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, blast 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 many traits at once. 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.
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” $\hat{k}$ from the $\text{LnP}(D)$ (Fig. 2). Thus, $\Delta \hat{k}$ (Evanno et al., 2005) was used to determine the $\hat{k}$ value. Fig. 2 indicated that when $\Delta \hat{k}$ 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).

**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 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**

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.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PH</td>
<td>2</td>
<td>umc1419</td>
<td>2.46</td>
<td>2.85</td>
<td>1.07E-02</td>
<td>3.50E-03</td>
<td>AD</td>
<td>5.09</td>
<td>bnlg389 *</td>
<td>3.56</td>
<td>3.64</td>
<td>3.50E-03</td>
<td>3.10E-03</td>
</tr>
<tr>
<td>PH</td>
<td>4.03</td>
<td>umc1550</td>
<td>2.80</td>
<td>2.10</td>
<td>8.70E-03</td>
<td>4.57E-02</td>
<td>ASI</td>
<td>1.06</td>
<td>umc1122 *</td>
<td>4.21</td>
<td>4.99</td>
<td>8.00E-03</td>
<td>3.20E-03</td>
</tr>
<tr>
<td>PH</td>
<td>4.04</td>
<td>umc1652</td>
<td>2.74</td>
<td>3.40</td>
<td>4.81E-02</td>
<td>2.16E-02</td>
<td>KRN</td>
<td>1.03</td>
<td>phi001 *</td>
<td>2.90</td>
<td>2.96</td>
<td>5.00E-03</td>
<td>4.60E-03</td>
</tr>
<tr>
<td>EH</td>
<td>1.05</td>
<td>umc1395</td>
<td>2.92</td>
<td>2.17</td>
<td>5.00E-03</td>
<td>3.38E-02</td>
<td>KRN</td>
<td>1.08</td>
<td>umc1446</td>
<td>3.14</td>
<td>4.11</td>
<td>4.80E-02</td>
<td>1.99E-02</td>
</tr>
<tr>
<td>LA</td>
<td>2</td>
<td>umc1419 *</td>
<td>2.50</td>
<td>3.02</td>
<td>9.70E-03</td>
<td>2.20E-03</td>
<td>KRN</td>
<td>5.01</td>
<td>phi024</td>
<td>4.89</td>
<td>3.91</td>
<td>3.40E-03</td>
<td>1.16E-02</td>
</tr>
<tr>
<td>LL</td>
<td>1.06</td>
<td>umc1035</td>
<td>3.11</td>
<td>2.49</td>
<td>5.90E-03</td>
<td>2.31E-02</td>
<td>KRN</td>
<td>7</td>
<td>umc1695</td>
<td>3.37</td>
<td>4.69</td>
<td>3.88E-02</td>
<td>1.17E-02</td>
</tr>
<tr>
<td>LL</td>
<td>2.09</td>
<td>bnlg1520</td>
<td>2.58</td>
<td>2.50</td>
<td>1.87E-02</td>
<td>2.27E-02</td>
<td>KRN</td>
<td>10.05</td>
<td>umc1506</td>
<td>3.93</td>
<td>3.03</td>
<td>3.00E-03</td>
<td>1.50E-02</td>
</tr>
<tr>
<td>LL</td>
<td>6.06</td>
<td>umc1859</td>
<td>4.04</td>
<td>2.39</td>
<td>1.40E-03</td>
<td>3.58E-02</td>
<td>CD</td>
<td>1.03</td>
<td>umc1397</td>
<td>12.60</td>
<td>4.21</td>
<td>6.26E-04</td>
<td>4.34E-02</td>
</tr>
<tr>
<td>LL</td>
<td>7.03</td>
<td>umc1593</td>
<td>2.36</td>
<td>3.26</td>
<td>4.74E-02</td>
<td>9.90E-03</td>
<td>CD</td>
<td>1.05</td>
<td>umc2025</td>
<td>2.69</td>
<td>5.27</td>
<td>3.66E-02</td>
<td>8.31E-04</td>
</tr>
<tr>
<td>LW</td>
<td>10.07</td>
<td>umc1185</td>
<td>2.11</td>
<td>2.64</td>
<td>1.98E-02</td>
<td>3.70E-03</td>
<td>EW</td>
<td>1.01</td>
<td>bnlg1014</td>
<td>2.63</td>
<td>2.18</td>
<td>1.70E-02</td>
<td>4.58E-02</td>
</tr>
<tr>
<td>BYC</td>
<td>4.08</td>
<td>bnlg2162</td>
<td>3.36</td>
<td>4.78</td>
<td>2.24E-02</td>
<td>4.00E-03</td>
<td>CW</td>
<td>5.07</td>
<td>bnlg1118</td>
<td>2.90</td>
<td>3.34</td>
<td>1.30E-02</td>
<td>8.80E-03</td>
</tr>
<tr>
<td>BYC</td>
<td>5.07</td>
<td>umc1118</td>
<td>2.46</td>
<td>2.62</td>
<td>3.12E-02</td>
<td>3.05E-02</td>
<td>CW</td>
<td>8.05</td>
<td>umc1562</td>
<td>3.24</td>
<td>2.35</td>
<td>6.60E-03</td>
<td>4.84E-02</td>
</tr>
<tr>
<td>BYC</td>
<td>6.01</td>
<td>phi077</td>
<td>3.20</td>
<td>3.24</td>
<td>4.55E-02</td>
<td>4.43E-02</td>
<td>CW</td>
<td>9.02</td>
<td>bnlg1401</td>
<td>2.17</td>
<td>2.65</td>
<td>3.39E-02</td>
<td>1.43E-02</td>
</tr>
<tr>
<td>BYC</td>
<td>8.06</td>
<td>umc1161 *</td>
<td>3.78</td>
<td>7.79</td>
<td>7.10E-03</td>
<td>2.30E-05</td>
<td>GW</td>
<td>1.01</td>
<td>bnlg1014</td>
<td>2.40</td>
<td>2.97</td>
<td>2.80E-02</td>
<td>8.50E-03</td>
</tr>
<tr>
<td>GLN</td>
<td>2.08</td>
<td>bnlg198</td>
<td>2.67</td>
<td>3.42</td>
<td>3.03E-02</td>
<td>8.70E-03</td>
<td>GW</td>
<td>6.02</td>
<td>umc1006</td>
<td>2.38</td>
<td>4.56</td>
<td>6.70E-03</td>
<td>1.34E-02</td>
</tr>
<tr>
<td>GLN</td>
<td>5.07</td>
<td>bnlg1118</td>
<td>2.71</td>
<td>2.97</td>
<td>1.88E-02</td>
<td>1.65E-02</td>
<td>HKW</td>
<td>1.04</td>
<td>umc1917</td>
<td>6.65</td>
<td>5.20</td>
<td>1.16E-02</td>
<td>2.52E-02</td>
</tr>
<tr>
<td>TD</td>
<td>1.02</td>
<td>umc1976</td>
<td>3.27</td>
<td>3.33</td>
<td>1.53E-02</td>
<td>1.42E-02</td>
<td>HKW</td>
<td>3.05</td>
<td>umc1307</td>
<td>2.75</td>
<td>2.51</td>
<td>1.74E-02</td>
<td>2.88E-02</td>
</tr>
<tr>
<td>AD</td>
<td>1.04</td>
<td>umc1917</td>
<td>9.39</td>
<td>4.26</td>
<td>2.90E-03</td>
<td>4.21E-02</td>
<td>HKW</td>
<td>7.03</td>
<td>bnlg1579</td>
<td>4.09</td>
<td>2.89</td>
<td>4.50E-03</td>
<td>2.75E-02</td>
</tr>
<tr>
<td>AD</td>
<td>1.1</td>
<td>phi308707</td>
<td>2.39</td>
<td>3.15</td>
<td>2.87E-02</td>
<td>5.70E-03</td>
<td>EL/GL</td>
<td>4.02</td>
<td>umc1288 *</td>
<td>7.44</td>
<td>4.97</td>
<td>1.00E-03</td>
<td>9.20E-03</td>
</tr>
<tr>
<td>AD</td>
<td>2.07</td>
<td>bnlg1045</td>
<td>2.17</td>
<td>2.42</td>
<td>3.90E-02</td>
<td>2.70E-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a: P<0.01 level.

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

<table>
<thead>
<tr>
<th>Trait</th>
<th>Locus</th>
<th>Bin</th>
<th>Putative QTL in literature previously</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASI</td>
<td>umc1122</td>
<td>1.06-1.07</td>
<td>Rainer Messmer et al. (2009)</td>
</tr>
<tr>
<td>HKW</td>
<td>umc1917</td>
<td>1.04</td>
<td>Rainer Messmer et al. (2009)</td>
</tr>
<tr>
<td>KRN</td>
<td>phi001</td>
<td>1.03</td>
<td>Wang (2009)</td>
</tr>
<tr>
<td>EW</td>
<td>bnlg2132</td>
<td>7</td>
<td>Wang (2009)</td>
</tr>
<tr>
<td>CW</td>
<td>bnlg1118</td>
<td>5.07</td>
<td>Meng Li. et al. (2010)</td>
</tr>
<tr>
<td>ED</td>
<td>bnlg2190</td>
<td>10.06</td>
<td>Meng Li. et al. (2009)</td>
</tr>
</tbody>
</table>
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 marker 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 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., 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
phenotypic data

Field experiments were performed at Taian, Shandong Province, in 2010 and 2011 (summer sowing). On an individual basis, 25 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), beld 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 (KRN), 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 MgCl₂, 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, Δk = ml(L(k + 1) - 2L(k) + L(k - 1))/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 MLM in TASSEL V2.1; the K + Q model were performed using MLM in TASSEL V2.1 (Yu, Pressoir 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 coincides with known major genes or QTL, some marker-trait 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.

Acknowledgments

This research was supported by the Natural Science Foundation of Shandong Province, China (Y2007D52), the Improved Variety Project of Shandong Province (2008 No.6) and the Foundation of innovation team for modern agricultural industry of Shandong province. We thank JiPing Zhao for giving constructive advice.

References

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


