrama and Eizenga (2008) reported that wild relatives (*Oryza* spp.) represented by ten different species had the highest PIC value (0.78), while US cultivars had the lowest value (0.39). The gene diversity values in this study averaged 0.4736, ranging from 0.0473 to 0.7618 (Supplementary Table 2). The US accessions had an average gene diversity of 0.43, ranging for a single marker from 0.03 to 0.86 (Agrama and Eizenga 2008). From these comparisons, the Chinese rice materials used in this study represented molecular diversity comparable to the US cultivars, but less diversity than the wild relatives. The neighbor-joining tree of 416 accessions based on Nei's (1983) genetic distance revealed that these rice materials could be divided into several groups (Fig. 2), representing extensive genetic and phenotypic diversity (Table 5). The diversity of these rice materials indicated that they are suitable for association mapping.

Population structure and LD in rice germplasm

Association mapping exploits existing LD in natural populations. Most association mapping strategies start by first inspecting the population to assess whether groups can be discerned within the population and then testing for marker–trait association within groups. Understanding the population structure is important to avoid identifying spurious associations between phenotype and genotype in association mapping (Pritchard and Rosenberg 1999; Pritchard et al. 2000a). The genetic structure of rice has previously been documented (Agrama et al. 2007; Garris et al. 2005; Glaszmann 1987; Ni et al. 2002; Zhang et al. 2007). Garris et al. (2005) detected 5 major groups from a diverse sample of 234 rice accessions including *indica*, *aus*, *tropical japonica*, *temperate japonica* and *aromatic*. Zhang et al. (2007) analyzed the genetic structure of rice in Guizhou province, China, and revealed that the rice germplasms in Guizhou could be divided into 7 subpopulations. Agrama et al. (2007) described the population structure among 103 rice accessions and detected 8 subpopulations with 123 SSR markers. Similarly, the present study revealed

7 subpopulations among 416 rice accessions with 100 SSRs (Fig. 1), and each subpopulation had 31–88 rice accessions (Table 2). The sample size used in this study is larger than used in the previous studies in rice (e.g., Garris et al. 2003, 2005; Agrama et al. 2007; Rakshit et al. 2007; Mather et al. 2007; Thomson et al. 2007). A larger sample size increases detection power and allows the quantification of the effects of more alleles.

The studies of Garris et al. (2003), Olsen et al. (2006), Mather et al. (2007) and Rakshit et al. (2007) indicated that LD decays at 1 cM or less in rice using DNA sequences, whereas other studies indicated that LD decays at 20–30 cM using SSR markers (Agrama et al. 2007; Agrama and Eizenga 2008). The pattern of LD revealed in this study is similar to that of the US accessions (Agrama and Eizenga 2008) and that of rice germplasm recently introduced into the US that have blast resistance (Agrama et al. 2007). The LD in this set of germplasm did not decay until 25–50 cM (Fig. 5). In maize, the level of genome-wide LD inferred by the SSR markers is much higher than that by the candidate genes (Remington et al. 2001). Remington et al. (2001) explained that the discrepancy could be due to chance alone or a higher percentage of SSR mutations than SNPs that arose during the development of regional maize subpopulations. LD in some *Arabidopsis* populations exceeds 50 cM (Nordborg et al. 2002). These studies suggest that the extent of LD varies among different genomic regions (Mather et al. 2007), different rice accessions studied (Agrama and Eizenga 2008) and different markers used. Thus, in large populations of autogamous species, the stretches of LD extending over several cM are expected. On the basis of the LD decay range in the present study, genome-wide LD mapping is possible using this set of rice materials.

Association mapping in rice

We performed structure-based association mapping for starch quality using both candidate gene mapping and genome-wide association mapping. In candidate gene mapping, the *Wx* and *SSIIa* genes could be identified as strongly associated with AAC and PT, respectively ($P < 4 \times 10^{-65}$). Significant association of *SS1* and *SSIIa* with AAC may be derived from their LD with the *Wx* gene. However, the significant association between *Wx* and PT was not a result of its LD with *SSIIa* gene, because *SS1* which is located between *Wx* and *SSIIa* is not associated with PT (Table 6). Many studies suggested that, in addition to amylopectin, amylose is an important factor determining the gelatinization temperature of starch. Previous studies have indicated that the starch branching enzyme gene 1 (*Sbe1*) is significantly associated with amylose content and pasting temperature (Bao et al. 2006a, b; Han et al. 2004). Using a recombinant inbred line population, Bao et al. (2008) revealed that the *Sbe1* has no relationship with the starch quality. In this study, *Sbe1* was shown to have poor association with AAC and PT, and the variation explained by *Sbe1* was less than 2%. Neither AAC nor PT was found to be associated with *Sbe3* gene (Table 5). In addition to the candidate genes mapping, genome-wide association mapping identified five and seven SSRs that were associated with AAC and PT, respectively. Among them, RM253 (Fan et al. 2005), RM276 (Wang et al. 2007) and RM484 (Fan et al. 2005) have been detected by QTL mapping. Two markers, RM253 and RM276 are closely linked to the *SSIIa* gene explaining 5–10% of the variation of PT, their associations with PT could be a result of LD with *SSIIa*.

It is worthy of note that structure-based association analyses will have little power to detect the effects of individual genes if population structure is found to explain too much of the variation (Flint-Garcia et al. 2005). For example, as plant height is highly correlated with population structure ($R^2 = 46.6\%$) (Table 5), functional alleles whose distributions coincide with population structure will not be detected when association models include population structure estimates. This problem can be seen as the trait and polymorphism associate very strongly when population structure is ignored, but the association disappears when structure is considered (Flint-Garcia et al. 2005). In these cases, alternative association populations would be more useful for evaluating the candidate polymorphisms.

In conclusion, we analyzed LD and investigated population structure present in 416 rice accessions collected mostly in China. Substantial phenotypic and genotypic diversity exist in this germplasm set, allowing for genome-wide association mapping and candidate gene mapping.

Acknowledgments The authors thank Yun Shen and Jian Chen for assistance in SSR marker genotyping. Financial support for this research was provided in part by the Program for New Century Excellent Talents in University (NCET-08-0484), National High Technology Development Project of China (2006AA10Z193), Zhejiang Provincial Natural Science Foundation of China (R3080016), Natural Science Foundation of China (30771327), and a Hong Kong Research Grants Council GRF award (7507/06M).

References

Agrama HA, Eizenga GC (2008) Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its wild relatives. Euphytica 160:339–355

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

Bao JS, Cai Y, Sun M, Wang GY, Corke H (2005) Anthocyanins, flavonols, and free radical scavenging activity of Chinese bayberry (*Myrica rubra*) extracts and their color properties and stability. J Agric Food Chem 53:2327–2332

Bao JS, Corke H, Sun M (2006a) Microsatellites, single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). Theor Appl Genet 113:1185–1196

Bao JS, Corke H, Sun M (2006b) Nucleotide diversity in *starch synthase IIa* and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.). Theor Appl Genet 113:1171–1183

Bao JS, Shen SQ, Sun M, Corke H (2006c) Analysis of genotypic diversity in the starch physicochemical properties of nonwaxy rice: apparent amylose content, pasting viscosity and gel texture. Starch/Stärke 58:259–267

Bao JS, Jin L, Xiao P, Shen SQ, Sun M, Corke H (2008) Starch physicochemical properties and their associations with microsatellite alleles of starch-synthesizing genes in a rice RIL population. J Agric Food Chem 56:1589–1594

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Doyle JJ (1991) DNA protocols for plants-CTAB total DNA isolation. In: Hewitt GM (ed) Molecular techniques in taxonomy. Springer, Berlin, pp 283–293

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Fan CC, Yu XQ, Xing YZ, Xu CG, Luo LJ, Zhang QF (2005) The main effects, epistatic effects and environmental interactions of QTLs on the cooking and eating quality of rice in a doubled-haploid line population. Theor Appl Genet 110:1445–1452

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flint-Garcia SA, Thuillet AC, Yu JM, 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

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

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

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

Han YP, Xu ML, Liu XY, Yan CJ, Korban SS, Chen XL, Gu MH (2004) Genes coding for starch branching enzymes are major contributors to starch viscosity characteristics in *waxy* rice (*Oryza sativa* L.). Plant Sci 166:357–364

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 43:520–526

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. Science 311:1936–1939

Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129

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

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879–889

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:2223–2232

Maurer HP, Knaak C, Melchinger AE, Ouzunova M, Frisch M (2006) Linkage disequilibrium between SSR markers in six pools of elite lines of a European breeding program for hybrid maize. Maydica 51:269–280

McCouch SR, 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

Nei M, Tajima FA, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170

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 P, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

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

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

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software. The University of Chicago Press, Chicago

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rakshit S, Rakshit A, Matsumura H, Takahashi Y, Hasegawa Y, Ito A, Ishii T, Miyashita NT, Terauchi R (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. Theor Appl Genet 114:731–743

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

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat Genet 37:1141–1146

Shen Y, Jin L, Xiao P, Lu Y, Bao JS (2009) Total phenolics, flavonoids, antioxidant capacity in rice grain and their relations to grain color, size and weight. J Cereal Sci 49:106–111

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thomson MJ, Septiningsih EM, Suwardjo F, Santoso TJ, Silitonga TS, McCouch SR (2007) Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. Theor Appl Genet 114:559–5568

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

Wang LQ, Liu WJ, Xu Y, He YQ, Luo LJ, Xing YZ, Xu CG, Zhang QF (2007) Genetic basis of 17 traits and viscosity parameters characterizing the eating and cooking quality of rice grain. Theor Appl Genet 115:463–476

Whitt SR, Buckler ES (2003) Using natural allelic diversity to evaluate gene function. In: Grotewald E (ed) Plant functional genomics: methods and protocols. Methods in molecular biology, vol 236. Humana Press, Inc., Totowa, NJ, pp 123–140

Xu Y, Beachell H, McCouch SR (2005) A marker-based approach to broadening the genetic base of rice in the USA. Crop Sci 44:1947–1959

Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat Genet 40:761–767

Zhang DL, Zhang HL, Wei XH, Qi YW, Wang MX, Sun JL, Ding L, Tang SX, Qiu ZE, Cao YS, Wang XK, Li ZC (2007) Genetic structure and diversity of *Oryza sativa* L. in Guizhou, China. Chin Sci Bull 52:343–351