Plant Science xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Genome-wide association study (GWAS) of resistance to head smut in maize

Ming Wang[a], Jianbing Yan[a], Jiuran Zhao[b], Wei Song[b], Xiaobo Zhang[a], Yannong Xiao[a], Yonglian Zheng[a,*]

[a] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, 430070 Wuhan, China
[b] Maize Research Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, PR China

ARTICLE INFO

Article history:
Received 5 April 2012
Received in revised form 9 June 2012
Accepted 3 August 2012
Available online xxx

Keywords:
MaizeSNP50
Genome-wide association study
Head smut
Zea mays

ABSTRACT

Head smut, caused by the fungus *Sphacelotheca reiliana* (Kühn) Clint, is a devastating global disease in maize, leading to severe quality and yield loss each year. The present study is the first to conduct a genome-wide association study (GWAS) of head smut resistance using the Illumina MaizeSNP50 array. Out of 45,868 single nucleotide polymorphisms in a panel of 144 inbred lines, 18 novel candidate genes were associated with head smut resistance in maize. These candidate genes were classified into three groups, namely, resistance genes, disease response genes, and other genes with possible plant disease resistance functions. The data suggested a complicated molecular mechanism of maize resistance against *S. reiliana*. This study also suggested that GWAS is a useful approach for identifying causal genetic factors for head smut resistance in maize.

1. Introduction

Maize head smut, caused by the fungus *Sphacelotheca reiliana* (Kühn) Clint, is a devastating disease that can lead to significant yield losses in most maize-growing regions of the world [1]. An annual yield loss of 0.3 million tons, which account for 10–15% of total yield, occurs in northern China [2]. Current disease control for field crops depends on chemical applications and field management practices. However, this strategy is time consuming, costly, and an environmental concern. Therefore, developing head smut-resistant cultivars has long been a favored method for disease control [3].

Conventional breeding for maize resistance against *S. reiliana* has met with a variety of difficulties such as the scarcity of resistant germplasm, making it an inefficient approach to trait-based selection [3]. Thus, maize breeders have made significant efforts in identifying molecular markers associated with resistance to improve breeding efficiency and shorten the breeding cycle. To date, many types of molecular markers, such as amplified fragment length polymorphisms, simple sequence repeats, and single nucleotide polymorphisms (SNPs), have been used in genetic linkage maps construction and quantitative trait locus (QTL) location. In recent years, studies on maize resistance against *S. reiliana* mainly focused on resistant QTL location. Using a population of 100 recombinant inbred lines (derived from Hi34 and TZil7) with 120 markers, Lu and Brewbaker [4] detected four QTLs on chromosomes 1, 3, 9, and 10. In the same year, Lübberstedt et al. [5] reported the location of three and eight QTLs in different chromosomes by genotyping a mapping population of 220 F_3 families derived from the 'D32' × 'D145' cross, in France and China, respectively. Genotyping a mapping population of 184 $F_{2:3}$ families (derived from Mo17 and Huangzao4), Gao [6] the location of six and eight QTLs in different year. With the 191 $F_{2:3}$ families produced from the cross of two Chinese elite inbreds (Mo17 and Huangzao4), Li et al. [7] detected five QTLs on chromosomes 1, 2, 3, 8, and 9 at one location, as well as five QTLs on chromosomes 1, 2, 3, 4, and 7 at another location. Based on the 68 BC_2 recombinants from the cross of 'Ji1037' and 'Huangzao4', the major resistance QTL, qHSR1, has been fine-mapped in bin 2.09 [8]. To the best of our knowledge, QTL or gene resistance to head smut has not been located by association analysis.

Association analysis based on linkage disequilibrium (LD) complements linkage mapping. This method generally consists of five steps: germplasm choice, estimation of population structure, trait evaluation, population sample genotype, and testing the genotypes and phenotypes for their associations [9]. Association analysis has been receiving unprecedented attention because of its advantages, including high resolution, cost efficiency, and non-requirement of pedigrees or crosses [10]. Moreover, genome-wide association studies (GWAS) are useful and powerful for genetic variations that underlie many important and complicated phenotypes in plants such as rice, *Arabidopsis*, and maize. These traits include disease resistance of *Arabidopsis thaliana* [11] and maize [12,13], flowering time of maize [14] and rice [15], carotenoid content of maize [16]. Hence, GWAS can reduce breeding program costs and time.

The first high-density 50 K SNP Array in maize has been recently developed, which offers a much higher resolution than previous arrays [17]. We initiated a genome-wide association study (GWAS)

* Corresponding author.
E-mail address: yonglianzheng@gmail.com (Y. Zheng).

0168-9452/$ – see front matter.
http://dx.doi.org/10.1016/j.plantsci.2012.08.004

using this SNP array to identify the chromosomal regions and specific SNPs that affect resistance levels in a commercial breeding population. Our objectives include the following: (1) to assess the extent of LD in the target population; (2) to identify the gene(s) or QTL(s) that significantly affect head smut resistance; and (3) to characterize those genes based on known function, and co-location of known QTLs for head smut resistance.

2. Materials and methods

2.1. Germplasm, experiment design and statistical analyses

A set of 144 inbred lines was used for whole-genome association mapping. The assembled lines included widely used parents of commercial hybrids in China. Two inbred lines from the International Maize and Wheat Improvement Center were also included. Phenotype was screened in two disease "hot spots", Gongzhuling, Jilin Province and Harbin, Heilongjiang Province in China during April to September in 2009 and 2010. In each plot, individuals were planted in a randomized complete block design with two replications, (three replications in Gongzhuling in 2010). The inbreds were grown in single 3 m rows 0.67 m apart, with a planting density of 45,000 plants/ha. Artificial inoculation and resistance scoring to head smut in the field were performed as described by Chen et al. [8]. The teliospores of *S. reiliana* were collected from the field in the previous growing season, stored in a cloth bag in a dry and well-ventilated environment, and then filtered prior to use. Seeds of the inbred lines were sown and covered with a mixture of soil and teliospore, at a ratio of 1000:1 [8]. The percentages of infected plants per plot were scored in either ears or tassels at the mature plant stage with completely expressed disease symptoms.

The best linear unbiased predictors for each line were calculated with the PROC MIXED procedure of SAS software. Broad sense heritability (H^2) was calculated in SAS as

$$H^2 = \frac{\delta_G^2}{[\delta_G^2 + (\delta_{GE}^2/n) + \delta_e^2/rn]}$$

where δ_G^2 is the genotypic variance, δ_{GE}^2 is the genotype × environment variance, δ_e^2 is the residual error variance, and n and r are the number of environments and replications, respectively [18].

2.2. Genotype and quality control

Young leaves from six random thirty-day seedlings per inbred line were bulked together and ground in liquid nitrogen. DNA was manually isolated from the leaves using a modified CTAB method [19]. DNA was quantified and genotyped using an Illumina BeadStation 500GX machine, following the manufacturer's instructions. The SNP content, selected from public and private sources, was characterized for genic representation, optimized physical spacing, diversity for 30 known diverse lines, and genetic distance. The MaizeSNP50 Genotyping BeadChip contains 56,110 markers spaced at a median distance of 40 kb. The Illumina BeadStudio genotyping software was used to assess gene clusters and call the data automatically. Each SNP, called by the Illumina software, was further re-checked manually and rescored if any error was detected in the clustering of homozygous and heterozygous groups [20]. This is due to the fact that the cluster separation from the BeadStudio software could be calculated on the degree of separation between two homozygous clusters from the heterozygous cluster, rather than that between the two homozygous groups, which are common in inbred lines with few heterozygotes. Thus, heterozygous SNPs were considered as missing data.

2.3. Population structure, relative kinship, and LD decay

The STRUCTURE software, with a Bayesian Markov Chain Monte Carlo model (MCMC), was used to estimate the population structure. Three runs of STRUCTURE were performed for each number of populations (k) (set from 1 to 10). Burn-in time and MCMC replication number were both set to 500,000 in each run. The true k value was determined by the log probability of the data (LnP(D)) and an ad hoc statistic delta k, based on the rate of change in LnP(D) between successive k values [21]. Principal component analysis (PCA) was also used to stratify the population structure. The PowerMarker software counted Nei's genetic distances [22] among given subgroups and created genetic distance matrices [23]. After double centering, distance matrices were used to obtain eigenvectors in NTSYS-pc Version 2.02 [24]. Using the SPAGeDi software, 9905 SNP loci with a minor allele frequency (MAF) ≥ 0.05 and a lower value of missing data of <0.1 were assessed for relative kinship [25]. All negative values from this software were set to zero. The parameter r^2 was used to estimate LD from data where the missing data were less than 20%. When assessing LD, a serial spacing between two loci on the same chromosome of 0.1 kb, 0.2 kb, 0.5 kb, 1 kb, 2 kb, 5 kb, 10 kb, 30 kb, 50 kb, 100 kb, 300 kb, 500 kb, 1 Mb, 3 Mb, 5 Mb, 40 Mb, 50 Mb, and 100 Mb were considered.

2.4. Model testing and genome wide association mapping

Six models referring to the population structure (Q, PCA) and kinship (K) were selected to correct for false positives. The general linear model (GLM) included the Q model, the PCA model, and a model that did not control for Q and PCA. The mixed linear model (MLM) comprised the K model, the Q + K model, and the PCA + K model. Both the GLM and MLM models were performed in TASSEL V2.1 [26]. Quantile–quantile plots were shown with a negative log value of the expected *P*-value from the genotype–phenotype association and the statistical *P*-value that deviated from the expected *P*-value and could give rise to false positive results. Both the quantile–quantile plots and Manhattan plots were drawn using the R package.

In this study, negative log (1/n) was used as a threshold based on the following reasons. First, the sample size is small and the statistical power is not enough. Second, the Bonferroni test (0.05/numbers of tests) criterion is typically very strict to be a threshold because GWAS is hypothesis generating [27]. The statistical threshold for GWAS was decreased to obtain the true associations for plants [28–31].

2.5. Association analysis of candidate genes

To investigate the causal polymorphisms of candidate genes, direct polymerase chain reaction (PCR) amplification and gene sequencing were employed, corresponding to the 5′- and 3′-end across a panel of 144 inbred lines. The sequencing primers are shown in Table S3. Sequence alignment for polymorphism identification was performed using the multiple sequence alignment program MUSCLE [32] and was refined manually using BioEdit [33]. Refined sequences were exported to Phylip [34] for further analysis. SNPs and InDels were identified in TASSEL V2.1 [26] for all polymorphisms present at a frequency of ≥0.05.

3. Results

3.1. Trait variation and heritability

The statistics of the phenotypic measurements of 144 maize lines for the GWAS are shown in Table S1. The observed value of susceptibility to head smut was within 0–83.83%, showing

Please cite this article in press as: M. Wang, et al., Genome-wide association study (GWAS) of resistance to head smut in maize, Plant Sci. (2012), http://dx.doi.org/10.1016/j.plantsci.2012.08.004

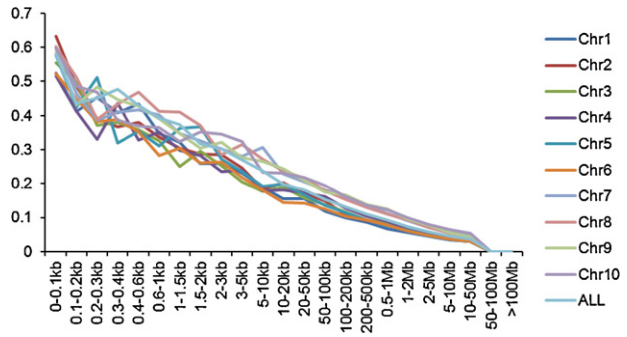

Fig. 1. LD across 10 chromosomes in maize. The *X*-axis indicates the physical distance between SNPs within the same chromosome and the *Y*-axis indicates the parameter r^2 of LD.

extensive variation. The broad sense heritability was calculated as approximately 88.7%. The $G \times E$ interaction was highly significant ($P < 0.001$).

3.2. SNP genotyping

Using the Illumina MaizeSNP50 BeadChip, genotypes were distinguished among the panel of 144 accessions. Excluding data with MAF <0.05 and missing data >20%, data from 45,868 SNPs were confirmed for further analysis. In addition, oligomer sequences from the 45,868 SNPs were blasted against the maize sequence database AGPV1 (http://www.maizesequence.org). Out of 45,868 oligomer sequences, 43,102 (92%) had a single megablast hit and could therefore be assigned to both a chromosome and a distinct position.

3.3. LD, population structure, relative kinship, and model testing

The distributions of r^2 between the different physical distances of each chromosome, as well as all the chromosomes, are presented in Fig. 1. The r^2 value sharply declined as the physical distance increased. LD decay for each chromosome is also different. In addition, the mean r^2 value varied among all chromosomes, ranging from ~100 kb to ~750 kb. The average r^2 for all chromosomes was estimated at ~200 kb, when the value of the cut off for r^2 was set to 0.1.

Population structure was assessed by running the STRUCTURE software for *K* values ranging from 1 to 10 on the entire panel using 28,791 high-quality SNP markers among the 45,868 SNPs. The different SNP sets for different analyses were due to reducing computing time. The value of LnP(D) increased continuously, with no obvious inflexion point. However, a significant increase was observed when delta *k* changed from 2 to 3 or 3 to 4 (Fig. S1). The population was divided into three subgroups based on the known pedigree and breeding history. A neighbor-joining (NJ) tree was also drawn, based on Nei's genetic distance (Fig. S2). The resulting NJ tree showed three divergent subgroups, in accordance with the results from population stratification.

Both the GLM and the MLM were used to perform the association analysis. Association results with the different models are shown as quantile–quantile plots of estimated $-\log 10(p)$ (Fig. 2). The results from the MLM (including K, PCA + K, and Q + K models) were better compared with those from the GLM model (including Q, PCA and no Q + PCA models). For resistance to head smut, the PCA + K and Q + K models performed much better than the K model with respect to false positive correction.

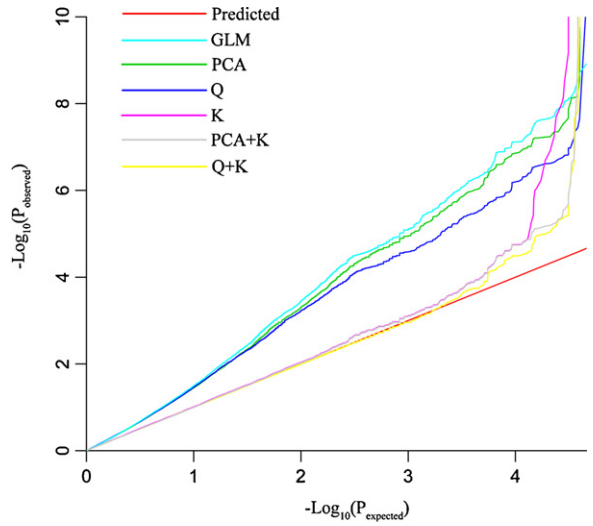

Fig. 2. Quantile–quantile plots of estimated $-\log 10(P)$ from association analysis using six methods in resistance to head smut. The red line is the expected line under the null distribution where there are few true marker associations. The observed *P* values are expected to nearly follow the expected *P* values. Deviations from expectation demonstrate that the statistical analysis may cause spurious associations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.4. Association mapping

Nineteen association signals were identified with $P < 2.18 \times 10^{-5}$ from the MLM in a genome wide scan. Among the 19 SNPs, eight were declared significant under the threshold of $P < 1.09 \times 10^{-6}$ (Fig. 3). The frequencies of SNP alleles for the 19 loci are shown in Table S2. The 19 SNPs explained 86.5% of the total phenotypic variation, ranging from 3.5% to 9.2%. Detailed information is summarized in Table 1. The 19 SNPs identified by association mapping were adjacent to or in 18 genes. Three significant SNPs were found within genes, and 16 from nearby genes. Several encode serine/threonine protein kinases known to be involved in plant disease response were detected among these genes. We also identified a gene encoding a leucine-rich repeat-containing protein, which may function in plant disease resistance pathways in response to a variety of external stimuli from pathogens [35].

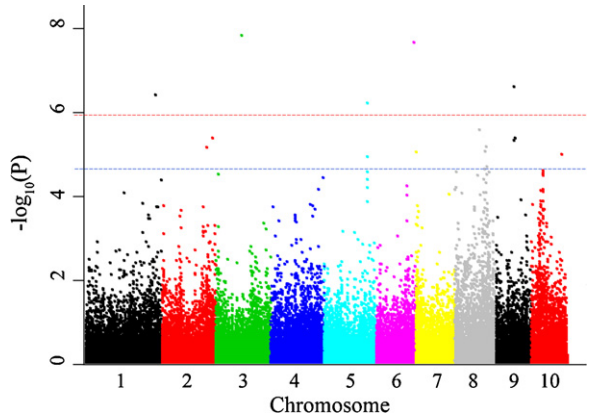

Fig. 3. Manhattan plots of a mixed linear model (MLM) for resistance to head smut. Plots above the blue horizontal dashed line show the genome-wide significance with a moderately stringent threshold of $-\log(1/45{,}868)$. Plots above the red horizontal dashed line show the genome-wide significance with stringent threshold of $-\log(0.05/45{,}868)$. The different colors indicate plots for different chromosomes, which follow the order: chromosome 1–chromosome 10. The plots with the $-\log 10(P)$ value above 8 were not shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: M. Wang, et al., Genome-wide association study (GWAS) of resistance to head smut in maize, Plant Sci. (2012), http://dx.doi.org/10.1016/j.plantsci.2012.08.004

Table 1
Physical positions of 19 SNPs significantly associated with head smut resistance and the predicted function or homology of adjacent candidate genes.

SNP	Chr.	Physical position (AGPv1 bp)	bin	r^2	*P*	Candidate genes	Function	Class	Previous studies
1	1	278884507	1.09	0.046	3.84E−07[a]	GRMZM2G300990	Serine/threonine protein kinase	Plant defense response	
2	2	56673949	2.04	0.043	5.54E−17[a]	GRMZM2G434669	Antifreeze	Plant defense response	
3	2	201359446	2.08	0.051	6.71E−06	GRMZM2G140231	Serine/threonine protein kinase	Plant defense response	Gao (2005)
4	2	219834173	2.09	0.093	4.05E−06	GRMZM2G166566	Basic-leucine zipper transcription factor	Plant defense response	Gao (2005), Chen et al. (2008), Li et al. (2008)
5	3	124139795	3.05	0.074	1.52E−08[a]	GRMZM2G137289	MADS-box	Plant defense response	Lübberstedt et al. (1999), Li et al. (2008)
6	4	220145226	4.08	0.036	6.31E−41[a]	GRMZM2G046816	Tubby-like	Plant defense response	
7	5	187810205	5.05	0.043	6.87E−06	GRMZM2G139858	Antifreeze	Plant defense response	
8	5	188373740	5.05	0.035	5.97E−07[a]	GRMZM2G312274	Auxin	Plant defense response	
9	6	163988192	6.06	0.047	2.25E−08[a]	AC195860.3_FG002	Unknown	Unknown	
10	7	3672618	7.01	0.038	9.00E−06	GRMZM2G465226	Pathogenesis-related protein	Disease resistance	Liet al. (2008)
11	8	103237278	8.03	0.036	1.39E−06	GRMZM2G047152	NB-ARC	Disease resistance	Lübberstedt et al. (1999)
12	8	131978501	8.05	0.045	4.86E−06	GRMZM2G075000	Actin cross link	Plant defense response	
13	8	136910174	8.05	0.039	3.74E−06	GRMZM2G164341	Antifreeze protein	Plant defense response	
14	8	136948482	8.05	0.039	1.16E−10[a]	GRMZM2G164341	Antifreeze protein	Plant defense response	
15	8	139906939	8.05	0.041	1.05E−05	GRMZM2G017603	NB-ARC	disease resistance	
16	9	84924940	9.03	0.037	4.58E−06	GRMZM2G443953	WD40 repeat	Plant defense response	Lübberstedt et al. (1999), Lu and Brewbaker (1999)
17	9	85140137	9.03	0.035	2.45E−07[a]	GRMZM2G383607	Unknown	Unknown	Lübberstedt et al. (1999), Lu and Brewbaker (2000)
18	9	90505475	9.03	0.050	4.24E−06	GRMZM2G348666	aminoacyl-tRNA ligase activity	Plant defense response	Lübberstedt et al. (1999), Lu and Brewbaker (2001)
19	10	137723733	10.05	0.037	1.01E−05	GRMZM2G117667	Lipase	Disease resistance	

r^2, proportion of phenotypic variance explained by SNP.
[a] Significant SNP-trait associations with a stringent threshold of −log(0.05/45,868).

Please cite this article in press as: M. Wang, et al., Genome-wide association study (GWAS) of resistance to head smut in maize, Plant Sci. (2012), http://dx.doi.org/10.1016/j.plantsci.2012.08.004

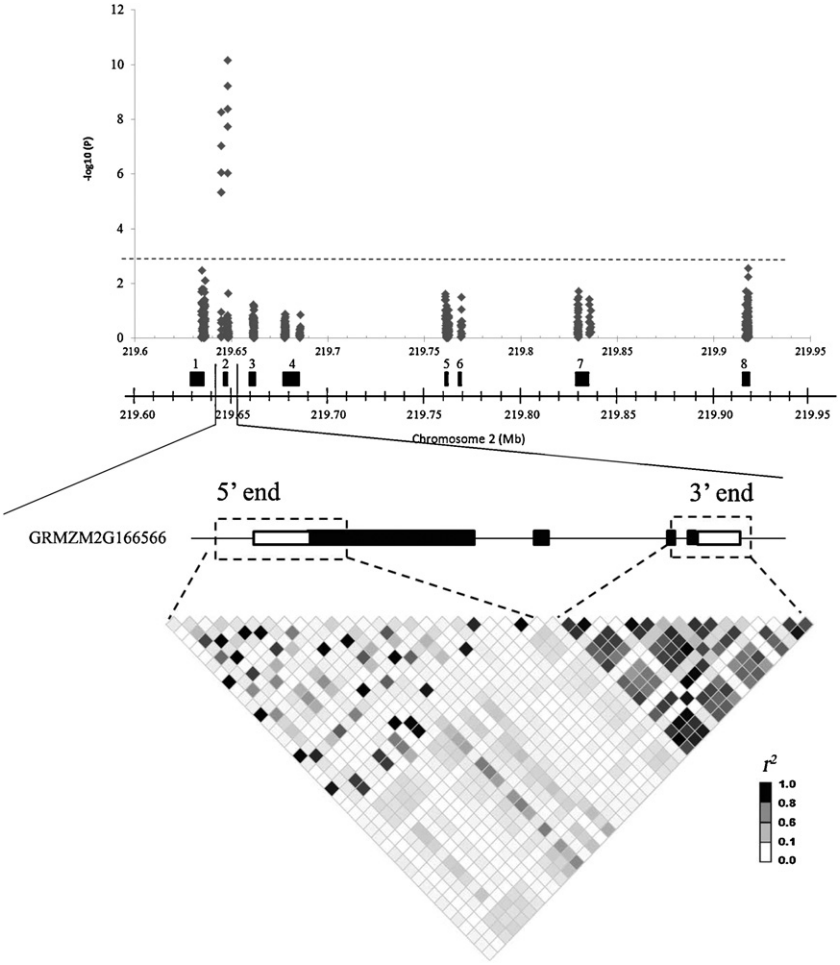

Fig. 4. Manhattan plots of mixed linear model (MLM) for resistance to head smut in the region expanding SNP 4–400 kb from both ends. Eight genes were GRMZM2G015933, GRMZM2G166566, GRMZM2G094955, GRMZM2G094978, GRMZM2G410975, GRMZM2G110894, GRMZM2G172101 and GRMZM2G052507, respectively. The diamonds show the result of the regional genome association. The plots above the black horizontal dashed line show the regional genome significance. The triangle indicates the significant result of genome-wide association in this region.

A tubby-like gene associated with the plant defense response was located in bin 4.08 [36]. Other genes involved in plant defense responses were identified, such as a MADS-box protein in bin 3.05, Auxin in bin 5.05, and a WD40 repeat-containing protein in bin 9.03. In particular, two genes encoding proteins with a nucleotide-binding site (NBS) commonly found among resistance (R) genes were identified on chromosome 8 [37]. Several other candidate genes identified by GWAS were related to stress response, such as several antifreeze genes responsive to cold damage.

Several studies indicate that a QTL located in bin 2.09 is responsible for head smut resistance [6–8]. We identified a significant signal of association (SNP 4) in this region through GWAS. Thus, candidate gene association in a region expanding SNP 4–400 kb from both ends was performed to ascertain whether a true association signal exists in bin 2.09. Eight candidate genes were predicted in this region (Fig. 4). To investigate the causal polymorphisms of these eight predicted genes within this 400 kb region, direct PCR amplification and sequencing of the 5′ and 3′ ends were employed. As a result, 1.6 Mb of sequence from a panel of 144 maize inbred lines was obtained, including 231 SNPs and 82 InDels. The association results of these candidate genes are shown in Fig. 4. Nine loci showed significant associations with head smut resistance. These nine loci all belonged to the gene GRMZM2G166566, which is annotated as a basic-leucine zipper transcription factor.

4. Discussion

In the present research, the level of LD decay was different among the ten chromosomes in maize, ranging from 0.5–1 Mb in chromosome 10 to 100–200 kb in chromosome 1. This trend is in accordance with previous studies [38]. Our results also showed an average LD decay of approximately 200 kb. In addition, assessing whether prior knowledge of the power of GWAS increasing the likelihood of meaningful data useful. The minimum number of markers required for a successful GWAS relies on the genome size and the rate of LD decay of the target germplasm. Using only 8590 loci in elite maize inbred lines, a gene associated with a trait concerning oleic content trait was uncovered [39]. To some degree, the maize SNP50 Array, with 56,110 SNPs partially derived from 19,540 genes, should be enough to scan GWAS for some special traits [16]. Genome wide trait-marker associations need to be selected using the optimum statistical method to eliminate false positive caused by complex patterns of the populations and relatedness in targeted populations [40]. Both GLM and MLM methods were employed. Previous reports indicated that the mixed model is superior to the general model [41,42]. In our study, the associated loci decreased to 19 when the mixed model (Q + K) was employed.

In the present study, 19 association loci were found to be distributed over 10 chromosomes. The single SNP can explain 9% of the phenotypic variation in maximum. A total number of 26 QTLs

Please cite this article in press as: M. Wang, et al., Genome-wide association study (GWAS) of resistance to head smut in maize, Plant Sci. (2012), http://dx.doi.org/10.1016/j.plantsci.2012.08.004

were identified from previous studies on maize head smut, but four QTLs had r^2 values greater than 10%. Several loci overlapped with previous studies [4–8]. Several studies have reported that a region in bin 2.09 harbors a QTL for head smut resistance, which corresponds to the SNP 4 location. We also identified a locus on chromosome 3 (bin 3.05). Interestingly, a cluster of R genes in the region 3L was found, including the *Rp3* [43], *mv1* [44], *wsm2* [45], *Scmv2* [46] genes, together with a QTL associated with resistance against *Gibberella zeae* infection [47] and to the European corn borer [48]. In addition, several associated loci exist on chromosome 8, one of which is located in the genomic region coinciding with linkage mapping-based results [5].

In many plants, disease resistance involves numerous genes and displays complex inheritance. The present GWAS study revealed 18 candidate genes that could be classified into three groups according to their predicted functions. The first group contains plant disease resistance R genes, such as GRMZM2G047152, which encode the protein with NBS domain. Most known R genes encode NBS proteins. In maize, an NBS-encoding gene, *RP1-D*, plays a role in resistance to rust (*Puccinia sorghi*) [49]. The second group contains genes for disease response. For example, GRMZM2G137289 contains a MADS box involved in plant disease response [50]. GRMZM2G44395 encodes a WD40 repeat domain associated with disease response [51]. The third group contains genes encoding proteins with domains homologous to other proteins that function in plant disease resistance. In this study, candidate genes overlapped with those identified in two previous studies in functions [12,13]. One of the 19 loci was selected to ascertain whether the true association signal existed. Our results proved that this signal was indeed associated with resistance to head smut.

There were two limitations in the panel used for the genome-wide association. First, the number of inbred lines was small, which could weaken the power of the association and could result in some loci being missed, especially those with small effects. It has been reported that a panel of 155 could obtain 59.2% of the quantitative genes related to 5% of the phenotypic variation [42]. Rare SNPs with large effects, untyped SNPs, and other structural variations (e.g., microsatellites, variable number tandem repeats, retro-element insertion, deletions, and duplication), could not be identified. New trends in combination association analysis with larger populations and bi-parental linkage mapping could help to resolve rare variant locations for traits [11,52].

In conclusion, the current study is the first to apply the Illumina MaizeSNP50 to GWAS for head smut resistance. Eighteen candidate genes were shown to be associated with the phenotypic variation of resistance to head smut disease. This research may serve as a basis for resistant genes or QTL cloning to understand further the complicated molecular mechanism of maize resistance against *S. reiliana*. To validate the significant association signals, further research is required. Further studies may focus on, deep association work (i.e., resequencing the rest of the gene from the panel), QTL mapping/fine-mapping, characterization of near-isogenic lines, gene knock-outs, complementation studies, larger sample sizes, association analysis in more diverse populations, and validation of findings in other maize lines and environments to identify other genetic factors contributing to phenotypic variations.

Acknowledgments

This project is supported by Beijing Agricultural Innovative Platform – Beijing Natural Science Fund Program (D08070500690802). The authors gratefully thank Pro. Zhenhua Wang from Northeast Agriculture University and Pro. Guoqing Tan from Jilin Academy of Agricultural Sciences for the phenotypic screening on head smut resistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2012.08.004.

References

[1] R.A. Frederiksen, Head smuts of corn and sorghum, Proc. Corn Sorghum Res. Conf. 32 (1977) 89–104.
[2] Q. Jin, J. Li, X. Zhang, G. Wang, S. Song, Y. Liu, L. Wang, Establishment IPM of system of corn diseases and pest insects in the spring corn belt, J. Maize Sci. 8 (2000) 84–88.
[3] Z. Wang, Y. Jiang, L. Wang, Y. Jin, X. Li, H. Shi, Research advance on Head smut, J. Maize Sci. 10 (2002) 61–64.
[4] X.W. Lu, J.L. Brewbaker, Molecular mapping of QTLs conferring resistance to *Sphacelotheca reiliana* (Kühn) Clint, Maize Genet. Coop. News Lett. 73 (1999) 36.
[5] T. Lübberstedt, X.C. Xia, G. Tan, X. Liu, A.E. Melchinger, QTL mapping of resistance to *Sporisorium reiliana* in maize, Theor. Appl. Genet. 99 (1999) 593–598.
[6] S. Gao, Inheritance and quantitative trait loci mapping of resistance to head smut caused by *Sphacelotheca reiliana* in Maize, PhD dissertation, Jilin University, Changchun, 2005.
[7] X.H. Li, Z.H. Wang, S.R. Gao, H.L. Shi, S.H. Zhang, M.L.C. George, M.S. Li, C.X. Xie, Analysis of QTL for resistance to head smut (*Sporisorium reiliana*) in maize, Field Crop Res. 106 (2008) 148–155.
[8] Y. Chen, Q. Chao, G. Tan, J. Zhao, M. Zhang, Q. Ji, M. Xu, Identification and fine-mapping of a major QTL conferring resistance against head smut in maize, Theor. Appl. Genet. 117 (2008) 1241–1252.
[9] C. Zhu, M. Gore, E.D. Buckler, J. Yu, Status and prospects of association mapping in plants, Plant Genome 1 (2008) 1–20.
[10] S. Myles, J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, E.S. Buckler, Association mapping: critical considerations shift from genotyping to experimental design, Plant Cell 21 (2009) 2194–2202.
[11] A. Nemri, S. Atwell, A.M. Tarone, Y.S. Huang, K. Zhao, D.J. Studholme, M. Nordborg, J.D.G. Jones, Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 10302–10307.
[12] J.A. Poland, P.J. Bradbury, E.S. Buckler, R.J. Nelson, Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 6893–6898.
[13] K.L. Kump, P.J. Bradbury, R.J. Wisser, E.S. Buckler, A.R. Belcher, M.A. Oropeza-Rosas, J.C. Zwonitzer, S. Kresovich, M.D. McMullen, et al., Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population, Nat. Genet. 43 (2011) 163–168.
[14] E.S. Buckler, J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, et al., The genetic architecture of maize flowering time, Science 325 (2009) 714–718.
[15] X. Huang, Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, W. Li, Y. Guo, L. Deng, et al., Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm, Nat. Genet. 44 (2011) 32–39. Q5
[16] C.E. Harjes, T.R. Rocheford, L. Bai, T.P. Brutnell, C.B. Kandianis, S.G. Sowinski, A.E. Stapleton, R. Vallabhaneni, M. Williams, et al., Natural genetic variation in Lycopene Epsilon Cyclase tapped for maize biofortification, Science 319 (2008) 330–333.
[17] G. Martin, The maize infinium 60K array – design, first data and potential applications, in: Abstr. Plant Animal Genome Conf, XVIII, San Diego, CA, 9–13 January, 2010, Available at: http://www.intl-pag.org/18/18i-illumina.html (verified 08.12.10).
[18] S.J. Knapp, W.W. Stroup, W.M. Ross, Exact confidence intervals for heritability on a progeny mean basis, Crop. Sci. 25 (1985) 192–195.
[19] M.A. Saghai-Maroof, K.M. Soliman, R.A. Jorgensen, R.W. Allard, Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 8014–8019.
[20] J. Yan, X. Yang, T. Shah, H. Sánchez-Villeda, J. Li, M.L. Warburton, Y. Zhou, J.H. Crouch, Y. Xu, High-throughput SNP genotyping with the GoldenGate assay in maize, Mol. Breeding 25 (2010) 441–451.
[21] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[22] M. Nei, Genetic distance between populations, Am. Nat. 106 (1972) 283–292.
[23] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, Bioinformatics 21 (2005) 2128–2129.
[24] F.J. Rohlf, NTSYS-Pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1, Exeter Software, New York, 2000.
[25] N. Patterson, A.L. Price, D. Reich, Population structure and eigenanalysis, PLoS Genet. 2 (2006) e190.
[26] J. Yu, G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, et al., A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2006) 203–208.
[27] J. Yang, T.A. Manolio, L.R. Pasquale, E. Boerwinkle, N. Caporaso, J.M. Cunningham, M. de Andrade, B. Feenstra, E. Feingold, et al., Genome partitioning of

genetic variation for complex traits using common SNPs, Nat. Genet. 43 (2011) 519–525.
[28] Y. Lu, S. Zhang, T. Shah, C. Xie, Z. Hao, X. Li, M. Farkhari, J.M. Ribaut, M. Cao, T. Rong, Y. Xu, Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 19585–19590.
[29] M.N. Weedon, H. Lango, C.M. Lindgren, C. Wallace, D.M. Evans, M. Mangino, R.M. Freathy, J.R.B. Perry, S. Stevens, et al., Genome-wide association analysis identifies 20 loci that influence adult height, Nat. Genet. 40 (2008) 575–583.
[30] G. Lettre, A.U. Jackson, C. Gieger, F.R. Schumacher, S.I. Berndt, S. Sanna, S. Eyheramendy, B.F. Voight, J.L. Butler, et al., Identification of ten loci associated with height highlights new biological pathways in human growth, Nat. Genet. 40 (2008) 584–591.
[31] D.F. Gudbjartsson, G.B. Walters, G. Thorleifsson, H. Stefansson, B.V. Halldorsson, P. Zusmanovich, P. Sulem, S. Thorlacius, A. Gylfason, et al., Many sequence variants affecting diversity of adult human height, Nat. Genet. 40 (2008) 609–615.
[32] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acid Res. 32 (2004) 1792–1797.
[33] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids. Symp. Ser. 41 (1999) 95–98.
[34] J. Felsenstein, PHYLIP – phylogeny inference package (version 3.2), Cladistics 5 (1989) 164–166.
[35] A.F. Bent, D. Mackey, Elicitors, effectors, and R genes the new paradigm and a lifetime supply of questions, Annu. Rev. Phytopathol. 45 (2007) 399–436.
[36] Y. Kou, D. Qiu, L. Wang, X. Li, S. Wang, Molecular analyses of the rice tubby-like protein gene family and their response to bacterial infection, Plant Cell Rep. 28 (2009) 113–121.
[37] T.K. Eitas, J.L. Dangl, NB-LRR proteins: pairs, pieces, perception, partners, and pathways, Curr. Opin. Plant Biol. 13 (2010) 472–477.
[38] J. Yan, T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, J. Crouch, Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers, PLoS One 4 (2009) e8451.
[39] A. Beló, P. Zheng, S. Luck, B. Shen, D.J. Meyer, B. Li, S. Tingey, A. Rafalski, Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize, Mol. Genet. Genomics 279 (2008) 1–10.
[40] Z. Zhang, E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, et al., Mixed linear model approach adapted for genome-wide association studies, Nat. Genet. 42 (2010) 355–360.
[41] X. Huang, X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, et al., Genome-wide association studies of 14 agronomic traits in rice landraces, Nat. Genet. 42 (2010) 961–967.
[42] X. Yang, J. Yan, T. Shah, M.L. Warburton, Q. Li, L. Li, Y. Gao, Y. Chai, Z. Fu, et al., Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection, Theor. Appl. Genet. 121 (2010) 417–431.
[43] S. Sanz-Alferez, T.E. Richter, S.H. Hulbert, J.L. Bennettzen, The Rp3 disease resistance gene of maize: mapping and characterization of introgressed alleles, Theor. Appl. Genet. 91 (1995) 25–32.
[44] R. Ming, J.L. Brewbaker, R.C. Pratt, T.A. Musket, M.D. McMullen, Molecular mapping of a major gene conferring resistance to maize mosaic virus, Theor. Appl. Genet. 95 (1997) 271–275.
[45] M.D. McMullen, M.W. Jones, K.D. Simcox, R. Louie, Three genetic loci control resistance to wheat streak mosaic virus in the maize inbred Pa405, Mol. Plant Microbe In. 7 (1994) 708–712.
[46] A.E. Melchinger, L. Kuntze, R.K. Gumber, T. Lübberstedt, E. Fuchs, Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm, Theor. Appl. Genet. 96 (1998) 1151–1161.
[47] M.E. Pè, L. Gianfranceschi, G. Taramino, R. Tarchini, P. Angelini, M. Dani, G. Binelli, Mapping quantitative trait loci (QTL) for resistance to *Gibberella zeae* infection in maize, Mol. Gen. Genet. 241 (1993) 11–16.
[48] C.C. Schön, M. Lee, A.E. Melchinger, W.D. Guthrie, W.L. Woodman, Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs, Heredity 70 (1993) 648–659.
[49] N. Collins, J. Drake, M. Ayliffe, Q. Sun, J. Ellis, S. Hulbert, T. Pryor, Cular characterization of the maize Rp1-D rust resistance haplotype and its mutants, Plant Cell 11 (1999), 1365-1176.
[50] S. Lee, Y.M. Woo, S.I. Ryu, Y.D. Shin, W.T. Kim, K.Y. Park, I.J. Lee, G. An, Further characterization of a rice AGL12 group MADS-box gene, OsMADS26, Plant Physiol. 147 (2008) 156–168.
[51] J.C. Jang, P. Leon, L. Zhou, J. Sheen, Hexokinase as a sugar sensor in higher plants, Plant Cell 9 (1997) 5–19.
[52] B. Brachi, N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, J. Bergelson, J. Cuguen, F. Roux, Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature, PLoS Genet. 6 (2010), e1000940.

Please cite this article in press as: M. Wang, et al., Genome-wide association study (GWAS) of resistance to head smut in maize, Plant Sci. (2012), http://dx.doi.org/10.1016/j.plantsci.2012.08.004