New Phytologist

Research

Genome-wide recombination dynamics are associated with phenotypic variation in maize

Qingchun Pan[1]*, Lin Li[1,2]*, Xiaohong Yang[3], Hao Tong[1], Shutu Xu[3], Zhigang Li[3], Weiya Li[3], Gary J. Muehlbauer[2,4], Jiansheng Li[3] and Jianbing Yan[1]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; [2]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA; [3]National Maize Improvement Center of China, China Agricultural University, Beijing 100193, China; [4]Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108, USA

Author for correspondence:
Jianbing Yan
Tel: +86 027 87280110
Email: yjianbing@mail.hzau.edu.cn

Received: *26 August 2015*
Accepted: *11 November 2015*

New Phytologist (2015)
doi: 10.1111/nph.13810

Key words: intergenic recombination, intragenic recombination, maize (*Zea mays*), phenotypic variation, recombination.

Summary

- Meiotic recombination is a major driver of genetic diversity, species evolution, and agricultural improvement. Thus, an understanding of the genetic recombination landscape across the maize (*Zea mays*) genome will provide insight and tools for further study of maize evolution and improvement.
- Here, we used *c*. 50 000 single nucleotide polymorphisms to precisely map recombination events in 12 artificial maize segregating populations. We observed substantial variation in the recombination frequency and distribution along the ten maize chromosomes among the 12 populations and identified 143 recombination hot regions.
- Recombination breakpoints were partitioned into intragenic and intergenic events. Interestingly, an increase in the number of genes containing recombination events was accompanied by a decrease in the number of recombination events per gene. This kept the overall number of intragenic recombination events nearly invariable in a given population, suggesting that the recombination variation observed among populations was largely attributed to intergenic recombination. However, significant associations between intragenic recombination events and variation in gene expression and agronomic traits were observed, suggesting potential roles for intragenic recombination in plant phenotypic diversity.
- Our results provide a comprehensive view of the maize recombination landscape, and show an association between recombination, gene expression and phenotypic variation, which may enhance crop genetic improvement.

Introduction

Recombination refers to the phenomenon of genomic exchange among chromatids, which includes crossover and noncrossover events in meiotic prophase I, leading to new alleles and new combinations of existing alleles (Villeneuve & Hillers, 2001; Hamant *et al.*, 2006; Mezard *et al.*, 2007). Together with DNA mutations, meiotic recombination is a key driving force in genome evolution and can enhance the genetic diversity of species (Zhang & Gaut, 2003; Meunier & Duret, 2004; Gaut *et al.*, 2007; Li *et al.*, 2007a,b; Kulathinal *et al.*, 2008; Kent *et al.*, 2012). In crops, recombination contributes substantially to breeding through the rearrangement of polymorphic sites, which leads to the opportunity to select new and improved allele combinations as well as the elimination of deleterious mutations (Wijnker & de Jong, 2008; Martin & Wagner, 2009).

Recombination can be measured as the ratio of the genetic distance in centimorgans to the physical genome size in megabase pairs (cM/Mb). It is an effective indicator of recombination both at the genome-wide scale, as the genome-wide recombination ratio (GWRR), and at the specific locus level, as the locus recombination ratio (LRR) (Dumont *et al.*, 2011). Recombination events are not uniformly distributed along the chromosomes, and recombination hot regions are frequently observed (Yao & Schnable, 2005; Baudat & De Massy, 2007; Smagulova *et al.*, 2011). In maize (*Zea mays*), a putative recombination hot region at the Bronze locus and anthocyanin 1 gene locus were reported to lie in an unusually gene-rich region (Fu *et al.*, 2002; Yao *et al.*, 2002). In addition to *cis*-acting genomic features that affect recombination frequency, the *POOR HOMOLOGOUS SYNAPSIS1* (*PHS1*), *RAD50* and *RAD51* genes were reported to affect recombination variation at the whole-genome scale, indicating that recombination frequency is under genetic control (Pawlowski, 2003; Pawlowski *et al.*, 2004; Li *et al.*, 2007a,b).

Ultra-high-density genotyping of large populations is expected to reveal new features and refine our understanding of recombination at the whole-genome scale (Yao *et al.*, 2002; Yang *et al.*, 2012). Rodgers-Melnick and colleagues employed genotyping-by-sequencing on 25 US nested association mapping (NAM) maize populations and nine Chinese recombinant inbred line

*These authors contributed equally to this work.

(RIL) populations, constructed ultra-high-density genetic maps and showed that recombination was stable and predictable, with some variations across different populations (Rodgers-Melnick *et al.*, 2015). As recombination is most likely to occur in the homologous chromatid regions (Haber *et al.*, 2004), the observation of stable recombination might be attributable to the common parent used in the population design (Rodgers-Melnick *et al.*, 2015). In another study, Bauer *et al.* (2013) genotyped two intermated RIL and 23 double haploid (DH) populations with a MaizeSNP50 array (Ganal *et al.*, 2011) and found that recombination in different maize subgroups exhibited different genome-wide patterns, suggesting intraspecific variation for recombination. However, fundamental questions concerning the kind of recombination (intergenic or intragenic) responsible for intraspecific variation and the potential phenotypic consequences of recombination are unanswered. Previous studies indicated that recombination tends to occur in intragenic regions (Fu *et al.*, 2002; Yao *et al.*, 2002). Meanwhile, several studies have refined the location of recombination and proposed that it is more likely to occur in the 5′ and 3′ untranslated regions (UTRs) of genes in humans, mice, and Arabidopsis (Myers *et al.*, 2005; Brick *et al.*, 2012; Choi *et al.*, 2013). Given the strong enrichment of recombination events within genes, we sought to examine the role that intragenic and intergenic recombination plays in determining the intraspecific recombination variation observed across different maize populations. We used the maize SNP50 array to construct high-resolution linkage maps of 12 segregating populations. We identified variation for recombination between populations with some consensus recombination hot regions and many population-specific hot regions. In addition, we partitioned the recombination events into intragenic and intergenic patterns. The comparison of intragenic and intergenic recombination across different populations and the haplotype-based association mapping analysis indicated that intergenic recombination contributes substantially to the intraspecific recombination variation, while intragenic recombination significantly correlates with gene expression levels and phenotypic variation.

Materials and Methods

Mapping populations

A total of 12 mapping populations, including one F_{10} RIL (ZONG3/YU87-1), one F_9 RIL (B73/BY804), nine F_6 RILs (BY815/KUI3, DAN340/K22, DE3/BY815, K22/CI7, K22/BY815, KUI3/B77, KUI3/SC55, YU87-1/BK and ZHENG58/SK), and one BC_2F_5 (MO17/X26-4), were derived from crosses among 16 maize inbred lines (B73, BY804, K22, CI7, D340, By815, De3, KUI3, B77, SC55, ZONG3, YU87-1, ZHENG58, SK, BK, and MO17) and one teosinte entry (Teo; X26-4; accession number: PI 566686; *Zea mays* ssp. *mexicana*) (Supporting Information Table S1). Among the 16 maize parents, SK is an inbred line selfed from a landrace in Peru, and BK is an inbred line selfed from a landrace in the USA, whereas the others are from a diverse maize-association panel that was previously described (Yang *et al.*, 2010). Each of the 11 RIL populations was derived from a single F_1 plant and was developed through self-pollination and single seed descent for at least six generations. The BC_2F_5 population was derived from a single F_1 individual from a cross between Mo17 and a wild teosinte (*Z. mays* ssp. *mexicana*, 566686). The single F_1 individual was then backcrossed with Mo17 twice, followed by self-pollination for five generations.

Genotyping and construction of genetic linkage maps

All 2478 lines in the 12 mapping populations, together with their parents, were genotyped using the Illumina MaizeSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), which contains 56 110 single nucleotide polymorphisms (SNPs) (Ganal *et al.*, 2011). SNP genotyping was performed on the Illumina Infinium SNP genotyping platform at the Cornell University Life Sciences Core Laboratories Center or at DuPont Pioneer Co. (Wilmington, DE, USA).

In each population, we used in-house Perl scripts to compare the genotypes between the RILs and their parents and calculated missing data, minor allele frequency and heterozygosity for each SNP, as well as the missing data and heterozygosity for each line. After quality control, a total of 2319 lines (Table S2) with missing data of $< 15.0\%$ and heterozygosity of $< 8.0\%$ were used for subsequent analyses. The SNPs with missing data of $< 10.0\%$ and heterozygosity of $< 10.0\%$ were used to construct the genetic linkage map.

We used a modified physical order method (Matise *et al.*, 2007) that integrated Carthagene software (De Givry *et al.*, 2005) in a Linux system with in-house Perl scripts to construct the genetic linkage map. First, polymorphic markers between two parents were selected and grouped. To form a group, the distance between the nearest two markers must be < 10 cM and the logarithm of odds value must be > 8. The 'Kosambi' map function was used to estimate the genetic distance between the two nearest markers. The polymorphic markers with identical genotypes in each group were merged using the 'markmerge' command. Second, each polymorphic marker was mapped *in silico* to the B73 reference genome v.2, and a draft map was obtained. Third, to test the marker order accuracy of the draft map, all of the markers were extracted (one marker each time) from the data set and placed in the retained data set using the 'buildfw' command, which is similar to the 'try' function in Mapmaker (Lander *et al.*, 1987), and joined using Perl scripts to determine if the marker would map back to the same position. Only the markers that could be remapped to the same positions were retained, and the new linkage map was constructed with fewer, but higher quality markers. This process was repeated three times to assure the authenticity of our results. Fourth, the unmapped markers and noncoordinated markers were remapped to the new map with the command 'buildfw'. Finally, we ran permutations of the order of neighboring markers and compared the likelihoods of the resulting maps using the 'flips' command, which is similar to 'ripple' in Mapmaker, with a window size of seven markers, to construct the final linkage map for each individual population (Fig. S1). The Perl script combined the Carthagene software's

construction linkage map code, raw genotypic data, and 12 constructed linkage maps, which can be downloaded from the websites http://www.maizego.org/Resources.html and https://github.com/panqingchun/linkage_map.

Estimation of recombination events

After constructing the linkage map, recombination events were estimated. First, we imputed the missing data according to the linkage map order. If the marker alleles located before and after the missing genotypic information came from the same parent, then we used the information from that parent to fill the gap. If the marker alleles before and after the missing genotypic information came from different parents, then we filled the gap using the nearest marker, based on genetic distances. After including the missing genotype data, we recalculated the recombination events using the filled linkage map. If two adjacent markers were originally homozygous and became heterozygous, we assumed that 0.5 recombination events had occurred, and if the flanking markers on both sides of the recombination breakpoint were homozygous and from different parents, we assumed that one recombination event had occurred.

Estimation of recombination rate and recombination hot regions

When we analyzed recombination rate, some markers with a physical position on one of the chromosomes were found at a genetic position on another chromosome. If we used such markers as the boundary markers of the analyzed genomic window, inaccurate estimation of local recombination rate would be obtained. Additionally, the estimation of recombination rate (cM/Mb) relies on accurate measurement of both physical and genetic distances in the genome. From the linkage map, we deleted markers with inconsistent orders in both the genetic and physical maps and summarized the recombination as cM/Mb using a 2-Mb window size with a step size of 1 Mb. To identify the recombination hot regions where recombination is more likely to occur than neutral expectation, we performed 1000 permutation tests using the total linkage map length and window size with the null hypothesis that every genomic region has equal chance of recombination. A global permutation threshold as the cut-off for recombination hot regions was obtained at a significance level of 0.05 (Li *et al.*, 2013b).

Identification of intragenic and intergenic recombination events

As both Chinese populations in our current study and 23 European double haploid (DH) populations (Bauer *et al.*, 2013) were genotyped using the same SNP50 array platform, these data sets were used to detect intragenic recombination events in these populations. For our 12 linkage maps, the occurrence of intragenic recombination was detected based on the genotypic variation and physical coordinates for pairs of markers located in the same gene (Fig. S2). First, by comparing the coordinates of polymorphic SNPs and the filtered gene set from the maize reference genome version 2 (AGPv2), we summarized the genes with at least one SNP. Then we obtained the genotype information for all the SNPs from genes with at least two SNPs. Finally, we checked all the genotypic variation in each population for any pairs of SNPs from the same gene. Intragenic recombination events were recorded if genotype alteration relative to the parental genotypes occurred in any pairs of intragenic SNPs. If the genotypes of one pair of genic SNPs were both homozygous and altered compared with the parental genotypes in a specific individual, two recombinations were recorded, while if one SNP was homozygous and one was heterozygous, one recombination was recorded. If there was more than one pair of SNPs located in the genic region, the highest number of recombination events for any of these SNP pairs was used to represent the overall intragenic recombination number for the gene. If the total number of intragenic recombination events was >35 in an entire family, these events were ignored because they might be attributable to the genotyping error by array. The proportion of genes with genic recombination (number of genes with intragenic recombination/number of genes with at least two genic SNPs) and average intragenic recombination (sum of maximum numbers of intragenic recombinations of any pair of genic SNPs per gene for all the genes/number of genes with intragenic recombinations) were computed for each Chinese and European population. SNPs that were not located in the genic regions were considered intergenic SNPs. Intergenic recombination was detected if two intergenic SNPs were in a 10-kb genomic region and had altered haplotypes in the populations. We eliminated nine DH populations (Bauer *et al.*, 2013) that had only a few genes ($<1.0\%$ harboring intragenic recombination events) and used the remaining 14 DH populations in this analysis. Of these 14 DH populations, 12 had a fairly low proportion of genes containing recombination events, whereas only two populations had intragenic recombination in $>20.0\%$ of their genes with at least two SNPs, comparable to the level of the observation in our study.

Association analyses between recombination event and phenotypic variation

To assay the genetic effects of intragenic recombination, we used the genic markers found in two parents and their progeny to classify the haplotypes in an association mapping panel of 508 inbred lines (Yang *et al.*, 2010). If intragenic recombination occurred, then four haplotypes, the two parental haplotypes and two recombination haplotypes, were produced. For example, if the parental haplotypes of two markers were AC and GT, then the recombination haplotypes would be AT and GC. We used the four haplotypes to search the genotypes of the 508 inbred lines and redefined the association population genotypes as 1, 2, 3, and 4 (Fig. S3). Then, we tested the association between the four haplotypes and the phenotypic variation. Population structure and kinship matrices from a previous study (Yang *et al.*, 2014) were also used. To determine whether intragenic recombination was associated with phenotypic variation, we conducted an association analysis using TASSEL software (Yu *et al.*, 2006) with mixed linear

model (MLM) on 21 agronomic traits including plant height (PH), ear height (EH), ear leaf width (ELW), ear leaf length (ELL), tassel main axis length (TMAL), tassel branch number (TBN), leaf number above the ear (LNAE), ear length (EL), ear diameter (ED), cob diameter (CD), kernel row number (KRN), kernel number per row (KNPR), 100 grain weight (GW), cob weight (CW), kernel width (KW), kernel thickness (KT), kernel length (KL), cob color (CC), days to silking (DTS), days to anthesis (DTA), days to heading (DTH) and oil composition in kernels (Yang *et al.*, 2010, 2014; Li *et al.*, 2013a).

To identify genes significantly associated with a trait, we used the $1/n$ and $0.05/n$ (n = total gene number) cut-offs for all 22 traits. If the P value was lower than this threshold, we classified the gene as being significantly associated with the phenotype. The CC phenotypic data of the B73/BY804 and DAN340/K22 populations were measured in Hubei Province, China in 2012. The functional annotations of the genes were downloaded from the maize genome sequence project (ftp://ftp.gramene.org/pub/gramene/maizesequence.org/).

Identification of the relationships between recombination and gene expression variation

To analyze the relationships between intragenic recombination and gene expression, the expression levels of genes containing recombination events in a natural population of 28 679 expressed genes in kernels at 15 d after pollination were obtained (Fu *et al.*, 2013). The expression levels of these genes were divided into different groups based on the recombination haplotype, which we defined in terms of the available markers, and used to correlate the relationship between gene haplotypes and gene expression values. The association analysis was performed using the MLM algorithm of the TASSEL software (Yu *et al.*, 2006). To obtain the P-value cut-off, we used the uniform Bonferroni-corrected thresholds at $a = 1$. The P-value cut-off was obtained by $1/n$, where n is the number of tests. We compared the rates of recombination in genes controlled by *cis*- and *trans*-acting genetic factors, which were detected by expression quantitative trait locus mapping (Fu *et al.*, 2013), with that of randomly selected genes containing recombination events.

Shannon entropy (H_g), a measure of tissue-specific gene expression (Schug *et al.*, 2005), ranges from zero for genes expressed in a single tissue to $\log_2$ (number of tissues) for genes expressed uniformly in all tissues considered. To test for a relationship between recombination events and gene expression levels, we extracted the genes in recombination hot regions and genes containing recombination events, calculated the tissue-specific Shannon entropy (H_g value) (Schug *et al.*, 2005), and compared the H_g values of recombination-related genes with those of randomly selected genes across 10 different tissues and stages of the maize inbred reference line B73.

Monte Carlo resampling tests

To test whether intragenic recombination is significantly associated with gene expression and phenotypic variation in maize, Monte Carlo resampling tests were used to assess the significance of differences in intragenic and intergenic recombination associated with the 22 phenotypic values and the *cis*-acting locus rates of intragenic recombination events between groups. Here, the mean significant intragenic recombination P value and intragenic *cis*-acting locus rate of the test group were compared with the significant intergenic recombination P value and intergenic *cis*-acting locus rate of a reference group of the same size that was randomly selected. For each reference gene, we used the same sample sizes and repeated the process 1000 times. The test group value was larger than that of the reference group, and the 0.95 ratio value indicates that the *cis*-acting locus rates of the test and reference groups are different (Hughes *et al.*, 2014).

Results

Construction of high-density linkage maps in maize

We constructed 12 segregating populations, including 11 RIL populations and one BC_2F_5 backcross population (Table S1). A total of 2478 segregating lines were obtained with a range of 194–217 lines per population (Fig. 1a; Table S2). These 12 populations were genotyped using the maize SNP50 array. Lines that had a high ratio of missing or heterozygous genotypes or were contaminated during the generation of the populations were excluded, and 2319 lines were selected and genotyped using 11 360 to 15 285 polymorphic markers for further analysis (Tables 1, S2). High-resolution genetic maps with an average of 13 064 polymorphic markers and an average length of 1879.3 cM were assembled for the 12 populations. The average genetic distance between two neighboring markers was 0.14 cM (Tables 1, S3).

Genome-wide landscape of recombination indicates genotype-specific and locus-specific variation in maize

We first estimated the average genome-wide recombination ratio (GWRR) over the 12 segregating populations based on the sequenced B73 reference genome (AGPv2). The GWRR ranged from 0.81 cM/Mb in the K22/BY815 population to 1.32 cM/Mb in the ZONG3/YU87-1 population, with an average of 0.91 ± 0.04 cM/Mb (Table 1). A bootstrapping analysis of different sample sizes in each RIL population showed that the recombination events and standard errors were quite stable and similar in each population but distinct between populations (Table S4), indicating that the differences in the GWRRs are population specific. In total, 92 562 unique recombination events were identified in the 12 populations (Table S5). An average of 39.8 recombination events per line was identified for the 12 segregating populations. However, the average recombination event number per plant varied from 27 to 53 across different populations, a nearly two-fold difference (Fig. S4).

To provide a panoramic view of recombination variation, we then applied a 2-Mb sliding window to estimate the locus recombination ratio (LRR) along each chromosome over all populations (Fig. 1b). The LRRs ranged from 0 (no recombination

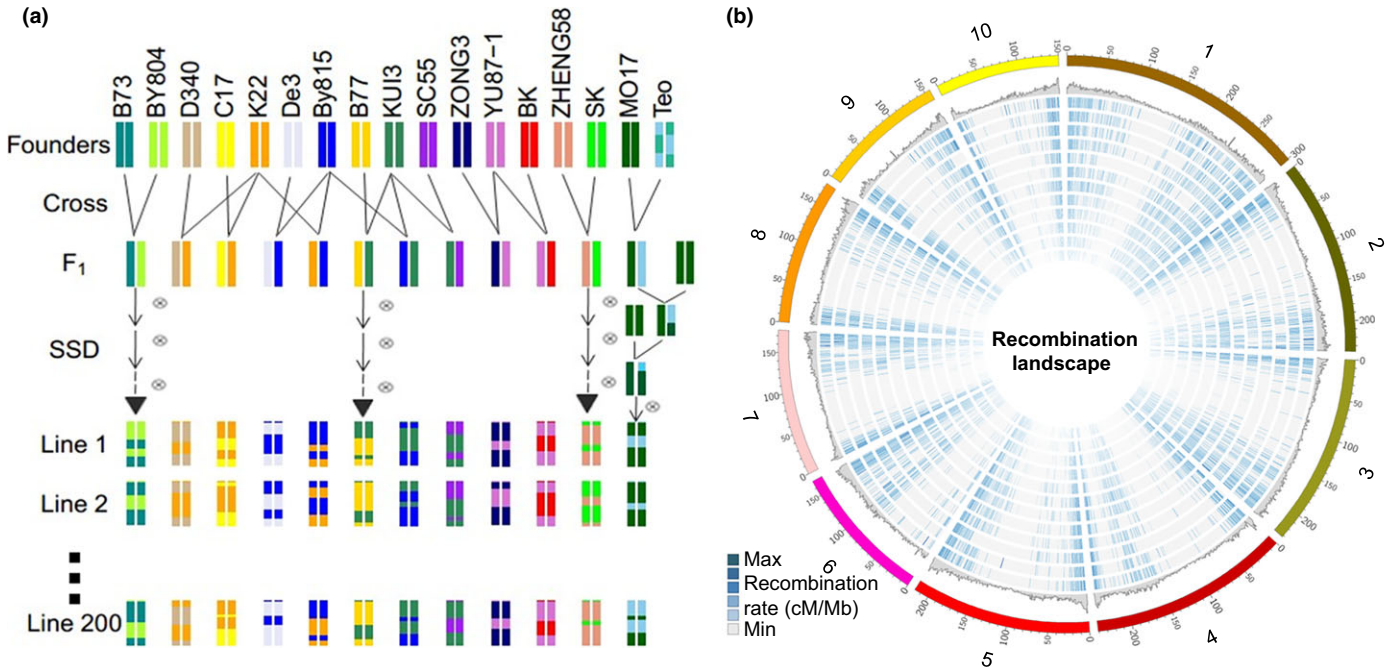

Fig. 1 The recombination landscape in maize. (a) Construction of 12 segregating populations used in this study. SSD, single seed descendent. (b) The genome-wide landscape of recombination (cM/Mb) across 12 segregating populations and the gene density (in gray). Lanes from inner to outer are MO17/X26-4, K22/BY815, DE3/BY815, BY815/KUI3, KUI3/SC55, K22/CI7, ZHENG58/SK, DAN340/K22, B73/BY804, ZONG3/YU87-1, KUI3/B77 and YU87-1/BK.

Table 1 Summary of linkage maps for 12 segregating maize populations

No.	Population	Number of lines	Map length (cM)	Number of unique bins[a]	Total number of markers	Recombination rate (cM/Mb)	Average genetic length (cM)[b]
1	B73/BY804	197	1790.2	2496	15 285	0.871	0.117
2	YU87-1/BK	165	1854.2	1979	11 384	0.902	0.163
3	K22/CI7	196	1719.7	2386	13 433	0.836	0.128
4	DAN340/K22	201	1698.4	2100	11 378	0.826	0.149
5	KUI3/B77	177	1744	2126	11 360	0.848	0.153
6	MO17/X26-4	191	1748	1282	12 390	0.850	0.141
7	ZHENG58/SK	204	1860.9	2486	13 703	0.905	0.135
8	ZONG3/ YU87-1	197	2716	3071	13 759	1.321	0.197
9	DE3/BY815	207	1806.4	2382	13 729	0.879	0.131
10	K22/BY815	207	1670.4	2263	13 603	0.812	0.122
11	BY815/KUI3	180	1984.8	2372	12 725	0.965	0.156
12	KUI3/SC55	197	1958.6	2683	14 024	0.953	0.139
	Average	193	1879.3	2302	13 064	0.914	0.144

[a]The number of bins that contain no recombination events. cM, centimorgan; Mb, megabase, corresponding to *c*. 1 million nucleotides.
[b]Average distance between successive markers.

events) to 15.40 cM/Mb along the maize chromosomes. Generally, pericentromeric regions had lower LRRs, whereas the telomere regions had higher LRRs. The LRR variation matched the variation in gene density along each chromosome (Fig. 1b), indicating strong locus-specific effects on recombination. This is consistent with the existence of recombination hot regions, relative to a neutral expectation. In total, 143 recombination hot regions of 2 Mb size with significantly higher levels of recombination than neutral expectation were identified at the whole-genome level (see the Materials and Methods section). These 143 recombination hot regions were distributed in 26 nonoverlapping genomic regions. The length of nonoverlapping recombination hot regions ranged from 2 to 6 Mb with an average distinct length of 3.46 Mb. Of these 143 recombination hot regions, 109 were identified in more than one population, and nine hot regions were detected consistently in at least five populations (Tables 2, S6), suggesting that genetic factors are involved in the formation of recombination hot regions. Interestingly, 34 recombination hot regions were detected in only one segregating population, thus exhibiting population specificity.

Table 2 Recombination hot regions identified in more than five maize populations

Chromosome	Marker Start position (Mb)	Marker End position (Mb)	Identified populations[a]	Number of SNPs[b]	Known genes[c]
3	1.02	2.94	2, 3, 8, 9, 10	24/25/21/12/23	
3	2.10	3.91	1, 2, 3, 5, 7, 9, 10, 12	28/19/33/18/19/27/27/26	
3	3.02	4.89	1, 2, 7, 9, 10	28/21/17/40/39	
5	1.02	2.92	1, 2, 3, 7, 8, 10	50/17/34/34/33/33	*glu1*, *rps4*
7	4.06	5.97	1, 2, 6, 9, 10, 12	35/15/36/35/28/16	
7	1.14	2.96	3, 6, 8, 9, 12	24/25/28/28/19	
7	2.02	3.82	3, 5, 6, 9, 10	14/6/13/11/9	*rs1*
8	171.01	172.97	1, 4, 6, 7, 10	32/42/24/35/29	*obf3.2*, *psei2*
10	147.00	148.99	1, 2, 6, 8, 9, 10	31/27/21/32/18/29	

[a]Population numbers as shown in Table 1.
[b]The number of single nucleotide polymorphisms (SNPs) in the region for each respective population.
[c]The detail information of these known genes is: *beta glucosidase1* (*glu1*), *ribosomal protein S4* (*rps4*), *rough sheath1* (*rs1*), Ocs-element binding factor 3.2 (*obf3.2*), *cysteine proteinase inhibitor ii* (*psei2*).

Widespread occurrence of intragenic recombination and substantial contribution of intergenic recombination variation in maize

To partition recombination, we first classified the SNPs as genic or intergenic, based on their physical positions relative to genic regions in the Working Gene Set annotation from the maize genome project (Schnable *et al.*, 2009). A majority (*c.* 70.0%) of the segregating SNPs were mapped to genic regions, resulting in 5252 to 6937 genes in the 12 segregating populations (Table S7). We found that 8725 genes (ranging from 1450 to 2151 in each population) contained at least two SNPs and could be used for the identification of intragenic recombination (Fig. 2a). In total, we identified 3030 genes containing recombination events in at least one segregating population, of which 1272 (42.1%) were uncovered in at least two populations. Interestingly, 289 of these 3030 genes that exhibited intragenic recombination were located in the 143 recombination hot regions. One gene (*GRMZM2G060866*) had intragenic recombination events in seven out of the 12 segregating populations. The YU87-1/BK population had the most genes (443) that contained recombination events, whereas the MO17/X26-4 population had the least (197) (Table S8). The proportion of genes with at least two SNPs containing recombination events ranged from 12.1% in the MO17/X26-4 population to 29.2% in the YU87-1/BK population, indicating possible genotype-specific variation. Moreover, for the genes exhibiting intragenic recombination, the average number of recombination events in each population ranged from 2.3 in the ZONG3/YU87-1 population to 6.3 in the MO17/X26-4 population (Fig. 2a).

Surprisingly, the proportion of genes having detectable recombination events was significantly negatively correlated with the average number of intragenic recombination events ($r = -0.74$; $P = 7.01\text{E-}03$), whereas this association was not present in the randomly selected intergenic regions across the 12 segregating populations ($r = 0.01$; $P = 0.96$) (Fig. 2b). However, such a negative correlation implies that the total number of intragenic recombination events tends to be stable given a particular population size. It also indicates that recombination variation across different genetic backgrounds is mainly determined by intergenic recombination rather than intragenic recombination (Fig. 2c). It is noteworthy that the BC_2F_5 population MO17/X26-4 was an outlier for intergenic recombination because of less recombination. However, even after ruling out this population, the positive correlation between intergenic recombination and the overall recombination variation was detected ($r = 0.53$; $P = 0.05$). To further validate the negative correlation, we used previously published genotypic data from 23 DH populations, which were also genotyped using the same maize SNP50 array (Bauer *et al.*, 2013). We consistently uncovered a similar negative correlation between the average number of recombination events in genic regions and the proportion of genes with intragenic recombination ($r = -0.66$; $P = 0.01$) (Fig. S5).

Intragenic recombination is significantly associated with gene expression and phenotypic variation in maize

As described in the previous section, we identified a substantial number of intragenic recombination events, of which 41.9% were localized to the same marker interval in at least two segregating populations (Fig. S6). After these intragenic recombination events, four haplotypes (two original parental haplotypes and two recombinant nonparental haplotypes) could be classified within a segregating population. As the frequency of the nonparental haplotypes within genes is very low in the segregating population, a natural population of 508 inbred lines that had exhibited a wider range of genetic variation was examined. This population was genotyped using the same maize SNP50 array, phenotyped for many agronomic traits (Yang *et al.*, 2010, 2014; Li *et al.*, 2013a), and used to evaluate the genetic effect of intragenic recombination on gene expression and phenotypic variation. In this natural population, the 2763 genes containing recombination events exhibited a higher frequency (> 5%) for each haplotype than in the segregating populations (Fig. S2).

To dissect the relationship between recombination and gene expression, the expression levels of 28 789 genes in whole kernels 15 d after pollination from 368 inbred lines (a subset of the 508 inbred lines), and across 10 different tissues of the

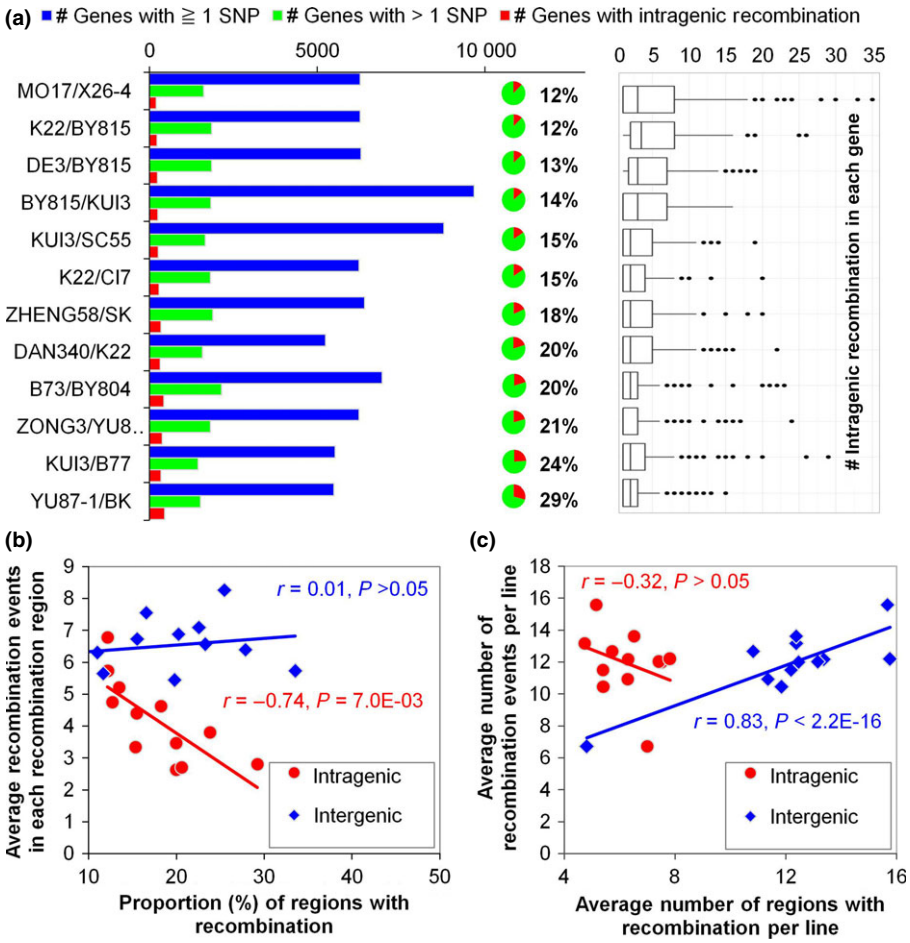

Fig. 2 Intragenic recombination and its comparison with intergenic recombination in maize. (a) The number of genes with intragenic recombination events in each segregating population. The blue, green, and red bars in the left panel represent the number of genes with at least one single nucleotide polymorphism (SNP), the number of genes with more than one SNP, and the number of genes containing intragenic recombination, respectively. The red-green pie charts in the center of the panel indicate the proportion of genes with intragenic recombination. The box plots in the right panel show the number of intragenic recombination events for each detectable gene in the corresponding populations in the left panel. (b) The relationship between the proportion of regions (genic regions in red and intergenic regions in blue) having recombination events and the average number of recombination events in each recombination region across 12 segregating populations. (c) The relationship between the average number of recombination events in each line across 12 populations and the average number of regions with recombination in each line in intragenic (red) and intergenic (blue) regions. The vertical line within the box plots is the median value of the number of intragenic recombination events. The 'dots' at the end of the boxplot represent outliers, which have values larger than the upper inner fence of the box plot.

maize reference inbred line B73 were investigated (Fu *et al.*, 2013; Li *et al.*, 2013b). By comparing the expression variation of genes in B73 located in recombination hot regions, including genes containing recombination events, as well as a set of randomly selected control genes, we found that 2763 genes undergoing intragenic recombination events were more likely to be expressed across the 10 different tissues than control genes ($P = 5.81E-05$; Fig. 3a). In addition, the genes located in recombination hot regions, including those in which intragenic recombination was observed, were more likely to be expressed at a higher level than control genes ($P < 2.20E-16$; Fig. 3b) in nine tissues, but not in pollen. A Shannon entropy analysis (Schug *et al.*, 2005) across different tissues indicated that genes located in recombination hot regions tended to be constitutively expressed ($P < 2.20E-16$; Fig. 3c). In our previous study, we obtained mRNA Sequencing (mRNA-Seq) data from 368 inbred lines from kernels 15 d after pollination, profiled the expression-level variation for 28 789 genes, and performed expression Quantitative Trait Locus (eQTL) mapping to dissect the genetic factors underlying gene expression-level variation (Fu *et al.*, 2013). We used these data to compare the proportion of genes containing recombination events controlled by *cis*-eQTLs with that of randomly selected genes with *cis*-eQTLs. We found that the genes containing recombination events had a higher association (40.8%) with *cis*-eQTLs than randomly selected genes (32.0%) ($P = 1.90E-03$). For the Monte Carlo resampling tests, genes lacking intragenic recombination events had a *cis*-eQTL rate of 35.4%, which was less than the rate of 40.8% for genes possessing recombination events (Fig. S7), indicating that intragenic recombination favored *cis* regulation rather than *trans* regulation (see the Materials and Methods section).

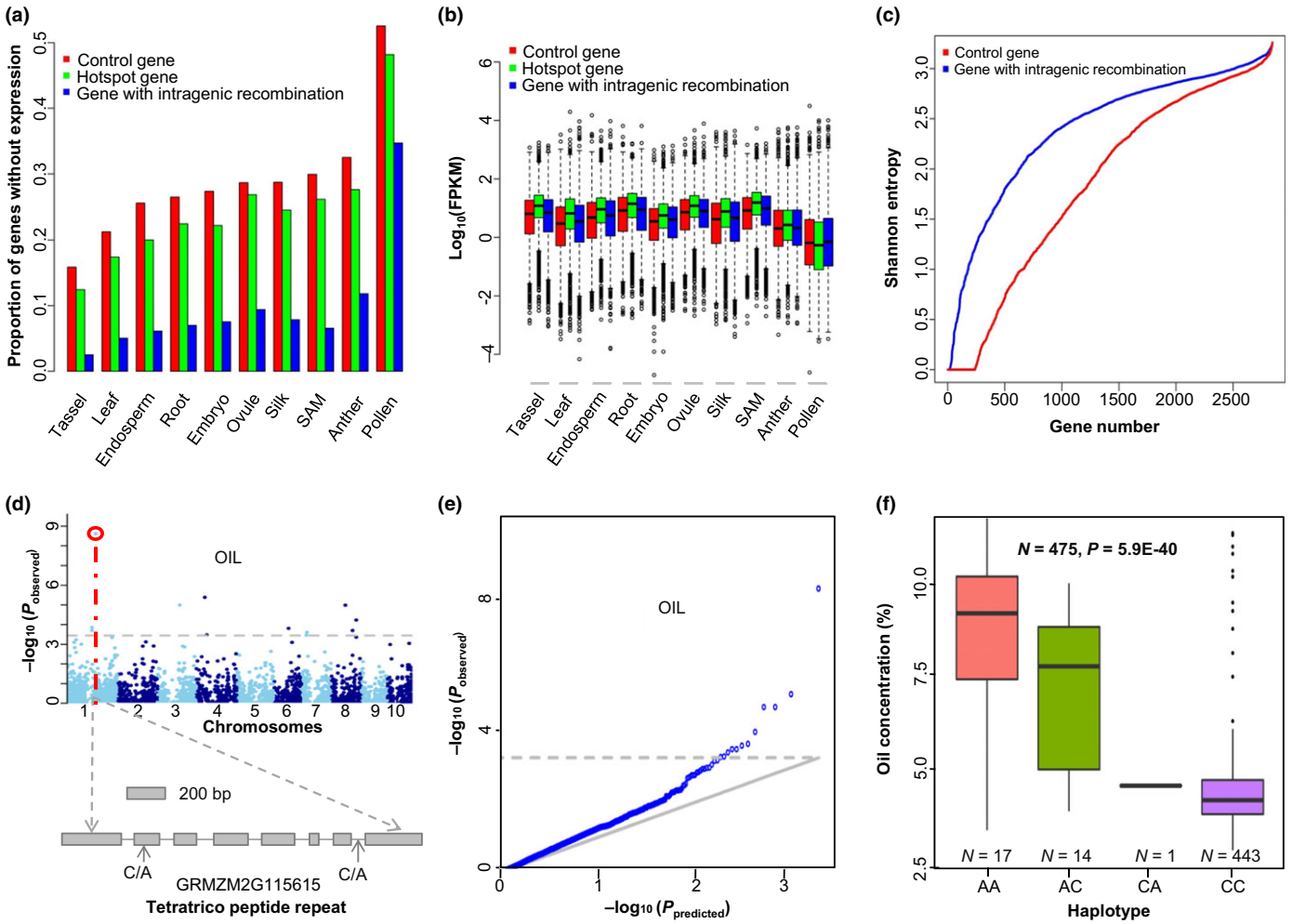

Fig. 3 Intragenic recombination events and their association with maize gene expression and phenotypic variations. (a) The proportion of genes without measurable expression across 10 different tissues in maize. (b) Variations in the expression levels of genes across 10 different tissues in maize. SAM, shoot apical meristem. The red, green and blue boxplots represent the expression level distribution of control genes, hotspot genes and genes with intragenic recombination across different tissues, respectively. The horizontal line in the boxplots represents the median value of the expression level. The black 'dots' in each boxplot are outliers, the values of which are either higher than the upper inner fence or lower than the lower inner fence of the boxplot. (c) The Shannon entropy cumulative curve of genes across 10 different tissues in maize. (d) Manhattan plot of the association results for kernel total oil concentration based on intragenic recombination haplotypes, showing the physical locations of the intragenic recombination haplotypes (*x*-axis), the $-\log_{10}(P$ values) from a mixed linear model (*y*-axis), and the Bonferroni-adjusted significance threshold (dashed horizontal line). The spike on chromosome 1 (red circle) was located near *GRMZM2G115615*. Arrows indicate the positions of single nucleotide polymorphism (SNP) markers that had recombination events between them in the target gene. (e) Quantile–quantile plot of the intragenic recombination–haplotype association result for oil concentration in maize kernels. The gray line shows the predicted $-\log_{10}(P)$ value, and the dashed gray line is the cut-off of the association study. (f) Variation in kernel total oil concentration for two parental genic haplotypes and two intragenic recombination haplotypes. AA and CC are the parental haplotypes. The horizontal line within the boxplots shows the median value of oil concentration. The black points are the outliers.

To determine whether intragenic recombination also affects phenotypic variation, 22 traits (see the Materials and Methods section) were used to perform a haplotype-based association mapping analysis in a natural population of 508 inbred lines (Fig. S3). For these traits, we identified 31 significant associations (0–10 genes with intragenic recombination per trait) with nine traits at a genome-wide significance level of $P < 1.81\text{E-}05$ ($0.05/n$; $n = 2763$) (Table S9). At a relaxed genome-wide significance level of $P < 3.61\text{E-}04$ ($1/n$), we identified significant association signals with 19 traits for 113 genes exhibiting recombination events. It is noteworthy that 28.3% of genes exhibiting intragenic recombination and associated with a trait had *cis*-eQTLs, which was significantly higher than the percentage of randomly selected genes controlled by *cis*-eQTLs without intragenic recombination ($P = 6.80\text{E-}03$; Table S9).

Using the total oil concentration of maize kernels as an example, we identified eight significant loci ($P < 3.61\text{E-}04$) from 2763 genes containing intragenic recombination (Fig. 3d). In our previous SNP-based association study, one gene (*GRMZM2G115615*), annotated as a tetratricopeptide repeat, was found to be significantly associated with total oil concentration ($P = 4.10\text{E-}08$) (Li *et al.*, 2013a). Two SNPs were identified within this gene, and the four resulting haplotypes differed significantly in phenotype ($P = 2.42\text{E-}09$, MLM; $P = 5.90\text{E-}40$,

ANOVA; Fig. 3d–f), suggesting that intragenic recombination produced alleles with altered kernel oil concentrations (Fig. 3f). A similar association signal potentially originating from intragenic recombination was observed for the pericarp color1 gene to alter cob color in maize (Fig. S8). These results implied that the intragenic recombination site is either the causal polymorphism or in linkage disequilibrium with the causal polymorphism(s). Furthermore, we compared the degree of phenotypic associations of genes containing recombination events to that of randomly selected genes using Monte Carlo resampling tests. Genes containing recombination events had a slightly, but significantly greater chance to show a phenotypic association than randomly selected genes ($P = 7.90E-03$; Fig. S9). Taken together, these results indicated that there was a potential effect of intragenic recombination on phenotypic variation.

Discussion

The landscape of genome-wide recombination frequency in maize

The re-assortment of functional alleles by recombination is the basis of plant and animal breeding. Recombination frequencies and the location of recombination events are important factors that influence breeder decisions and the progress of breeding programs (Wijnker & de Jong, 2008). Given the remarkable variation in recombination among different genetic backgrounds (Bauer *et al.*, 2013), our 12 high-density genetic maps from 17 diverse parental lines provide a comprehensive landscape of maize recombination. The understanding of this genome-wide recombination landscape could aid in the construction of populations with a higher frequency of recombination events in specific regions of the genome. In addition, the 143 recombination hot regions identified in this study could be explored in specific breeding materials using marker-assisted selection to maximize recombination during population development and increase gains from selection (Holland, 2004; Wijnker & de Jong, 2008). The genes located in these recombination hot regions could be easily recombined in a short round of breeding. Notably, 34 recombination hot regions were detected in one population and 24 hot regions in two populations, which not only suggests the population specificity of recombination events but also indicates that cautious parental selection for population construction should be carried out to increase the chance of recombination in specific regions.

In this study, there are some limitations in our data set and analyses. First, our RIL populations were constructed following several rounds of meiosis, which precludes us from directly measuring recombination events and probably resulted in removal of some recombination events in the recovered homologous genomic regions during the generation of these populations. Thus, our data set may underestimate the actual number of crossover events that occurred. Second, even though the maize SNP50 array was used to construct the genetic maps, the resolution did not reach the nucleotide scale of whole-genome sequencing. Third, > 70.0% of the SNPs were located in genic regions, whereas a small fraction of SNPs were from intergenic regions, which would also lead to the underestimation of the number of intergenic recombination events. Fourth, the maize SNP50 was designed based on B73, not the maize pangenome, and thus genes and regions not present in B73 were not part of this analysis. For example, Mo17 and B73 share only *c.* 50% of their sequence, which may explain the low recombination frequency observed in the Mo17-derived population. Fifth, the number of individuals studied was relatively small to detect the recombination events in pericentromeric regions.

Intergenic recombination mainly affects the recombination variation across different genetic backgrounds

Recombination not only varies across the whole genome with recombination hot regions and cold regions (Myers *et al.*, 2005) but also is extremely variable across different genetic backgrounds (Bauer *et al.*, 2013). Recombination across whole genomes has been exhaustively studied and may be associated with specific DNA sequences (Xu *et al.*, 1995; Fu *et al.*, 2002; Steiner *et al.*, 2009; Hellsten *et al.*, 2013). Relative to recombination variation across the whole genome, recombination variation among different genetic backgrounds has not been well explored. Bauer *et al.* (2013) used 23 DH populations to identify the intraspecific variation of recombination in maize. However, they did not classify the types of recombination that contributed to the intraspecific variation or determine the phenotypic consequences of different recombination types (intragenic and intergenic) because of a limited population size (half of our population size). Thus, the basic scientific question concerning what kind of recombination contributes substantially to intraspecific variation is still unanswered. Here, we partitioned recombination variation into intragenic and intergenic patterns and compared the differences between these two types of recombination across different genetic backgrounds. We found that, although recombination occurred frequently in genic regions (Dooner & He, 2014; Li *et al.*, 2015), the total number of intragenic recombination events was relatively invariable across different genetic backgrounds (Chinese and European populations), whereas intergenic recombination was positively correlated with the total number of recombination events across different populations. Our results indicate that intergenic recombination might be the major driver of intraspecific variation in recombination in maize.

Given the dramatic genomic structural variation among maize inbreds (Tenaillon *et al.*, 2001; Springer *et al.*, 2009; Lai *et al.*, 2010), the observation of relative invariance of intragenic recombination and variation in intergenic recombination may be related to structural variation. This is likely in view of the observation that the majority of recombination events are in homologous chromosomal regions (Haber *et al.*, 2004). Different segregating populations are derived from crosses of different inbreds, or different genetic backgrounds, where extensive genomic structural variation would exist among different populations. Genes, the basic functional elements, exhibit higher conservation among different inbreds/segregating populations than intergenic regions where abundant transposons are located

(Schnable *et al.*, 2009). Thus, recombination within genic regions would be quite even, while recombination in intergenic regions would vary dramatically depending on the extent of genomic variation among different genetic populations. The number of intergenic recombination events significantly correlated with the genetic diversity of the two parents in each population ($r = 0.45$; $P = 0.02$), indicating the important contribution of genetic diversity to the intraspecific recombination variation across different populations.

Genic recombination affects gene expression and phenotypic variation

Recombination is more likely to occur at the beginning and end of genes rather than within the coding sequence showing polarity of meiotic recombination (Thijs & Heyting, 1998; Dooner & He, 2014), indicating a degree of protection for genomic elements during recombination events. By contrast, regions that flank genes, including promoters, are reshuffled to produce genome diversity. Recombination events occurring in a genic region, including gene truncations, could lead to functional mutations. However, recombination in the regulatory regions may lead to quantitative variation in expression level, timing, or tissue specificity of a given gene product, generating novel regulatory variants during crop domestication and improvement (Otto & Barton, 1997). Many studies have demonstrated that regulatory variation, rather than coding variation, is the primary driver of evolution (Fullerton *et al.*, 2001; Nachman, 2001; Doebley *et al.*, 2006; Carrol, 2008; Hawkins *et al.*, 2014). Superimposing our sites of recombination with our previous eQTL mapping results (Fu *et al.*, 2013), we found that mapped *cis*-eQTLs overlapped genes containing recombination events significantly more often than randomly selected genes.

Genetic recombination contributes to genomic diversity through meiotic crossovers and noncrossovers (Villeneuve & Hillers, 2001; Hamant *et al.*, 2006; Mezard *et al.*, 2007). The bulk of the genetic variation in plant and animal breeds was created through a history of recombination and selection (Choi & Henderson, 2015). As expected, strong associations between novel haplotypes generated by intragenic recombination events and phenotypic variations were observed. Using a strict genome-wide significance level, we identified 113 loci affecting 19 agronomic traits, including some well-known genes identified in previous studies (Coe, 2001; Li *et al.*, 2013a). Meanwhile, we utilized high-density markers in 12 segregating populations and identified the possible molecular consequences of recombination – gene expression level variation, which could somehow explain how recombination can lead to phenotypic variation in plants and animals. Our results enhance our understanding of the role of genomic recombination in plant and animal breeding.

Acknowledgements

We thank all the students and colleagues from J.Y.'s present and former laboratories at China Agricultural University and X.Y. and J.L.'s laboratories at China Agricultural University for help with population development. This research was supported by the National Natural Science Foundation of China (31525017, 31123009 and 31222041) and the National Hi-Tech Research and Development Program of China (2012AA10A307).

Author contributions

J.Y. designed and supervised this study. X.Y., S.X., Q.P., Z.L., W.L., J.L. and J.Y. developed the populations and extracted the DNA. Q.P., L.L. and H.T. performed the data analysis. G.J.M. contributed materials. Q.P., L.L. and J.Y. prepared the manuscript; X.Y. and G.J.M. helped to edit the manuscript.

References

Baudat F, De Massy B. 2007. *Cis*-and *trans*-acting elements regulate the mouse Psmb9 meiotic recombination hotspots region. *PLoS Genetics* **3**: e100.

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W *et al.* 2013. Intraspecific variation of recombination rate in maize. *Genome Biology* **14**: R103.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**: 642–645.

Carrol SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.

Choi K, Henderson IR. 2015. Meiotic recombination hotspots – a comparative view. *Plant Journal* **83**: 52–61.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FC *et al.* 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nature Genetics* **45**: 1327–1336.

Coe EH Jr. 2001. The origins of maize genetics. *Nature Review Genetics* **2**: 898–905.

De Givry S, Bouchez M, Chabrier P, Milan D, Schiex T. 2005. CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* **21**: 1703–1704.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Dooner HK, He L. 2014. Polarized gene conversion at the *bz* locus of maize. *Proceedings of the National Academy of Sciences, USA* **111**: 13918–13923.

Dumont BL, White MA, Steffy B, Wiltshire T, Payseur BA. 2011. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Research* **21**: 114–125.

Fu H, Zheng Z, Dooner HK. 2002. Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proceedings of the National Academy of Sciences, USA* **99**: 1082–1087.

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du XM, Peng ZY *et al.* 2013. RNA sequencing reveals the complex regulatory network in the maize kernel. *Nature Communications* **4**: 2832.

Fullerton SM, Carvalho AB, Clark AG. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Molecular Biology and Evolution* **18**: 1139–1142.

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J *et al.* 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**: e28334.

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. *Nature Reviews Genetics* **8**: 77–84.

Haber JE, Ira G, Malkova A, Sugawara N. 2004. Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* **359**: 79–86.

Hamant O, Ma H, Cande WZ. 2006. Genetics of meiotic prophase I in plants. *Annual Review of Plant Biology* **57**: 267–302.

Hawkins JS, Delgado V, Feng L, Carlise M, Hugo K, Dooner HK, Bennetzen JL. 2014. Variation in allelic expression associated with a recombination hotspot in *Zea mays*. *Plant Journal* **79**: 375–384.

Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. 2013. Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences, USA* **110**: 19478–19482.

Holland JB. 2004. Implementation of molecular markers for quantitative traits in breeding programs-challenges and opportunities. In: *New directions for a diverse planet. Proceedings for the 4th International Crop Science Congress Regional Institute*, Gosford, Australia www cropscience org au/icsc2004.

Hughes TE, Langdale GA, Kelly S. 2014. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Research* **24**: 1348–1355.

Kent CF, Minaei S, Harpur BA, Zayed A. 2012. Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proceedings of the National Academy of Sciences, USA* **109**: 18012–18017.

Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. 2008. Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. *Proceedings of the National Academy of Sciences, USA* **105**: 10051–10056.

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M *et al.* 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* **42**: 1027–1030.

Lander ES, Green P, Abrahamson J. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N *et al.* 2013a. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics* **45**: 43–50.

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

Li J, Hsia AP, Schnable PS. 2007b. Recent advances in plant recombination. *Current Opinion in Plant Biology* **10**: 131–135.

Li L, Petsch K, Shimizu R, Liu SZ, Xu WZ, Ying K, Yu JM, Scanlon MJ, Schnable PS, Timmermans MCP *et al.* 2013b. Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genetics* **9**: e1003202.

Li X, Li L, Yan J. 2015. Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. *Nature Communications* **6**: 6648.

Martin OC, Wagner A. 2009. Effects of recombination on complex regulatory circuits. *Genetics* **183**: 673–684.

Matise TC, Chen F, Chen WW, De La Vega FM, Hansen M, He CS, Hyland FC, Kennedy GC, Kong X, Murray SS *et al.* 2007. A second-generation combined linkage–physical map of the human genome. *Genome Research* **17**: 1783–1786.

Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Molecular Biology and Evolution* **21**: 984–990.

Mezard C, Vignard J, Drouaud J, Mercier R. 2007. The road to crossovers: plants have their say. *Trends in Genetics* **23**: 91–99.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**: 321–324.

Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. *Trends in Genetics.* **17**: 481–485.

Otto SP, Barton NH. 1997. The evolution of recombination: removing the limits to natural selection. *Genetics* **147**: 879–906.

Pawlowski WP. 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.

Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li C, Li Y, Buckler ES. 2015. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences, USA* **112**: 3823–3828.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.* 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112–1115.

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biology* **6**: R33.

Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* **472**: 375–378.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H *et al.* 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genetics* **5**: e1000734.

Steiner WW, Steiner EM, Girvin AR, Plewik LE. 2009. Novel nucleotide sequence motifs that produce hotspots of meiotic recombination in Schizosaccharomyces pombe. *Genetics* **182**: 459–469.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences, USA* **98**: 916–926.

Thijs H, Heyting C. 1998. Polarity of meiotic recombination in the bronze locus of maize. *Plant Cell* **10**: 645–648.

Villeneuve AM, Hillers KJ. 2001. Whence meiosis? *Cell* **106**: 647–650.

Wijnker E, de Jong H. 2008. Managing meiotic recombination in plant breeding. *Trends in Plant Science* **13**: 640–646.

Xu X, Hsia AP, Zhang L, Nikolau BJ, Schnable PS. 1995. Meiotic recombination break points resolve at high rates at the 5′end of a maize coding sequence. *Plant Cell* **7**: 2151–2161.

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen WW, Liu J, Li JS, Yan JB. 2014. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genetics* **10**: e1004573.

Yang S, Yuan Y, Wang L, Li J, Wang W, Liu H, Chen JQ, Hurst LD, Tian D. 2012. Great majority of recombination events in Arabidopsis are gene conversion events. *Proceedings of the National Academy of Sciences, USA* **109**: 20992–20997.

Yang X, Yan J, Shah T, Warburton M, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y *et al.* 2010. Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theoretical and Applied Genetics* **121**: 417–431.

Yao H, Schnable PS. 2005. Cis-effects on meiotic recombination across distinct *a1-sh2* intervals in a common *Zea* genetic background. *Genetics* **170**: 1929–1944.

Yao H, Zhou Q, Li J, Smith H, Yandeau M, Nikolau BJ, Schnable PS. 2002. Molecular characterization of meiotic recombination across the 140-kb multigenic *a1-sh2* interval of maize. *Proceedings of the National Academy of Sciences, USA* **99**: 6157–6162.

Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

Zhang L, Gaut BS. 2003. Does recombination shape the distribution and evolution of tandemly arrayed genes (TAGs) in the *Arabidopsis thaliana* genome? *Genome Research* **13**: 2533–2540.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Flow chart of maize linkage map construction.

Fig. S2 Map of associations between maize genic recombination haplotypes and phenotypic values for two markers in a single gene and a single phenotypic trait.

Fig. S3 Manhattan (left) and quantile–quantile (right) plots for 21 maize agronomic traits and one oil trait based on genic recombination induced haplotype association mapping in 508 maize inbred lines.

Fig. S4 Distribution of family recombination events in 12 maize populations.

Fig. S5 Negative correlation between the average number of recombination events in each gene and the proportion of genes with genic recombination in maize DH populations.

Fig. S6 The percentage of intragenic recombination genes in 12 segregating maize populations.

Fig. S7 Relationships between genic recombination and gene expression levels.

Fig. S8 Recombination in the gene closest to *p1*, which is significantly associated with maize cob color variation.

Fig. S9 Relationships between genic recombination and 22 maize phenotypes.

Table S1 Information on the 12 segregating maize populations constructed in our study

Table S2 Summary of the segregating lines in the 12 maize populations used for our genome-wide recombination analysis

Table S3 Information on the 12 high-density maize linkage maps constructed in our study

Table S4 Bootstrapping analysis of recombination frequencies in different maize subpopulations

Table S5 The number of recombination events along each chromosome across 12 segregating maize populations

Table S6 Recombination hot regions in 12 segregating maize populations

Table S7 Number of polymorphic markers and genic recombination events detected in 12 maize populations

Table S8 Number of genes containing recombination events detected in 12 maize populations

Table S9 Genic recombination associated with 22 maize phenotypes and gene expression levels

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.