Association analysis of important agronomical traits of maize inbred lines with SSRs

Qianqian Zhang, Chenglai Wu, Fengyang Ren, Yan Li, Chunqing Zhang*

**State Key Laboratory of Crop Biology, Shandong Agricultural University,
Tai'an 271018 Shandong, China**

*Corresponding author: cqzhang@sdau.edu.cn

Abstract

The genetic markers of important traits are evaluated in order to improve the maize inbred lines. Ninety-four maize inbred lines were used to assess the genetic and phenotypic diversity and make association analysis of 26 agronomical traits with 204 genome-wide SSR markers, which were divided into five subpopulations by a model based population structure analysis. The population consisted of 94 maize inbred lines, presented high genetic diversity and significant linkage disequilibrium (LD), and could be used in the detection of genome-wide SSR marker-phenotype association. Although a total of 106 loci were associated with the trait of the mean results of two years at $P<0.01$ level, thirty-nine association loci were detected with an MLM association analysis model to existing significant association ($P<0.05$) with 17 traits in two years, simultaneously, in which there were three loci associated with PH, four loci with AD, five loci with KRN, three loci with HKW, etc. Five association loci were new discovery, which were bnlg2162, bnlg1118, phi077, umc1161 with BYC, and bnlg1118 with GLN. The strongest association loci were umc1917 with AD and HKW ($P<0.01$), umc2025 with CD ($P<0.0001$), etc. The number of associated loci detected on chromosome 1 was thirteen, which was more than chromosome 2 and 5(5), and more than chromosome 4(4), etc. The above results were useful for genetic improvement and molecular maker-assisted breeding in maize.

Keywords: genetic diversity, population structure, marker-phenotypic, association analysis, agronomical traits.
Abbreviations: LD, linkage disequilibrium; Q, population structure; K, kinship; GLM, general linear model; MLM, mixed linear model; PIC, polymorphism information content; MAF, major allele frequency; PH, plant height; EH, ear height; LA, leaf angle above ear; LL, leaf length at the ear; LW, leaf width at the ear; BYC, bract leaf length; LN, leaf number; GLN, the number of green leaves at mature; TD, days to tasseling; AD, days to anthesis; SD, days to silking; MD, days to mature; EL, ear length; ED, ear diameter; BL, bald length; EW, ear weight; CW, cob weight; CD, cob diameter; KRN, the number of kernel rows; RKN, the number of kernels per row; GW, grain weight per ear; EL, embryo length; GL, grain length; HKW, hundred kernels weight; TAI, tasseling to-anthesis internal days; ASI, anthesis to-silking internal days; KRO, kernel ratio; EL/GL, embryo length ratio of grain length.

Introduction

Maize (*Zea mays* L.) is one of the most important crops in the world, serving as a source of food, feed and fuel. To address increased demands globally due mainly to rapid population growth, energy insufficiencies and environmental issues, it is essential to improve maize productivity and quality through efficient breeding programs (Tester and Langridge 2010). Most important traits related to yield in cereals are quantitatively inherited and are difficult to investigate. Compared with linkage analysis, association mapping has a number of advantages that include shorter research time, higher mapping resolutions and investigation of a greater number of alleles (Yu and Buckler 2006). Because association mapping is a powerful and, thus, is a widely used approach for identifying the genes or loci that affect the phenotypic variations, a number of association mapping studies have been conducted to investigate the causal variants associated with many important traits, including flowering time (Camus-Kulandaivelu et al., 2006; Ducrocq et al., 2009; Pressoir et al., 2009), kernel starch (Wilson et al., 2004), maysin synthesis (Szalma et al., 2005), forage quality (Andersen et al., 2007), carotenoid content (Harjes et al., 2008; Yan et al., 2010), kernel oil (Belo et al., 2008) and kernel size (Li et al., 2010a and 2010b). The details have

been reviewed recently by Yang et al. (2010a and 2010b) and Yan et al. (2011). Maize is a desirable crop for association mapping due to its great genetic diversity and rapid linkage disequilibrium (LD) decay. Indeed, a large-scale maize QTL/association mapping population (nested association mapping, NAM) has been constructed to dissect the genetic basis of many quantitative traits with great power (Yu et al., 2008; Buckler et al., 2009; McMullen et al., 2009). This association panel comprises 5,000 inbred lines that mainly came from crosses using a common parent, B73 and crossed with each of the 25 diverse founder lines. At present, SSRs are the most widely used markers in maize research. SNPs are also very popular molecular markers as they can be reliably applied on a large scale of linkage analysis, association studies and they are highly amenable for automation. However, SNP markers are not as informative as SSRs because they have a biallelic nature (Rosenberg et al., 2003; Liu et al., 2005) and one has to increase the SNPs in order to gain the same information (Hamblin et al., 2007). Many factors, for example, population structure, sample size and frequency of specific alleles, may affect accuracy of association analysis and may influence the ability to detect false positive associations. Among a number of models

proposed and used for minimizing the false-positive of association, the population structure (Q), kinship (K) and a combination of Q and K (Q + K) methods have been considered to be superior to those conventional linear models in association analyses (Yu and Buckler 2006). Although it is difficult to determine which significance level is acceptable in a given association study, many methods can verify if the identified polymorphisms are, indeed, significantly associated with the target trait. For example, a P value may be improved by adding additional individuals to the same panel or confirmed in independent panels of germplasm, thereby, increasing the researcher's confidence in the marker-trait association. Finally, based on the genetic diversity information provided by SSR markers and adaptation data obtained from the filed experiments, 94 lines and 26 phenotypic traits were chosen for the present study. In our study, we found the more valuable marker-trait association loci through finding the same associated loci of the two year period. Then we compared the same association loci with association results of the mean results of two years at the same significant level, which could delete some false positive results and made our results more valuable. The objectives of our research were to (1) assess the phenotypic and genetic diversity of our association panel; (2) investigate the population structure among the inbred lines; (3) association analysis of agronomical traits with markers and compared with the predecessors' QTL positioning results.

Results

Phenotypic variations of measured quantitative traits

Extensive phenotypic variations were observed for all the measured quantitative traits in this maize panel, as shown by the descriptive statistics in Table S2. Correlation coefficient analysis was conducted for each trait of 2010 and 2011. Specifically, significant correlations at *P<0.01* and *P<0.05* levels were found with 19 traits and six traits (e.g., TAI, MD, RKN, ED, KRO and EL/GL), respectively, suggesting the strong genetic impact. Bad Length (BL) was the only trait that did not show significant difference at either level, indicating the role of environmental factors in determining this trait. Data of the two years' traits were averaged, then we used SPSS to descriptive the statistics. Cob weight, with an average of 23.97 g/cob, showed the largest variation (6.40-fold), ranging from 7.56 to 48.37 g/cob. Whereas, days to mature (MD, day), with an average of 97 days, exhibited the least variation (1.23-fold), ranging from 87 to 107 days. Because of such extensive phenotypic and genetic variations, it is clear that we should be able to conduct the association analysis by using this association panel.

Genetic diversity of inbred lines

To evaluate the genetic diversity of the 94 inbred lines, we used 204 SSRs which are randomly and uniformly distributed within the whole maize genome. High polymorphism of these SSRs within the 94 inbred lines was detected: 1, 460 alleles with an average of 6.30 alleles/locus, ranging from 2-16 alleles. As shown in Fig. 1a, the major allele frequency (MAF) exceeding 40% value was within 0.3-0.6, with the mean of 0.4622, ranging from 0.1579 to 0.8579. Whereas, the polymorphic information content (PIC) exceeding 40% value was within 0.5-0.8, with the mean of 0.6095, ranging from 0.2141 to 0.8738 (Fig. 1b). Meanwhile, the average genetic diversity was 0.6597 with the range between 0.2438-0.8849.

Population structure and relative kinship

In order to understand the genetic structure of the association analysis population, a model-based approach in the STRUCTURE software was used to subdivide each inbred line into the corresponding subgroup. As the STRUCTURE software overestimates the number of subgroups for inbred lines, and it is difficult to choose the "correct" k from the LnP(D) (Fig. 2). Thus, Δk (Evanno et al., 2005) was used to determine the k value. Fig. 2 indicated that when Δk was five, the model-based subgroups were rather consistent with the known pedigrees of the inbred lines. The five subgroups matched the five major germplasm, Reid, Sipingtou (SPT), Luda Red Cob (LRC), PB and BSSS (Iowa Stiff Stalk Synthetic maize population germplasm), most of which were developed in China (Fig. 3). The Q value of five groups was listed in Table S1 of all the inbred lines. Of all inbred lines, 73.40% were assigned into the corresponding subgroups, and the remaining ones were categorized into the "mixed" subgroups based on their Q values (Table S1). Relative kinship within this inbred population were evaluated based on the analyses of the 204 SSR markers, and all the pairwise values ranged from 0 to 1, with a mean of 0.01104 (Fig. 4). Approximately 60% of the pairwise estimates were close to 0, indicating that there was no close genetic relationship within these lines (Fig. 4).

Linkage disequilibrium for pairwise markers

In the entire collection under investigation 1, 580 (7.94%) of the 19, 900 intrachromosomal marker pairs showed a significant level of LD (*P<0.01*). The mean of r^2 for all pairs was 0.015 and the mean of LD value (D´) for all statistically significant loci pairs was 0.286. As expected, the LD of these breeding germplasm was considerably higher than that had been previously reported (Yan et al., 2009). In the 94 inbred lines, 63.05% linked pairwise SSR loci were in significant LD at the 0.01 level. Overall, linkage was the main factor resulting in the pair-wise SSR loci with significant LD in the entire sample.

Association analysis between traits and SSRs

Effect of population groupings on association analysis

The prerequisite for association analysis was the characterization of population structure within our new set of inbred lines using the software package STRUCTURE 2.3.1 (Pritchard et al., 2000). The association analyses based on the Q matrix of our seven groups (one to seven) of the new inbred lines were conducted using the means of our two years of observations. The distribution of the P values of the 26 traits from these seven groups was shown in Fig. 5. Five groups corresponding to the optimal group subdivided results gave the highest P values, with their 95% confidence interval for estimated mean of 0.4604, ranging from 0.4482 to 0.4715 (Fig. 5). Whereas, one group had the lowest P values, with the estimated mean value of 0.4433, ranging from 0.4321 to 0.4545. Thus, the incorporation of Q corresponding to the optimal subdividing results influenced the marker-trait associated results.

Association analysis of the results individual year

For each trait make association analysis for two years, respectively, with mixed linear model (MLM).

Table 1. Same marker-trait results of association analysis for each individual year.

Trait	Bin	Locus	F_Marker 2010	F_Marker 2011	P_Marker 2010	P_Marker 2011
PH	2	umc1419	2.46	2.85	1.07E-02	3.50E-03
PH	4.03	umc1550	2.80	2.10	8.70E-03	4.57E-02
PH	4.04	umc1652	2.74	3.40	4.81E-02	2.16E-02
EH	1.05	umc1395	2.92	2.17	5.00E-03	3.38E-02
LA	2	umc1419 [a]	2.50	3.02	9.70E-03	2.20E-03
LL	1.06	umc1035	3.11	2.49	5.90E-03	2.31E-02
LL	2.09	bnlg1520	2.58	2.50	1.87E-02	2.27E-02
LL	6.06	umc1859	4.04	2.39	1.40E-03	3.58E-02
LL	7.03	umc1593	2.36	3.26	4.74E-02	9.90E-03
LW	10.07	bnlg1185	2.11	2.64	1.98E-02	3.70E-03
BYC	4.08	bnlg2162	3.36	4.78	2.24E-02	4.00E-03
BYC	5.07	bnlg1118	2.46	2.62	3.12E-02	3.05E-02
BYC	6.01	phi077	3.20	3.24	4.55E-02	4.43E-02
BYC	8.06	umc1161 [a]	3.78	7.79	7.10E-03	2.30E-05
GLN	2.08	bnlg198	2.67	3.42	3.03E-02	8.70E-03
GLN	5.07	bnlg1118	2.71	2.97	1.88E-02	1.65E-02
TD	1.02	umc1976	3.27	3.33	1.53E-02	1.42E-02
AD	1.04	umc1917	9.39	4.26	2.90E-03	4.21E-02
AD	1.1	phi308707	2.39	3.15	2.87E-02	5.70E-03
AD	2.07	bnlg1045	2.17	2.42	3.90E-02	2.70E-02
AD	5.09	bnlg389 [a]	3.56	3.64	3.50E-03	3.10E-03
ASI	1.06	umc1122 [a]	4.21	4.99	8.00E-03	3.20E-03
KRN	1.03	phi001 [a]	2.90	2.96	5.00E-03	4.60E-03
KRN	1.08	umc1446	3.14	4.11	4.80E-02	1.99E-02
KRN	5.01	phi024	4.89	3.91	3.40E-03	1.16E-02
KRN	7	umc1695	3.37	4.69	3.88E-02	1.17E-02
KRN	10.05	umc1506	3.93	3.03	3.00E-03	1.50E-02
CD	1.03	umc1397	12.60	4.21	6.26E-04	4.34E-02
CD	1.05	umc2025	2.69	5.27	3.66E-02	8.31E-04
EW	1.01	bnlg1014	2.63	2.18	1.70E-02	4.58E-02
CW	5.07	bnlg1118	2.90	3.34	1.30E-02	8.80E-03
CW	8.05	umc1562	3.24	2.35	6.60E-03	4.84E-02
CW	9.02	bnlg1401	2.17	2.63	3.39E-02	1.43E-02
GW	1.01	bnlg1014	2.40	2.97	2.80E-02	8.50E-03
GW	6.02	umc1006	2.38	4.56	6.70E-03	1.34E-02
HKW	1.04	umc1917	6.65	5.20	1.16E-02	2.52E-02
HKW	3.05	umc1307	2.75	2.51	1.74E-02	2.88E-02
HKW	7.03	bnlg1579	4.09	2.89	4.50E-03	2.75E-02
EL/GL	4.02	umc1288 [a]	7.44	4.97	1.00E-03	9.20E-03

a: $P<0.01$ level.

Table 2. Trait-associated loci and QTL reported previously.

	Association in this study		Putative QTL in literature previously
Trait	Locus	Bin	
ASI	umc1122	1.06-1.07	Rainer Messmer et al. (2009)
HKW	umc1917	1.04	Rainer Messmer et al. (2009)
KRN	phi001	1.03	Wang (2009)
EW	bnlg2132	7	Wang (2009)
CW	bnlg1118	5.07	Ming Lu. et al. (2010)
ED	bnlg2190	10.06	Meng Li. et al. (2009)

The association analysis conducted in 2010 resulted in 138 SSR loci, all of which were associated with the 26 traits (*P<0.01*), while that in 2011 gave rise to 96 SSR loci that were strongly associated with the 26 traits (*P<0.01*). However, compared with the same association locus that were used for association analyses in both 2010 and 2011, only six loci showed similar association levels with several traits (e.g., umc1419 with LA, umc1161 with BYC, bnlg389 with AD, umc1122 with ASI, phi001 with KRN and umc1288 with EL/GL). When considering *P<0.05* level, 417 SSR loci and 347 SSR loci were associated with all 26 traits in our 2010 and 2011 study, respectively. This result indicated that the same association locus reached 39 (Fig. 6 and for the *P* marker, see Table 1).

Traits with strong association with markers

MLM model was used to make association analysis each individual year, a total of 39 associated loci associated with 17 traits were identified in two years at p<0.05 level, simultaneously, which were distributed on all of the ten chromosomes. Some important traits with strong association with markers have been detected. Umc1419, umc1550 and umc1652 were associated with PH. Umc1917, phi308707, bnlg1045 and bnlg389 were associated with AD. Five loci were associated with KRN, phi001, umc1446, phi024, etc. Umc1917, umc1307 and bnlg1579 were associated with HKW. The number of associated loci detected on chromosome 1 was thirteen, which was more than chromosome 2 and 5(5), and more than chromosome 4(4), etc. Of these loci, umc1917 located in bin 1.04 was associated with AD and HKW. Bnlg1014 located in bin 1.01 was associated with EW and GW. Umc1397 located in bin 1.03 and umc2025 located in bin 1.05 were associated with CD. Bnlg1118 located in bin 5.07 was associated with BYC, GLN and CW. Phi077 and umc1161 were associated with BYC. Umc1695 and umc1506 were associated with KRN. Bnlg1579 was associated with HKW. Umc1562 located in bin 8.05 and bnlg1401 located in bin 9.02 were associated with CW.

Association analysis based on the meaning of the two years' phenotypic data

The association analysis with the mixed linear model (MLM) indicated that 106 SSR loci were strongly associated with the 26 agronomical traits and such association significance reached the *P<0.01* level (Table S3). The P_maker was listed in Table S3. Data listed in Table S3 also suggested that one maker could be associated with more than one trait. For example, each of these three loci, namely umc2215, umc1917 and bnlg2190, was associated with (four traits/locus) EW, GW, TD, AD; HKW, TD, AD, SD; and CD, TD, AD, SD, respectively. Meanwhile, loci of bnlg2162, umc1457, umc1562, and bnlg1401 were associated with (three traits/locus) EL, CW, BYC; LW, TD, AD; EL, EW, GW and TD, AD, SD, respectively. 318 associated SSR loci were detected at *P<0.05* level based on the meaning of the two years of data.

Discussion

Genetic diversity and population structure of the inbred lines

A suitable association mapping panel should encompass as much phenotypic and molecular diversity as can be reliably
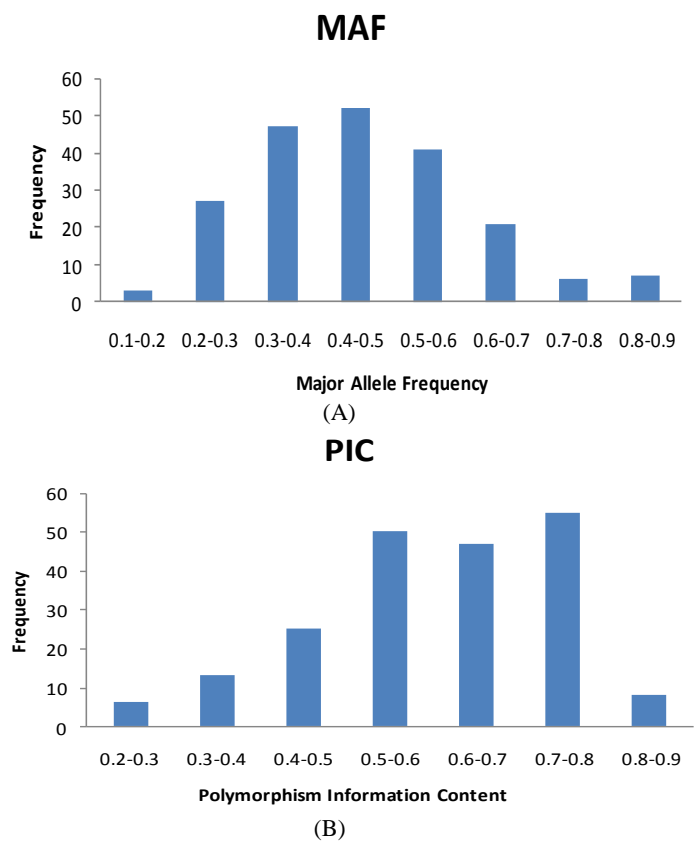
Fig 1. Frequency distribution of MAF (A) and PIC (B) for all of 204 SSR markers of all the inbred lines.
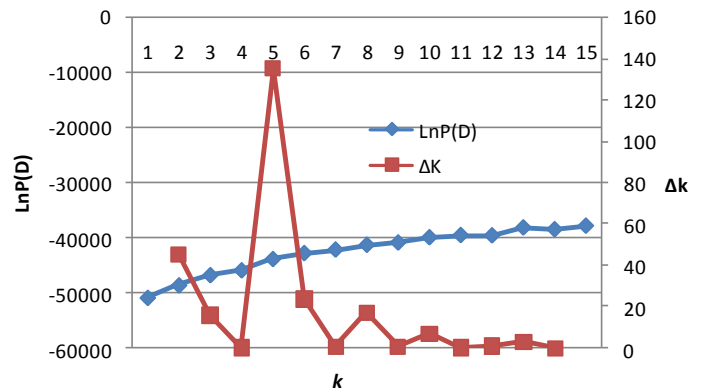
Fig 2. Model-based cluster membership for 94 lines in five groups.
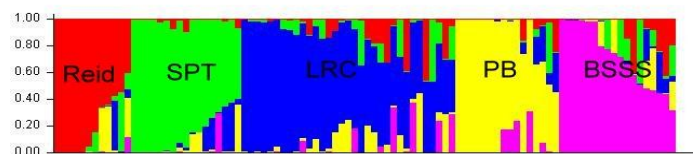
Fig 3. Population structure of 94 individuals based on 204 SSRs. Red: Reid (Ye478); Green: SPT (Huangzaosi); Blue: LRC (Dan340); Yellow: PB (P138); Pink: BSSS (B73).

measured in a common environment (Flint-Garcia et al., 2005). The genetic diversity of our association panel represented by 94 Chinese inbred lines is higher than, or equal to, those that had been reported so far, with the exception of those described by Taramino and Tingey (1996), and Liu et al. (2003). The present study used 204 pairs of SSR markers and subdivided the association panel into five

groups (Reid, SPT, LRC, PB and BSSS) using the model-based cluster method. Data from our genetic diversity and population structure analysis revealed that this association panel showed a diverse genetic variation and, therefore, could be used for the association analysis.

Linkage disequilibrium analysis and association analysis

When using an association panel to uncover a variant for quantitative traits of interest, a primary consideration should be the power of this panel, namely, the probability of detecting the causal variant. Studies of power evaluations have suggested that population size is one of the most fundamental decisions when identifying associations (Spencer et al., 2009). Marker density is another determinant for increasing the power of association analysis (Mackay et al., 2009), especially for GWAS. It is often related to the LD pattern of a maize association panel at the genome-wide level. Among factors influencing LD, linkage was the major cause for LD of SSR loci. In our study, we used 204 pairs of SSR markers among the 10 chromosomes, the LD level of the markers were high and could be used for association analysis. Many models were used to minimize the false-positive of association analysis. It seems that a K matrix incorporated into the K model was sufficient to minimize false-positive associations, consistent with other model simulations and comparisons (Yu and Buckler 2006; Stich et al., 2008; Zhu and Yu 2009; Yang et al., 2010b). Similarly, the Q + K model can also reduce the false positives, and in fact, such a combined model has proven to be better than either the K matrix or the Q matrix alone (e.g., traits of flowering time, ear height and ear diameter) (Yang et al., 2010a). Using GLM and MLM analyses, the mean values of our two consecutive years' observations revealed that 145 and 106 associated loci were significant at $P<0.01$ level, respectively, whereas 407 and 318 associated loci were significant at P<0.05 level, respectively. Our study suggests that the K + Q model can eliminate some false-positive associations.

The same association loci based on two-individual years' results

The results of our association analyses conducted in 2010 and 2011 suggested that 39 same association loci were found to be significant at P<0.05 level. False positives of 90.9% and 89.0% of the same association loci resulted from 2010 and 2011, respectively, were removed. In order to improve the power of the results of association analysis, we compared the 39 same association loci with the mean of two years of association loci, 17 same association loci were selected at $P<0.01$ level. Finally, 25 same associated loci were found at P<0.05 level. We thought that they are the more valuable associated loci, on the basis of our experiment. A number of agronomical traits related to yield have been mapped or found to be associated, but limited information is available on traits such as GLN and BYC. Of the 25 associated loci, we found that four loci, bnlg2162, bnlg1118, phi077 and umc1161 were strongly associated with BYC, while one locus, bnlg1118 was associated with GLN.

Our association results compared with others QTL positioning results

Some of the SSR markers used in the present investigation had been described previously by others in their linkage mapping or association mapping studies (Table 2). For example, it had been reported that umc1122 (Messmer et al.,
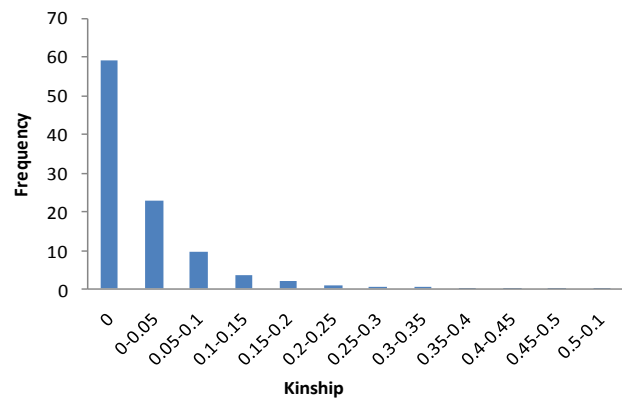
Fig 4. Distribution of pairwise relative kinship estimates between 94 maize inbred lines.
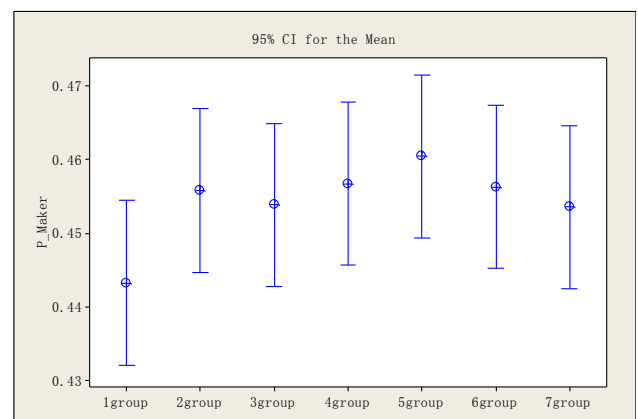
Fig 5. Interval plots of P values for the number of groups from one to seven subdivided by structure software.

2009), umc1917 (Messmer et al., 2009), bnlg1118 (Lu et al., 2010) and bnlg2190 (Li, et al., 2010) were linked to ASI, HKW, CW, and ED, respectively. Similarly, Wang (2009) constructed an integrated map of QTL for Grain yield and its related traits in maize. Our study indicated that the phi001 locus was associated with KRN. A similar result was reported by Wang (2009) who suggested that the QTL, bnlg2132, was strongly related to EW-another important yield trait. Furthermore, other loci associated with the yield traits were also found in the present study: e.g., bnlg2180 was associated with the KRO (in the consensus map, this locus showed association with a kernel number/row). In our study, there are many associated loci with the mean results of two years at $P<0.01$ level in the tenth chromosome, but the previous mapping results were little, which may have made a supplement to the association mapping progress.

Materials and methods

Plant materials

Table S1 listed 94 maize inbred lines and described their detailed pedigree information. Of these inbred lines, 65 were developed by domestic or international researchers and have been utilized extensively in the current maize hybrid breeding programs in China. Whereas, the remaining 29 inbred lines were derived from our breeding team. All field experiments were performed in the summers of 2010 and

1

82.8	bnlg1014 EW GW*
85.2	umc1071
89.9	umc2215
133.6	umc1166
160.6	umc1976 TD
165.8	umc2226
226.4	umc1397 CD*
257.4	umc1479
259.1	bnlg439
272.0	phi001 KRN*
273.6	bnlg2180
290.4	phi109275
374.8	umc1917 AD*HKW
417.0	umc2025 CD*
457.0	umc1124
471.7	umc1395 EH
517.0	umc1590
587.0	umc1035 LL*
599.9	umc1122 ASI
714.4	umc1147
718.5	bnlg1564
781.6	umc1446 KRN
806.5	umc2240
839.3	phi011
927.4	phi308707 AD
936.1	umc1774
1007.8	phi120
1014.9	umc1681
1103.0	phi064

2

0.0	umc1419 PH LA
3.1	umc2246
28.1	phi96100
56.7	phi098
65.7	bnlg1017
92.8	umc1823
141.6	umc1756
197.8	umc1845
225.3	umc1555
235.4	bnlg1621
250.1	umc1024
339.3	umc1454
373.5	bnlg1036
379.2	nc003
438.3	mmc0271
456.1	bnlg198 GLN
457.2	bnlg1045 AD*
599.1	bnlg1520 LL

3

2.0	umc2105 umc2101
28.2	umc1970
77.0	bnlg1144
127.3	bnlg1904
129.4	bnlg1447
159.0	phi036
190.8	bnlg1452
192.4	umc1025
270.4	umc1750
280.4	umc1773
296.1	phi053
312.8	umc1102
313.4	bnlg1035
339.6	umc1307 HKW
439.7	bnlg1350
511.5	bnlg197
567.6	umc1489
759.9	phi047
760.9	bnlg1496

4

2.9	phi072
23.4	umc2148
107.4	umc1288 EL/GL
114.0	umc1294
158.8	umc2280
185.2	umc1550 PH*
228.4	umc1652 PH
252.5	umc1662
254.0	nc005 phi079
284.5	umc1896
371.4	umc1299
452.9	umc1808
462.5	bnlg1444
475.7	bnlg2162 BYC*
535.5	umc1856
554.1	umc2188
731.6	phi076
737.8	umc1058

5

72.1	phi024 KRN*
79.2	umc1478
263.1	umc1274
267.5	umc2296
285.7	phi109188
305.2	umc2066
312.6	bnlg1287
328.5	bnlg2323
332.1	bnlg1074
368.4	umc1221
383.8	umc1482
389.9	mmc0081
396.6	umc1822
402.2	umc2164
470.1	umc1019
476.6	mmc0481
488.4	umc1752
493.5	umc1680
536.6	phi048
590.4	bnlg1118 BYC*GLN*CW*
641.4	umc1225
674.5	bnlg389 AD*

6

9.1	phi075
23.2	bnlg161
72.7	bnlg1371
98.6	umc1133
107.3	bnlg1538
121.1	phi077 BYC*
125.0	umc1006 GW
127.1	umc1186
145.3	umc1178
203.2	umc1857
211.5	nc010 umc1014
339.6	umc1805
391.4	umc1859 LL*
502.9	umc2165
534.6	umc1653
564.2	phi123

7

0.0	umc2177
45.0	umc1695 KRN*
53.3	bnlg2132
122.4	umc1066
155.8	umc1016
180.5	bnlg1094
244.3	umc1983
286.3	bnlg1808
300.0	umc1015
302.9	bnlg1305
322.7	bnlg1579 HKW*
405.5	umc1936
410.5	umc1710
451.2	umc1593 LL*
466.8	umc1944
471.4	umc1708
499.9	Dupssr13
611.5	phi116

8

10.1	umc1359
48.0	umc1414
245.7	bnlg1863
245.9	umc2366
284.6	umc1457
295.3	phi014
353.3	umc1562 CW*
370.0	bnlg1812
413.2	umc1960
422.8	umc1161 BYC*
494.7	umc1268 bnlg1834
547.2	umc2218
571.5	phi015 phi080
608.1	umc1663

9

11.8	bnlg1724
17.7	umc2084
24.3	umc1867
80.3	phi028
109.0	umc1647
142.6	bnlg244
147.5	bnlg1401 CW*
191.7	phi022
225.4	phi065
226.3	bnlg127
256.8	mmc0051
257.6	umc1267
285.8	umc1107
321.6	umc1519
327.8	bnlg1270
333.1	umc1094
340.4	umc1357
342.0	umc1231
344.8	umc1657
489.9	umc1366

10

30.9	phi041
91.4	umc1152
183.8	umc1367
187.0	umc1381
195.4	umc2016
217.8	bnlg1712
260.5	phi062
269.6	umc2163
272.2	umc1053
284.7	bnlg1526
295.9	umc1115
335.5	bnlg1250
344.8	umc1506 KRN*
352.9	umc2043
392.5	umc2122
410.6	umc1993
412.3	bnlg2190
444.8	umc1196
449.3	bnlg1677
466.4	bnlg1839
483.7	bnlg1450
526.4	bnlg1518
550.9	bnlg1185 LW*
578.1	umc1645

Fig 6. Thirty-nine same marker-trait loci at P <0.05 level based on two individual years. Genetic map showing the marker positions and estimated map distances based on the IBM2 2008 Neighbors Frame 1. Asterisks indicated 25 same marker-trait loci (compared the 39 same association results with the mean of two years of results at P <0.05 level).

2011 at our Experimental Farm, Shandong Agricultural University, Taian, Shandong Province, P. R. China.

Phenotypic data

Field experiments were performed at Taian, Shandong Province, China, in 2010 and 2011 (summer sowing). On an individual basis, 26 agronomical traits were measured or further calculated. The measured traits were plant height (PH, cm), ear height (EH, cm), leaf angle above ear (LA,°), leaf length at the ear (LL, cm), leaf width at the ear (LW, cm), bract leaf length (BYC, cm), leaf number (LN), the number of green leaves at mature (GLN), days to tasseling (TD, day), days to anthesis (AD, day), days to silking (SD, day), days to mature (MD, day), ear length (EL, cm), ear diameter (ED, cm), bald length (BL, cm), ear weight (EW, g), cob weight (CW, g), cob diameter (CD, cm), the number of kernel rows (KRN), the number of kernels per row (RKN), grain weight per ear (GW, g), embryo length (EL, cm), grain length (GL, cm), and hundred kernels weight (HKW, g). The calculated

variables were tasseling to-anthesis internal days (TAI), anthesis to-silking internal days (ASI), kernel ratio (KRO) and embryo length ratio of grain length (EL/GL).

Genotypic data

DNA extraction and SSR genotyping

For DNA extraction, bulk of leaf samples representing at least six individual plants were used, and the extraction protocol was modified (CTAB) according to Murray and Thompson (1980). PCR amplification reactions were performed using 75 ng sample DNA in 25 μL of DNA amplification system, containing 0.5 μM of each primer pair, 100 μM of dNTPs, 2.5 μL 1×Taq polymerase buffer, 1.5 mM MgCl$_2$, and 0.75 U Taq DNA polymerase. The SSR reactions were carried out using Touchdown PCR program. PCR products were separated on a 9% non-denaturing PAGE gel in 1×TBE buffer and stained using the silver method (Creste et al., 2001).

Two hundred and four pairs of SSR primers with high polymorphism rate and even distribution throughout the maize genome were used to genotype all of the 94 lines. All the SSR primers and the bin values were searched in MaizeGDB (http://www. Maizegdb.org). Most of the SSR repeat motifs and sequences were obtained from MaizeGDB (Table S4).

Genetic structure analysis

Microsatellite profiles were scored reflecting either the presence (1) or absence (0) of clear bands. Powermarker version 3.25 (Liu and Muse, 2005) was used to calculate allele number, gene diversity, polymorphism information content (PIC) and major allele frequency (MAF). The model-based program STRUCTURE 2.3.1 (Pritchard et al., 2000; Falush et al., 2003) was used to infer population structure using 204 SSRs. Five independent runs were performed, setting the number of populations (k) from 1 to 15, burn in time and MCMC (Markov Chain Monte Carlo) replication number both to 100, 000, and a model for admixture and correlated allele frequencies. The k value was determined by LnP(D) in STRUCTURE output and an ad hoc statistic Δk based on the rate of change in LnP(D) between successive k (Evanno et al., 2005). Both LnP(D) in STRUCTURE output and its derived Δk were used to determine the k value. Since the distribution of L(k) did not show a clear mode of the true k,

$$\Delta k = m(|L(k+1) - 2L(k) + L(k-1)| / s[L(k)]$$

was used to show a clear peak to represent the true value of k (Evanno et al., 2005). Lines with membership probabilities of ≥ 0.70 and <0.70 were assigned to match clusters and to represent a mixed group, respectively. Relative kinship between individuals was inferred based on the molecular markers, which represent the approximate identity between any two given individuals (Yu and Buckler 2006). We used SPAGeDi (Hardy and Vekemans 2002) to estimate the kinship coefficients based on 204 SSR markers. The Q matrix corresponding to the highest Δk and the kinship matrix (K) were adopted for association analysis.

Linkage disequilibrium estimation and association analysis

Linkage disequilibrium analysis was further performed for the 204 polymorphic SSRs, with the dedicated procedure of the TASSEL software, using 1, 000 permutations (http://www. Maizegenetics.net/). Pairs of loci were considered to be in significant LD if P was < 0.01. The significance of pairwise LD (P values) among all possible intrachromosomal and genome-wide comparisons for the 204 loci was also evaluated with the rapid permutations test. The loci were considered to be in significant LD if $P<0.01$. The estimated genetic distance (cM) among the loci was inferred from the public IBM2 2008 Neighbors Frame 1. The Q model were performed using GLM in TASSEL V2.1; the K + Q model were performed using MLM in TASSEL V2.1 (Yu, Pressior et al., 2006; Bradbury et al., 2007). Association analysis was conducted for 26 traits with 204 SSRs. The observed P values were used to estimate the SSR-trait associations between marker and trait.

Conclusion

We used the high genetic diversity association panel to make association analysis with the important agronomical traits which made a great contribution to yield improvement. Then, we took advantage of annual field trials to detect the same associated loci; the method could delete some false positive results rather than using the mean value of traits to make association analysis. With a mix of the kinship and Q matrix model, many loci were detected that coincide with known major genes or QTL, some marker-traits loci were not reported previously, indicating the power of this association panel. The role of these regions will need to be further investigated. Additionally, potential novel loci were identified that may help to better understand the architecture of complex genetic traits. Association mapping based on LD may have broad applications in maize genetics and selective breeding.

References

Abdellatif KF, Khidr YA (2010) Genetic diversity of new maize hybrids based on SSR markers as compared with other molecular and biochemical markers. J Crop Sci Biotech 13(3): 139-145.

Andersen JR, Zein I, Wenzel G, Krutzfeldt B, Eder J, Ouzunova M, Lubberstedt T (2007) High levels of linkage disequilibrium and associations with forage quality at a phenylalanine ammonia-lyase locus in European maize (*Zea mays* L.) inbreds. Theor Appl Genet 114: 307-319.

Belo A, Zheng PZ, Luck S, Shen B, Meyer DJ, Li BL, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Mol Genet Genomics 279: 1-10.

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

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, SFlint-Garcia (2009) The genetic architecture of maize flowering time. Science 325: 714-718.

Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. Genetics 172: 2449-2463.

Creste S, Tulmann A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide. Plant Mol Biol 19: 299-306.

Ducrocq S, Giauffret C, Madur D, Combes V, Dumas F, Jouanne S, Coubriche D, Jamin P, Moreau L, Charcosset A (2009) Fine mapping and haplotype structure analysis of a major flowering time quantitative trait locus on maize chromosome 10. Genetics 183: 1555-1563.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620.

Falush D, Stephens MS, Pritchard JK (2003) Inference of population structure using multilocus genotype data linked loci and correlated allele frequencies. Genetics 164:

1567-1587.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44: 1054-1064.

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PloS One 2(12): e1367.

Hardy OJ, Vekemans X (2002) Spagedi: a versatile computer program to analyse spatial. Mol Ecol Notes 2: 618-620.

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319: 330-333.

Li, M, Guo XH, Zhang M, Wang XP, Zhang GD, Tian YC, Wang ZL (2010) Mapping QTLs for grain yield and yield components under high and low phosphorus treatments in maize (*Zea mays* L.). Plant Science 178: 454-462.

Li Q, Li L, Yang XH, Warburton ML, Bai GH, Dai JR, Li JS, Yan JB (2010a) Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight. BMC Plant Biol 10: 143.

Li Q, Yang XH, Bai GH, Warburton ML, Mahuku G, Gore M, Dai JR, Li JS, Yan JB (2010b) Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development. Theor Appl Genet 120: 753-763.

Liu KJ, Goodman MM, Muse SV, Smith JS, Buckler ES, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165: 2117-2128.

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

Liu NJ, Chen L, Wang S, Oh C, Zhao HY (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genet 6(Suppl 1): S26.

Lu M, Xie CX, Li XH, Hao ZF, Li MS, Weng JF, Zhang DG, Bai L, Zhang SH (2010) Mapping of quantitative trait loci for kernel row number in maize across seven environments. Mol Breeding 28: 143-152.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10: 565-577.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C (2009) Genetic properties of the maize nested association mapping population. Science 325: 737-740.

Messmer R, Fracheboud Y, Banziger M, Vargas M, Stamp P, Ribaut JM (2009) Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theor Appl Genet 119: 913-930.

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

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breeding 2: 225-238.

Pressoir G, Brown PJ, Zhu W, Upadyayula N, Rocheford T, Buckler ES, Kresovich S (2009) Natural variation in maize architecture is mediated by allelic differences at the PINOID co-ortholog barren inflorescence2. Plant J 58: 618-628.

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

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73: 1402-1422.

Spencer CC, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 5: e1000477.

Stich B, Mohring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. Genetics 178: 1745-1754.

Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110: 1324-1333.

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39: 277-287.

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327: 818-822.

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

Wang BT, Wu JY, Ding JQ, Xi ZY (2009) Map integration of QTLs for grain yield and its related traits in maize. Acta Agron Sin 35: 1836-1843.

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene associations. Plant Cell 16: 2719-2733.

Yan JB, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PloS One 4: e8451.

Yan JB, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang XH, Skinner DJ, Fu ZY, Mitchell S, Li Q, Fernandez MGS, Zaharieva M, Babu R, Fu Y, Palacios N, Li JS, DellaPenna D, Brutnell TP, Buckler ES, Warburton ML, Rocheford T (2010) Rare genetic variation at *Zea mays crtRB1* increases *β*-carotene in maize grain. Nat Genet 42: 322-327.

Yan JB, Warburton M, Crouch J (2011) Association mapping for enhancing maize (L.) genetic improvement. Crop Sci 51: 433-449.

Yang XH, Gao SB, Xu ST, Zhang ZX, Prasanna BM, Li L, Li JS, Yan JB (2010a) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breeding 28: 511-526.

Yang XH, Yan JB, Trushar S, Warburton ML, Li Q, Li L, Gao YF, Chai YC, Fu ZY, Zhou Y, Xu ST, Bai GH, Meng YJ, Zheng YP, Li JS (2010b) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121: 417-431.

Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17: 155-160.

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203-208.

Yu JM, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539-551.

Zhu CS, Yu JM (2009) Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182: 875-888.