Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.)

Wen Gui Yan · Yong Li · Hesham A. Agrama ·
Dagang Luo · Fangyuan Gao · Xianjun Lu ·
Guangjun Ren

Abstract Stigma and spikelet characteristics play
an essential role in hybrid seed production. A mini-
core of 90 accessions developed from USDA rice
core collection was phenotyped in field grown for
nine traits of stigma and spikelet and genotyped with
109 DNA markers, 108 SSRs plus an indel. Three
major clusters were built upon Rogers' genetic
distance, indicative of indicas, and temperate and
tropical japonicas. A mixed linear model combining
PC-matrix and K-matrix was adapted for mapping
marker-trait associations. Resulting associations were
adjusted using false discovery rate technique. We
identified 34 marker-trait associations involving 22
SSR markers for eight traits. Four markers were
associated with single stigma exsertion (SStgE), six
with dual exsertion (DStgE) and five with total
exsertion. RM5_Chr1 played major role indicative of
high regression with not only DStgE but also SStgE.
Four markers were associated with spikelet length,
three with width and seven with L/W ratio. Numerous
markers were co-associated with multiple traits that
were phenotypically correlated, i.e. RM12521_Chr2
associated with all three correlated spikelet traits. The
co-association should improve breeding efficiency
because single marker could be used to assist
breeding for multiple traits. Indica entry 1032
(cultivar 50638) and japonica entry 671 (cultivar
Linia 84 Icar) with 80.65 and 75.17% of TStgE,
respectively are recommended to breeder for improv-
ing stigma exsertion.

Abbreviations

Chr	Chromosome
DStgE	Dual stigma exsertion
FAO	Food and Agriculture Organization of the United Nations

Wen Gui Yan, Yong Li and Hesham A. Agrama have
contributed equally to this work.

W. G. Yan (✉)
Dale Bumpers National Rice Research Center, United
States Department of Agriculture, Agricultural Research
Service, 2890 Highway 130 East, Stuttgart, AR 72160,
USA
e-mail: Wengui.Yan@ars.usda.gov

Y. Li · D. Luo · F. Gao · X. Lu · G. Ren
Sichuan Academy of Agricultural Sciences, No. 20
Jingjusi Road, 610066 Chengdu, Sichuan, China

G. Ren
e-mail: rgj80@hotmail.com

H. A. Agrama
Rice Research and Extension Center, University
of Arkansas, 2890 Highway 130 East, Stuttgart,
AR 72160, USA

FDR	False discovery rate
GA	Gibberellin
LD	Linkage disequilibrium
MLM	Mixed linear model
PCA	Principal components analysis
PCR	Polymerase chain reaction
PIC	Polymorphism information content
QTL	Quantitative trait loci
RFLP	Randomized fragment length polymorphism
SpkL	Spikelet length
SpkL/W	Spikelet length/width ratio
SpkW	Spikelet width
SSR	Single sequence repeat
SStgE	Single stigma exsertion
StgL	Stigma length
StgW	Stigma width
StgL/W	Stigma length/width ratio
TStgE	Total stigma exsertion
UPGMA	Unweighted pair group method using arithmetic average
USDA	United States Department of Agriculture

Introduction

Rice is an important staple food crop for more than half of the world's population. As such, increased rice production is vital to global food security. The development of hybrid rice technology has been a key contribution to meeting the challenge presented by an ever increasing population. Hybrid rice exhibits a yield advantage of 15–20% (or more than one ton of paddy rice per hectare) over the best traditional varieties in a large-scale production worldwide (Xu 2003; FAO 2004). Being a self-pollinated crop, commercial production of hybrid seed plays a key role in successful implementation of hybrid rice. Anther dehiscence or pollination and spikelet-flowering in rice occur more or less simultaneously so male sterility has to be adapted to the female parents to prevent self-pollination and secure cross-pollination (Virmani 1994). However, side effects of the male sterility usually create barriers for cross-pollination of the female parent including incomplete exsertion of panicle which prevents access to about 20% of the spikelets and the failure of about 20% of spikelets to open at all (Yan and Li 1987; Tian 1991).

Thus up to 40% of spikelets may not be available for pollination and subsequent seed production if gibberellin (GA_3) technology is not adapted (Yuan 1981; Yuan and Fu 1995). Furthermore, flowering in the male sterile (e.g., female) parents typically occurs from 9 a.m. to 7 p.m. while pollen donor (e.g., male) parents usually undergo dehiscence from 11 a.m. to 3 p.m. (Yan and Li 1987; Tian 1991). The asynchronous flowering results in many spikelets not being available for cross-pollination. Spikelets which have stigmas that remain out of the spikelet (e.g., exserted) after the flowering have an extended opportunity to be cross-pollinated. Among these factors, stigma exsertion is especially emphasized as a component increasing the opportunity of pollination (Kato and Namai 1987). Exserted stigmas remain viable up to 6 days with a decrease of 20% of seed set from cross-pollination per day (Yan and Li 1987; Xu and Shen 1988). As a result, stigma exsertion including single and dual stigmas and other stigma traits play important roles in hybrid seed production and receive consistent attention from rice researchers (Virmani and Athwal 1973; Virmani 1994; Yan and Li 1987; Tian 1991; Yuan 1981; Yuan and Fu 1995; Li et al. 2001; Uga et al. 2003a, b; Xu 2003; Miyata et al. 2007; Sidharthan et al. 2007).

The small size of rice spikelets and the large effect of environmental effects on flowering in rice contribute to the difficulty in traditional selection-based breeding. As a result, development of DNA markers associated with desirable floral traits for breeding programs has received increasing attention lately. Using a segregating population derived from Dongxiang wild rice (*Oryza rufipogon*) and *indica* cultivar Guichao 2 (*O. sativa*), two quantitative trait loci (QTL) for stigma exsertion were identified using random fragment length polymorphism (RFLP) markers (Li et al. 2001). These loci, qPEST-5 and qPEST-8, have an additive effect of 0.063 and 0.086 and contribute 10.9 and 11.4% of variance to stigma exsertion, respectively. Similarly using the population between *O. rufipogon* and *O. sativa* and RFLP markers, Uga et al. (2003b) found two stigma exsertion QTLs qRES-5 and qRES-10 on chromosome 5 and 10 accounting for 9.8 and 24.8% of phenotypic variance, respectively. Using a segregating population derived from *indica* (IR24) and *japonica* (Koshihikari) sub-species in *O. sativa*, a major QTL qES3 on chromosome 3 was identified

using RFLP markers, which explains 31.63% of the total phenotypic variance and the IR24 allele at the QTL increased stigma exsertion by 20.1% (Miyata et al. 2007). Previous studies have collectively identified QTL markers for about 35% of total phenotypic variations and the remaining portion needs further effort. For the first time we are using association mapping strategy, germplasm collection with diversified genetic variability and simple sequence repeat (SSR) markers to identify DNA markers for marker-assisted selection in hybrid-rice breeding programs.

Association mapping identifies QTLs by examining the marker-trait associations, and enables researchers to use modern genetic technologies to exploit natural diversity and locate valuable genes in the genome (Zhu et al. 2008). Population structure is an important component in association mapping analyses because it can reduce both type I and type II errors between molecular markers and traits of interest in an autogamous species, e.g., rice (*O. sativa*) and barley (*Hordeum vulgare*) (Agrama et al. 2007; Agrama and Eizenga 2008; Breseghello and Sorrells 2006; Yu et al. 2006a). A mixed linear model (MLM) approach was found effective in removing the confounding effects of the population substructure in association mapping (Yu et al. 2006a) by using both the population structure information (Q-matrix) and pair-wise relatedness coefficients-'kinship' (K-matrix). The MLM or Q + K model works better than either the K model or Q model alone, demonstrated in a global set of 95 highly structured Arabidopsis population (Yu et al. 2006a; Zhao et al. 2007). Previous studies proved that (1) K-matrix alone is sufficient if a kinship is estimated as a proportion of shared haplotypes for each pair of individuals; (2) the replacement of Q-matrix from the computational intensive structure analysis (Falush et al. 2003, 2007; Pritchard et al. 2000) with principal components (PC) matrix from more robust PC analysis (Price et al. 2006) performs similarly to the MLM proposed by Yu et al. (2006a), thus suggesting a potential for replacement; (3) removing of the confounding effects has possibility to remove true associations with biological effect strongly correlated with population structure, so a caution needs to be taken; and (4) in a highly structured population, the causations with major effect have good chance to be expected, and larger samples and adequate marker densities are needed for genome-wide dissection of most traits of interest segregating in an association mapping population (Zhao et al. 2007). The PC matrix approach proved to have similar power to capture population structure as STRUCTURE, and provides a formal way of assigning statistical significance to population subdivision (Patterson et al. 2006). Price et al. (2006) found that incorporating the outcome of population structure contained in the PC-matrix increases the power to detect true marker-trait association.

There are specific statistical approaches to estimate the false-positive level of the obtained strong (*P*-value) associations (control for Type I error) such as a false discovery rate (FDR) (Benjamini and Yekutieli 2001). The way that FDR technique controls the proportion of false positives to observed positive *P*-values is described by García (2004). Type I error rate effects can be decreased by adjustment of the *P*-values with the FDR, so combination of the MLM and FDR reduces the risk of finding false marker-trait associations. As a result, the MLM that combines both the PC-matrix and K-matrix was adapted for our objective to map the molecular markers associated with stigma and spikelet characteristics. The resulting marker-trait associations were then adjusted using the FDR.

Materials and methods

Materials and phenotyping

The United States Department of Agriculture (USDA) rice core collection consists of 1,794 accessions from 114 countries, which was developed by a stratified random sampling from the whole collection of more than 18,000 accessions (Yan et al. 2007). Ward's method (Ward 1963) of hierarchical cluster analysis was used for grouping those of *Oryza sativa* L. species in the 1,794 accessions based on the 14 phenotypic traits described by Yan et al. (2007). The Ward minimizes the variance within groups by a classification strategy to allow maximum distances between each pair of accessions located in different groups (Franco et al. 2006). The distances were used as a measure of similarity among accessions to form the clusters. Ninety accessions (e.g., mini-core collection) were then selected from the hierarchical groups to represent the sub-clustering within each cluster. In

addition to the mini-core collection, four accessions, three indicas Guichao No. 2, Aijaonante and IR 24 and one japonica Koshihikari, were included in the analysis using their published information of both stigma exsertion and molecular markers (Li et al. 2001, Xiong et al. 1999, Miyata et al. 2007).

Stigma and spikelet characteristics were observed for the sampled 90 accessions at the experiment station of Sichuan Academy of Agricultural Sciences, Chengdu, China in 2007. Two local male-sterile lines and eight maintainers for the male sterility that are popularly grown in China were included in the observation as checks, but they were not included in genotypic analysis in the US because of Chinese property right regulation covering DNA of germplasm. Single row plot including 12 plants, 20 cm square in spacing by transplanting for each accession was arranged by a randomized complete block design with three replications. At the peak flowering time (12 a.m. to 1 p.m.), a single panicle from five different plants was selected from each plot, resulting in 15 panicles sampled for each germplasm accession. All the spikelets with open lemma and palea, indicative of flowering on the day, were marked. Five spikelets from each panicle whose lemma and palea stayed open were removed and kept in a tube containing FAA (70% of ethanol:acetic acid:formalin = 90:5:5) for 24 h or more. Spikelet length and width, and stigma length and width were then measured with a micrometer under a stereomicroscope (Fig. 1). For each accession, measurements were taken from a total of 75 spikelets. In the late afternoon (after 3 p.m.) when flowering is complete for the day, the spikelet number with exserted stigma in one side (single stigma exsertion, SStgE), and two sides (dual stigma exsertion, DStgE), and no stigma in either sides (NStgE) among the marked spikelets in each of 15 sampled panicles were recorded for the 15 sampled panicles (Fig. 1). Then, these counts were converted to:

$$\mathrm{SStgE}(\%) = [\mathrm{SStgE}/(\mathrm{SStgE} + \mathrm{DStgE} + \mathrm{NStgE})] \times 100,$$

$$\mathrm{DStgE}(\%) = [\mathrm{DStgE}/(\mathrm{SStgE} + \mathrm{DStgE} + \mathrm{NStgE})] \times 100,$$

Total stigma exsertion (TStgE)(%)
$$= \mathrm{SStgE}(\%) + \mathrm{DStgE}(\%) \text{ and}$$

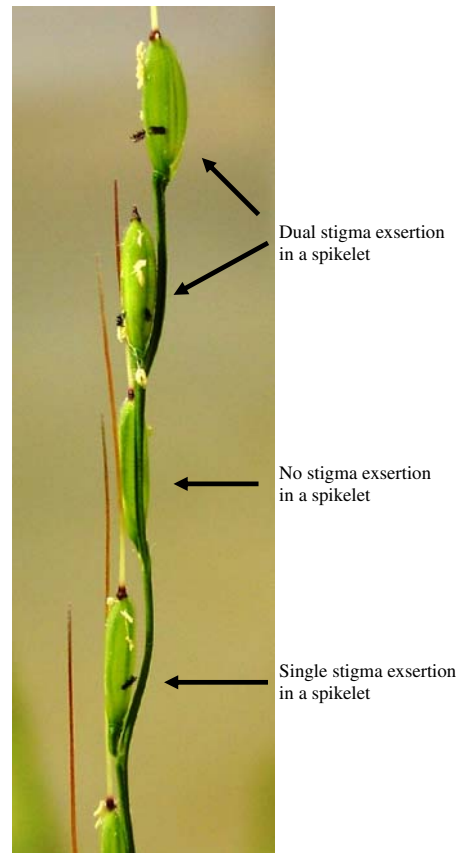$$\mathrm{NStgE}(\%) = 100 - \mathrm{TStgE}(\%).$$

Fig. 1 Display of rice stigma and spikelet characteristics investigated in the study

Genotyping

Total genomic DNA of each accession was extracted using a rapid alkali extraction procedure (Xin et al. 2003) and genotyped with 109 DNA markers, 108 simple sequence repeat (SSR) plus an indel. The markers cover the entire rice genome, especially in those chromosome regions previously reported for stigma exsertion. The genotyping was conducted in the genomic lab of USDA-ARS, Dale Bumpers National Rice Research Center. The SSR markers were obtained from the Gramene database (http://www.gramene.org/) except AP5652-1 and AP5652-2 that were developed in house from BAC AP5652. The indel marker at the *Rc* locus (Rid 12) is responsible for bran color in rice (Brooks et al. 2008). PCR amplifications were performed according to Eizenga et al. (2006). For each marker, forward

primers were labeled with either 6FAM, NED or Hex (Applied Biosystems, Foster City, CA, USA or Integrated DNA Technologies, Coralville, IA, USA). DNA amplifications were performed using MJ Research Tetrad thermocyclers (Waltham, MA, USA). PCR products were pooled based on color and size range of amplified fragments (typically three markers per run along with ROX-labeled size standard), and the DNA was denatured by heating samples at 94°C for 5 min. The samples were separated on an ABI Prism 3730 DNA Analyzer according to the manufacturer's instructions (Applied Biosystems). Fragments were sized and binned into alleles using GeneMapper v. 3.7 software (Applied Biosystems).

Statistical analysis and association mapping

Variance analysis for stigma and spikelet characteristics was conducted using mean of five plants in each replication. Genetic variation in terms of percentage to total variation was used as an estimate of heritability in the broad sense (Virmani and Athwal 1973). The estimated heritability simply represents the magnitude of genetic variation in the studied mini-core collection in relation to the total variation. Polymorphism information content (PIC) was used to measure allele diversity at a locus (Smith et al. 1997). Number of alleles per locus and heterozygosity were calculated to quantify the genetic variation within the 90 accessions in the mini-core collection plus four reference cultivars. Rogers genetic distances (RD; Rogers 1972) were computed among these genotypes and cluster analysis was performed on the Rogers's distances using the unweighted pair-group method with an arithmetic mean (UPGMA). PowerMarker software (Liu and Muse 2005) was used to calculate these parameters mentioned above.

The model-based program STRUCTURE (Pritchard et al. 2000) was used to infer population structure using a burn-in of 50,000, run length of 100,000, and a model allowing for admixture and correlated allele frequencies. A model-based clustering algorithm was applied to identify subgroups with distinctive allele frequencies. This procedure, implemented in STRUCTURE, places genotypes into K clusters, where K is chosen in advance but can be varied across independent runs of the algorithm (Agrama

et al. 2007). STRUCTURE was applied for ten independent runs with an assumption of 'independent allele frequencies' for each K value from 1 to 11. Runs with the highest LnP(D) probability were reported for each K. Graphical outputs from STRUCTURE were produced to visualize the best number of clusters.

Principal components analysis (PCA) was conducted using 109 markers and GenAlex 6.1 (Peakall and Smouse 2006) software to structure the mini-core collection and generate a PC-matrix. FSTAT (Goudet 1995) was used to estimate pairwise Fst values among the resulting structural populations. Analysis of molecular variances (AMOVA) was conducted using GenAlex 6.1 as well (Peakall and Smouse 2006). The first two PCs were included in the MLM analysis (TASSEL 2009). Pair-wise relatedness coefficients—'kinship' (K-matrix) (Ritland 1996) among individuals in the collection were estimated with SPAGeDi software (Hardy and Vekemans 2002). DNA markers associated with the nine stigma and spikelet characteristics were mapped using TASSEL software (Bradbury et al. 2007). The MLM or PC + K model (Yu et al. 2006a) was implemented in TASSEL as described in Henderson's notation (Bradbury et al. 2007): $y = Xb + Zu + e$, where y is the vector of observations; b is an unknown vector containing fixed effects including genetic marker and population structure (PC); u is an unknown vector of random additive genetic effects from multiple background QTL for germplasm accessions; X and Z are the known design matrices; and e is the unobserved vector of random residuals. Both PC- and K-matrices were used in association mapping to control spurious associations. The squared correlation coefficient (r^2) was used to estimate linkage disequilibrium (LD) between each pair of marker loci (Pritchard and Przeworski 2001) using TASSEL (Bradbury et al. 2007), and the estimated r^2 frequency was plotted.

Sharpened "step-up False Discovery Rate" (hereafter "FDR") significance levels were calculated using SAS 9.1 "PROC multtest" (Benjamini and Yekutieli 2001; Reiner et al. 2003). Markers exhibiting a P-value less than 0.001 ($-\log_{10}$FDR > 3.0) were considered significantly associated with a phenotypic trait.

Results

Stigma and spikelet characteristics

The 90 accessions in the mini-core collection developed from the USDA rice core collection originated from 13 regions including 47 countries. Detailed information of their name, origin, date of introduction and scores of the traits involved is presented in the Supplementary Table in comparison with the checks of local male-sterile lines and maintainers. Means, ranges and broad sense heritability of spikelet length (SpkL), width (SpkW) and L/W ratio (SpkL/W), stigma length (StgL), width (StgW) and L/W ratio (StgL/W), single stigma exsertion (SStgE), dual stigma exsertion (DStgE) and total stigma exsertion (TStgE) and their correlations among the 90 accessions are presented in Table 1 and their frequency distributions are shown in Fig. 2. Variance analysis revealed significant results for all the nine traits ($P < 0.0001$ identically), while variation due to replication was not significant ($P = 0.0658$–0.7970) for any trait among the 90 accessions. Heritability in broad sense ranged from 56.1% for SpkW to 77.3% for SpkL.

Fourteen accessions (16%) had significantly greater DStgE (5.38–32.31% with entry 671 from Romania on the top) than the maintainer (0.79%) (Supplementary Table; Fig. 2). TStgE ranged from 65.94 to 80.65% among entry 671, 913, 1032, 1049 and 1456 from five countries including China, and they were higher than the maintainer (49.15%). SStgE was 72.48% in entry 1032 and 62.22% in entry 1456, more than the maintainer (48.35%). Entry 1032 (cultivar '50638') from Guyana (PI 346441) had the highest SStgE and TStgE, which were 24.13 and 31.50 percent points higher than the average of eight local maintainers, and 36.77 and 34.72 percent points higher than the average of two local male sterile (MS) lines, respectively. Entry 671 (cultivar 'Linia 84 Icar') had the second highest TStgE (75.17%) because of its highest DStgE among 90 accessions of germplasm. TStgE was highly and positively correlated with both SStgE ($r = 0.96$) and DStgE ($r = 0.69$), but these three stigma exsertions were not correlated with stigma length or width.

Wider stigma in 27 accessions and longer stigma in 13 accessions than the checks were observed (Table 1; Fig. 2). Stigma was the longest in entry 571

(1.50 mm) (cultivar 'Sesia') from Portugal and the widest in entry 572 (0.97 mm) (cultivar 'Triomphe') from Morocco. Thirteen accessions had longer and two had wider spikelet than the checks. Spikelet was the longest in entry 571 (10.54 mm) from Portugal and the widest in entry 667 (5.70 mm) (cultivar 'Arpa Shaly Mestnyj') from Uzbekistan. SpkL was correlated negatively with SpkW ($r = -0.24$) and positively with SpkL/W ($r = 0.85$), and further positively with StgL ($r = 0.21$), SStgE ($r = 0.28$) and TStgE ($r = 0.23$).

Molecular markers and genetic structure

Marker analysis of the mini-core and reference accessions using 109 molecular markers resulted in the detection of a total of 1,066 alleles. Numbers of alleles ranged from 2 (at locus RM338_Chr3, RM507_Chr5 and RM455_Chr7) to 30 (RM18700-2_Chr5) with an average of 9.8 ± 0.5 alleles per locus. Heterozygosity was very low as expected in autogamous crops with an average of 0.013, but the highest reached 0.066 for RM44_Chr8. Polymorphic information content (PIC) measures the probability that two randomly chosen alleles from a population are distinguished. The PIC averaged 0.672 ± 0.017 ranging from 0.055 (AP5625-1) to 0.935 (RM18700-2 and RM3525) with a major distribution between 0.319 and 0.894. Eighty-nine markers (82%) were highly informative (PIC > 0.5), 18 (16%) reasonably informative ($0.5 > \text{PIC} > 0.25$) and 2 (2%) slightly informative (PIC < 0.25).

Three major clusters were obtained using the unweighted pair-group (UPGMA) analysis based on the Rogers' genetic distance for the 90 accessions plus four reference cultivars (Fig. 3). Two dimensional principal components analysis (PCA) resulted in three clusters (Fig. 4) as well, largely corresponding to the grouping resulted from UPGMA cluster analysis. The PCA provides another perspective on the relationships among the accessions, and generally supports results obtained by the UPGMA. The first (49.2%) and second (20.2%) axes of PCA explained a relatively high percentage (69.4%) of the total variation. Accessions 307, 951 and 1386 were not assigned to any cluster in the PCA, but were instead placed intermediately among the clusters 1, 2 and 3 in the UPGMA. Both the PCA and UPGMA grouped the indica reference cultivars Guichao No2 (Li et al.

Table 1 Stigma and spikelet characteristics and their correlations among 90 accessions as a mini-core collection developed from the USDA rice core collection in comparison with local male sterile (MS) lines and maintainers

	Spikelet length (SpkL) (mm)	Spikelet width (SpkW) (mm)	Spikelet L/W ratio (Spk L/W)	Stigma length (StgL)	Stigma width (StgW) (mm)	Stigma L/W ratio (Stg L/W) (mm)	Single stigma exsertion (S StgE) (%)	Dual stigma exsertion (DStgE) (%)	Total stigma exsertion (T StgE) (%)
Mean ± SE	7.93 ± 0.10	4.68 ± 0.04	1.71 ± 0.03	1.06 ± 0.02	0.48 ± 0.03	2.26 ± 0.05	18.88 ± 1.59	2.43 ± 0.58	21.31 ± 1.94
Minimum	6.38	3.76	1.23	0.69	0.28	1.36	0.00	0.00	0.00
Maximum	10.54	5.70	2.42	1.50	0.97	3.40	72.48	32.31	80.65
Heritability (%)	77.3	56.1	69.2	56.4	70.5	74.2	76.6	75.1	76.7
8 Local maintainers	8.08	5.01	1.62	1.09	0.48	2.03	48.35	0.79	49.15
2 Local MS lines	8.15	5.15	1.59	1.18	0.44	3.00	35.71	10.22	45.93
Correlations									
SpkW	−0.24*								
SpkL/W	0.85***	−0.71***							
StgL	0.21*	−0.16	0.24*						
StgW	<0.01	0.41***	−0.21	0.43***					
StgL/W	0.12	−0.58***	0.39***	0.24*	−0.74***				
SStgE	0.28**	0.11	0.15	0.07	0.12	−0.12			
DStgE	<0.01	0.13	−0.07	0.12	0.19	−0.12	0.47***		
TStgE	0.23*	0.13	0.10	0.09	0.15	−0.14	0.96***	0.69***	

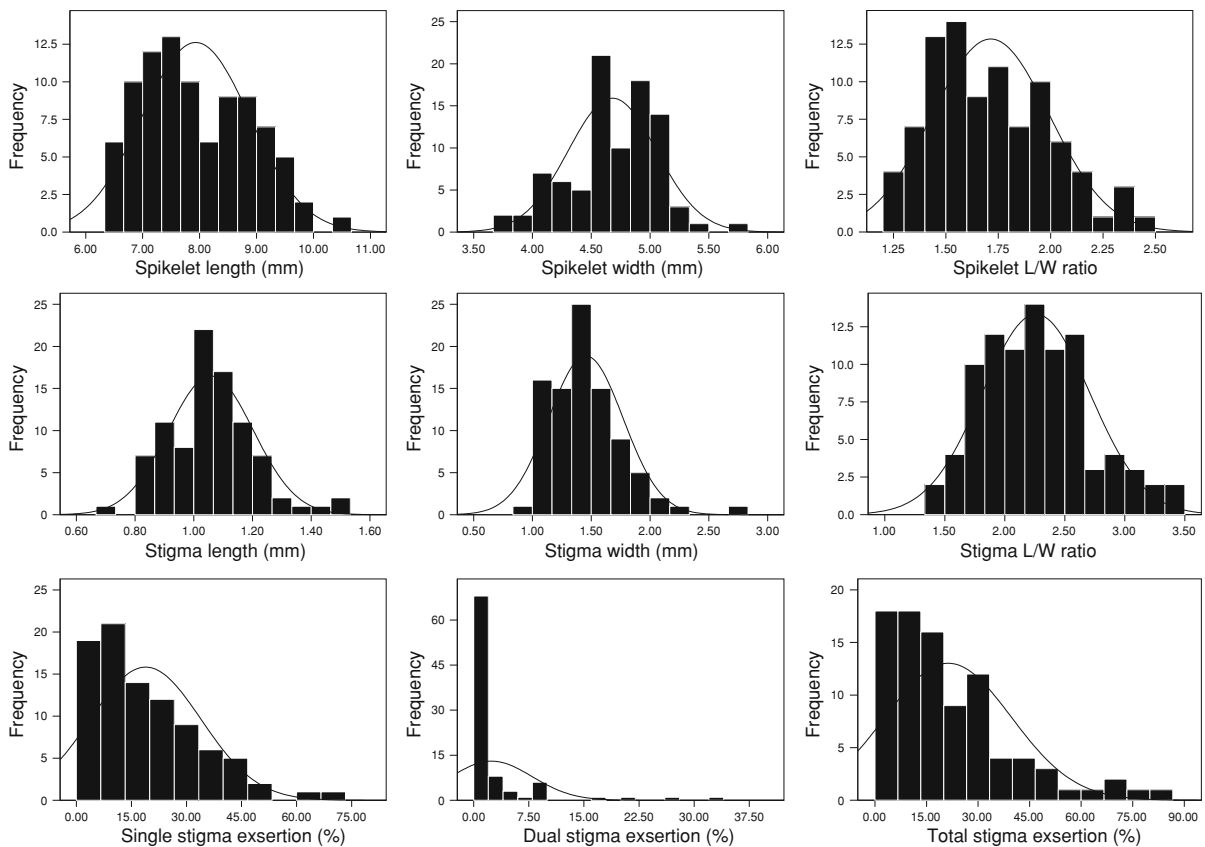*, ** and *** significant at 0.05, 0.01 and 0.001 probability level, respectively

Fig. 2 Frequency distribution of stigma and spikelet characteristics among 90 accessions as a mini-core collection sampled from the USDA rice core collection

2001), Aijiaonante (Xiong et al. 1999) and IR24 (Miyata et al. 2007) in cluster 1, while japonica Koshihikari (Miyata et al. 2007), temperate japonica 718 (PI 584588), and 1321 (PI 439625) (Garris et al. 2005) were placed in cluster 3. Two accessions 259 (PI 303646) and 888 (PI 199542) in Cluster 2 was classified as tropical japonica (Garris et al. 2005). This grouping suggests that cluster 1 accessions are indicas while cluster 2 and 3 accessions correspond to tropical and temperate japonicas, respectively.

Entry 1032 with the highest SStgE and TStgE was clustered in indica group, matching up with the tropical climate where it originated, Guyana of South America. Winner for the highest DStgE and second highest TStgE, entry 671 from Romania, was placed in temperate japonica cluster. Entry 571 from Portugal having the longest stigma was clustered in tropical and entry 572 from Morocco having the widest stigma in temperate japonica group, respectively.

Analysis of population structure using STRUC-TURE software provides correct number of subpopulations for association mapping with no prior information about the three clusters found by the PCA or UPGMA and the posterior probabilities of kinship (K) tested from 1 to 11. Population structure analysis revealed that LnP(D) estimates increased with increase of K up to $K = 4$ and then decreased steadily, suggesting the genetic diversity should be structured into four subpopulations in this sampled mini-core collection. The four Ks structuration corresponded to the genetic populations revealed by both the UPGMA (Fig. 3) and PCA (Fig. 4). Indica, tropical japonica and temperate japonica cultivars were structured in K1, K2 and K3, respectively. Accessions 951 and 1386 located intermediately between K1 indica and K2 tropical japonica in the UPGMA (Fig. 3) were grouped to K4 subpopulation. This agreement validated the results from the PCA

Fig. 3 UPGMA tree based on Rogers's genetic distance for 90 accessions (Core entry number used in the chart) and reference cultivars (*highlighted*) genotyped with 109 molecular markers
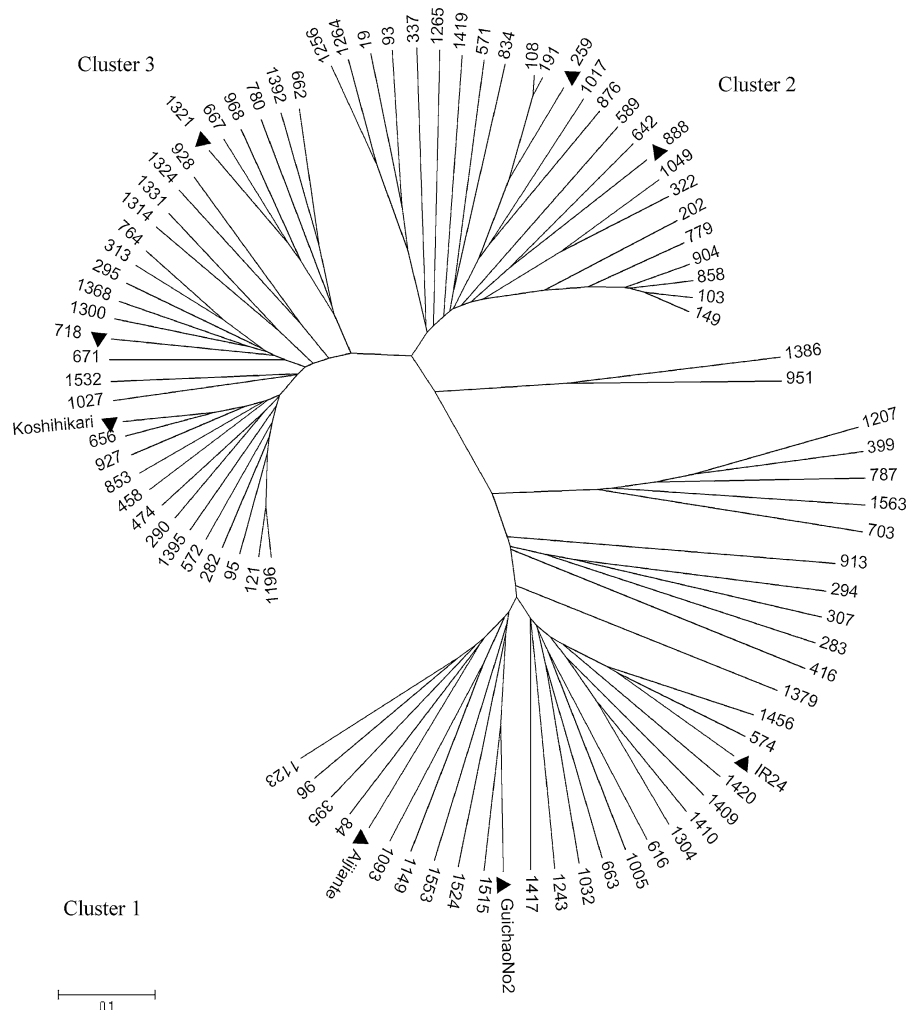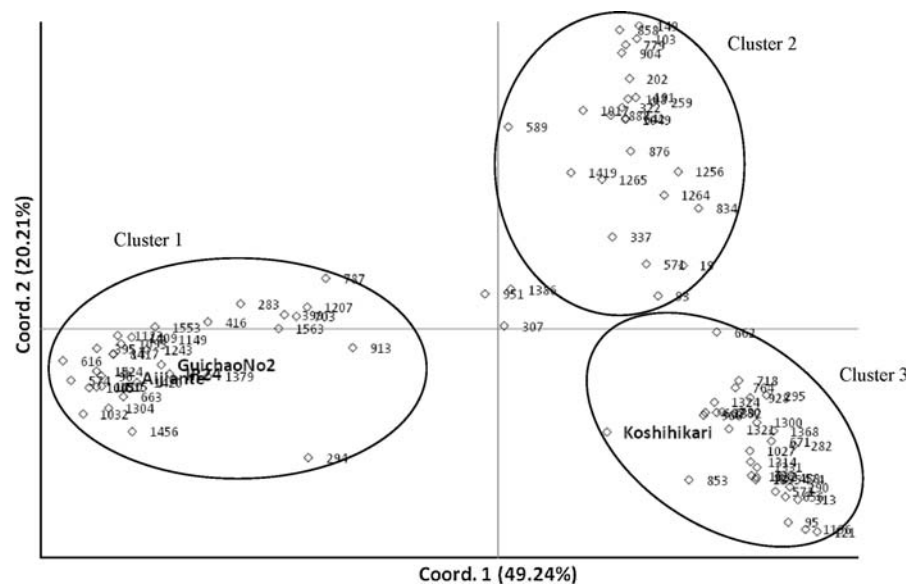

Fig. 4 Principal components analysis (*PCA*) for 90 accessions as a mini-core collection (Core entry number used in the chart) and reference cultivars genotyped with 109 molecular markers

including population structure and PC-matrix for following association mapping.

Among four subpopulations, Fst of K1 indica was 0.316 with K2 tropical japonica, 0.362 with K3 temperate japonica and 0.206 with K4 others. K2 had a Fst of 0.172 with K3 and 0.279 with K4. Fst was 0.318 between K3 and K4. Analysis of molecular variance resulted in 40.1% and 59.9% of total variation among and within four subpopulations, respectively.

Linkage disequilibrium and association mapping

Extent of genome-wide linkage disequilibrium (LD) was evaluated through pairwise comparisons among the 109 marker loci. The LD pattern included a total of 4,046 estimates expressed by r^2 averaged

0.0835 ± 0.0013 ranging from 0.0004 to 0.7663. Among these estimates, 312 (7.7%) showed LD scores as r^2 higher than 0.20. The r^2 value corresponding to the 95th percentile of the distribution was 0.0152 and this value was used as a population-specific threshold for r^2 as evidence of genetic linkage. Up to 92% of the LD parameter r^2 were significant ($P < 0.05$) among the 4,046 pairwise comparisons of the 109 markers (Fig. 5), indicative of a validation for further analysis of association mapping.

Based on the criterion P-value less than 0.001 with 109 markers for nine traits, we identified 45 significant marker-trait associations involving 33 markers (Fig. 6). After false discovery rate (FDR) correction, 34 of these associations involving 22 markers remained significant at $P < 0.001$
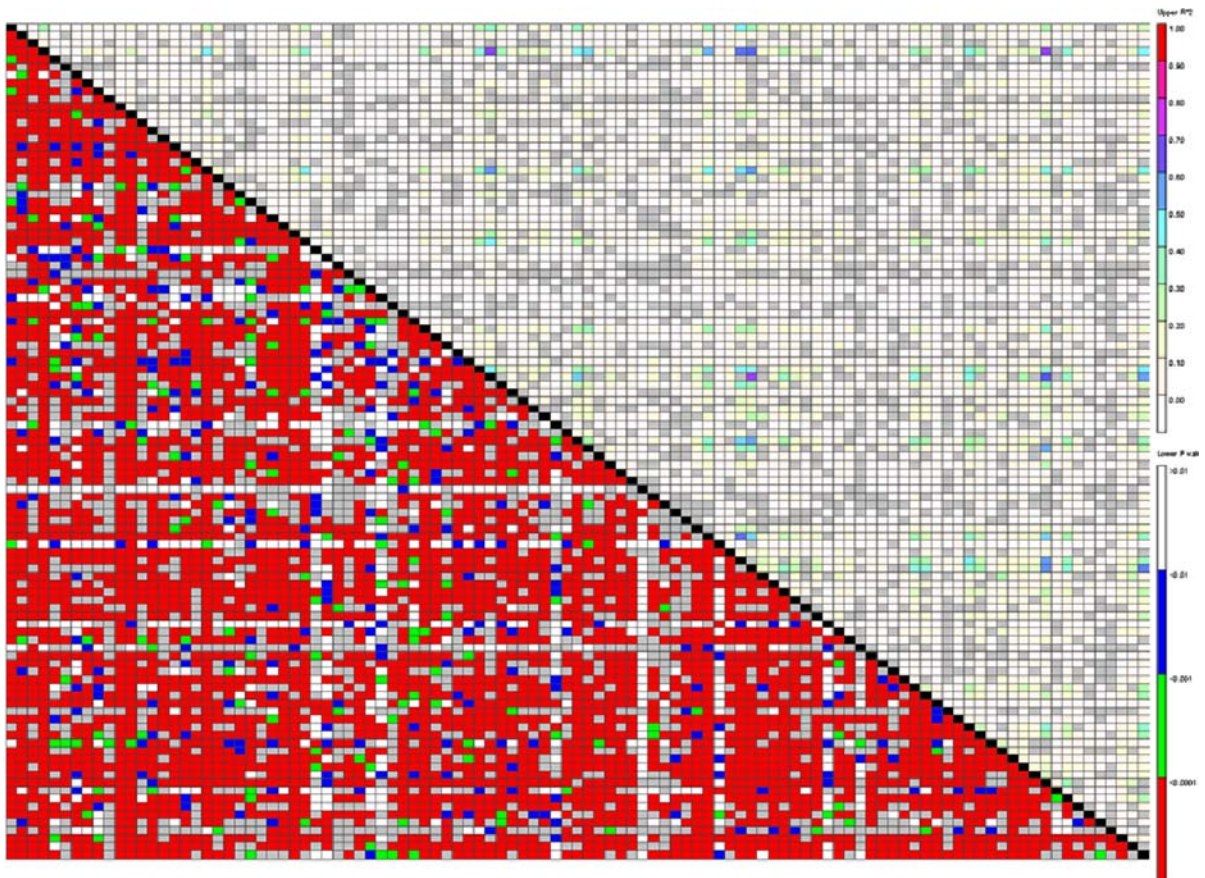

Fig. 5 Linkage disequilibrium (*LD*) patterns among 90 accessions as a mini-core collection and reference cultivars genotyped with 109 molecular markers. The squared correlation coefficients (r^2) for each pair of markers are presented in the *upper triangle* and their corresponding tests in the *lower triangle*: *white* $P > 0.05$, *blue* $0.05 > P > 0.01$, *green* $0.01 > P > 0.001$ and *red* $P < 0.001$

($-\log_{10}$FDR > 3.0) (Table 2). Among 34 marker-trait associations, dual stigma exsertion (DStgE) with RM7203_Chr11 (8.7 cM) had the highest—$\log_{10}$FDR value of 5.91 resulting in the lowest probability of 1.22E-06 (Table 2). Regression (R^2) was the highest (0.308) between StgL and RM474_Chr10 (0.8 cM), but lowest (0.042) between DStgE and RM455_Chr7 (78.9 cM).

There were four markers significantly associated with single stigma exsertion (SStgE) and RM5_Chr1 (98.5 cM) had the highest regression (0.169) (Fig. 6; Table 2). DStgE was significantly associated with six markers with regressions ranging from 0.042 to 0.300. Markers RM5, RM3642_Chr1 (102.3 cM) and RM44_Chr8 (46.9 cM) had high regression, 0.300, 0.287 and 0.220, respectively with DStgE. Total sigma exsertion (TStgE) was significantly associated with five markers and their regressions ranged from 0.055 for RM284 to 0.112 for RM25669. We identified 15 marker-trait associations involving 10 SSR markers distributed on eight chromosomes for three traits of stigma exsertion. Five markers were synchronously associated with two correlative exsertion traits: RM105 and RM25669 with both SStgE and TStgE, RM178 and RM455 with both DStgE and TStgE, and RM5 with both SStgE and DStgE (Tables 1, 2).

SSR markers RM1334_Chr3 (83.5 cM) and RM474_Chr10 (0.8 cM) were associated with stigma length (StgL) with high regressions, 0.296 and 0.308, respectively (Fig. 6; Table 2). Three markers were significantly associated with stigma L/W ratio. No marker had significant association with stigma width (StgW).

Four markers were associated with spikelet length (SpkL), and two on chromosome 2 had about equivalent regression (0.160) (Fig. 6; Table 2). RM3400_Chr3 (83.3 cM) had the highest regression (0.212) with SpkL. Seven markers were associated with spikelet L/W ratio (SpkL/W). Two markers on chromosome 2, RM12521 and RM3865, had high regression with SpkL/W. Three markers were associated with spikelet width (SpkW) and all had high regression (about 0.2). In total, the 14 marker-trait associations included seven SSR markers on six chromosomes accounted for three spikelet traits, five synchronously associated with two correlative traits: RM224 and RM271 with both SpkW and SpkL/W, respectively and RM312, RM3400 and RM3865 with

both SpkL and SpkL/W, respectively, and one marker RM12521 with all three correlative traits of spikelet: SpkL, SpkW and SpkL/W (Tables 1, 2).

Discussion

Observation technique for stigma exsertion

Stigma exsertion is affected by environmental conditions, e.g., water stress (Yu et al. 2006b), so that the number of sampled panicles in observation should be reasonable for reliable estimate of a genotype. Among previous studies, Yan and Li (1987) observed the greatest number per genotype, 24 panicles for stigma exsertion, followed by Yu et al. (2006a, b)18 panicles, Virmani and Athwal (1973) 15 panicles, Miyata et al. (2007) 2 panicles and Uga et al. (2003b) 27 spikelets. All these observations were made on the panicles when all spikelets finished flowering except Yan and Li (1987) and Uga et al. (2003b). Flowering from the beginning to the end in a panicle lasts 5–7 days (Yan and Li 1987; Tian 1991). The exserted stigma is fragile and can be easily damaged by environmental conditions, e.g., wind, water stress, physical interruption, etc. during the flowering period. Observation of stigma exsertion on the flowering day for those spikelets flowering on that day only should avoid those damages and improve data accuracy. Therefore, data generated in the present study using this observation technique for 15 panicles should accurately describe a genotype for stigma exsertion and build a solid foundation for the association mapping. In consideration of environmental effects, observation on different weather, clear, cloudy and rainy in multi-location should be more accurate to describe stigma exsertion.

Molecular markers identified for stigma exsertion in rice

We identified two markers associated with stigma exsertion on chromosome 1 (RM5 and RM3642), and 8 (RM44 and RM284) and one each on chromosome (Chr) 5, 6, 7, 9, 10 and 11. Markers associated with stigma exsertion have never been reported on Chr1, 7 and 9. Two associated markers on Chr2 reported by Li et al. (2001, 2003), six on Chr3 reported by Li et al. (2003), Miyata et al. (2007) and Yamamoto et al.
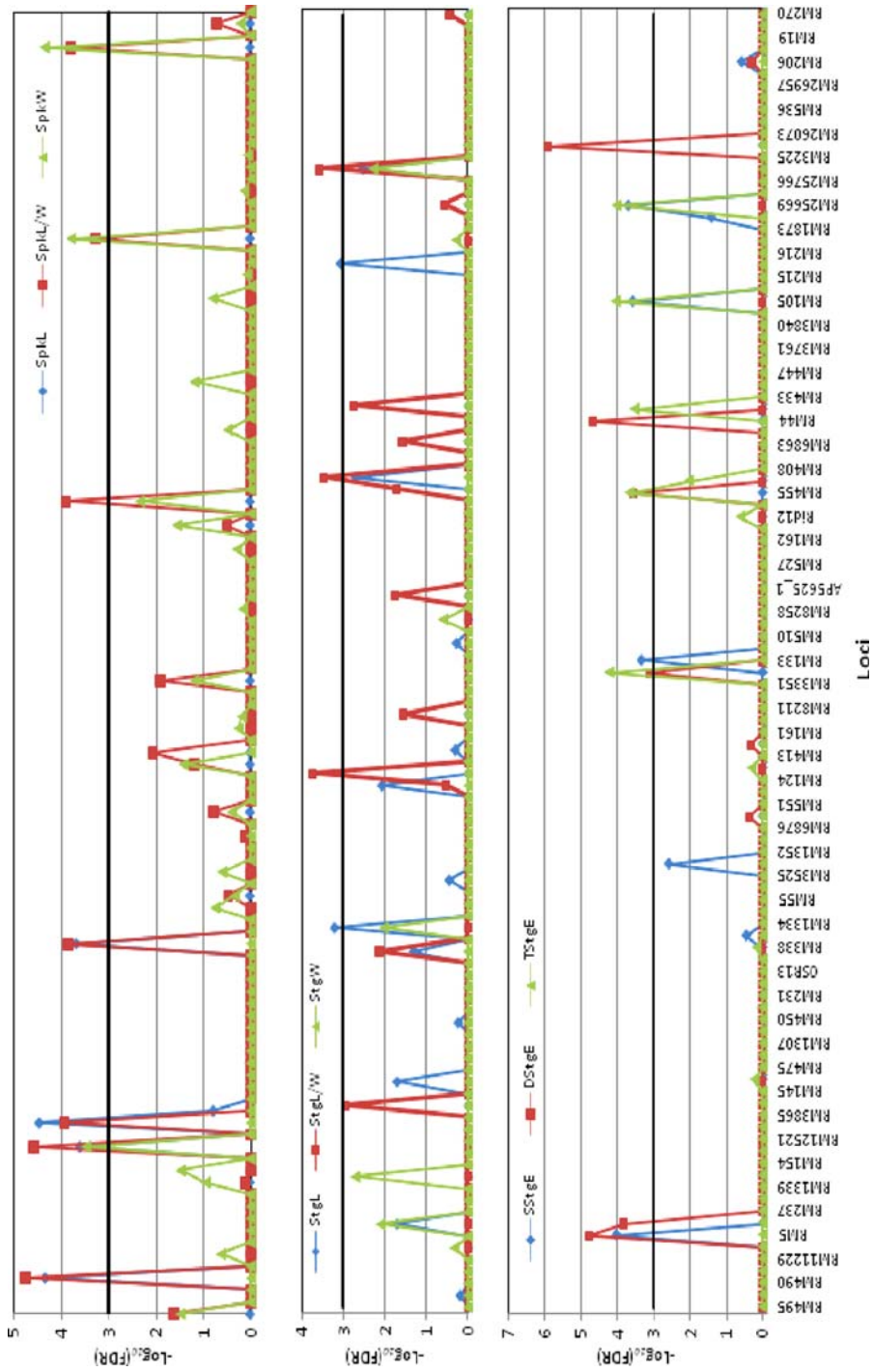
Fig. 6 Marker-trait associations tested with false discovery rate (−log₁₀FDR > 3) corrected from *P* values with 90 accessions as a mini-core collection and reference cultivars genotyped with 109 molecular markers for percentage of single (SStgE), dual (DStgE) and total stigma exsertion (TStgE), stigma length (StgL), width (StgW) and L/W ratio (StgL/W) and spikelet length (SpkL), width (SpkW) and L/W ratio (SpkL/W)

Table 2 Marker-trait associations with P-value less than 0.001 and $-\log_{10} P$ greater than 3.0, their equivalent false discovery rate probability (FDR) and $-\log_{10}$ FDR and marker-trait regression coefficient (R^2) and marker position (cM) on chromosome (Chr) derived from 109 markers and 90 accessions as a mini-core collection in rice

Trait	Chr	Position	Marker	P value	FDR	R^2	$-\log_{10} P$	$-\log_{10}$ FDR
Spikelet length	1	48.7	RM312	4.20E−06	4.54E−05	0.081	5.38	4.34
	2	17.1	RM12521	2.31E−05	2.45E−04	0.156	4.64	3.61
	2	19.0	RM3865	3.10E−06	3.38E−05	0.165	5.51	4.47
	3	83.3	RM3400	1.90E−05	2.03E−04	0.212	4.72	3.69
Spikelet L/W ratio	1	48.7	RM312	1.61E−07	1.75E−05	0.088	6.79	4.76
	2	17.1	RM12521	2.44E−07	2.64E−05	0.201	6.61	4.58
	2	19.0	RM3865	1.10E−06	1.18E−04	0.151	5.96	3.93
	3	83.3	RM3400	1.40E−05	1.41E−04	0.189	4.85	3.85
	7	67.0	RM11	1.24E−05	1.26E−04	0.158	4.91	3.90
	10	35.7	RM271	5.42E−05	5.37E−04	0.087	4.27	3.27
	11	75.7	RM224	1.60E−05	1.60E−04	0.138	4.80	3.80
Spikelet width	2	17.1	RM12521	3.10E−05	3.25E−04	0.207	4.51	3.49
	10	35.7	RM271	1.50E−05	1.62E−04	0.199	4.82	3.79
	11	75.7	RM224	4.03E−06	4.39E−05	0.197	5.39	4.36
Stigma length	3	83.5	RM1334	5.40E−06	5.89E−04	0.296	5.27	3.23
	10	0.8	RM474	7.50E−06	8.10E−04	0.308	5.12	3.09
Stigma L/W ratio	5	3.0	RM507	1.70E−06	1.85E−04	0.061	5.77	3.73
	7	96.9	RM118	3.10E−06	3.32E−04	0.075	5.51	3.48
	10	71.4	RM484	2.43E−06	2.62E−04	0.084	5.61	3.58
Single stigma exsertion	1	98.5	RM5	8.59E−06	9.19E−05	0.169	5.07	4.04
	6	0.5	RM133	2.35E−04	4.48E−04	0.078	3.63	3.35
	9	7.8	RM105	2.44E−05	2.60E−04	0.121	4.61	3.58
	10	55.2	RM25669	1.81E−06	1.97E−04	0.111	5.74	3.70
Dual stigma exsertion	1	98.5	RM5	1.52E−07	1.64E−05	0.300	6.82	4.78
	1	102.3	RM3642	1.41E−06	1.49E−04	0.287	5.85	3.83
	5	104.4	RM178	7.70E−06	8.01E−04	0.113	5.11	3.10
	7	78.9	RM455	2.59E−06	2.72E−04	0.042	5.59	3.57
	8	46.9	RM44	1.99E−07	2.13E−05	0.220	6.70	4.67
	11	8.7	RM7203	1.12E−08	1.22E−06	0.127	7.95	5.91
Total stigma exsertion	5	104.4	RM178	5.74E−06	6.08E−05	0.064	5.24	4.22
	7	78.9	RM455	2.03E−06	2.19E−04	0.071	5.69	3.66
	8	78.5	RM284	2.80E−06	3.00E−04	0.055	5.55	3.52
	9	7.8	RM105	8.31E−07	9.06E−05	0.093	6.08	4.04
	10	55.2	RM25669	9.08E−06	9.53E−05	0.112	5.04	4.02

(2003), and one on Chr12 reported by Yamamoto et al. (2003) were not verified by this study. Associated marker RM178 on Chr5 in this study was about 25 cM away from the one identified by Uga et al. (2003b) and further away from the markers reported by Yamamoto et al. (2003) and Li et al. (2001). Our results revealed RM133 on the beginning of Chr6 associated with stigma exertion, but Yamamoto et al.

(2003) reported a QTL on the other end of the chromosome. Yamamoto et al. (2003) reported a QTL on the top of Chr8 that was far away from RM44 (46.9 cM) and RM284 (78.5 cM) identified in this study. However, RM25669 on Chr10 revealed in this study was close to associated marker reported by Uga et al. (2003b). RM7203 on Chr11 in this study was less than 5 cM away from the associated marker

identified by Yamamoto et al. (2003). In summary, we identified four SSR markers distributing on three new chromosomes (1, 7 and 9), four markers in different regions from previous studies on three chromosomes (5, 6 and 8) and two markers in similar regions with previous studies on two chromosomes (10 and 11). The addition of new chromosomes and new regions increases descriptive power for QTLs associated with stigma exsertion in rice.

Trait–trait correlation and marker–trait association in breeding

Correlation describes the degree to which several variables or traits vary together. Exsertion of dual stigmas (DStgE) doubles pollen receptivity, thereby doubling the probability of outcrossing for production of hybrid seed in comparison with the exsertion of one out of two stigmas in rice. As a result, we distinguished the DStgE from SStgE in the experiment like others (Yan and Li 1987; Yu et al. 2006b). For this consideration, Miyata et al. (2007) scored the stigma exsertion based on the flowered glumes. In other words, they counted the DStgE twice. It is not clear if others did this separation (Virmani and Athwal 1973; Uga et al. 2003a, b; Xiong et al. 1999; Yamamoto et al. 2003). However, over 60% of the mini-core accessions had zero DStgE (Supplementary Table), which brought about total stigma exsertion (TStgE) as sum of DStgE and SStgE. The three parameters describe stigma exsertion in different ways, so they are logically correlated to each other (Table 1). Rice is classified into short, medium and long grain types according to kernel length to width ratio and each type is related to certain cooking quality in the US (Mackill and McKenzie 2003). Due to this consideration, we measured spikelet length (SpkL) and width (SpkW) and then converted them to the ratio (SpkL/W) in this study. Obviously, the three spikelet traits are correlated to each other (Table 1).

Association mapping is to detect the degree to which gene and trait or genotype and phenotype vary together in a sampled population on the basis of linkage disequilibrium (Zondervan and Cardon 2004). When a molecular marker is associated with a phenotypic trait, it should associate with others that highly correlate with this trait in theory. The present study identified five molecular makers that were co-associated with two correlated exsertions of stigma among TStgE, DStgE and SStgE. Furthermore, five markers were co-associated with two spikelet traits and RM12521_Chr2 was associated with all three, SpkL, SpkW and SpkL/W. These correspondences between trait–trait correlation and trait–marker association further validated the resulting association mapping of stigma and spikelet characteristics.

In breeding program, there is high chance for a trait to come along with the subject trait when they are tightly correlated. In marker-assisted breeding, one marker co-associated with multiple traits that correlate each other can be used to identify all these traits for selection, which obviously improves breeding efficiency. Total stigma exsertion (TStgE) is the exsertion sum of single (SStgE) and dual (DStgE) stigma. We can use RM5 to assist the selection for improvement of stigma exsertion because it is highly associated with either SStgE ($R^2 = 0.169$) or DStgE ($R^2 = 0.300$) (Fig. 6; Table 2). Similarly with breeding for long grain cultivar, RM12521 can be solely used for breeding assistance because it was highly associated with either SpkL ($R^2 = 0.156$) or SpkW ($R^2 = 0.207$) or SpkL/W ($R^2 = 0.201$) and SpkL is highly and positively correlated with mature grain length ($r = 0.79$). Furthermore, the breeding assisted with RM12521 may result in an elongation of stigma because of the positive correlation between SpkL and stigma length ($r = 0.21$). However, breeder should pay special attention if the traits of interest are negatively correlated.

Heritability and breeding germplasm for stigma exsertion

In rice, the inheritance of stigma exsertion is quantitative with large dominant effects followed by additive effect, which results in a high heritability (Li and Chen 1985) up to 90% (Virmani and Athwal 1973). In our study, heritability in broad sense was over 75% for three exsertions of stigma (Table 1). Therefore, improvement of stigma exsertion to help trap more pollen dispersed from male parent and overcome the barrier of pollination caused by differences in flowering time between parents is feasible. The feasibility validates the effort to develop molecular markers for effective improvement of seed production in hybrid rice. The resulting SSR markers significantly associated with stigma exsertion in this study should well serve the improvement.

Two cultivars outstanding in stigma exsertion identified in present study could be used to improve hybrid seed production for either indica or japonica rice. Indica entry 1032 (cultivar 50638) is a long grain type (L/W of de-hulled kernel = 3.3) with light brown kernel and 78 cm in height, matured in 130 days and had high amylose content (25.2%) and low gelatinization temperature (ASV = 7) in Arkansas, US. Further, entry 1032 is moderately resistant to sheath blight (*Rhizoctonia solani*) and highly resistant to blast (*Pyricularia grisea*) during vegetative and productive stages in field and in laboratory for seven races, IB1, IB33, IB49, IC17, IE1, IG1 and IH1. This indicates that entry 1032 can be used as germplasm for improvement of not only seed production but also disease resistance in hybrid rice. However, japonica entry 671 (cultivar Linia 84 Icar) is medium grain type (L/W = 2.2) with light brown kernel and 110 cm in height, matured in 98 days and had intermediate amylose content (15.8%) and low gelatinization temperature (ASV = 6.3) in Arkansas, US. Adversely with entry 1032, entry 671 is highly susceptible to sheath blight and to blast during both stages in field and in laboratory for the races except IH1. Germplasm accessions described in this study are managed in the USDA-ARS Genetic Stocks—Oryza Collection, available for the public through http://www.ars.usda.gov/Main/docs.htm?docid=8318, free of charge and restrictions for research purposes.

References

Agrama HA, Eizenga GC (2008) Molecular diversity and genome-wide linkage disequilibrium pattern in worldwide rice and its wild relatives. Euphytica 160:339–355. doi:10.1007/s10681-007-9535-y

Agrama HA, Eizenga GC, Yan WG (2007) Association mapping of yield and its components in rice cultivars. Mol Breed 19:341–356. doi:10.1007/s11032-006-9066-6

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. doi:10.1214/aos/1013699998

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. doi:10.1093/bioinformatics/btm308

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177. doi:10.1534/genetics.105.044586

Brooks SA, Yan WG, Jackson AK, Deren CW (2008) A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele: *Rc-g*. Theor Appl Genet 117:575–580. doi:10.1007/s00122-008-0801-8

Eizenga GC, Agrama HA, Lee FN, Yan WG, Jia YL (2006) Identifying novel resistance genes in newly introduced blast resistant rice germplasm. Crop Sci 46:1870–1878. doi:10.2135/cropsci2006.0143

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. doi:10.1111/j.1471-8286.2007.01758.x

FAO (Food and Agriculture Organization of the United Nations) (2004) Rice is life. International Year of Rice, Rome, Italy

Franco J, Crossa J, Warburton ML, Taba S (2006) Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. Crop Sci 46:854–864. doi:10.2135/cropsci2005.07-0201

García LV (2004) Escaping the bonferroni iron claw in ecological studies. Oikos 105:657–663. doi:10.1111/j.0030-1299.2004.13046.x

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

Goudet J (1995) Fstat version 1.2: a computer program to calculate Fstatistics. J Hered 86(6):485–486

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. doi:10.1046/j.1471-8286.2002.00305.x

Kato H, Namai H (1987) Floral characteristics and environmental factors for increasing natural outcrossing rate for F1 hybrid seed production of rice *Oryza sativa* L. Jpn J Breed 37:318–330

Li T, Chen YW (1985) Genetics of stigma exsertion in rice. Rice Genet Newsl 2:84–85

Li C, Sun CQ, Mu P, Chen L, Wang XK (2001) QTL analysis of anther length and ratio of stigma exsertion, two key traits of classification for cultivated rice (*Oryza sativa* L.) and common wild rice (*O. rufipogon* Griff.). Acta Genetica Sin 28(8):746–751 (Chinese with English abstract)

Li WH, Dong GJ, Hu XM, Teng S, Guo LB, Zheng DL, Qian Q (2003) QTL analysis for percentage of exserted stigma in rice (*Oryza sativa* L.). Acta Genetica Sin 30:637–640

Liu K, Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Mackill DJ, McKenzie KS (2003) Origin and characteristics of US rice cultivars. In: Smith CW, Dilday RH (eds) Rice origin, history, technology and production. John Wiley & Sons Inc., New York, pp 87–100

Miyata M, Yamamoto T, Komori T, Nitta N (2007) Marker-assisted selection and evaluation of the QTL for stigma exsertion under japonica rice genetic background. Theor Appl Genet 114:539–548. doi:10.1007/s00122-006-0454-4

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:e190. doi:10.1371/journal.pgen.0020190

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel population genetic software for teaching and research. Mol Ecol Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. doi:10.1038/ng1847

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans models data. Am J Hum Genet 69:1–14. doi:10.1086/321275

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

Reiner A, Yakubov R, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19:368–375. doi:10.1093/bioinformatics/btf877

Ritland K (1996) Estimators for pairwise relatedness and inbreeding coefficients. Genet Res 67:175–186

Rogers JS (1972) Measures of genetic similarity and genetic distance. In: Studies in genetics VII. University Texas Publication 7213, Austin, Texas, pp 145–153

Sidharthan B, Thiyagarajan K, Manonmani S (2007) Cytoplasmic male sterile lines for hybrid rice production. J Appl Sci Res 3:935–937

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (Zea mays L.) comparisons with data from RFLPs and pedigree. Theor Appl Genet 95:163–173. doi:10.1007/s001220050544

TASSEL (2009). User manual, trait analysis by association, evolution and linkage. www.maizegenetics.net/tassel

Tian DC (1991) Cultivation for out-crossing in rice-principles and techniques to achieve high yield in hybrid seed production. Sichuan Press of Sci & Tec, Chengdu (Chinese)

Uga Y, Fukuta Y, Ohsawa R, Fujimura T (2003a) Variations of floral traits in Asian cultivated rice (Oryza sativa L.) and its wild relatives (O. rufipogon Griff.). Breed Sci 53:345–352. doi:10.1270/jsbbs.53.345

Uga Y, Fukuta Y, Cai HW, Iwata H, Ohsawa R, Morishima H, Fujimura T (2003b) Mapping QTLs influencing rice floral morphology using recombinant inbred lines derived from a cross between Oryza sativa L. and Oryza rufipogon

Griff. Theor Appl Genet 107:218–226. doi:10.1007/s00122-003-1227-y

Virmani SS (1994) Heterosis and hybrid rice breeding. Monogra Theor Appl Gent 22. Springer-Verlag

Virmani SS, Athwal DS (1973) Genetic variability in floral characteristics influencing outcrossing in Oryza sativa L. Crop Sci 13:66–67

Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244. doi:10.2307/2282967

Xin Z, Velten JP, Oliver MJ, Burke JJ (2003) High-throughput DNA extraction method suitable for PCR. BioTechniq 34:820–826

Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang QF (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between Oryza sativa and O. rufipogon. Theor Appl Genet 98:243–251. doi:10.1007/s001220051064

Xu YB (2003) Developing marker-assisted selection strategies for breeding hybrid rice. In: Janick J (ed) Plant breeding reviews ISBN0–471-35421-X 23:73–174. Wiley, New York

Xu YB, Shen ZT (1988) Receptivity of exserted stigma. Int Rice Res Newsl 13(3):7–8

Yamamoto T, Takemori N, Sue N, Nitta N (2003) QTL analysis of stigma exsertion in rice. Rice Genet Newsl 20:33–34

Yan WG, Li SF (1987) Study on out-crossing characteristics among male sterile lines containing same nucleus in rice. Hybrid Rice 4:8–11 (Chinese)

Yan WG, Rutger JN, Bryant RJ, Bockelman HE, Fjellstrom RG, Chen MH, Tai TH, McClung AM (2007) Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci 47:869–878

Yu J, Pressoir G, Briggs W, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland J, Kresovich S, Buckler ES (2006a) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi:10.1038/ng1702

Yu XQ, Mei HW, Luo LJ, Liu GL, Zou GH, Hu SP, Li MS, Wu JH (2006b) Dissection of additive, epistatic and Q × E interaction of quantitative trait loci influencing stigma exsertion under water stress in rice. Acta Genetica Sin 33(6):542–550. doi:10.1016/S0379-4172(06)60083-8

Yuan LP (1981) Hybrid rice. United States patent and trademark office 4305225

Yuan LP, Fu XQ (1995) Technology of hybrid rice production. Food and Agriculture Organization of the United Nations, Rome, Italy

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 31:e4. doi:10.1371/journal.pgen.0030004

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1(1):5–20. doi:10.3835/plantgenome2008.02.0089

Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. Nat Rev Genet 5:89–100. doi:10.1038/nrg1270