Theor Appl Genet (2014) 127:621–631
DOI 10.1007/s00122-013-2246-y

ORIGINAL PAPER

# Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping

**Xun Wu · Yongxiang Li · Yunsu Shi · Yanchun Song · Tianyu Wang · Yubi Huang · Yu Li**

Received: 15 January 2013 / Accepted: 19 November 2013 / Published online: 18 December 2013


**Abstract** To investigate the genetic structure of Chinese maize germplasm, the MaizeSNP50 BeadChip with 56,110 single nucleotide polymorphisms (SNPs) was used to genotype a collection of 367 inbred lines widely used in maize breeding of China. A total of 41,819 informative SNPs with minor allele number of more than 0.05 were used to estimate the genetic diversity, relatedness, and linkage disequilibrium (LD) decay. Totally 1,015 SNPs evenly distributed in the genome were selected randomly to evaluate the population structure of these accessions. Results showed that two main groups could be determined i.e., the introduced germplasm and the local germplasm. Further, five subgroups corresponding to different heterotic groups, that is, Reid Yellow Dent (Reid), Lancaster Sure Crop (Lancaster), P group (P), Tang Sipingtou (TSPT), and Tem-tropic I group (Tem-tropic I), were determined. The genetic diversity of within subgroups was highest in the Tem-Tropic I and lowest in the P. Most lines in this panel showed limited relatedness with each other. Comparisons of gene diversity showed that there existed some conservative genetic regions in specific subgroups across the ten chromosomes, i.e., seven in the Lancaster, seven in the Reid, six in the TSPT, five in the P, and two in the Tem-Tropical I. In addition, the results also revealed that there existed fifteen conservative regions transmitted from Huangzaosi, an important foundation parent, to its descendants. These are important for further studies since the outcomes may provide clues to understand why Huangzaosi could become a foundation parent in Chinese maize breeding. For the panel of 367 elite lines, average LD distance was 391 kb and varied among different chromosomes as well as in different genomic regions of one chromosome. This analysis uncovered a high natural genetic diversity in the elite maize inbred set, suggesting that the panel can be used in association study, esp. for temperate regions.

Communicated by A. Charcosset.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-013-2246-y) contains supplementary material, which is available to authorized users.

X. Wu · Y. Huang (✉)
Maize Institute, Sichuan Agricultural Unversity, Ya'an 625014, Sichuan, China
e-mail: yubihuang@sina.com

X. Wu
e-mail: wuxunyong@126.com

X. Wu · Y. Li · Y. Shi · Y. Song · T. Wang · Y. Li (✉)
Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: liyu03@caas.cn

X. Wu
Nanchong Academy of Agricultural Sciences, Nanchong 637000, Sichuan, China

## Introduction

Maize (*Zea mays* L.) is an important cross-pollinated crop, with an extremely high level of natural genetic variation at both phenotype and genotype levels (Yan et al. 2010). Tracing back to the history of hybrid maize breeding, it has undergone different stages: (1) From the 1930s to the 1950s, parents of maize hybrids were mainly inbred lines derived directly from landraces (Lu and Bernardo 2001) and had a relatively wide genetic basis; (2) From the 1950s to the 1980s, parents of maize hybrids were mainly derived from crosses among inbred lines (Troyer 1990). Compared with the former stage, the genomic diversity of maize hybrids developed at this stage became narrower

under artificial selection; and (3) Since the 1980s, due to the privatization of seed industry and the dependence on high-yielding target of maize breeding, elite commercial hybrids have been widely developed and their parents have become the main source of new inbred lines (Troyer 1999). Consequently, a lot of elite maize inbred lines were derived from only a few progenitors (Heerwaarden et al. 2012). An investigation showed that 78 % of 381 hybrids were derived from only eight widely used inbred lines (Troyer 1999), leading to an expected narrow genomic diversity. Linkage disequilibrium (LD) is the base of association mapping which can help understanding the genetic basis of complex traits (Yan et al. 2010). Therefore, fine genetic characterization of maize germplasm for genetic diversity and population structure is important for germplasm enhancement and commercial breeding to broaden the genetic basis.

Chinese maize inbred lines are mainly derived from local landraces and introduced germplasm (Li and Wang 2010). The inbred lines from local landraces showed higher adaptation, and Huangzaosi has been widely used in most maize ecological regions of China. Since the 1970s to the 1980s, a number of maize materials were introduced from other countries, e.g., Mo17 from the U.S. The popular pattern of heterosis utilization in China for single-cross hybrids is local germplasm × introduced germplasm. About 65.7 % of important hybrids were derived from this pattern (Zeng 1990; Wang et al. 1997). During the history of single-cross breeding, some important foundation parents in Chinese maize breeding were formed, e.g., Huangzaosi, Dan340, Zi330, Mo17, Ye478 etc., Especially, Huangzaosi showed the highest utilization frequency in Chinese maize breeding. Using Huangzaosi or its descendants as parental lines, more than 70 descended inbred lines and 80 important hybrids were released (Li and Wang 2010), with the total planting area of these hybrids exceeding more than 10 million ha 15 years ago (Li 1997). The very important line, "Chang7-2", one parent of the maize hybrid "Zhengdan 958" (Zheng58 × Chang7-2) which has been grown in an extensive area about 35 million ha in China (Weng et al. 2011) is also one of Huangzaosi's descendant [(Huangzaosi × Wei95) × S901]. Another representative line "92-8", one parent of the maize hybrid Xundan20 (Xun9058 × 92-8) which has been grown in an extensive area about 0.13 billion ha is Chang7-2's descendant (Chang7-2 × 5237) (5237 is an inbred derived from the combination of Huangzaosi × Dan340). Previous research had documented the changes of genetic components among maize lines including Ye478, Zheng58, 5003 and 8112 (Lai et al. 2010). But the knowledge about the genetic differences among Huangzaosi's descendants is still unknown. Thus, it is important to understand the genetic changes in Huangzaosi's descendants during artificial selection, which will be informative not only for uncovering the genetic basis of maize foundation parents' formation, but also for genetic improvement of maize germplasm in the future.

Over the world, more than 47,000 accessions of maize germplasm are stored in the genebanks (Yan et al. 2009), among which about 21,000 accessions of maize germplasm are collected and stored in the Chinese National Genebanks, including landraces, inbred lines and improved populations. Although these maize germplasm have been well phenotypically characterized and documented in the past decades, most of these germplasm have not yet been characterized at the genomic level. Previous marker-based studies have addressed genetic diversity, population structure and genetic relationships using different panels and limited number of markers. Five main heterotic groups were reported, that is, the Reid with representative of Ye478, the Lancaster with representative of Mo17, the TSPT with representative of Huangzaosi, the P with representative of Shen137, and the Tem-Tropic I with representative of Lu28 (Lu et al. 2009; Wang et al. 2008; Yan et al. 2009; Yang et al. 2010b). Meanwhile, to facilitate analysis of larger panels, some subsets with minimum samples but representing maximum diversity of original maize collections were established. For example, a core collection of maize inbred lines including 242 inbreds from a larger collection of 3,258 accessions was constructed (Li et al. 2005). This set was used in the investigation of genetic diversity and population structure with SSR markers (Wang et al. 2008; Yu et al. 2007).

LD is the basis of association mapping, which can help understanding the genetic basis of complex traits (Yan et al. 2010). Recently, single nucleotide polymorphisms (SNPs) markers have become an important genotyping marker system in maize because of its high throughput and low cost (Yan et al. 2010). For instance, using 1,536 SNPs, Yan et al. (2009) found that LD decay distance to reach a $R^2$ of 0.1 ranged from 1 to 10 kb among ten chromosomes of maize in a diverse panel including 632 inbred lines from temperate, tropical, and subtropical regions. Lu et al. (2011) investigated a panel with 447 diverse inbred lines and pointed out that tropical germplasm contained higher genetic diversity than that of temperate germplasm, and average LD decay distance in the tropical germplasm (5–10 kb) was significantly smaller than that in the temperate germplasm (10–100 kb). However, the GoldenGate Assay containing only 1,536 SNPs makes the valuable information limited although the genetic diversity could be basically understood (Yang et al. 2010b). More recently, the MaizeSNP50 BeadChip with 56,110 SNPs was developed to satisfy the need of high throughput genotyping in maize genetic research (Ganal et al. 2011). Using this BeadChip, Weng et al. (2011) reported that, in 284 diverse maize inbred lines, the LD decay distance ranged from 25.4 to 29.2 kb among ten chromosomes, with an average LD distance of 27.7 kb.

Riedelsheimer et al. (2012) reported that in 285 dent inbred lines, average LD was estimated to be larger (500 kb). In a set of North American maize lines, the genomic history of maize was marked by a steady increase in genetic differentiation and LD (Heerwaarden et al. 2012). These studies showed that the BeadChip with 56,110 SNPs is able to analyze genetic diversity, population structure of maize germplasm and LD feature in the genome, which will greatly speed up the identification and utilization of new alleles in crop improvement.

In the present study, 367 elite maize inbred lines, mostly from the temperate region in the world, were selected and genotyped using the BeadChip, among which 43 Huangzaosi-related lines were included. The objectives were to finely estimate the genetic diversity, population structure, relative kinship, and LD decay of the accessions, to evaluate whether this panel is suitable for association mapping and to elucidate the genetic differentiation among Huangzaosi's descendants during artificial selection in maize commercial breeding.

## Materials and methods

### Plant material

The maize panel used in this study contained two sets of inbred lines, i.e., one from the core established previously by Li et al. (2005), including 242 diverse accessions which were historically used in maize breeding, and the other one collected recently from research institutions or companies including 125 elite inbred lines, most of which are widely used in current maize breeding. The details of the lines are listed in Table S1.

### SNP genotyping and quality evaluation

The MaizeSNP50 BeadChip with 56,110 SNPs was used to genotype this panel of maize germplasm. Those SNPs were evenly distributed across the maize genome based on B73 reference sequence (www.illumina.com/maizeSNP50). When maize seedlings were 1 month old, leaves of five plants were sampled as a bulk to extract genomic DNA according to the modified CTAB procedure (Saghai-Maroof et al. 1984). DNA quality checking and genotyping was accomplished in the Emei Tongde Company (Beijing) according to the Infinium® HD assay ultra-protocol guide (Illumina). A total of 55,126 of 56,110 SNPs were called successfully among the 367 lines. SNPs with missing rate of more than 20 % and minor allele number (MAF) of <0.05 were excluded from the genotyping dataset. The source sequences of remaining SNPs were indentified through BlastN search against the reference genome sequence of B73 (RefGen-V1 http://www.maizegdb.org/). SNPs with ambiguous physical position or multiple blast-hits were excluded from the genotyped dataset. After that, 41,819 high-quality SNPs with average marker density of 1 SNP/50 kb were used in further analysis. Most of the high-quality SNPs were distributed widely across the genome, and 53.73 % of markers showed a distance of <10 kb between the neighboring markers (Figure S1).

### Analysis of population structure

Totally 1,015 SNPs with high genetic diversity, low missing rate and even distribution across the genome were selected to estimate the population structure of the 367 lines using the model-based approach (Wang et al. 2008). Structure V2.3.3 software (Hubisz et al. 2009) was run with k (the number of populations) from 1 to 10, with five runs for each $k$ with a burn-in period of 10,000 and 10,000 replications. The *ad hoc* statistic delta $K$ ($\Delta K$) was used to determine the number of clusters (Evanno et al. 2005). Outputs of Structure were integrated using CLUMPP software (Jakobsson and Rosenberg 2007). Lines with membership probabilities of more than 0.5 were assigned to corresponding clusters (Liu et al. 2012).

### Analysis of relative kinship

Marker-based relative kinship between inbred lines $i$ and $j$ was estimated as: $F_{ij} = (Q_{ij} - Q_{\mathrm{m}})/(1 - Q_{\mathrm{m}})$, where $Q_{ij}$ is the probability of identity by state for random genes from i and j, and $Q_{\mathrm{m}}$ is the average probability of identity by state for genes coming from random individuals in the population from which $i$ and $j$ are issued. All the calculations were done using TASSEL 3.0.124 package with 41,819 SNPs with MAF >0.05 and missing data <20 %. Value of zero indicated that there was no more relationship than expected between two random individuals. Kinship coefficients from 0 to 0.1 indicated that there was weak similarity. Kinship coefficients from 0.1 to 0.5 indicated that there was intermediate similarity. Kinship coefficients from 0.5 to 1 indicated that there was strong similarity (Yang et al. 2010a).

### Analysis of genetic diversity

The total number of alleles, gene diversity (GD) and polymorphism information content (PIC) were analyzed using PowerMarker V3.25 (Liu and Muse 2005). GD was defined as the probability that two randomly chosen alleles from the population differ and estimated as: $\hat{D}_{\mathrm{l}} = \left(1 - \sum_{u=1}^{k} \hat{p}_{\mathrm{l}u}^2\right) \Big/ \left(1 + \frac{1+f}{n}\right)$, in which $P_{\mathrm{l}u}$ referred

to the frequency of the $u$th allele, $f$ the inbreeding coefficient, and $n$ the sample size. PIC was estimated as: $\hat{\mathrm{PIC}}_l = 1 - \sum_{u=1}^{k} \tilde{P}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2\tilde{P}_{lu}^2 \tilde{P}_{lv}^2$, in which $P_{lv}$ referred to the frequency of the $v$th allele. GD was compared among different genetic regions on one chromosome as well as different subgroups. Conservative regions displaying low diversity in a given subgroup were defined when GD was less than '$\widehat{\mathrm{GD}}$ (average GD of one chromosome in given subgroup) − 0.5 × SD (Standard Deviation for the chromosome in given subgroup)'. To compare the genetic diversity among different subgroups, four independent randomly selected sets of 19 lines were re-sampled based on simulated annealing method (Liu and Muse 2005), with 1,000 permutations. Computations were made using SAS 9.1 software (Galecki et al. 2004). To evaluate the genetic differentiation among Huangzaosi's descendants, 43 Huangzaosi-related lines were selected, among which 40 inbred lines were derived from Huangzaosi based on the pedigree information, and the other three lines, i.e., Tang Sipingtou, Huangzaosi and Wutang448, were derived from the Chinese landrace "Tang Sipingtou". Genetic components were pair-compared using 44,927 SNPs without missing rate in Huangzaosi and with missing rate <0.2 among the other 42 inbred lines. The window size was set to 50 SNPs and the step length was set to 1. Genetic regions were defined to be originated from Huangzaosi where more than 80 % SNPs were the same with those of Huangzaosi. Genetic regions were considered to be conservative when they existed in more than 60 % of Huangzaosi's descendants.

### Analysis of linkage disequilibrium

LD between SNPs for each chromosome was measured using squared Pearson correlation coefficient ($r^2$) between vectors of SNP alleles following Hill and Weir (Hill and Weir 1994). The cutoff value of $r^2$ was set to be 0.1 (Yan et al. 2009). TASSEL 3.0.124 package was run with a 50 kb slide-window which determines the width of the window on one side of the start site, and LD was calculated for sites within the window of sites surrounding the current site (Bradbury et al. 2007), using 41,819 SNPs with MAF >0.05 and missing data <20 %. The spacing between two loci on the same chromosome was segmented in distance bins of 50 kb and the average LD was assessed for each bin (Riedelsheimer et al. 2012).

## Results

### Population structure and relative kinship

The most significant peak of $\Delta K$ was observed when $k = 2$ (Fig. 1a), suggesting that the entire set of maize germplasm could be divided into two groups. Based on the sources, pedigrees and knowledge of breeding history, the two groups corresponded to local germplasm and introduced germplasm (Fig. 1). The local germplasm group included 44 lines, mainly being the inbred lines derived from Chinese landraces such as Huangzaosi, and their descendants. The introduced germplasm group included 323 lines, mainly being the inbreds introduced directly from other countries or improved from those germplasm. Furthermore, the second significant peak of $\Delta K$ was observed when $k$ was equal to 5 (Fig. 1a), suggesting that the entire set of maize germplasm could be clustered further into five subgroups (Fig. 1). Considering the sources, pedigrees and breeding history, the five subgroups corresponded to five heterotic groups, i.e., Reid, Tem-tropic I, P, TSPT and Lancaster. The Reid contained 31 lines with representatives of B73 and Ye478; the Tem-Tropic I contained 155 lines with representative of Suwan1611, CML125-2 and Lu28; the P group contained 19 lines with representative of P138 and Shen137; the TSPT contained 35 lines with representative of Huangzaosi; and the Lancaster contained 33 lines with representative of Mo17. The other lines with membership probabilities <0.5 were appointed into one cluster called the mix group (Table S1).

Estimation of relative kinships showed that 94.97 % of paired relative kinship ranged from 0.05 to 0.28, 0.17 % of paired relative kinship equaled to 0, 0.94 % of paired relative kinship ranged from 0 to 0.05, and the remaining ranged from 0.30 to 0.50 (Fig. 2). This analysis indicated that a weak or various relative kinship existed in this collection of elite maize germplasm, which was consistent with known pedigree/sources of the 367 lines.

### Genetic diversity

For the entire panel, a total of 83,638 alleles were detected. The average GD was 0.364, ranging from 0.095 to 0.500, and the average PIC was 0.291, ranging from 0.091 to 0.375 (Table 1). The introduced germplasm group showed a higher diversity than the local germplasm group with GD of 0.367 (0.032–0.500) and 0.234 (0.000–0.500), respectively; PIC of 0.293 (0.031–0.375) and 0.196 (0.000–0.375), respectively (Table 1). To estimate the effect of sample size on GD, 44 lines were selected randomly from the introduced germplasm group to make the sample size equal to that in the local germplasm group. The results of comparison confirmed that the introduced germplasm group had a higher genetic diversity than the local germplasm group with an average GD of 0.358 (±0.005) and a PIC of 0.287 (±0.004). Furthermore, the Tem-Tropic I subgroup showed the highest diversity, with GD of 0.358 (0.000–0.500) and PIC of 0.287 (0.000–0.375). The next was the Lancaster subgroup with GD of 0.322 (0.000–0.500) and PIC of

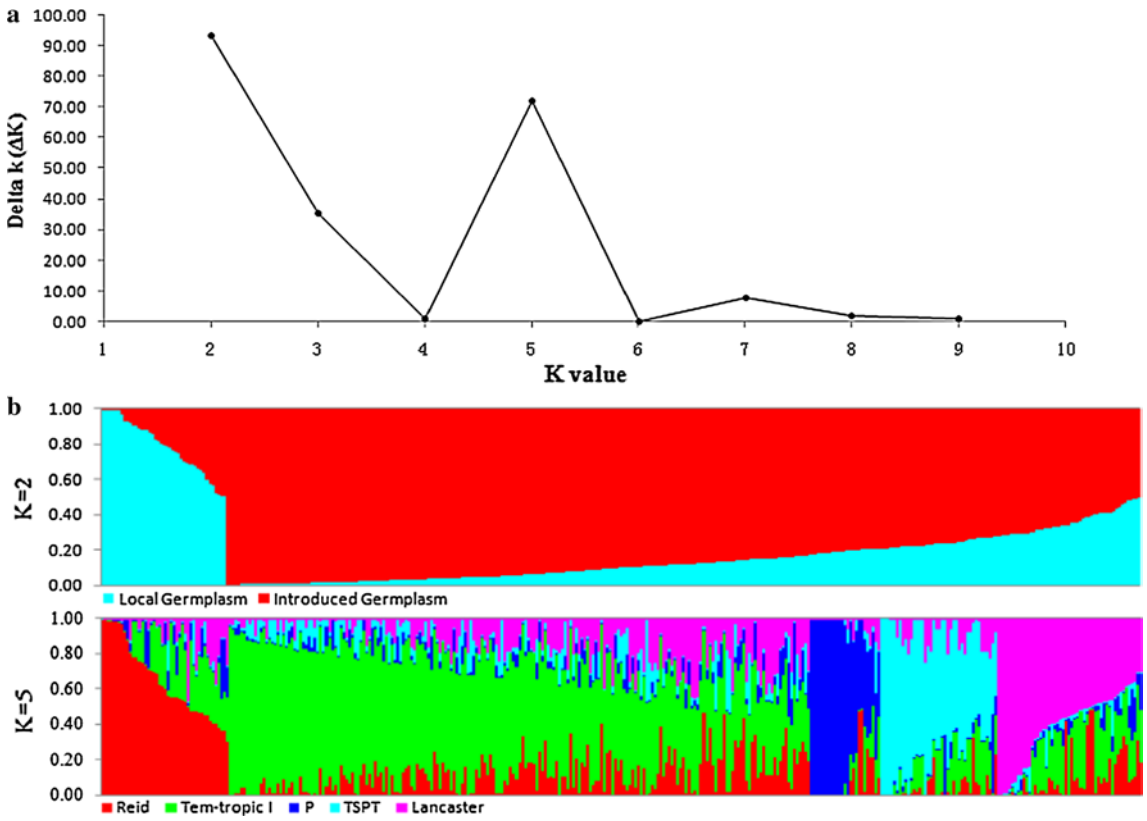


**Fig. 1** Estimation of the Ln (probability of data). Delta *K* was calculated from $k = 2$ to $k = 9$ (**a**). **b** displays the population structure of the 367 lines using 1,015 SNPs

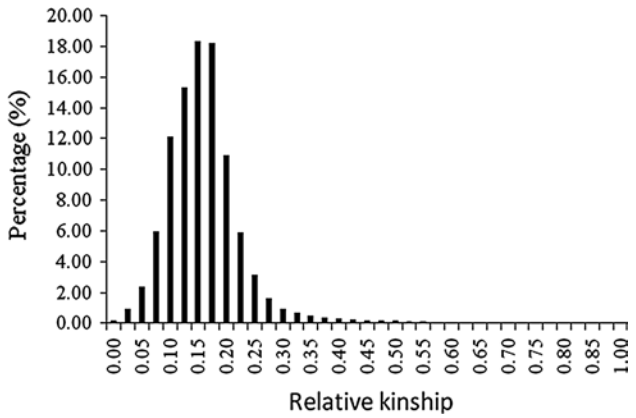


**Fig. 2** Distribution of paired relative kinship evaluated between the 367 lines. Values were calculated through TASSEL 3.0.124 package using 41,819 SNPs

0.257 (0.000–0.375). The Reid subgroup was more diverse than the TSPT, and the P was the least diverse subgroup (Table 1). Based on the re-sampling strategy, 19 lines were selected randomly in the four subgroups except the P which had only 19 lines. The results of multiple comparisons further confirmed the similar trend mentioned above (Table 2).

In addition, the GD varied among different genetic regions of a given chromosome in the subgroups. In general, the GD was lower near the centromeres when compared with other regions. Among the five subgroups, there existed some conservative regions with GD $< \widehat{\text{GD}} - 0.5 \times \text{SD}$. For example, there were seven conservative regions in the Lancaster, seven in the Reid, six in the TSPT, five in the P, and two in the Tem-Tropical I (Fig. 3, Fig. S2). These regions were distinct with each other among the five subgroups and distributed on different chromosomes.

## Genetic components transmitted from Huangzaosi to its descendants

According to the genotyping data of 44,927 SNPs, the proportion of genetic components which were the same as those in Huangzaosi among the 42 Huangzaosi-related lines varied from 1.35 % in the line 785 to 93.79 % in the line 72–125. The inbred lines 196, 5237 and Xi502, which were directly derived from the combination of Huangzaosi × Dan340, contained 62.10, 44.04, and 58.94 % of the genetic components transmitted from Huangzaosi across the genome, respectively. H21 (from Huangzaosi × H84) and Chang7-2 (from Huangzaosi × Wei95 × S901) contained

**Table 1** Genetic diversity of different groups/subgroups clustered using 41,819 SNPs

| Group | No. of lines | Gene diversity (GD) | PIC |
|---|---|---|---|
| Entire panel | 367 | 0.364 (0.091–0.500)[a] | 0.291 (0.091–0.375) |
| Introduced germplasm | 323 | 0.367 (0.032–0.500) | 0.293 (0.031–0.375) |
| Introduced germplasm (re-sampled) | 44 | 0.358 (±0.005) | 0.287 (±0.004) |
| Local germplasm | 44 | 0.234 (0.000–0.500) | 0.196 (0.000–0.375) |
| Lancaster | 33 | 0.322 (0.000–0.500) | 0.257 (0.000–0.375) |
| P | 19 | 0.209 (0.000–0.500) | 0.167 (0.000–0.375) |
| Reid | 31 | 0.296 (0.000–0.500) | 0.238 (0.000–0.375) |
| Tem-tropic I | 155 | 0.358 (0.000–0.500) | 0.287 (0.000–0.375) |
| TSPT | 35 | 0.211 (0.000–0.500) | 0.178 (0.000–0.375) |

[a] Numbers in the parenthesis refer to variations of GD and PIC

**Table 2** Multiple comparison of gene diversity (GD) and PIC among five subgroups with 19 lines

| Groups | GD | PIC |
|---|---|---|
| Tem-Tropic I | 0.341 (±0.002)$^{A}$* | 0.273 (±0.001)$^{A}$ |
| Lancaster | 0.306 (±0.002)$^{B}$ | 0.244 (±0.001)$^{B}$ |
| Reid | 0.294 (±0.005)$^{C}$ | 0.236 (±0.003)$^{C}$ |
| TSPT | 0.212 (±0.005)$^{D}$ | 0.176 (±0.004)$^{D}$ |
| P | 0.209$^{D}$ | 0.167$^{D}$ |

Numbers in the parenthesis refer to standard deviation of GD and PIC

* Different letters showed significant difference at 0.01 level

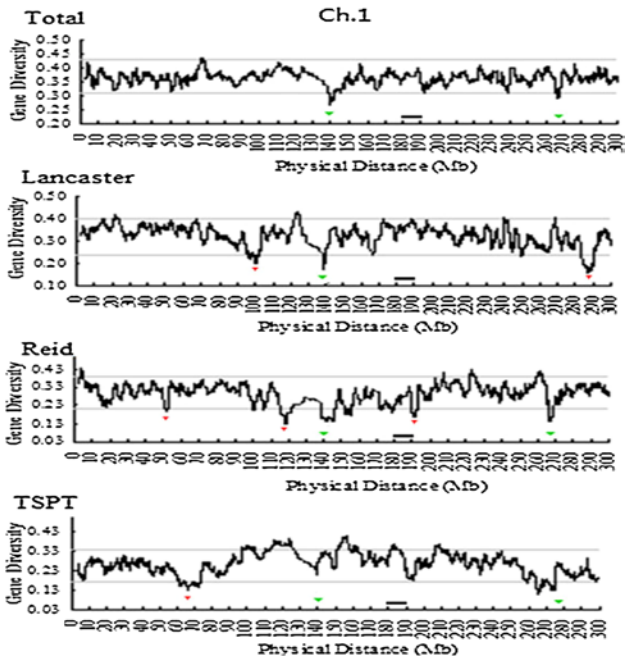


**Fig. 3** Comparison of GD among different subgroups for Chrom. 1. *Inverted red triangle* refers to conservative regions defined with GD < $\widehat{\text{GD}} - 0.5 \times \text{SD}$ in one specified subgroup. *Inverted green triangle* refers to conservative regions common in five subgroups. *Pink coloured rectangle* refers to centromeric regions (Color figure online)

40.81 and 58.56 % of the genetic components transmitted from Huangzaosi, respectively. Lx9801, Lx03-2 and 92-8, which were derived from Huangzaosi's descendants, contained 64.59, 55.96 and 58.16 % of the genetic components transmitted from Huangzaosi, respectively. About 58 % of the genetic components in the inbred Tang Sipingtou, which was derived from the same parental landrace as Huangzaosi, were the same as that of Huangzaosi; but Wutang448 had only 8.32 % of the genetic components to be the same as Huangzaosi (Table 3). Meanwhile, we detected 15 conservative genetic regions which appeared in more than 60 % of Huangzaosi's descendants. They were distributed on Chrom. 1, 2, 3, 4, 5, 6, 8 and 10 (Fig. 4).

### Linkage disequilibrium

This analysis indicated that, the average LD decay distance is 391 kb in the panel of 367 lines, varies from 265 to 598 kb across ten chromosomes (Chrom.), with 265 kb on Chrom.1, 598 kb on Chrom.10 (Table 4; Fig. 5). In the introduced group of 323 lines, the averaged LD distance is 328 kb, varies from 272 kb (Chrom.1) to 595 kb (Chrom.10). The average LD distance of introduced group is shorter than that in the original panel of 367 lines (Table 4). This trend was found on the Chrom.2, 5, 8 and 10, whereas the opposite was found on Chrom.1, 3, 4, 6 and 9. Chrom. 9 showed equal LD decay between the two samples.

## Discussion

### Change of population structure during maize breeding

Maize populations are subjected to migration and drift, to both natural selection and artificial selection during breeding history and subsequently diverse germplasm has been formed (Pressoir and Berthaud 2004). Therefore, it is important to obtain a complete picture of population

**Table 3** Proportion of genetic components on different chromosomes which were same as those of Huangzaosi among the Huangzaosi-related inbred lines

| Name | Chr.1 | Chr.2 | Chr.3 | Chr.4 | Chr.5 | Chr.6 | Chr.7 | Chr.8 | Chr.9 | Chr.10 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 72-125 | 76.64 | 98.62 | 94.82 | 92.24 | 100.00 | 99.10 | 99.28 | 100.00 | 92.13 | 85.07 | 93.79 |
| Huangyesi | 81.32 | 89.62 | 95.91 | 93.74 | 75.51 | 95.41 | 90.43 | 100.00 | 78.38 | 95.28 | 89.56 |
| HR962 | 82.15 | 100.00 | 95.71 | 54.51 | 79.53 | 94.23 | 96.42 | 100.00 | 88.41 | 50.78 | 84.17 |
| Ji35 | 81.14 | 98.98 | 69.60 | 75.35 | 22.70 | 92.18 | 90.86 | 40.58 | 35.58 | 87.33 | 69.43 |
| Jing7 | 61.72 | 53.11 | 92.80 | 38.09 | 88.46 | 93.03 | 67.09 | 83.50 | 25.89 | 80.60 | 68.43 |
| Bai197 | 73.90 | 94.48 | 56.60 | 51.63 | 62.59 | 20.85 | 53.19 | 96.94 | 60.95 | 91.08 | 66.22 |
| Dhuang212 | 49.54 | 76.24 | 82.26 | 47.01 | 87.74 | 86.44 | 58.01 | 37.80 | 58.73 | 87.54 | 67.13 |
| Wenhuang31413 | 64.38 | 43.94 | 87.58 | 59.76 | 84.80 | 63.24 | 81.56 | 91.27 | 28.20 | 38.14 | 64.29 |
| Baiyesi | 51.11 | 91.61 | 71.26 | 67.99 | 77.81 | 35.67 | 37.13 | 95.11 | 32.50 | 82.21 | 64.24 |
| 444 | 58.71 | 53.84 | 83.73 | 69.23 | 98.24 | 80.40 | 51.40 | 32.60 | 36.16 | 70.76 | 63.51 |
| Ji444 | 58.37 | 52.74 | 83.67 | 68.88 | 98.24 | 79.75 | 50.65 | 32.16 | 36.01 | 70.55 | 63.10 |
| Lx9801 | 36.36 | 98.09 | 45.21 | 92.76 | 77.06 | 83.96 | 32.61 | 60.29 | 41.30 | 78.28 | 64.59 |
| Jing7huang | 49.13 | 53.61 | 90.91 | 30.39 | 89.17 | 65.40 | 51.19 | 85.05 | 29.49 | 83.76 | 62.81 |
| 196 | 33.52 | 96.66 | 67.77 | 74.01 | 77.22 | 55.94 | 43.60 | 64.86 | 14.11 | 93.33 | 62.10 |
| Tang Sipingtou | 83.92 | 55.98 | 59.06 | 59.01 | 61.06 | 46.49 | 44.88 | 96.10 | 30.17 | 43.04 | 57.97 |
| Tianya4 | 69.51 | 32.50 | 93.69 | 35.58 | 67.27 | 36.49 | 72.15 | 91.84 | 19.10 | 77.12 | 59.52 |
| Jing404 | 34.21 | 63.90 | 50.19 | 92.76 | 80.86 | 77.15 | 32.77 | 57.80 | 33.27 | 78.13 | 60.10 |
| Xi502 | 32.51 | 99.02 | 69.14 | 55.26 | 71.71 | 55.62 | 34.08 | 64.97 | 13.90 | 93.21 | 58.94 |
| Ji853 | 50.01 | 53.37 | 68.27 | 37.95 | 39.67 | 84.45 | 57.93 | 84.37 | 71.80 | 60.49 | 60.83 |
| Huangyesi3 | 31.62 | 91.00 | 63.05 | 43.67 | 90.83 | 61.30 | 68.03 | 49.99 | 53.01 | 26.01 | 57.85 |
| Chang7-2 | 42.87 | 75.56 | 20.84 | 78.31 | 57.98 | 37.74 | 22.11 | 97.88 | 69.90 | 82.42 | 58.56 |
| Lx03-2 | 45.32 | 82.85 | 32.19 | 85.47 | 72.87 | 46.52 | 37.08 | 59.40 | 28.97 | 68.96 | 55.96 |
| 92-8 | 40.24 | 75.06 | 19.85 | 77.94 | 56.96 | 37.28 | 22.00 | 97.88 | 73.40 | 80.96 | 58.16 |
| Wu314 | 42.32 | 70.27 | 78.67 | 51.15 | 39.43 | 61.52 | 37.50 | 34.56 | 16.57 | 61.93 | 49.39 |
| K12 | 47.83 | 73.08 | 63.11 | 29.58 | 28.95 | 41.62 | 24.11 | 93.10 | 44.62 | 38.53 | 48.45 |
| Jing24 | 12.66 | 52.88 | 77.57 | 40.85 | 63.21 | 52.50 | 22.40 | 64.63 | 30.93 | 47.37 | 46.50 |
| Luyuan133 | 61.09 | 46.43 | 95.12 | 45.94 | 9.13 | 28.40 | 6.31 | 96.78 | 22.29 | 11.45 | 42.29 |
| 5237 | 20.46 | 77.95 | 72.45 | 42.65 | 62.64 | 28.75 | 24.06 | 53.18 | 3.44 | 54.83 | 44.04 |
| Yue89E4-2 | 36.75 | 26.53 | 40.98 | 41.44 | 57.77 | 32.09 | 69.29 | 82.33 | 37.95 | 30.12 | 45.52 |
| H21 | 27.88 | 57.24 | 38.88 | 69.92 | 27.29 | 46.41 | 44.56 | 44.76 | 23.89 | 27.23 | 40.81 |
| Shuang741 | 21.04 | 62.90 | 24.53 | 60.81 | 45.33 | 23.12 | 62.10 | 47.66 | 28.35 | 28.82 | 40.47 |
| S001 | 29.14 | 55.20 | 61.44 | 17.07 | 32.73 | 54.03 | 38.06 | 38.59 | 17.84 | 35.61 | 37.97 |
| Q1261 | 8.24 | 47.38 | 37.43 | 12.24 | 44.66 | 39.22 | 30.66 | 83.74 | 23.77 | 58.00 | 38.53 |
| Xun928 | 18.21 | 45.18 | 13.93 | 24.59 | 22.59 | 32.14 | 18.71 | 82.22 | 63.25 | 53.24 | 37.41 |
| Ye515 | 14.54 | 16.39 | 72.38 | 39.04 | 34.46 | 10.33 | 6.36 | 8.05 | 28.11 | 52.54 | 28.22 |
| DH40 | 5.07 | 6.11 | 36.38 | 12.67 | 28.15 | 6.20 | 4.65 | 4.47 | 2.06 | 26.74 | 13.25 |
| 897 | 6.61 | 13.58 | 13.53 | 16.87 | 11.58 | 8.34 | 13.07 | 7.97 | 8.03 | 2.35 | 10.19 |
| Wutang448 | 1.79 | 7.80 | 15.50 | 9.91 | 4.41 | 0.96 | 11.60 | 15.48 | 7.13 | 8.59 | 8.32 |
| 8723 | 5.69 | 7.39 | 3.73 | 11.07 | 8.13 | 12.82 | 7.14 | 6.33 | 9.04 | 4.93 | 7.63 |
| Ji854 | 4.20 | 9.00 | 9.21 | 9.73 | 7.78 | 4.59 | 6.17 | 9.20 | 3.54 | 6.67 | 7.01 |
| Ji856 | 2.47 | 8.39 | 9.07 | 7.16 | 3.51 | 3.88 | 3.10 | 5.20 | 7.07 | 7.77 | 5.76 |
| 785 | 0.03 | 1.77 | 0.52 | 2.09 | 1.84 | 0.93 | 0.19 | 1.67 | 1.66 | 2.77 | 1.35 |
| Average | 39.39 | 57.53 | 56.16 | 48.25 | 53.62 | 48.01 | 41.06 | 59.53 | 34.07 | 53.92 | |

structure for different maize germplasms, which could be informative for prediction of hybrid performance. Major heterotic groups such as Reid, Lancaster, and European Flint have been established from traditional populations or landraces in the world. First-generation inbred lines developed from these populations were then intercrossed to generate second and further generation inbred lines (Bernardo 1990; Charcosset and Essioux 1994). In Chinese maize breeding, two main groups of local germplasm and introduced germplasm have been formed based on the sources of inbred lines or hybrids/populations (Zeng 1990; Wang et al. 1997). Further, five major subgroups, i.e., Reid,

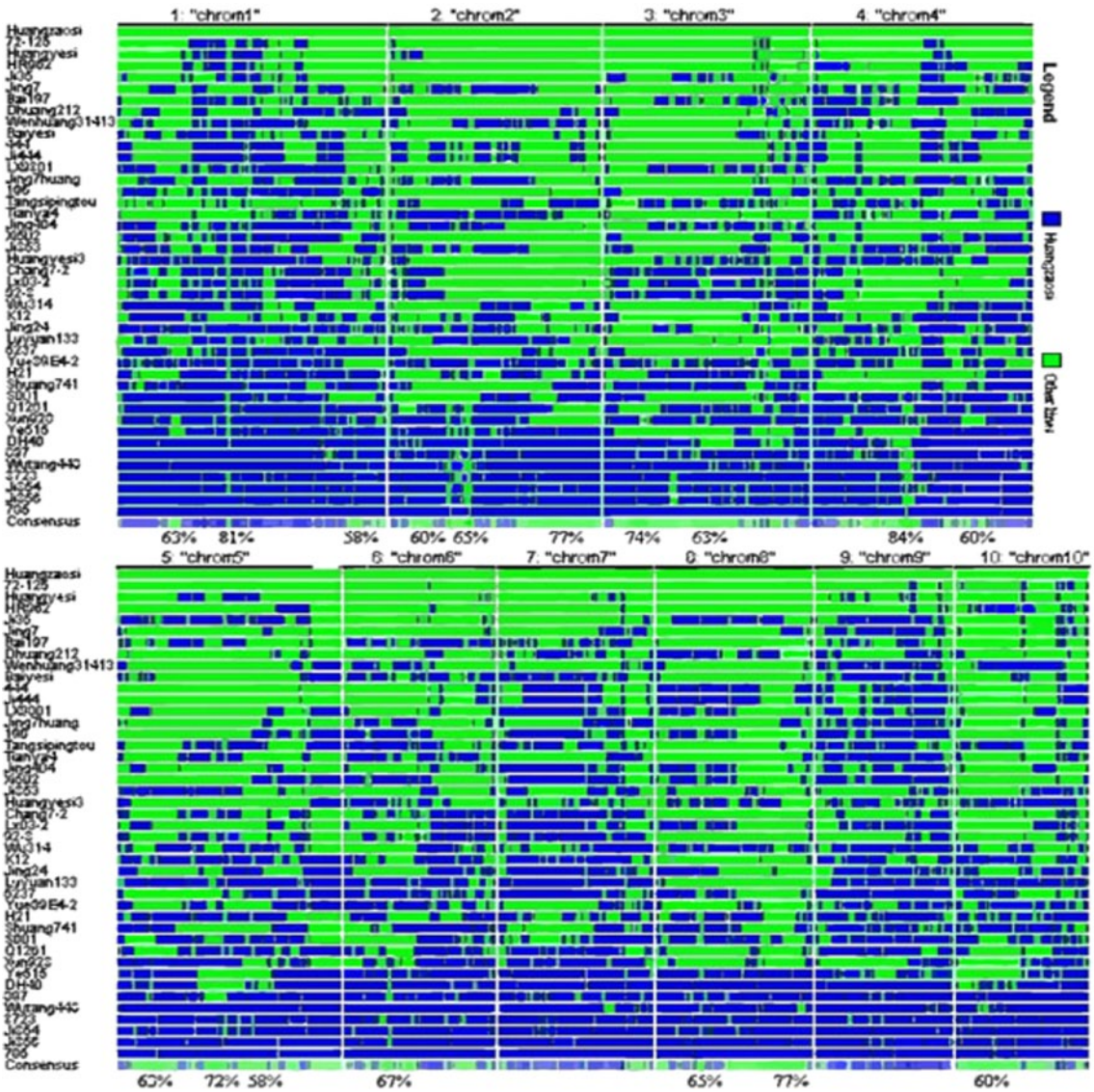


**Fig. 4** Comparison of genetic components using 44,927 SNPs among 43 Huangzaosi-related lines, of which 40 inbred lines were descended from Huangzaosi, and other three inbred lines, i.e., Tang Sipingtou, Wutang448 and Huangzaosi, were developed from Chinese landrace "Tang Sipingtou". A window size was set to 50 SNPs and the step long was set to 1. Genetic regions being filled with green refer to those came from Huangzaosi, where more than 80 % of SNPs were the same as those of Huangzaosi. Totally 15 conservative genetic regions, which existed in more than 60 % of Huangzaosi's descendants, were detected

Lancaster, P, TSPT and Tem-Tropic I, have been established during the breeding of single-cross hybrids (Li et al. 2005; Liu et al. 2012; Lu et al. 2009; Wang et al. 2008; Yan et al. 2009). The present study found that two groups of germplasm could be clustered based on 1,015 SNPs, i.e., the local germplasm and the introduced germplasm. Some elite lines, which have been widely used in Chinese maize breeding historically and currently, were developed from Chinese landraces, most of which were appointed into the local germplasm group (Li and Wang 2010). For example, Huangzaosi was derived from the local landrace "Tang Sipingtou", and later some lines were developed from Huangzaosi. Actually, further analysis showed that most lines in the group of the local germplasm were appointed into the TSPT subgroup. This may partly explain why the popular pattern of heterosis utilization in China for single-cross hybrids is local germplasm × introduced germplasm.

On the other side, a lot of inbred lines such as B73 and Mo17 were introduced from the U.S., and some inbred lines, e.g., Ye478, Zheng58 and Sheng137, were generated from the intercross of these introduced lines or developed from introduced germplasm such as Pioneer hybrids e.g., 78599. Most of the inbreds were appointed into the group of introduced germplasm. In addition, the introduced

**Table 4** Average LD decay of the ten chromosomes with cutoff value of $r^2 > 0.1$

| Chrom. | Total | Introduced group |
|---|---|---|
| 1 | 265 | 272 |
| 2 | 345 | 328 |
| 3 | 368 | 372 |
| 4 | 368 | 474 |
| 5 | 335 | 331 |
| 6 | 338 | 342 |
| 7 | 426 | 426 |
| 8 | 503 | 489 |
| 9 | 474 | 476 |
| 10 | 598 | 595 |
| Average | 391 | 328 |

'Total' presents the original panel of 367 elite lines. 'Introduce group' presents another panel of 323 lines based on the analysis of population structure results. This table showed that the population admixture could affect LD decay. Otherwise, LD varied across different chromosomes

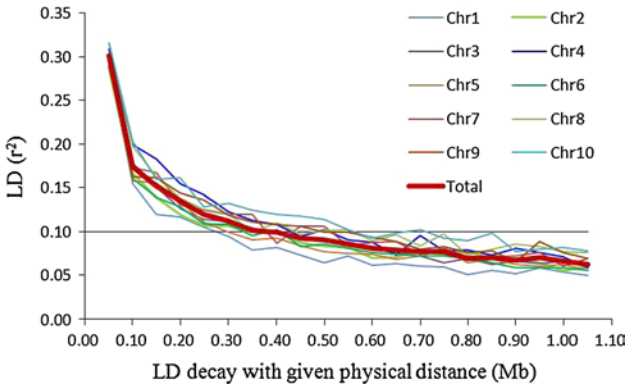


**Fig. 5** Picture of whole-genome LD in the panel of 367 lines. LD over chromosomes is given in physical distance of 50 kb. The value of $r^2 = 0.1$ was reached at 391 kb and varied among different chromosomes

germplasm was further divided into the other three subgroups, i.e., Reid, Lancaster and P, which was coincident with previous studies with different markers and samples (Li et al. 2005; Liu et al. 2012; Lu et al. 2009; Wang et al. 2008; Yan et al. 2009).

Interestingly, 155 of 367 lines were appointed into the Tem-Tropic I subgroup which had a very high genetic diversity. One of the major reasons is that in the past 20 years tropical maize germplasm was introduced into temperate regions and crossed with temperate germplasm (mainly inbred lines), and then a number of new inbreds and hybrids have been developed and released. In addition, the Tem-Tropic I subgroup included a few inbreds which were developed from Chinese landraces, e.g., Wuxi205, Jiao51, Lu28, Fanrong2, GB etc., However, the reason why some temperate landrace-derived lines were clustered into this subgroup is not clear although it was postulated that they contained pedigrees of tropical/subtropical germplasm in their progenitors. These results of population structure analysis could provide important information about this panel to be used in association mapping and breeding in the future.

## Genetic differentiation among Huangzaosi's descendants

According to the literature, Huangzaosi had the highest utilization frequency historically and currently in Chinese maize breeding, not only in inbred development but also in hybrid release, because Huangzaosi has a lot of advantages, such as short growth duration, resistance to maize dwarf mosaic virus, and high general combining ability etc., (Li 1997; Li and Wang 2010). In the present study, 40 Huangzaosi descendants and two Huangzaosi-related lines were included in the panel. Actually, Huangzaosi, Wutang448 and Tang Sipingtou were derived from the same landrace "Tang Sipingtou". Tang Sipingtou contained 57.97 % of genetic components which were the same as those of Huangzaosi, ranging from 30.17 % (Chrom. 9) to 96.10 % (Chrom. 8) (Table 3). Interestingly, only 8.32 % of genetic components in Wutang448 were the same as those in Huangzaosi, which could be explained by diverse genetic background in the landrace "Tang Sipingtou" and differences of artificial selection during inbred development by different breeders. Huangyesi was developed from the combination of Huangzaosi × Yejihong, and three backcrosses were followed using Huangzaosi as the recurrent parent. The results showed that Huangyesi contained 89.56 % of genetic component transmitted from Huangzaosi. However, it could be pointed out that some Huangzaosi's descendants, e.g., 785, Ji856 and 8723, contained only <10 % of Huangzaosi's genetic component and thus they were not appointed to the TSPT subgroup (Table 3). Probably during their breeding processes, Huangzaosi was just taken as the donor parent for a few favorable alleles. The results in this analysis revealed that the genetic components in Huangzaosi's descendants transmitted from Huangzaosi varied greatly.

The results obtained in the present study also indicated that the feature of chromosomal conservation among Huangzaosi's descendents varied among the chromosomes. Chrom. 8 showed the highest conservation, with 59.53 % of genetic components among Huangzaosi descendants which were the same as those of Huangzaosi, and two conservative regions existed in about 65 % of Huangzaosi's descendants. Chrom. 9 showed the lowest conservation, where only 34.07 % of genetic components were the same as those of Huangzaosi and no conservative regions were found among most of Huangzaosi descendants (Table 3; Fig. 4).

Totally 15 conservative regions among Tang Sipingtou and its descendents were found (Fig. 4). The results suggested that they are important genomic regions for Huangzaosi-derived lines which constitute the most widely used heterotic group in China. More research should be done to establish links between these genomic regions and phenotypes (esp. high general combining ability) to understand why Huangzaosi can become a foundation parent in maize breeding in China.

### Genetic diversity assessment

In the present study, the 367 maize lines had a higher GD (0.364) when compared with the GD (0.32) across three collections of maize inbred lines: the first is the sample set including 259 globally diverse lines (Hamblin et al. 2007), the second included 770 global maize lines from six countries (Lu et al. 2009), and the third included 1,537 elite lines (Inghelandt et al. 2010). The GD value obtained in this study was slightly lower when compared with the GD (0.39) across 527 diverse lines obtained by Yang et al.(2010a).

This study also indicated that the GD values varied across different genetic regions even on the same chromosome. Significantly, the GD in the regions near the centromeres was the lowest, which was consistent with the low recombination in the centromere region. Meanwhile, some conserved genetic regions with lower GD were found for different subgroups. More than five conservative genetic regions were found for each of four subgroups, i.e., Lancaster, Reid, TSPT, and P. However, the GD of the Tem-Tropical I was higher than that of the other subgroups and only two conservative regions were found across the ten Chromosomes. The results implied that different genetic regions had undergone different selection pressure during breeding processes to form different heterotic groups. In addition, the results could explain why Lancaster, Reid, TSPT and P were clustered unequivocally in different studies when using different maize collections and markers (Li et al. 2005; Liu et al. 2012; Lu et al. 2009; Wang et al. 2008; Yan et al. 2009).

### Effect of sample on linkage disequilibrium

LD is affected by recombination, genetic drift, selection within population, population admixture, and relatedness (Remington et al. 2001). Previous studies have evaluated different LD extents in germplasm collections and provided informative proofs for association mapping. In this study, the average LD decay according to physical distance was less rapid in the original panel of 367 lines than in the introduced group of 323 lines, with an average distance to reach $R^2 = 0.1$ of 391 and 328 kb, respectively (Table 4). This may be caused by the greater population admixture in the original panel. Also, the panel of 367 lines contained 44 local maize germplasm which consisting mostly of Huangzaosi and its descendants, the increasing relatedness would weaken the LD decay.

LD decay varied across ten chromosomes as well as in different genetic regions of one chromosome (Table 4; Fig. 5), which may be caused by different LD patterns among different chromosomes as well as different genetic regions on one chromosome; this result was similar to some previous studies (Jung et al. 2004; Yan et al. 2010).

When compared with previous studies, the average LD decay of 391 kb observed in this study was larger than 27.7 kb in 284 diverse lines reported by Weng et al. (2011), and much larger than 5–10 kb in the global maize collection of 632 lines worked by Yan et al. (2009). However, LD level in this study was <500 kb previously reported by Jung et al. (2004) using 192 elite lines and Riedelsheimer et al. (2012) using 285 diverse lines, respectively. It was also much <2.74 cM in 1,537 commercial maize germplasm reported by Inghelandt et al. (2011), because in the maize genome, 1 cM averagely corresponds to 1,460 kb (Civardi et al. 1994). Considering that maize genome is about 2,400 Mb (Yan et al. 2009; Wang et al. 2012) our analysis suggests that 56,000 SNP markers can be utilized for a first genome-wide association mapping in this panel of 367 elite lines.

**Acknowledgments** This work was partly supported by the Ministry of Science and Technology of China (2011CB100100, 2011DFA30450), National Natural Science Foundation (U1138304), CAAS (Innovation Program) and the Ministry of Agriculture of China (2011-G15, Baozhong Project). We are grateful to Dr. Alain Charcosset for language correcting and suggestions on data analyses. We also thank anonymous reviewers for suggestions to improve the quality of this manuscript.

## References

Bernardo R (1990) Methods used in developing maize inbreds. Maydica 35:1–16

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

Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 89:336–343

Civardi L, Xia YJ, Edwards KJ, Schnable PS, Nikolau BJ (1994) The relationship between genetic and physical distances in the cloned al-sh2 interval of the *Zea mays* L. genome. Proc Natl Acad Sci USA 91:8268–8272

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

Galecki AT, Wolfinger RD, Linares OA, Smith MJ, Halter JB (2004) Ordinary differential equation PK/PD models using the SAS macro NLINMIX. J Biopharm Stat 14:483–503

Ganal MW, Durstewitz G, Polley A, Be'rard A, Buckler ES, Charcosset A, Clarke JD, Graner E, Hansen M, Joets J, Paslier ML, McMullen M, Montalent P, Rose M, Schön CC, Sun Q, Walter H, Martin O, Falque M (2011) A large Maize (*Zea mays* L.) SNP

genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6(12):e28334

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:e1367

Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. Proc Natl Acad Sci USA 109:12420–12425

Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet 54(4):705–714

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. Theor Appl Genet 109:681–689

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer Nathan M, Yang H, Wang J, Dai J, Schnable Patrick S, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–1030

Li S (1997) The development and application of maize inbred line "Huangzaosi". Beijing Agric Sci 15:19–21 (in Chinese)

Li Y, Wang T (2010) Germplasm base of maize breeding in China and formation of foundation parents. J Maize Sci 18:1–8 (in Chinese)

Li Y, Shi Y, Cao Y, Wang T (2005) Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. Genet Resour Crop Evol 51:845–852

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Liu Z, Wu X, Liu H, Li Y, Li Q, Wang F, Shi Y, Song Y, Song W, Zhao J, Lai J, Li Y, Wang T (2012) Genetic diversity and population structure of important Chinese maize inbred lines revealed by 40 core simple sequence repeats (SSRs). Scientia Agricultura Sinica 45:2107–2138 (in Chinese)

Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. Theor Appl Genet 103:613–617

Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120:93–115

Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Prakash AB, Rong T, Xu Y (2011) Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS ONE 6:e24861

Pressoir G, Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92:95–101

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81:8014–8018

Troyer AF (1990) A retrospective view of corn genetic resources. J Hered 81:17–24

Troyer AF (1999) Background of US hybrid corn. Crop Sci 39:601–626

Wang Y, Wang Z, Wang Y (1997) Studies on the heterosis utilizing models of main maize germplasms in China. Scientia Agricutural Sinica 30:16–24 (in Chinese)

Wang R, Yu Y, Zhao J, Shi Y, Song Y, Wang T, Li Y (2008) Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. Theor Appl Genet 117:1141–1153

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci 196:125–131

Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, Zhang D, Bai L, Zhang S, Li X (2011) Genome-wide association study Identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) Inbred Lines. PLoS ONE 6:e29229

Yan J, 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 J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch J, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breeding 25:441–451

Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J (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 X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J (2010b) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121:417–431

Yu Y, Wang R, Zhao J, Shi Y, Song Y, Wang T, Li Y (2007) Genetic diversity and structure of the core collection for maize inbred lines in China. Maydica 52:181–194

Zeng S (1990) The maize germplasm base of hybrid in China. Scientia Agricultura Sinica 23:1–9 (in Chinese)