

Research Article

Association Analysis of Arsenic-Induced Straighthead in Rice (*Oryza sativa* L.) Based on the Selected Population with a Modified Model

Xiaobai Li,^{1,2} Biaolin Hu,³ Xuhao Pan,⁴ Ning Zhang,¹ and Dianxing Wu¹

¹State Key Lab of Rice Biology, IAEA Collaborating Center, Zhejiang University, Hangzhou, Zhejiang, China

²Zhejiang Academy of Agricultural Sciences, Hangzhou, China

³Rice Research Institute, Jiangxi Academy of Agricultural Sciences and Nanchang National Sub-Center for Rice Improvement, Nanchang, Jiangxi, China

⁴Institute of Tobacco Research of Chinese Academy of Agricultural Sciences, Qingzhou, Shandong, China

Correspondence should be addressed to Ning Zhang; zhangningzju@126.com

Received 19 January 2017; Accepted 4 June 2017; Published 25 July 2017

Academic Editor: Momiao Xiong

Copyright © 2017 Xiaobai Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A rice physiological disorder makes mature panicle keep erect with empty grains termed as “straighthead.” Straighthead causes yield losses and is a serious threat to rice production worldwide. Here, a new study of association mapping was conducted to identify QTL involved in straighthead. A subset of 380 accessions was selected from the USDA rice core collection and genotyped with 72 genome-wide SSR markers. An optimal model implemented with principle components (PCs) was used in this association mapping. As a result, five markers were identified to be significantly associated with straighthead. Three of them, RM263, RM169, and RM224, were consistent with a previous study. Three markers, RM475, RM263, and RM19, had a resistant allele associated with a decrease in straighthead rating (straighthead rating ≤ 4.8). In contrast, the two other marker loci RM169 and RM224 had a few susceptible alleles associated with an increase in straighthead rating (straighthead rating ≥ 8.7). Interestingly, RM475 is close to QTL “*qSH-2*” and “*As5*” with straighthead resistance, which was reported in two studies on linkage mapping of straighthead. This finding adds to previous work and is useful for further genetic study of straighthead.

1. Introduction

A physiological disorder of rice (*Oryza sativa* L.) exhibits a typical symptom with sterile florets and distorted palea and lemma [1]. The panicles bear unfilled grains and stay upright, while empty hulls are often distorted into a crescent or parrot-beak shape. That is why people call it by “straighthead.” In an extreme case, yield losses can approach 100%. The first report of straighthead dates from 1912 in southern states of the United States (US), including Arkansas, Louisiana, and Texas [2]. Besides, straighthead has been reported in Portugal [3], Thailand [4], Japan [5], Australia [6], and Argentina [7]. Breeding resistant cultivars is an efficient strategy to overcome straighthead. Marker assisted selection (MAS) will accelerate the breeding process. Association mapping

is promising for identifying causative polymorphisms for complex traits [8] because it takes advantage of accumulated historic recombination events in natural populations and can evaluate multiple alleles simultaneously [9, 10]. Diverse collections (population) have been extensively used for association mapping. However, the discrepancy exists in allele frequencies of many diverse populations and statistic models, which can dramatically influence the power for mapping the associated alleles [11]. Therefore, it is necessary to recheck the marker alleles associated with the QTL in detail. In this study, we optimized a model of association mapping and selected a subset of accessions with good phenotypic replication to map the genes responsible for straighthead in rice. The population for association mapping was developed from the USDA rice core collection as previously described [12, 13].

2. Materials and Methods

2.1. Plant Materials. A total of 990 accessions were selected from the USDA rice core collection [13] for straighthead evaluation based on maturity and plant status in a 2002 field test [12]. The field was treated with the arsenic containing herbicide monosodium methanearsonate (MSMA). MSMA can induce the symptom of straighthead and thus is commonly used for evaluating straighthead in rice. These accessions were planted in a randomized complete block design with three replications. As a control two straighthead resistant cultivars and two straighthead susceptible cultivars [1] were included in the center of each tier containing 99 plots. The resistant cultivars were Zhe733 and Jing185 and susceptible cultivars were “Cocodrie” and “Mars.” Straighthead was rated separately for early, intermediate, and late maturity groups of the accessions in August, September, and October, respectively. Because more than 80% of these accessions were highly susceptible to straighthead, 380 accessions were selected for association mapping.

2.2. Phenotyping and Genotyping. Straighthead was visually rated at maturity based on floret fertility or sterility and panicle emergence from the flag leaf sheath, with nine levels from resistance to susceptibility [1]. DNA was extracted and genotyped with one indel and 70 SSR markers covering the entire rice genome. Sixty-eight of these markers were obtained from <http://www.gramene.org>. The other two (AP5652-1 and AP5652-2) were developed from the BAC AP5652 in house as described by Li et al. [15]. PCR and electrophoresis were carried out according to the procedures [15].

2.3. Statistical Analysis. The model-based program *INSTRUCT* [16] was used to infer population structure using a burn-in of 50,000 runs, Markov chain Monte Carlo to 50,000 iterations, and a model allowing for admixture and correlated allele frequencies. *INSTRUCT* eliminates the assumption of Hardy–Weinberg equilibrium within clusters [16]. The population structure was graphically displayed using *Distribute* [17]. The number of groups (K) was set from 1 to 10, with 5 independent runs each, to identify the K with the highest log likelihood. The most probable structure number (K) was decided based on log probability- $\ln P(D)$ and deviance information criteria (DIC). An accession was assigned to a single group with its inferred maximum ancestry (Q) from the group. To validate the genetic structure and test for different models, principal components analysis (PCA) was performed with *NTSYSpc* software version 2.11V [18]. The PCA was performed on the correlation matrix and summarizes the major patterns of variation in a multilocus data set. The first three principal components were used to visualize the dispersion of accessions. Genetic distance was calculated using Nei’s genetic distance [19]. The best fit model was selected for mapping after model comparison. The model comparison were conducted as described by Li et al. [15, 20]. In order to reduce the risk of false marker-trait associations with high P inflation, a false-discovery rate (FDR) was calculated using MULTTEST procedure in SAS v9.2 [21].

TABLE 1: Identification of marker associated with straighthead responding to MSMA. Marker-straighthead associations detected with PCA model at P and q FDR value < 0.0001 and their position (cM) on chromosome (Chr) derived from 71 markers and 380 accessions in a diverse rice collection.

Maker	Chr	P value	q FDR
RM475*	2	6.80×10^{-8}	6.65×10^{-7}
RM263	2	1.58×10^{-5}	6.18×10^{-5}
RM169	5	2.95×10^{-6}	1.45×10^{-5}
RM224	11	1.77×10^{-7}	1.15×10^{-6}
RM19	12	1.28×10^{-8}	2.51×10^{-7}

*RM475 was also associated with straighthead resistance in linkage mapping of the F_2 population derived from *Zhe733/R312* [14].

Effects of allele at marker loci associated with straighthead were compared for least square means (LSMEANS option of PROC MIXED).

3. Results

In the collection, MSMA in the field induced straighthead with symptoms of distorted florets in the partially emerged panicles and no seed set at all for the susceptible check cultivars. As a result, the straighthead score in resistant check Zhe733 was categorized as “class I” and the susceptible checks Cocodrie and Mars were categorized as “class III.” Based on the straighthead score, the 380 accessions had 37 accessions classified into rating of 1–4, 168 into rating of 4.1–6.9, and 175 into rating of 7.0–9.0.

The set of 72 markers with a genome-wide distribution detected a total of 677 alleles in the collection of 380 accessions. The average number of alleles per locus was 9.54 ranging from 2 (RM507, RM338, RM455, and Rid12) to 30 (RM11229). Polymorphic information content (PIC) varied from 0.08 for AP5625-1 to 0.91 for RM11229 among these markers, with an average of 0.60.

Five model-based groups were identified by *INSTRUCT* and accessions were assigned into their corresponding groups according to their Q (ancestry index) (Figure 1(a)). The first three components of PCA containing 56.36% variation among the 380 accessions also exhibited a similar pattern of genetic structure (Figure 1(b)). All these approaches concluded that the five-group structure sufficiently explained genetic variation in the collection. Relative performance assessed by BIC among six models (naive, kinship, PCA, PCA + kinship, Q , and Q + kinship) for straighthead is present in Figure 2. The five dimensions of PCA were determined based on BIC and applied in PCA and PCA + kinship models. The PCA model had the smallest BIC score among the six models; thus it was selected as the best fit model for association mapping of straighthead.

Using the selected model, five marker loci were identified to be significantly associated with straighthead at the 0.01% level of q FDR (Table 1). The allelic effect for each associated loci could be estimated with the mean phenotypic value using LSMEANS statement in PROC MIXED. Comparisons among alleles were tested with option PDIF of LSMEANS

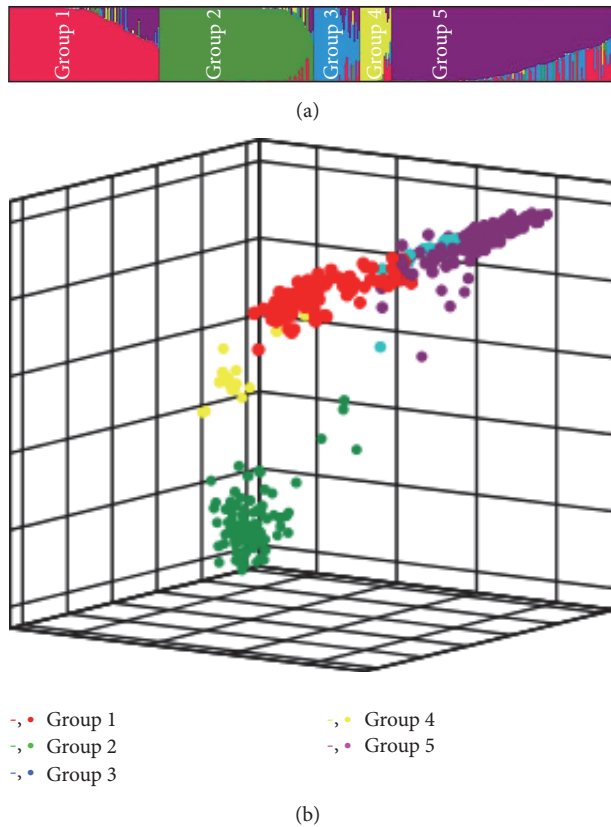


FIGURE 1: The genetic structure of 380 accessions for the analysis models of association mapping. (a) Estimated group structure is partitioned into five colored groups that represent the individual estimated levels of the five clusters and (b) principal component analysis (PCA) shows the accessions' pattern of spatial distribution. Each color represents one of the five groups indicated by the ancestry index. Each “-” or “•” represents one accession in (a) or (b), respectively.

statement individually for each marker. Allelic comparisons of the five marker loci displayed variation of their effects among alleles at the same locus (Figure 3). Three markers, RM475, RM263, and RM19, had a specific allele associated with a decrease in straighthead rating (straighthead rating ≤ 4.8) compared to the other alleles at their respective loci. In contrast, the two other marker loci RM169 and RM224 had a few alleles associated with an increase in straighthead rating (straighthead rating ≥ 8.7) compared to other alleles at the same locus. Three of the five markers were the same as those identified by Agrama and Yan [10]. However, RM475 with allele 194bp, which had the largest effect (3.3) for straighthead resistance among all the alleles (Figure 3), had not been reported previously [10]. Interestingly, RM475 was also identified to link to straighthead in a cross population [14].

4. Discussion

4.1. Model Comparison and Association Mapping. The simulations of Type I error estimated by BIC and visual plots of the observed and expected P values for all models suggest three

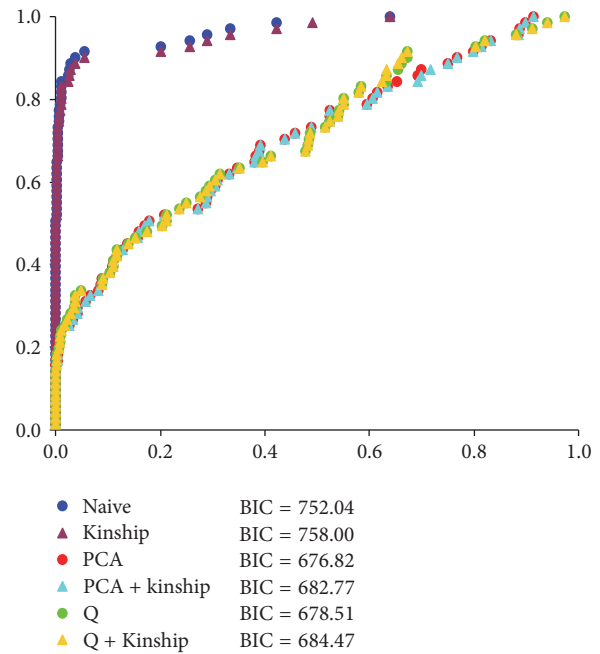


FIGURE 2: The selection of the best fit model for association mapping of straighthead. Comparative plots of observed versus expected P values and Bayesian information criterion (BIC) for six different association mapping models using 71 markers among 380 accessions. With an assumption that these markers are unlinked to the polymorphisms controlling straighthead, methods that appropriately control Type I error should show a uniform distribution of P values and have low BIC scores. The PCA model was selected as the best fit model for association mapping of straighthead due to the lowest BIC score and a uniform distribution of P values.

important points (Figure 2). First, the four models (Q, PCA, Q + kinship, and PCA + kinship) that account for population structure have fairly low BICs, which help control spurious associations and Type I error. Furthermore, naive and kinship models which do not account for population structure have very high BICs and an elevated risk of increased Type I errors. Second, kinship via ancestral relationship makes no improvement to the model as indicated by the highest BIC among the six models. PCA + kinship and Q + kinship models have higher BIC values than PCA and Q models, respectively. Shao et al. [22] has observed that the Q + kinship model performed similarly to the Q in their rice populations. The similar result has been observed in two another rice studies [15, 20]. However, kinship does improve the accuracy of association mapping results in studies with humans and cross pollinated crops [23, 24]. In our study, kinship among accessions may have less effect due to the low complexity of relatedness that results from the restricted gene flow of rice's self-reproductive mode and high genetic diversity represented in our rice panel [20, 25, 26]. Third, the PCA model with five dimensions performs better than the Q model based on five groups according to its smaller BIC. PCA is a fast and effective way to diagnose population structure [27, 28] and can handle a large number of markers and correct for subtle population stratification without being

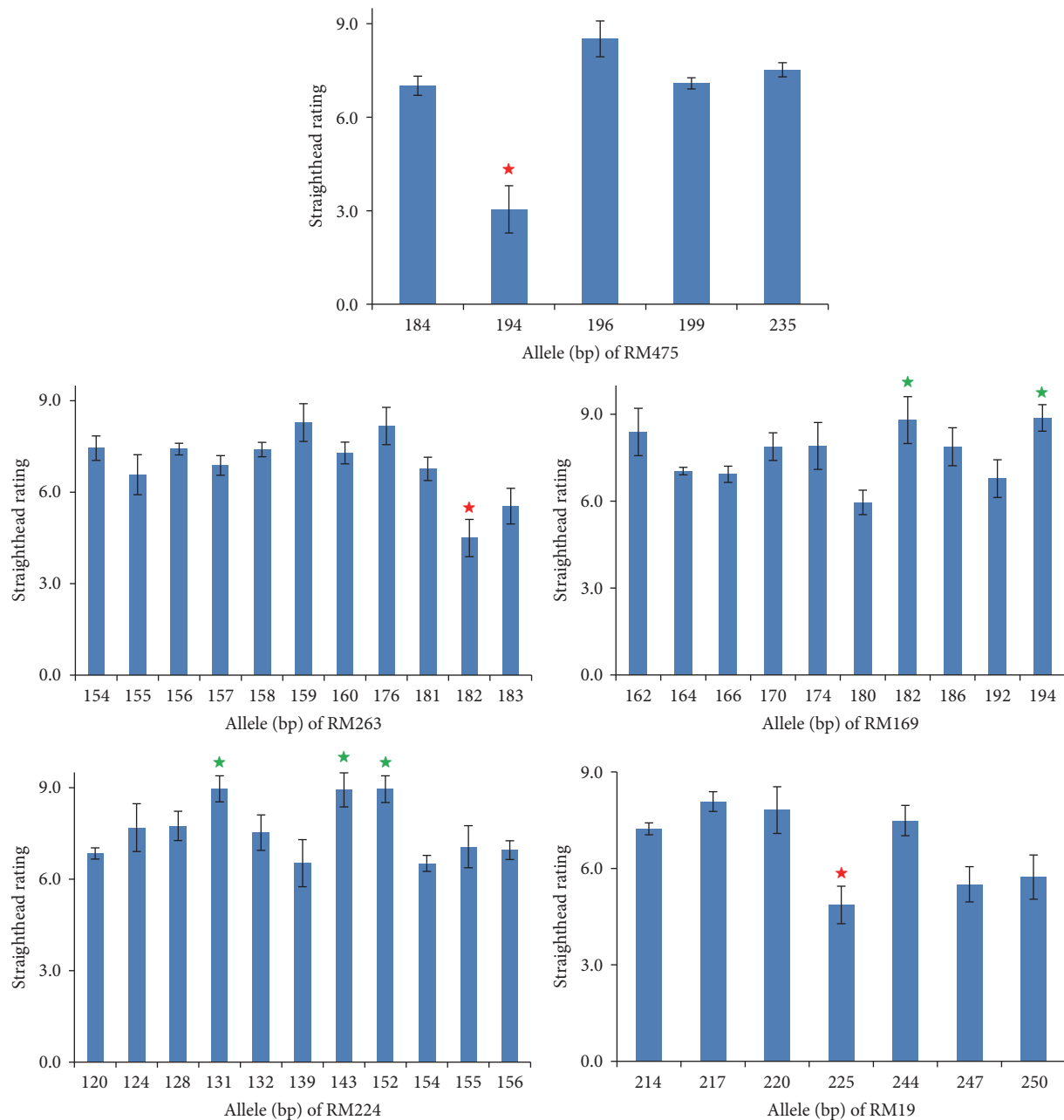


FIGURE 3: Deep analysis of the markers significantly associated with straighthead in terms of alleles. Alleles involved in straighthead resistance as shown by a decrease in effect of straighthead rating are indicated with a red star. Alleles involved in susceptibility of straighthead with an increased effect on straighthead rating are shown with a green star.

restricted by Hardy–Weinberg equilibrium [29–31]. These features of PCA may explain its increased performance over the Q model to control Type I error in our rice collection. Due to varying genetic composition, model selection should be used to determine the best model to use for association mapping with a specific population [15, 20].

Using the PCA model for our population, we identified five markers associated with straighthead (Table 1). Three of them, RM263, RM169, and RM224, were consistent with a previous study by Agrama and Yan [10] where seven markers were reported to associate with straighthead. This study

identified an additional novel marker associated with straighthead resistance, RM475, which was confirmed by linkage mapping with the F_2 population of Zhe733/R312 [14]. The 194 bp allele of RM475 was associated with the highest level of straighthead resistance as indicated by its strong effect (Figure 3). The reasons for different results between the present study and Agrama and Yan's [10] association mapping may be that (1) the distribution of allele frequencies and linkage disequilibrium (LD) may be substantially different between the two populations for mapping because they were built independently and consisted of different lines. Different

allele frequencies and LD would result in different results of association mapping [32, 33]; and (2) in order to control the spurious association or Type I error, we compared six models based on the BIC method and identified the best fit model “PCA” to map with our data. Therefore, the present study improves the results of previous study through the use of more stringent testing standards. Additionally this study validates previous findings through the use of a new rice panel composed of different accessions compared to the previous study and confirms the results of the novel marker identified in the study through the use of conventional linkage mapping with a segregating population.

4.2. Comparison of QTLs in Association Mapping with the Previous Ones. Occasionally markers linked to specific QTL are not always identified in association mapping. For example, while RM44 near *qSH-8* on Chr 8 was included in the association study it did not yield a significant association. *qSH-8* is a major straighthead QTL reported in two previous studies [14, 34]. This could be due to a low level of polymorphism of RM44. RM44 was monomorphic between the parents Zhe733 and R312 and thus could not be used in previous linkage mapping study [14, 34]. The polymorphism information content (PIC) of RM44 (0.28) was much lower than the average (0.45) over the other 70 markers within group 2. Group 2 had 25 resistant accessions accounting for 67.6% of the resistant accessions in this diverse population, including resistant Zhe733 with the 101 bp allele of RM44. However in group 2, 20 resistant accessions had the 101 bp allele, 25 moderately resistant accessions had the 101 bp allele, and 40 susceptible accessions had the 101 bp allele as well. Similarly, RM284 close to *qSH-8* was not associated with straighthead. These results suggest that the low levels of polymorphism within group 2 could decrease the power of QTL identification during association mapping.

On the other hand, the power of structure-based association analysis to detect the effects of individual genes is limited when population structure is found to explain too much of the variation [35]. By estimating the variation of allele frequencies (θ) at each locus among the groups in the diverse collection [36], we found that θ of locus RM284 near *qSH-8* was higher (0.54) than the average of other 70 markers (0.44), which indicated that the allele frequencies of RM284 were different from one group to another. In other words, the distributing pattern of RM284 alleles was in accordance with population structure. The accordance also can compromise the potential of marker to associate with gene(s)/QTL when population structure is taken in account in a model of association mapping. The similar result is also observed in another study [37]. In these cases, alternative populations for association mapping need to be evaluated for the candidate polymorphisms [37]. Additionally, an increase of marker coverage (i.e., high density SNP coverage) is a good option to increase the likelihood of polymorphic marker(s) and decrease the impact of population structure.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The project was supported by Rice Breeding Project (2016YFD0101801 and 2016C02052-6), the Open Fund of the State Key Laboratory of Rice Biology (160207), and the Doctoral Initiating Project of Innovation Fund of Jiangxi Academy of Agricultural Sciences, China (no. 20161CBS0002).

References

- [1] W. Yan, R. H. Dilday, T. H. Tai, J. W. Gibbons, R. W. McNew, and J. N. Rutger, “Differential response of rice germplasm to straighthead induced by arsenic,” *Crop Science*, vol. 45, no. 4, pp. 1223–1228, 2005.
- [2] J. S. Collier, *Rice Blight*, Illinois Agricultural Experimental Station Circular 156, University of Illinois, Urbana, Ill, USA, 1912.
- [3] J. M. A. Cunha and J. E. Baptista, *Estudo da branca do arroz. I. Combated a doenca Agronomia lusit*, vol. 20, 17-64, 1958.
- [4] P. Weerapat, “Straighthead disease of rice suspected in southern Thailand,” *IRRN*, vol. 4:67, 1979.
- [5] Y. Takeoka, Y. Tsutsui, and K. Matsuo, “Morphogenetic alterations of spikelets on a straighthead panicle in rice,” *Japanese Journal of Crop Science*, vol. 59, no. 4, pp. 785–791, 1990.
- [6] B. W. Dunn, G. D. Batten, T. S. Dunn, R. Subasinghe, and R. L. Williams, “Nitrogen fertiliser alleviates the disorder straighthead in Australian rice,” *Australian Journal of Experimental Agriculture*, vol. 46, no. 8, pp. 1077–1083, 2006.
- [7] W. G. Yan, F. Correa, A. Marin, J. Marassi, X. Li, and J. Re, “Comparative study on induced straighthead in the U.S. with natural straighthead in the U.S. with natural straighthead in Argentina,” in *Proceedings of the 33rd Rice Technical Working Group conference*, Biloxi, Mississippi, Miss, USA, 2010.
- [8] B. Stich, J. Möhring, H.-P. Piepho, M. Heckenberger, E. S. Buckler, and A. E. Melchinger, “Comparison of mixed-model approaches for association mapping,” *Genetics*, vol. 178, no. 3, pp. 1745–1754, 2008.
- [9] H. A. Agrama, G. C. Eizenga, and W. Yan, “Association mapping of yield and its components in rice cultivars,” *Molecular Breeding*, vol. 19, no. 4, pp. 341–356, 2007.
- [10] H. A. Agrama and W. G. Yan, “Association mapping of straighthead disorder induced by arsenic in *Oryza sativa*,” *Plant Breeding*, vol. 128, no. 6, pp. 551–558, 2009.
- [11] K. J. F. Verhoeven, G. Casella, and L. M. McIntyre, “Epistasis: Obstacle or advantage for mapping complex traits?” *PLoS ONE*, vol. 5, no. 8, Article ID e12264, 2010.
- [12] W. G. Yan, J. N. Rutger, H. E. Bockelman, and T. H. Tai, “Germplasm accessions resistant to straighthead in the USDA rice core collection,” in *Arkansas Agricultural Experiment Station*, University of Arkansas, Arkansas, Ark, USA.
- [13] W. Yan, J. N. Rutger, R. J. Bryant et al., “Development and evaluation of a core subset of the USDA rice germplasm collection,” *Crop Science*, vol. 47, no. 2, pp. 869–878, 2007.
- [14] X. Li, W. Yan, H. Agrama et al., “Genetic analysis of genetic basis of a physiological disorder “straighthead” in rice (*Oryza sativa* L.),” *Genes and Genomics*, vol. 38, no. 5, pp. 453–457, 2016.
- [15] X. Li, W. Yan, H. Agrama et al., “Mapping QTLs for improving grain yield using the USDA rice mini-core collection,” *Planta*, vol. 234, no. 2, pp. 347–361, 2011.

- [16] H. Gao, S. Williamson, and C. D. Bustamante, "A markov chain monte carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data," *Genetics*, vol. 176, no. 3, pp. 1635–1651, 2007.
- [17] N. A. Rosenberg, "DISTRUCT: a program for the graphical display of population structure," *Molecular Ecology Notes*, vol. 4, no. 1, pp. 137–138, 2004.
- [18] F. Rohlf, "FNTSYS-PC numerical taxonomy and multivariate analysis system ver 2.11L," *Applied Biostatistics*, 2000.
- [19] M. Nei, "The theory and estimation of genetic distance," N. E. Morton, Ed., University Press of Hawaii, Honolulu, Hawaii, USA, 1973.
- [20] X. Li, W. Yan, H. Agrama et al., "Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.)," *PLoS ONE*, vol. 7, no. 1, Article ID e29350, 2012.
- [21] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.
- [22] Y. Shao, L. Jin, G. Zhang, Y. Lu, Y. Shen, and J. Bao, "Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice," *Theoretical and Applied Genetics*, vol. 122, no. 5, pp. 1005–1016, 2011.
- [23] J. Yu, G. Pressoir, W. H. Briggs et al., "A unified mixed-model method for association mapping that accounts for multiple levels of relatedness," *Nature Genetics*, vol. 38, no. 2, pp. 203–208, 2006.
- [24] C. Zhu and J. Yu, "Nonmetric multidimensional scaling corrects for population structure in whole genome association studies," *Genetics*, vol. 182, no. 3, pp. 875–888, 2009.
- [25] X. Li, W. Yan, H. Agrama et al., "Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection," *Genetica*, vol. 138, no. 11, pp. 1221–1230, 2010.
- [26] W. Yan, H. Agrama, M. Jia, R. Fjellstrom, and A. McClung, "Geographic description of genetic diversity and relationships in the USDA rice world collection," *Crop Science*, vol. 50, no. 6, pp. 2406–2417, 2010.
- [27] N. Patterson, A. L. Price, and D. Reich, "Population structure and eigenanalysis," *PLoS Genetics*, vol. 2, p. e190, 2006.
- [28] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nature Genetics*, vol. 38, no. 8, pp. 904–909, 2006.
- [29] T. W. T. C. C. Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, Article ID 661678, pp. 661–678, 2007.
- [30] D. J. Hunter, P. Kraft, K. B. Jacobs et al., "A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer," *Nature Genetics*, vol. 39, no. 7, pp. 870–874, 2007.
- [31] L. A. Kiemeny, "Re: Genome-wide Association Study of Prostate Cancer Identifies a Second Risk Locus at 8q24," *European Urology*, vol. 52, no. 3, pp. 920–921, 2007.
- [32] L. R. Cardon and J. I. Bell, "Association study designs for complex diseases," *Nature Reviews Genetics*, vol. 2, no. 2, pp. 91–99, 2001.
- [33] K. A. B. Goddard, P. J. Hopkins, J. M. Hall, and J. S. Witte, "Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms five populations," *American Journal of Human Genetics*, vol. 66, no. 1, pp. 216–234, 2000.
- [34] X. Pan, Q. Zhang, W. Yan et al., "Development of genetic markers linked to straighthead resistance through fine mapping in rice (*Oryza sativa* L.)," *PLoS ONE*, vol. 7, no. 12, p. e52540, 2012.
- [35] S. A. Flint-Garcia, A.-C. Thuillet, J. Yu et al., "Maize association population: a high-resolution platform for quantitative trait locus dissection," *Plant Journal*, vol. 44, no. 6, pp. 1054–1064, 2005.
- [36] A. Rasamivelona, K. A. Gravois, and R. H. Dilday, "Heritability and genotype x environment interactions for straighthead in rice," *Crop Science*, vol. 35, no. 5, pp. 1365–1368, 1995.
- [37] L. Jin, Y. Lu, P. Xiao, M. Sun, H. Corke, and J. Bao, "Genetic diversity and population structure of a diverse set of rice germplasm for association mapping," *Theoretical and Applied Genetics*, vol. 121, no. 3, pp. 475–487, 2010.