Theor Appl Genet (2016) 129:1775–1784
DOI 10.1007/s00122-016-2739-6

ORIGINAL ARTICLE

# Analysis of recombination QTLs, segregation distortion, and epistasis for fitness in maize multiple populations using ultra-high-density markers

**Chunhui Li[1] · Yongxiang Li[1] · Yunsu Shi[1] · Yanchun Song[1] · Dengfeng Zhang[1] · Edward S. Buckler[2] · Zhiwu Zhang[2] · Yu Li[1] · Tianyu Wang[1]**

Received: 8 March 2016 / Accepted: 4 June 2016 / Published online: 5 July 2016


**Abstract**

***Key message*** **Using two nested association mapping populations and high-density markers, some important genomic regions controlling recombination frequency and segregation distortion were detected.**

*Abstract* Understanding the maize genomic features would be useful for the study of genetic diversity and evolution and for maize breeding. Here, we used two maize nested association mapping (NAM) populations separately derived in China (CN-NAM) and the US (US-NAM) to explore the maize genomic features. The two populations containing 36 families and about 7000 recombinant inbred lines were evaluated with genotyping-by-sequencing. Through the comparison between the two NAMs, we revealed that segregation distortion is little, whereas epistasis for fitness is present in the two maize NAM populations. When conducting quantitative trait loci (QTL) mapping for the total number of recombination events, we detected 14 QTLs controlling recombination. Using high-density markers to identify segregation distortion regions (SDRs), a total of 445 SDRs were detected within the 36 families, among which 15 common SDRs were found in at least ten families. About 80 % of the known maize gametophytic factors (*ga*) genes controlling segregation distortion were overlapped with highly significant SDRs. In addition, we also found that the regions with high recombination rate and high gene density usually tended to have little segregation distortion. This study will facilitate population genetic studies and gene cloning affecting recombination variation and segregation distortion in maize, which can improve plant breeding progress.

Communicated by N. de Leon.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-016-2739-6) contains supplementary material, which is available to authorized users.

✉ Yu Li
liyu03@caas.cn

✉ Tianyu Wang
wangtianyu@263.net

[1] Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2] Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA

## Introduction

The maize genome is a source of abundant genetic diversity and phenotypic variation (Buckler et al. 2006). Extensive single nucleotide polymorphism (SNP), small insertions–deletions (indels) and larger structure variations in the form of copy-number variations, presence/absence variations and movement of transposable elements have undoubtedly left a profound impact on the genome (Schnable et al. 2009; Chia et al. 2012). Understanding the genomic features, including recombination, segregation distortion, and epistasis, would be beneficial to revealing the genetic diversity and evolutionary mechanisms in maize.

Recombination is a foundation of genetic diversity, and also a crucial component of plant breeding (Li et al. 2007a). Combining favorable alleles into a single line is a main objective of plant and animal breeding. However, the identification of positive alleles is limited by the time and the number of recombination events. Consequently, understanding the genetic mechanism of recombination variation would help to produce high recombinogenic lines and improve efficiency of plant breeding. Recombination

does not randomly occur in the maize genome (Schnable et al. 2009). Lai et al. (2010) reported that 30–50 recombination events were expected for each line derived from two parents via continuous selfing in maize. Previous studies have used different segregating populations to identify the number of recombination events and study the variation in recombination frequencies across the genome in maize (Pan et al. 2012; Farkhari et al. 2011; McMullen et al. 2009; Tulsieram et al. 1992; Bauer et al. 2013). However, genetic factors controlling the number of recombination events occurred on per chromosome or in per line are poorly understood. To our knowledge, only two literatures have used quantitative trait loci (QTL) mapping method to determine the genomic regions controlling recombination frequency in maize segregating populations (Li et al. 2009; Esch et al. 2007).

Segregation distortion (SD) is widespread in plant and animal populations, and is increasingly considered as a potentially powerful force in the evolution of recombination and reproductive isolation (Taylor and Ingvarsson 2003). In plant, Mangelsdorf and Jones (1926) first reported the phenomenon of SD in maize. To date, five gametophytic factors (*ga*) genes affecting SD have been identified in maize (http://www.maizegdb.org/). If a gene that leads to SD is segregating in a mapping population, markers closely linked to the gene would also tend to show SD. A number of segregation distortion loci (SDL) and segregation distortion regions (SDRs) were identified using individual maize segregating populations (Fu et al. 2006; Lu et al. 2002; Casa et al. 2000; Falque et al. 2005; Sharopova et al. 2002). However, reliable results for SDL and SDRs are hard to obtain using single mapping population. Analysis of SD in multiple populations would be beneficial to finding common SDRs and to identifying genes that cause SD in these regions. McMullen et al. (2009) used a nested association mapping (NAM) population, including 5000 recombination inbred lines from 25 families that were created in the US (US-NAM) and 1106 SNPs to analyze the SD in maize. However, the conclusion obtained in the US-NAM population argued for little SD.

Interaction of alleles at different loci, i.e., epistasis effects, is particularly important in evolutionary research, such as reproductive isolation, inbreeding depression, and the evolution of sex (Mallet 2001). Traditionally, epistasis is always defined with respect to a specific phenotype and inferred on the basis of dependence of genotypic values between loci for that phenotype (Wade et al. 2001). However, epistasis for fitness (fitness epistasis) should also have a genomic signature (Phillips 2008; Bomblies et al. 2007; Payseur and Hoekstra 2005). Fitness epistasis will occur when the fitness value of an allele at one locus depends upon alleles at one or more other loci (Wade et al. 2001). Previous studies have suggested that epistasis for fitness might function in the evolution of sex and recombination (Kimura and Maruyama 1966; Kondrashov 1988; Charlesworth 1990). McMullen et al. (2009) used linkage disequilibrium method to detect no fitness epistasis in the US-NAM population. However, Corbett-Detig et al. (2013) found fitness epistasis is widespread within species, such as *Drosophila*, *Arabidopsis*, and maize.

In the present study, another NAM population was independently developed in China (CN) by crossing 11 diverse parents with a common parent HUANGZAOSI (HZS). These parents of the CN-NAM population were selected from different heterotic groups widely used in Chinese maize breeding (Li and Wang 2010). Through natural and artificial selection over time, the Chinese maize germplasm is well adapted in numerous ecological regions of China and have substantial genetic differences from foreign germplasm (Jiao et al. 2012). A parallel analysis of the genomic features of CN-NAM would provide direct evidence for validating the conclusion of US-NAM. In addition, a combination of the CN-NAM and the US-NAM populations is supposed to provide larger genomic resources for dissecting the genomic features in maize.

The CN-NAM and US-NAM populations have been genotyped using genotyping-by-sequencing (GBS) technology, and subsequently, ultra-high-density genetic maps based on GBS were constructed (Li et al. 2015). The high-density genetic maps provide more accurate results for understanding the genetic mechanisms of recombination, SD, and fitness epistasis in the huge genomic resources.

The objective of the present study is to investigate the genomic features of the two NAM populations including recombination, SD, and fitness epistasis by comparison. We also used high-density molecular markers to identify genetic factors controlling recombination variation and SD in the two NAMs.

## Materials and methods

### Plant material

The Chinese nested association mapping population (CN-NAM) is composed of about 2000 recombination inbred lines (RILs) derived from the crosses of the common parent HUANGZAOSI (HZS) with each of 11 diverse inbred lines: K12, YE478, ZHENG58, HUOBAI, QI319, WEIFENG322, LV28, HUANGYESI3, DUO229, PA405, and MO17 (Li et al. 2013). The US nested association mapping population (US-NAM) is consisted of about 5000 RILs derived from crossing B73 with 25 diverse inbred lines: B97, CML52, CML69, CML103, CML228, CML247, CML277, CML322, CML333, Hp301, Il14H, Ki3, Ki11, Ky21, M37W, M162W, Mo18W, MS71, NC350, NC358,

Oh43, Oh7B, P39, Tx303, and Tzi8 (Yu et al. 2008). In total, the two NAM populations contain 36 families and about 7000 RILs.

### Genetic maps

The CN-NAM and US-NAM populations have been genotyped using genotyping-by-sequencing (GBS) technology (Elshire et al. 2011; Glaubitz et al. 2014). The GBS data sets for CN-NAM and US-NAM are available from http://www.cgris.net/maize/data/ and http://www.panzea.org, respectively. Construction of the recombination bin maps using GBS data for each of the 36 families has been previously described (Li et al. 2015). Briefly, after obtaining the raw GBS data of each family, SNPs with minor allele frequency (MAF) <0.05 and within the same tag (64 bp) were filtered out. Drafts of parental genotypes were obtained with the assistance of low-coverage parental HZS sequence using a maximum parsimonious inference of recombination (MPR) method implemented in an R package MPR (Xie et al. 2010). High-quality SNPs were refined after removing low-quality SNPs through Bayesian inference method of the MPR package. The genotypes of each family line were determined using a hidden Markov model method, with heterozygotes set to missing. Then, recombination breakpoints were identified between two different genotype blocks with the same genotype. According to recombination breakpoints, the genotypic maps for all lines of each family were aligned and split into recombination bins. Resulting recombination bins were then treated as a genetic marker for linkage map construction of individual family in the R/qtl package (Broman et al. 2003). The total number of bins and the total length of genetic maps for each family are available in Online Resource 1.

After inferring parental genotypes and obtaining high-quality SNPs for each family, polymorphic markers were chosen to construct joint recombination bin maps for CN-NAM and US-NAM, respectively. Resulting bins were then used as a genetic marker for constructing composite genetic maps in the JoinMap 4.0 software across the 11 families and the 25 families, respectively. The composite genetic maps for CN-NAM and US-NAM were constructed through using 4932 and 5296 bins, respectively (Li et al. 2015).

### Identification and QTL analysis of recombination events

Recombination events (REs) were measured by counting breakpoints between stretches of marker alleles from one parent to the other parent in the RIL mapping data with markers ordered according to their map position. The total number of recombination events accumulated in individual RIL was used as a quantitative trait.

The method of inclusive composite interval mapping (ICIM) was used to detect the additive QTL for the total number of recombination events in each of the 36 families in the QTL ICIMapping software ver. 3.2 (Wang et al. 2012). In ICIM, the $P$ values for entering a variable (PIN) and removing a variable (POUT) were set at 0.001 and 0.002, and the scanning step was set to 1.0 cM. The LOD threshold was determined by a 1000 permutation test.

### Analysis of segregation distortion

For each family, the proportion of HZS or B73 alleles was calculated and segregation distortion was tested by $\chi^2$ analysis against the 1:1 expectation of RIL population. For CN-NAM and US-NAM, the proportions of HZS and B73 alleles were counted and $\chi^2$ test was performed to determine whether the marker ratio significantly differed from the expectation of 1:1, at $P < 0.05$ ($df = 1$). A region containing three or more closely linked markers exhibiting significant segregation distortion at the 0.05 level was defined as a segregation distortion region (SDR).

### Epistasis for fitness

We used two different statistical methods to search for potential epistatic interactions based on pairs of mapped markers on different chromosomes within each of the 36 NAM families. One method was inter-chromosomal linkage disequilibrium (LD) tests conducted in the Plink software (Purcell et al. 2007). $R^2$ was used as a measurement of LD. The other was $\chi^2$ tests for detecting genotype ratio distortion (GRD) developed by Corbett-Detig et al. (2013). Briefly, GRD was detected by computing a $\chi^2$ test between each pair of alleles on different chromosomes. To ensure that significant allelic pairs were not physical linked, GRD was restricted to inter-chromosomal allelic pairs. Moreover, pairwise allele comparisons for which any allele's frequency with less than 0.05 were excluded. Statistical significance were assessed via a $\chi^2$ test and a 5 % false discovery rate (FDR) was used to correct significance level. Finally, we reported pairwise alleles for which at least three adjacent markers were in local linkage and also showed significant GRD.

## Results

### Recombination events

High-density markers allow the identification of nearly all recombination events in each RIL. The two NAM populations captured a total of about 197,863 recombination events, including 54,000 in CN-NAM and 143,863 in

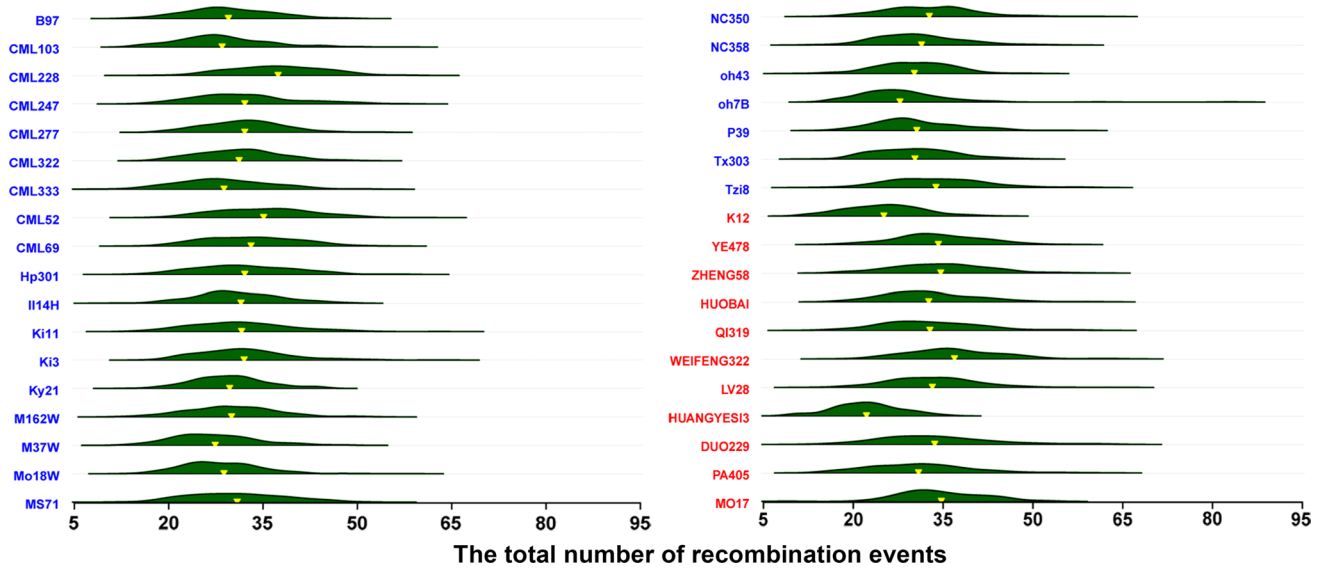


**Fig. 1** Frequency distribution of the total number of recombination events within the US-NAM families and the CN-NAM families. *Yellow triangle* represents mean number of recombination events in a family. *Blue font on the left side* indicates the US-NAM families; *red* indicates the CN-NAM families

**Table 1** QTL detected for the total number of recombination events in single family

| Family | Chr | Left marker position (Mb)[a] | Right marker position (Mb)[b] | LOD | PVE (%)[c] | Add[d] |
|---|---|---|---|---|---|---|
| CML228 | 6 | 147.512 | 147.897 | 3.6 | 8.3 | 2.2 |
| CML228 | 8 | 146.154 | 147.292 | 3.6 | 8.3 | 2.2 |
| CML69 | 7 | 32.478 | 33.089 | 5.2 | 15.7 | −3.3 |
| CML69 | 7 | 38.895 | 39.182 | 7.8 | 17.3 | 3.0 |
| CML69 | 9 | 120.365 | 123.239 | 3.6 | 7.2 | 2.0 |
| M37 W | 3 | 172.670 | 173.515 | 3.1 | 7.5 | 1.8 |
| MS71 | 4 | 200.103 | 200.368 | 2.6 | 6.2 | 1.9 |
| NC358 | 9 | 147.958 | 148.207 | 3.2 | 7.8 | 2.0 |
| K12 | 1 | 267.907 | 268.736 | 4.5 | 9.2 | −2.1 |
| K12 | 3 | 5.584 | 5.846 | 2.8 | 5.4 | −1.5 |
| K12 | 4 | 21.914 | 32.498 | 6.1 | 12.3 | −2.2 |
| ZHENG58 | 1 | 102.949 | 107.182 | 2.6 | 7.7 | −2.2 |
| QI319 | 2 | 219.901 | 220.144 | 2.8 | 8.8 | −2.5 |
| Mo17 | 6 | 113.421 | 114.782 | 2.5 | 8.6 | 2.0 |

Positive represents allele effect from common parent B73 or HZS; negative represents allele effect from diverse parents

[a] Physical location of the left-side marker of the identified QTL

[b] Physical location of the right-side marker of the identified QTL

[c] Phenotypic variation explained by QTL

[d] Estimated additive effect of QTL

US-NAM, with an average of about 31 crossovers per RIL. This is close to the value of 30–50 recombination events that is expected for each line derived from two parents via continuous selfing (Lai et al. 2010). The frequency distribution of the total number of recombination events for each RIL varied substantially among the 36 families, ranging from 9 to 83 crossovers. The total number of recombination events within individual family showed a typical distribution of a quantitative trait (Fig. 1).

By QTL mapping of the total of recombination events, we detected 14 QTLs within 9 of the 36 families (Table 1, Online Resource 2). These QTLs were distributed across

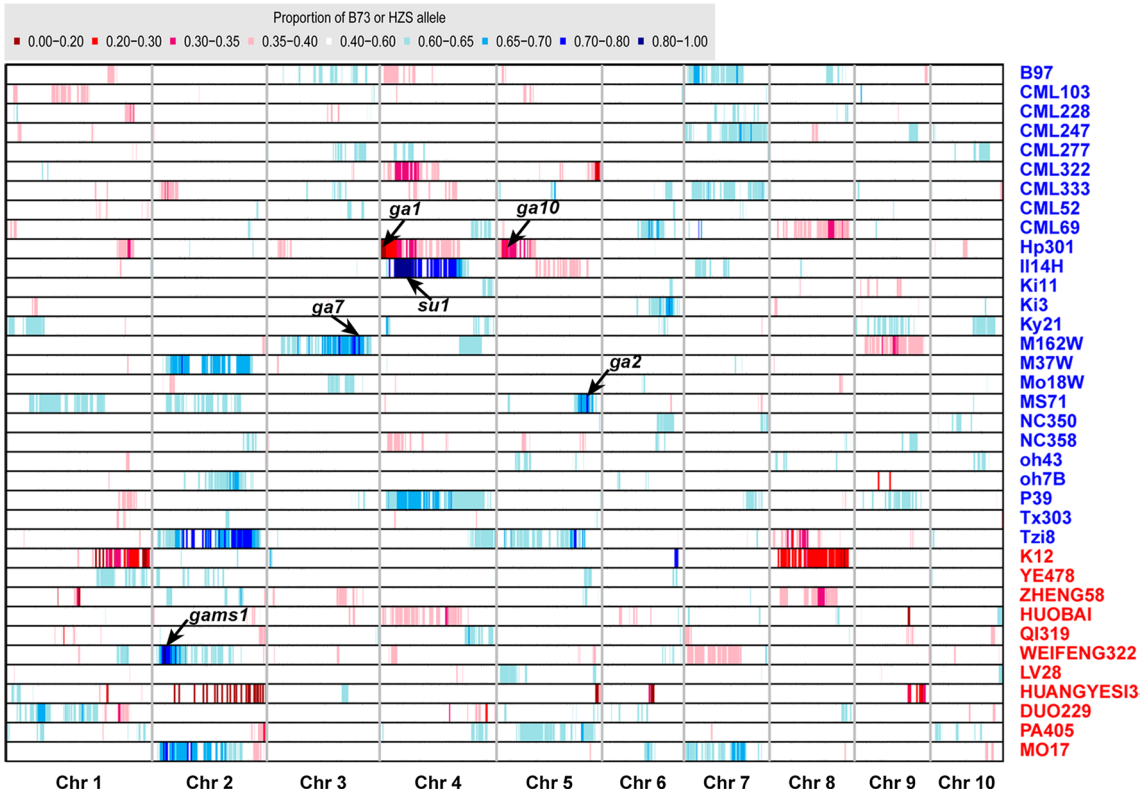


**Fig. 2** Segregation distortion within the 36 families of the two NAM populations. *Each row* represents one family. Blue font on the *right side* indicates the US-NAM families; *red* indicates the CN-NAM families. *Gray vertical lines* indicate the physical positions of chromosome boundaries for chromosomes 1 to 10 from *left* to *right*. The proportion of B73 or HZS alleles for a marker is indicated by the *color scale*

the genome, except for chromosomes 5 and 10. Phenotypic variation explained by individual QTL ranged from 5.4 to 17.3 %.

## Segregation distortion analysis

We attempted to use the high-density markers data to determine segregation distortion locus and regions and to identify genetic factors affecting segregation distortion in the two NAM families. Within CN-NAM families, 17 % of the markers exhibited segregation distortion at $P < 0.05$, 9 % at $P < 0.01$, and 4 % at $P < 0.001$; within US-NAM families, 18 % of the markers exhibited segregation distortion at $P < 0.05$, 8 % at $P < 0.01$, and 4 % at $P < 0.001$ (Fig. 2). Segregation distortion regions (SDRs) varied among families and chromosomes. A total of 445 significant SDRs ($P < 0.05$) were found within the 36 families, ranging from 6 to 20, with the averaged 12 SDRs per family (Online Resource 3). No SDR was detected in all families. However, 15 common SDRs were found in at least ten families; one common SDR in particular was located in the 251.440–254.559 Mb region on chromosome 1 in 17 families (Online Resource 4). A portion of the significant SDRs could be explained by known genetic factors (Fig. 2). For example, five gametophytic factors (*ga*) genes have been identified in maize (http://www.maizegdb.org/), four of which could correspond to the four most significant SDRs in some families. *Gametophytic male sterile-1* (*gams1*) (SariGorla et al. 1996), which caused segregation distortion on the short arm of chromosome 2, coincided with the two most significant SDRs in the WEIFENG322 ($P = 5.2 \times 10^{-17}$) and in the MO17 family ($P = 2.6 \times 10^{-5}$), respectively. *Sugary1* (*su1*) (Nass and Crane 1970), which reduced germination vigor, perhaps causes the most significant distortion on chromosome 4 in the sweet corn Il14H family and a modest distortion in the P39 family.

Segregation distortion was further analyzed through using genetic markers of the composite genetic maps across the CN-NAM families and the US-NAM families, respectively. CN-NAM and US-NAM had 40 and 57 % markers exhibiting segregation distortion at $P < 0.05$, respectively (Fig. 3).

When considering the relationship between the segregation distortion ratio and recombination rate and gene

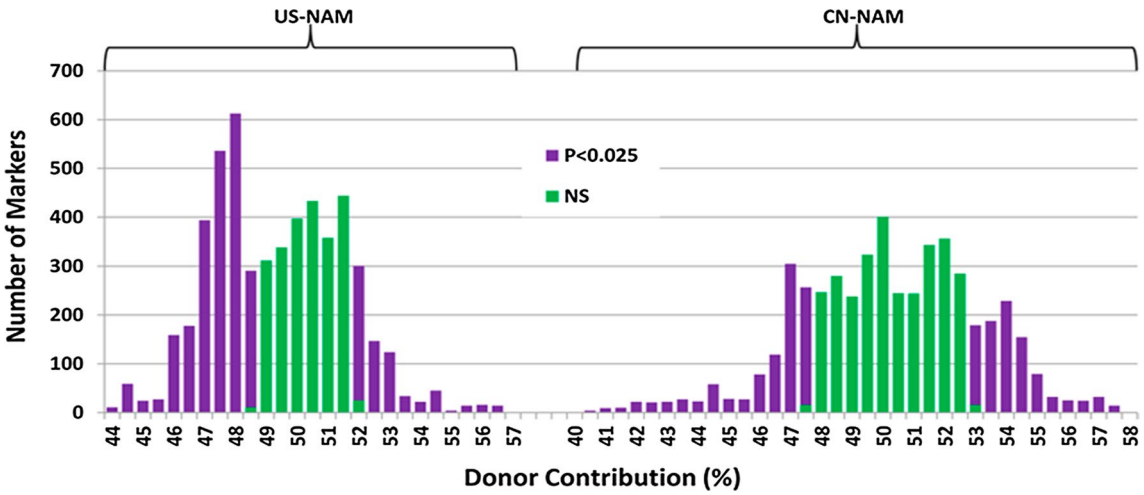


**Fig. 3** Distribution of donor contribution for the 5296 markers in US-NAM and the 4932 markers in CN-NAM. Donor contribution represents genome proportion from the common parents B73 allele or HZS allele. Markers show significant segregation distortion at one-tailed $P < 0.025$. *NS* represents markers show no significant segregation distortion

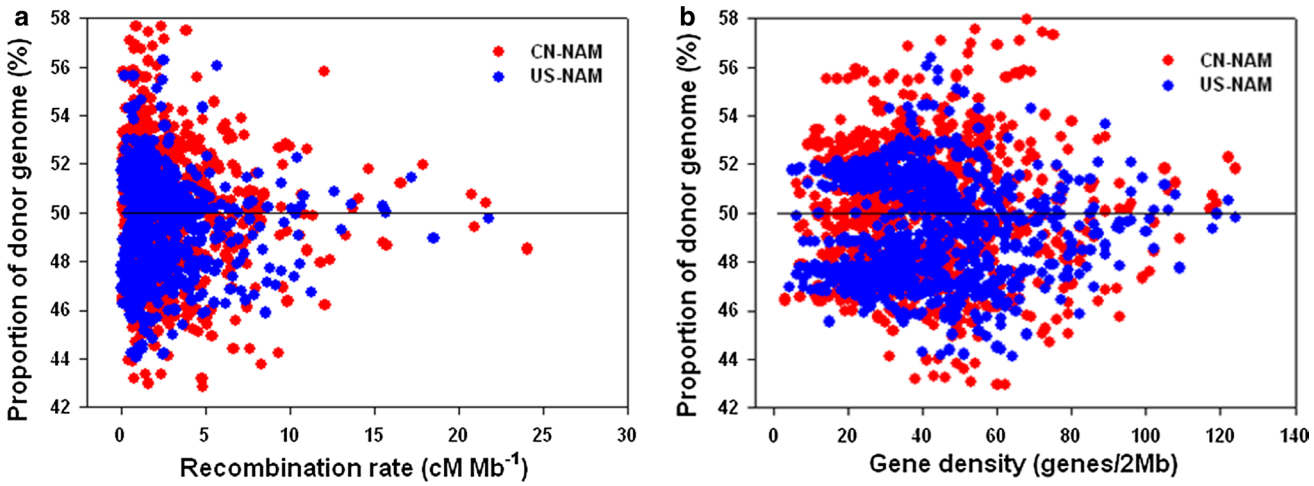


**Fig. 4** The relationship between segregation distortion and (**a**) recombination rate and (**b**) gene density on the whole genome. *Y*-axis represents the proportion of donor parent genome (HZS in CN-NAM and B73 in US-NAM). *Horizontal black lines* in A and B represent that the proportion of donor parent genome is 50 %. Recombination rate was calculated by dividing the genetic distance (cM) between two markers by the physical distance (Mb). When comparing the average recombination rate of two NAM populations, a 2-cM window size was used to calculate the average recombination rate. Gene density was counted by a 2-Mb window size

density, we found that the regions with high recombination rate and high gene density usually tended to have little segregation distortion in the two NAM populations (Fig. 4).

## Epistasis for fitness

Two-locus linkage disequilibrium (LD) was estimated for all pairs of markers within each of the 36 NAM families. A total of 53.8 million tests were performed for the mapped markers on separate chromosomes within the 36 families, the highest $r^2$ value being only 0.18, which was less than the expected highest $r^2$ value 0.22 obtained by randomly shuffling 100 times for each marker (Fig. 5). However, when we detected genotype ratio distortion (GRD) as a sign of epistasis (Corbett-Detig et al. 2013), seven instances of inter-chromosomal GRDs were found in the two NAM families (Table 2). Through the comparison of two statistical methods, we found that GRD method was more effective for detecting epistasis for fitness, and the epistasis between two-loci on different chromosomes is present in the two NAM populations.

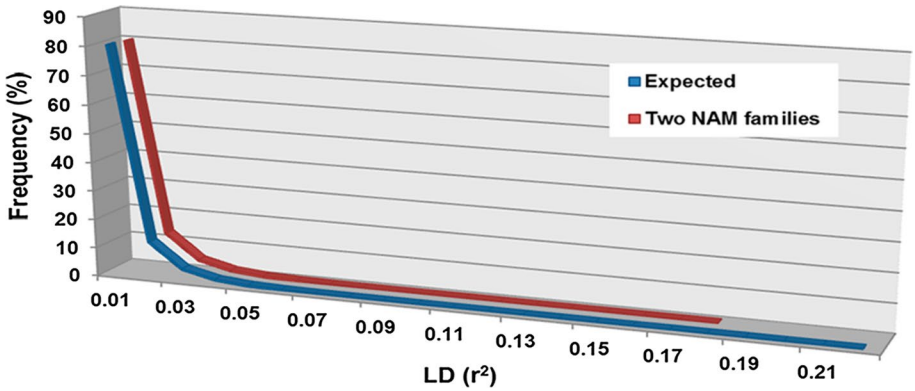


**Fig. 5** The frequency of inter-chromosomal linkage disequilibrium (LD) within the 36 families. The frequency distribution of LD based on pairs of mapped markers (*red*) and expected markers (*blue*) on different chromosomes within each of the 36 NAM families. Mapped markers represent genetic markers of linkage map for each NAM family. Expected markers were obtained by randomly shuffling 100 times for each mapped marker. Allele frequency of mapped markers and expected markers were fixed

**Table 2** List of significant inter-chromosomal genotype ratio distortion (GRD) within the 36 families

| Family | Marker 1[a] | Chr 1 | Position 1[b] | Marker 2[a] | Chr 2 | Position 2[b] | Number of RIL counted | Chi square | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| CML247 | Bin0567 | 1 | 231.419 | Bin4124 | 8 | 15.765 | 193 | 28.46 | 9.57E−08 |
| Ki11 | Bin1840 | 3 | 189.573 | Bin4493 | 8 | 170.787 | 189 | 29.87 | 4.61E−08 |
| Tx303 | Bin1774 | 3 | 174.763 | Bin4056 | 7 | 172.646 | 180 | 25.09 | 5.47E−07 |
| QI319 | Bin2776 | 5 | 198.798 | Bin3190 | 6 | 147.888 | 143 | 26.94 | 2.09E−07 |
| QI319 | Bin0078 | 1 | 14.098 | Bin3866 | 8 | 24.433 | 143 | 27.99 | 1.22E−07 |
| WEIFENG322 | Bin3936 | 8 | 102.620 | Bin4633 | 10 | 8.825 | 151 | 31.60 | 1.90E−08 |
| HUANGYESI3 | Bin0320 | 1 | 121.599 | Bin4509 | 9 | 144.969 | 160 | 24.49 | 7.48E−07 |

[a] A pair of markers showing significant two-locus GRD

[b] Physical positions in RefGen_v2 of markers with significant two-locus GRD

## Discussion

The genomic properties analysis of the US-NAM population argued for little segregation distortion and no epistasis for fitness (McMullen et al. 2009). In this study, CN-NAM provides a direct evaluation for this argument, since the common parents of the two NAM populations were chosen on the basis of a similar strategy. HZS and B73 were very important and widely deployed elite inbred lines in the history of maize breeding in China and the United States, respectively. The 11 diverse parents in CN-NAM were representative members in the Chinese heterotic groups (Li and Wang 2010), and the 25 different parents in US-NAM were chosen to maximally capture the genetic diversity of 302 maize inbred lines from around the world (McMullen et al. 2009). Since the two sets of NAM were constructed based on different germplasm sets, the comparisons of genetic features of the two NAM populations help validate the conclusion obtained in US-NAM, and further understand maize genetic diversity.

In the present research, we used the total number of recombination events as a quantitative trait to conduct QTL mapping. Although there were few reports for genetic dissection of this new trait, a minimum of one obligatory recombination per chromosome arm or chromosome occurs during meiosis as a requirement for proper chromosome segregation (Pardo-manuel de villena and Sapienza 2001). RIL populations were proper tools for QTL mapping for entire genome recombination frequency (Esch et al. 2007). The $F_2$ individuals themselves are not informative for the recombination variation, because the gametes that generate an $F_2$ individual were derived from the same $F_1$ individual, and therefore, all $F_2$ individuals experienced the same recombination factors. However, when $F_2$ individuals were continually selfed or intercrossing to generate RIL lines, the genetic differences on recombination frequency were segregating and fixed in individual RIL (Esch et al. 2007). Thus, the number of recombination events for RIL individuals was accumulated and had different recombination frequency among different RILs.

We detected 14 QTLs controlling the total number of recombination events in 36 RIL populations. The number of QTL detected in the single population was relatively low. Several limitation factors may explain the low power of our approach. First, less than linkage disequilibrium is present between the recombination QTL/gene (s) and molecular markers, since the number of recombination events is calculated after several generations of selfing. Second, not all recombination events are scorable, due to the proportion of scorable recombination events rapidly reduced during the successive inbreeding (Bauer et al. 2013). Third, differences between male and female recombination frequencies were not reflected in the number of recombination events. Although our approach had low power, we still detected QTL with large genetic effects explaining more than 17 % of total phenotypic variation. Moreover, two QTLs on chromosome 3 were located in bin 3.03 of the K12 family and in bin 3.05 of the M37 W family. Esch et al. (2007) also used the intermated B73 × Mo17 (IBM) population to identify two QTLs for the recombination events in bin 3.03 and bin 3.05. These genomic regions seem to be very important for recombination variation under different genetic backgrounds.

With the availability of complete genome sequence and thousands of molecular markers, an increasing number of studies have addressed the underlying mechanisms that are responsible for the difference of recombination frequency in plants, especially *Arabidopsis*. Two general classes of genetic factors, *cis* and *trans*, contribute to variation in the distribution of recombination events (Li et al. 2007a). *Cis*-acting factors are closely linked to the intervals where they modify recombination frequency. For example, some reported genes including *PHS1*, *RAD50*, and *RAD51* are *Cis*-acting to affect recombination frequency and the distribution of recombination breakpoints in maize (Pawlowski et al. 2003, 2004; Li et al. 2007b). *Trans*-acting factors, including chromatin remodelers, recombination machinery protein, and other functional proteins, are opposite to *cis*-acting factors. In the present study, the majority of the QTLs detected appeared to be *trans*-acting factors. Two QTLs on chromosome 1 in the ZHENG58 family and K12 family approached to two candidate genes GRMZM2G114707 and GRMZM2G001869, respectively. The QTL on chromosome 2 in the QI319 family was close to the position of the candidate gene AC235011.1_FG012. In addition, the QTL on chromosome 9 in the CML69 family was close to the position of the candidate gene GRMZM2G035417. These candidate genes are predicted to be involved in homologous recombination and non-homologous end-joining in the process of repairing double-stranded breaks. Therefore, *trans*-acting proteins might play important role in regulating the global number of recombination events in maize.

Segregation distortion, common in segregating populations, can be caused by gametic or zygotic selection due to a number of physiological or nuclear genetic factors (Liu and Qu 2008; Matsushita et al. 2003). Previous researches have suggested that the proportion of distorted markers was high in individual maize mapping population (Casa et al. 2000; Falque et al. 2005; Sharopova et al. 2002). In this study, segregation distortion could be observed in the 36 families, but the proportion of the markers exhibiting segregation distortion was low with the average of 17 % in the CN-NAM families and 18 % in the US-NAM families at $P < 0.05$. This value was similar to that identified using low-density markers in the US-NAM families (McMullen et al. 2009). Therefore, we confirmed that segregation distortion was little within individual NAM families. When analyzing segregation distortion across the CN-NAM families and the US-NAM families, we found that 40 % of the markers in the CN-NAM population and 57 % in the US-NAM population exhibited segregation distortion at $P < 0.05$. These proportions were higher than the average in individual NAM families. These distorted markers in the CN-/US-NAM population maybe show consistent segregation distortion favoring or disfavoring the common parent HZS/B73 allele compared with diverse parental alleles. Distorted markers, unevenly distributed on all chromosomes, were clustered into some specific regions. Only some of the most significant SDRs could be overlapped with the known gametophytic factors genes, which suggested that there might be other unidentified factors responsible for segregation distortion.

In the present study, we found that the regions with more segregation distortion generally have low recombination rate. One possible reason is the expression of genetic load, conserved both in common parents and other parents, which tends to accumulate in low recombination regions, via lethal recessive alleles (Chia et al. 2012). Although maize was domesticated ~10,000 years from its wild progenitor, teosinte (Hufford et al. 2012), the deliberate cultivation of maize inbreds was only started from the last century. Hence, a great number of recessive deleterious alleles are thought to reside within the maize genome. These recessive deleterious mutations in fitness genes or loci closely linked to fitness genes can lead to a reduction in the frequencies of recombination in the process of selfing. Reduced recombination leads to an accumulation of the same alleles in the low recombination regions. Finally, a high level of segregation distortion is expected to detect in the low recombination regions.

Epistasis for fitness was observed in *Arabidopsis* (Malmberg et al. 2005) and rice (Mei et al. 2005). Preferred allele combination may be observed in maize. We used high-density makers with none of heterozygous sites and different statistical methods to reanalyze the epistasis effects of

the US-NAM families, and used the CN-NAM families to validate the results of the US-NAM families. The presence or absence of linkage disequilibrium (LD) has been used to indicate the presence or absence of epistasis for fitness (Otto and Feldman 1997; Felsenstein 1965). When surveying two-locus LD for all pairs of markers on separate chromosomes in each family for both CN-NAM and US-NAM, we found no evidence for particular combinations of two-locus alleles being selected for or against, which was consistent with the results obtained by low-density marker data including heterozygous sites for US-NAM (McMullen et al. 2009). Corbett-Detig et al. (2013) have developed a new approach with high statistical power for detecting genotype ratio distortion (GRD) as a signature of fitness epistasis. In addition, they found that GRD is present in the US-NAM families using low-density marker data including heterozygous sites. When using the same method on the basis of high-density marker data to detect GRD, we also found that GRD is present in both of the two NAM populations. All of GRDs detected by high-density markers were consistent with the results obtained using a low density markers in the US-NAM families. Hence, these results further supported that epistasis for fitness is present in both NAM populations.

**Author contribution statement** CL analyzed the data and drafted the manuscript. YL participated in the study design and analyzed the data. YS, YS, and DZ provided genotype information. ESB and ZZ provided valuable research ideas. YL and TW conceived of the study, managed the project design and coordination, collected data, and helped to draft the manuscript.

**Acknowledgments** This work was partly supported by the Ministry of Science and Technology of China (2011DFA30450, 2011CB100100), China National Natural Science Foundation (91335206), CAAS (Innovation Program), US National Science Foundation under the Plant Genome Research Program (PGRP) (Grant Nos DBI-0820619 and IOS-1238014) and the Basic Research to Enable Agricultural Development (BREAD) project (ID: IOS-0965342), as well as by the USDA-ARS.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflicts of interest.

## References

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W, Altmann T, Flament P, Melchinger AE, Menz M, Moreno-Gonzalez J, Ouzunova M, Revilla P, Charcosset A, Martin OC, Schon CC (2013) Intraspecific variation of recombination rate in maize. Genome Biol 14(9):R103

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D (2007) Autoimmune response as a mechanism for a Dobzhansky–Muller-type incompatibility syndrome in plants. PLoS Biol 5(9):1962–1972

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7):889–890

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Curr Opin Plant Biol 9(2):172–176

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) The MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97(18):10083–10089

Charlesworth B (1990) Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet Res 55(3):199–221

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai JS, Li M, Liu X, Lu YL, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhajarvi T, Rong TZ, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang ZW, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang GY, Xu YB, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44(7):803–807

Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. Nature 504(7478):135–137

de Pardo-manuel villena F, Sapienza C (2001) Recombination is proportional to the number of chromosomes arms in mammals. Mamm Genome 12:318–322

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379

Esch E, Szymaniak JM, Yates H, Pawlowski WP, Buckler ES (2007) Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. Genetics 177(3):1851–1858

Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A (2005) Linkage mapping of 1454 new maize candidate gene Loci. Genetics 170(4):1957–1966

Farkhari M, Lu YL, Shah T, Zhang SH, Naghavi MR, Rong TZ, Xu YB (2011) Recombination frequency variation in maize as revealed by genomewide single-nucleotide polymorphisms. Plant Breed 130(5):533–539

Felsenstein J (1965) The effect of linkage on directional selection. Genetics 52:349–363

Fu Y, Wen TJ, Ronin YI, Chen HD, Guo L, Mester DI, Yang YJ, Lee M, Korol AB, Ashlock DA, Schnable PS (2006) Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics 174(3):1671–1683

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9(2):e90346

Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44(7):812–815

Kimura M, Maruyama T (1966) The mutational load with epistatic gene interactions in fitness. Genetics 54(6):1337–1351

Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336(6198):435–440

Lai JS, Li RQ, Xu X, Jin WW, Xu ML, Zhao HN, Xiang ZK, Song WB, Ying K, Zhang M, Jiao YP, Ni PX, Zhang JG, Li D, Guo XS, Ye KX, Jian M, Wang B, Zheng HS, Liang HQ, Zhang XQ, Wang SC, Chen SJ, Li JS, Fu Y, Springer NM, Yang HM, Wang JA, Dai JR, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42(11):1027–1030

Li Y, Wang TY (2010) Germplasm base of maize breeding in China and formation of foundation parents. J Maize Sci 18(5):1–8

Li J, Hsia AP, Schnable PS (2007a) Recent advances in plant recombination. Curr Opin Plant Biol 10(2):131–135

Li J, Harper LC, Golubovskaya I, Wang CR, Weber D, Meeley RB, McElver J, Bowen B, Cande WZ, Schnable PS (2007b) Functional analysis of maize RAD51 in meiosis and double-strand break repair. Genetics 176:1469–1482

Li L, Li Q, Wang LB, Zhang ZX, Li JS, Yan JB (2009) Genetic analysis of QTL affecting recombination frequency in whole genome of maize and rice. Sci Agric Sin 42:2262–2270

Li CH, Li YX, Sun BC, Peng B, Liu C, Liu ZZ, Yang ZZ, Li QC, Tan WW, Zhang Y, Wang D, Shi YS, Song YC, Wang TY, Li Y (2013) Quantitative trait loci mapping for yield components and kernel-related traits in multiple connected RIL populations in maize. Euphytica 193(3):303–316

Li CH, Li YX, Bradbury PJ, Wu X, Shi YS, Song YC, Zhang DF, Rodgers-Melnick E, Buckler ES, Zhang ZW, Li Y, Wang TY (2015) Construction of high-quality recombination maps with low-coverage genomic sequencing for joint linkage analysis in maize. BMC Biol 13:78

Liu JJ, Qu LJ (2008) Meiotic and mitotic cell cycle mutants involved in gametophyte development in *Arabidopsis*. Mol Plant 1(4):564–574

Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105(4):622–628

Mallet J (2001) Epistasis and the evolutionary process. Science 291(5504):602

Malmberg RL, Held S, Waits A, Mauricio R (2005) Epistasis for fitness-related quantitative traits in arabidopsis thaliana grown in the field and in the greenhouse. Genetics 171(4):2013–2027

Mangelsdorf PC, Jones DF (1926) The expression of mendelian factors in the gametophyte of maize. Genetics 11(5):423–455

Matsushita S, Iseki T, Fukuta Y, Araki E, Kobayashi S, Osaki M, Yamagishi M (2003) Characterization of segregation distortion on chromosome 3 induced in wide hybridization between indica and japonica type rice varieties. Euphytica 134(1):27–32

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2005) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110(4):649–659

Nass H, Crane P (1970) Effect of endosperm mutants on germination and early seedling growth rate in maize (*Zea mays* L.). Crop Sci 10(2):139–140

Otto SP, Feldman MW (1997) Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor Popul Biol 51(2):134–147

Pan QC, Ali F, Yang XH, Li JS, Yan JB (2012) Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. PLoS One 7(12):e52777

Pawlowski WP, Golubovskaya IN, Cande WZ (2003) Altered nuclear distribution of recombination protein RAD51 in maize mutants suggests the involvement of RAD51 in meiotic homology recognition. Plant Cell 15:1807–1816

Pawlowski WP, Golubovskaya IN, Timofejeva L, Meeley RB, Sheridan WF, Cande WZ (2004) Coordination of meiotic recombination, pairing, and synapsis by PHS1. Science 303:89–92

Payseur BA, Hoekstra HE (2005) Signatures of reproductive isolation in patterns of single nucleotide diversity across inbred strains of mice. Genetics 171(4):1905–1916

Phillips PC (2008) Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9(11):855–867

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575

SariGorla M, Ferrario S, Villa M, Pe ME (1996) *gaMS-1*: a gametophytic male sterile mutant in maize. Sex Plant Reprod 9(4):216–220

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH Jr (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48(5–6):463–481

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica 117(1):27–35

Tulsieram LCW, Morris R, Thomas-Compton M, Eskridge K (1992) Analysis of genetic recombination in maize populations using molecular markers. Theor Appl Genet 84(1):65–72

Wade MJ, Winther RG, Agrawal AF, Goodnight CJ (2001) Alternative definitions of epistasis: dependence and interaction. Trends Ecol Evol 16(9):498–504

Wang J LH, Zhang L, Li C, Meng L (2012) QTL ICIMapping software. Isbreeding, Beijing. http://www.isbreeding.net. Accessed 1 July 2016

Xie WB, Feng Q, Yu HH, Huang XH, Zhao Q, Xing YZ, Yu SB, Han B, Zhang QF (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA 107(23):10578–10583

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551