

Intragenic Meiotic Crossovers Generate Novel Alleles with Transgressive Expression Levels

Sanzhen Liu,^{†,1,2} James C. Schnable,^{†,3} Alina Ott,^{†,‡,2} Cheng-Ting “Eddy” Yeh,² Nathan M. Springer,⁴ Jianming Yu,² Gary Muehlbauer,⁵ Marja C.P. Timmermans,⁶ Michael J. Scanlon,⁷ and Patrick S. Schnable^{*,2}

¹Department of Plant Pathology, Kansas State University, Manhattan, KS

²Department of Agronomy, Iowa State University, Ames, IA

³Department of Agriculture and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

⁴Department of Plant and Microbial Biology, Microbial and Plant Genomics Institute, University of Minnesota, Saint Paul, MN

⁵Department of Agronomy and Plant Genetics, Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN

⁶Center for Plant Molecular Biology, University of Tuebingen, Tuebingen, Germany

⁷Department of Plant Biology, Cornell University, Ithaca, NY

[†]These authors contributed equally to this work.

[‡]Present address: Roche Sequencing Solutions, 500 S Rosa Road, Madison, WI

*Corresponding author: E-mail: schnable@iastate.edu.

Associate editor: Juliette de Meaux

Abstract

Meiotic recombination is an evolutionary force that generates new genetic diversity upon which selection can act. Whereas multiple studies have assessed genome-wide patterns of recombination and specific cases of intragenic recombination, few studies have assessed intragenic recombination genome-wide in higher eukaryotes. We identified recombination events within or near genes in a population of maize recombinant inbred lines (RILs) using RNA-sequencing data. Our results are consistent with case studies that have shown that intragenic crossovers cluster at the 5' ends of some genes. Further, we identified cases of intragenic crossovers that generate transgressive transcript accumulation patterns, that is, recombinant alleles displayed higher or lower levels of expression than did nonrecombinant alleles in any of ~100 RILs, implicating intragenic recombination in the generation of new variants upon which selection can act. Thousands of apparent gene conversion events were identified, allowing us to estimate the genome-wide rate of gene conversion at SNP sites (4.9×10^{-5}). The density of syntenic genes (i.e., those conserved at the same genomic locations since the divergence of maize and sorghum) exhibits a substantial correlation with crossover frequency, whereas the density of nonsyntenic genes (i.e., those which have transposed or been lost subsequent to the divergence of maize and sorghum) shows little correlation, suggesting that crossovers occur at higher rates in syntenic genes than in nonsyntenic genes. Increased rates of crossovers in syntenic genes could be either a consequence of the evolutionary conservation of synteny or a biological process that helps to maintain synteny.

Key words: gene regulation, evolutionary novelty, gene conversion.

Introduction

Adaptation, speciation, and selection (both natural and artificial) depend upon both independent assortment of whole chromosomes, as well as meiotic recombination's ability to create novel haplotypes within individual chromosomes, via crossovers and/or gene conversion (GC; Muller 1964; Gaut et al. 2007; Lu et al. 2012). These haplotypes can be novel combinations of alleles carried by the same chromosome and/or recombinant alleles of individual genes with novel functionalities, for example, by shuffling promoter or enhancer sequences or protein domains of varying efficiencies. In addition, recombination is essential for identifying associations between genic and phenotypic variation (Salome et al. 2012). In many organisms, meiotic recombination events are not uniformly distributed across the genome (Myers

et al. 2005; Giraut et al. 2011; Pan et al. 2011). As such, the cumulative effects of recombination over evolutionary time have left signatures that can be observed in existing populations. Favorable haplotype blocks can be maintained over evolutionary time in chromosomal regions that experience low rates of recombination (Wijnker et al. 2012). Conversely, low rates of recombination can reduce the probability of such favorable haplotype blocks being created in the first place. Low rates of recombination also contribute to linkage drag, decreasing the efficiency of both natural and artificial selection (e.g., breeding) causing selective sweeps (Brinkman and Frey 1977).

Although double-strand breaks are widely distributed across the maize genome (He et al. 2017), meiotic recombination events per se are generally suppressed in pericentromeric regions (Gore et al. 2009; Liu et al. 2009;

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Rodgers-Melnick et al. 2015). The lack of recombination between loci can contribute to heterosis via pseudo-overdominance (Schnable and Springer 2013). The reduced rate of recombination in pericentromeric regions may contribute to the accumulation of deleterious alleles in these regions, and regions with higher rates of recombination tend to have reduced genetic load (Rodgers-Melnick et al. 2015).

Studies in maize indicate that meiotic recombination occurs at high frequency in nonrepetitive genic regions (Fu et al. 2001; Rodgers-Melnick et al. 2015). Case studies of individual genes have suggested that meiotic recombination events occurring within genes (intragenic recombination) occur more often in 5' ends of some maize genes (Patterson et al. 1995; Xu et al. 1995; Choi et al. 2013; Shilo et al. 2015), more often in 3' ends of some other genes (Eggleston et al. 1995), or show no obvious preference (Dooner and Martínez-Férez 1997). A study of maize tetrads showed that cross-overs were more likely to occur in the 5' end of genes, followed by the 3' end and the middle of genes (Li et al. 2015). However, this study was limited to a handful of genes due to the small number of tetrads observed. More global analyses conducted in the Arabidopsis (Choi et al. 2013; Shilo et al. 2015) and human (Kong et al. 2010) genomes reached similar conclusions. The phenotypic importance of intragenic recombination has been shown in studies in multiple species but has so far been limited to indirect associations or single genes (Freeling 1976; Hagblom et al. 1985; McDowell et al. 1998; Ogasawara et al. 1998). Alleles derived via intragenic recombination can have higher expression in segregating populations and be more likely to be present in genomic association intervals for phenotypic traits than randomly selected genes in a maize diversity panel (Pan et al. 2016).

Many plant genomes have undergone evolutionarily recent whole genome duplication events (Jiao et al. 2011). Maize underwent a recent polyploidization event between 5 and 12 Ma that is not shared by sorghum (Swigonova et al. 2004) and that was followed by fractionation. Both the maize and sorghum genomes have been sequenced (Paterson et al. 2009; Schnable et al. 2009), making comparisons between these two species an excellent model to assess the effects of genome duplication. Maize contains two subgenomes relative to the unduplicated sorghum, termed maize1 and maize2. After the whole genome duplication, many maize genes were lost apparently via intrachromosomal nonhomologous recombination (Woodhouse et al. 2010). As a consequence of differential loss, the maize1 subgenome shares more genes with sorghum than does maize2 (Schnable et al. 2011). Genes retained at syntenic (ancestral, collinear) positions have distinct genomic properties as compared with nonsyntenic genes. For example, syntenic genes are more likely to be expressed and have lower levels of methylation than nonsyntenic genes (Eichten et al. 2011), consistent with mutation analyses which suggest that mutants in syntenic genes are more likely to yield a visible phenotype than mutations in nonsyntenic genes (Schnable and Freeling 2011). In addition, functional differentiation is also observed between the maize subgenomes. Maize1 genes

explain more phenotypic variation than maize2 genes as assessed via mutant and association analyses (Schnable and Freeling 2011; Renny-Byfield et al. 2017), which corresponds to higher maize1 gene expression than maize2 though no difference in methylation (Eichten et al. 2011).

In this study, we identified 176,279 polymorphic SNPs from RNA-seq data on 105 RILs, 7,854 recombination breakpoint intervals (RBIs), of which 848 are intragenic crossovers, and 3,014 apparent gene conversions (aGCs). Importantly, we demonstrated that crossovers within these genes can generate alleles with novel levels of transcript accumulation. Global GC events enabled us to provide the first genome-wide estimate for the rate of GC in maize. Finally, we demonstrated that syntenic genes are enriched for recombination relative to nonsyntenic genes, which could be either a consequence of the evolutionary conservation of synteny across maize haplotypes or a process that helps to maintain synteny.

Results

Genotyping the IBM Population by RNA-Seq and De Novo Construction of a Genetic Map

The intermated B73 and Mo17 recombinant inbred lines (IBM RILs) were developed by crossing the inbred lines B73 and Mo17, followed by several generations of random mating and multiple generations of self pollination (Lee et al. 2002). Using RNA-seq data from vegetative apices (Barbazuk et al. 2007), a set of high confidence SNPs ($N = 176,279$) that segregated among the RILs was identified (Materials and Methods). 162,356 of these SNPs are in annotated genes, providing a polymorphism rate of 1.3 SNPs per kb of genic space.

The SNP genotypes of all the RILs were subjected to segmentation (Olshen et al. 2004) to obtain a minimum set of genetic markers describing all crossovers observed within the RIL population (fig. 1A–C, Materials and Methods). A total of 7,856 segmental markers representing single genetic markers genotyped as the B73, the Mo17, or a recombinant haplotype were identified (supplementary fig. S1, Supplementary Material online). These markers were used to generate a genetic map (supplementary table S1, Supplementary Material online), which was shown to exhibit a high level of consistency with the physical map (correlation coefficient > 0.999). Segments in a particular RIL where a crossover event occurred are called RBIs ($N = 7,574$; fig. 1D). Given their size and the numbers of consistent flanking markers (Materials and Methods), RBIs are presumed to be the consequence of crossovers (fig. 1). The numbers of detected crossovers in each RIL range from 60 to 125 (fig. 2B), with an average of 82, which is similar to previous estimates in the IBM RILs (Fu et al. 2006; Esch et al. 2007). Some markers within segments have a parental genotype that does not agree with the evidence from the rest of the segment and therefore represent aGCs (fig. 1E).

Chromosomal Level Distribution of Crossovers

Overall, the genetic lengths of the ten maize chromosomes are highly positively correlated with the physical lengths of chromosomes (correlation coefficient = 0.93; supplementary

fig. S2, Supplementary Material online). Consistent with previous genetic maps of the IBM RILs and other maize populations, the middles of all chromosomes show low crossover frequency, whereas the arms of chromosomes exhibit high crossover frequency (supplementary fig. S3, Supplementary

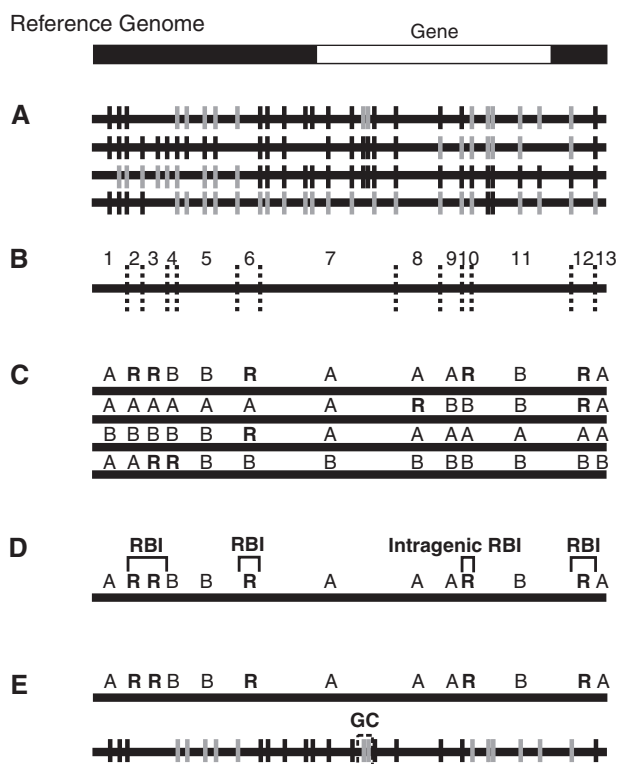


Fig. 1. Genotyping, segmentation, recombination breakpoint intervals, and gene conversions in the IBM RILs. (A) SNPs (vertical bars) were genotyped as B73-like (black) or Mo17-like (gray) in each RIL (horizontal bars). (B) Segmental markers were defined as genomic regions that originate from a single parent in each RIL. (C) Segments were genotyped as B73-like (A), Mo17-like (B), or recombinant (R). (D) Recombination breakpoint intervals (RBIs) are individual or multiple adjacent recombinant segments. RBIs that occur within a gene are intragenic RBIs. (E) Individual markers within a segment may disagree with the segmental genotype. These markers are potentially gene conversions.

Material online; Esch et al. 2007; Liu et al. 2009; Rodgers-Melnick et al. 2015). 72% of all crossovers occur in the ~30% of the genome that exhibits a recombination frequency >1 cM/Mb across generations.

The Relationship between Crossovers and Synteny

Many studies suggest that crossovers occur more often in genic versus nongenic regions (Brown and Sundaresan 1991; Civardi et al. 1994; Fu et al. 2001; Yao et al. 2002; Rodgers-Melnick et al. 2015). As a consequence of the RNA-seq strategy used for genotyping, most of our markers are located within genes. Hence, we were able to focus our investigations on differences in crossover frequency between classes of genes. A set of genes that are syntenic between maize and sorghum—that is, retained in ancestral, collinear positions in both species—has been defined (Schnable et al. 2011). We analyzed crossovers and synteny for 1 Mb bins genome-wide. The numbers of base pairs in each bin that could be assigned to syntenic genes, nonsyntenic genes or nongenic sequences was determined and compared with the crossover frequency. Crossover frequencies of bins are more strongly correlated with the size of the syntenic gene space within the bin (fig. 3A, Spearman's Rank correlation $\rho = 0.58$) than with the size of the nonsyntenic gene space (fig. 3B, Spearman's Rank correlation $\rho = 0.27$), suggesting that crossovers occur at higher rates in syntenic genes than in nonsyntenic genes.

Pericentromeric regions contain more nonsyntenic genes (Schnable et al. 2012). The enrichment of crossovers in syntenic regions could potentially be a result of gross chromosomal arrangements elsewhere in the genome, rather than a function of the properties of syntenic regions. To control for this possibility, the relationship between crossovers and synteny was assessed separately in pericentromeric (25 kb flanking the centromere) and nonpericentromeric regions. The correlations between synteny and crossovers in nonpericentromeric regions were very similar to the genome-wide correlations: $\rho = 0.52$ versus 0.58 for syntenic genes and $\rho = 0.25$ versus 0.27 for nonsyntenic genes. In contrast, within pericentromeric regions there was no discernable correlation between crossovers and either syntenic or nonsyntenic genes

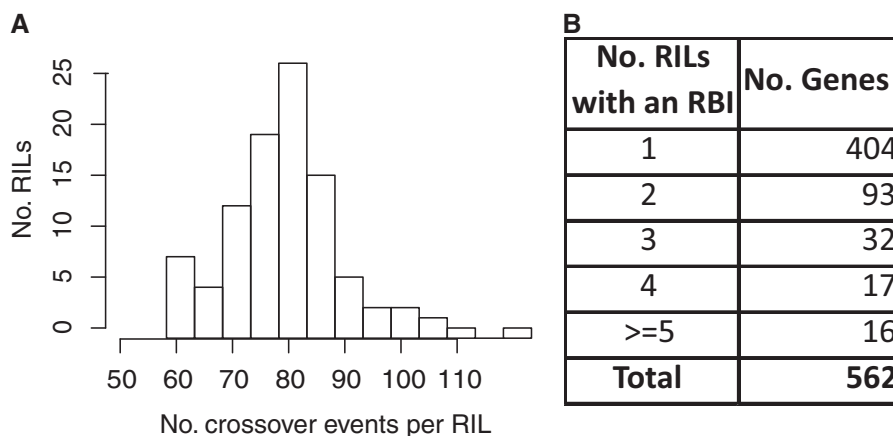


Fig. 2. Crossovers in the RILs. (A) The distribution of the number of crossovers per RIL ($N = 105$). (B) The distribution of RBIs in genes based on the number of RILs affected by an RBI in a particular gene.

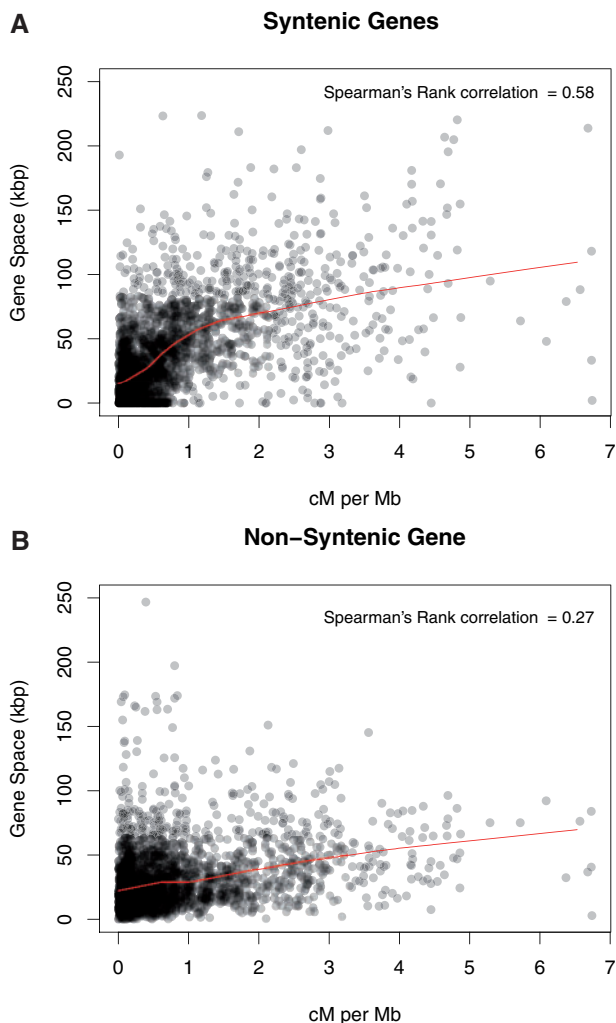


Fig. 3. Synteny and crossover frequency. The genome was divided into nonoverlapping 1 Mb bins (black dots, $N = 2,064$ for each panel) and the amount of sequence from syntenic and nonsyntenic genes was determined. The crossover frequency across each bin was estimated in cM per Mb. (A) Bins containing more syntenic genes have a stronger positive correlation with crossover frequency. (B) Bins containing more nonsyntenic genes have a reduced correlation with crossover frequency.

($\rho = -0.10$ and $\rho = 0.05$, respectively). We therefore conclude that the negative correlations of both recombination and syntenic gene density with pericentromeric identity are not sufficient to explain the correlation between these two characteristics of genomic regions.

Relative to sorghum, maize is an ancient tetraploid (Swigonova et al. 2004) and any retained syntenic orthologous genes between these species can be classified as belonging to either the maize1 or maize2 subgenome based on their genomic positions (Schnable et al. 2011). While there are differences between the subgenomes, such as a bias towards higher gene expression in maize1, higher gene loss across maize lines from maize2 (Schnable et al. 2011), and a larger proportion of phenotypic variation explained by maize1 singleton genes (Renny-Byfield et al. 2017), no difference in crossover frequency was detected between the two subgenomes (supplementary fig. S4, Supplementary Material online,

Spearman's Rank correlation $\rho = 0.56$ for maize1 and $\rho = 0.57$ for maize2).

Distribution of Crossovers among and within Genes

Previous case studies of individual genes showed that many genes are hotspots for crossovers (Brown and Sundaresan 1991; Civardi et al. 1994; Eggleston et al. 1995; Patterson et al. 1995; Xu et al. 1995; Dooner and Martínez-Férez 1997). To understand the distribution of crossovers among and within genes, we first identified RBIs encompassed by a single gene. The size of RBIs is dependent on the regional density of SNP markers; hence the estimated sizes of most RBIs are expected to be larger than the actual sizes of these RBIs. The 7,574 RBIs detected in this study range in size from 1 bp to ~ 30 Mb, with a median of 104.6 kb. Maize genes, as defined in this study (Materials and Methods) have a median length of 4.5 kb, and 99% genes are < 30 kb. Even so, across all RILs, 848 RBIs were located within a gene that was extended 1 kb upstream and downstream of the gene, including 793 unique RBIs encompassed by 561 nonredundant genes (fig. 1D, supplementary table S2, Supplementary Material online, Materials and Methods). Most RBIs (7,535/7,574) overlap with a gene, indicating few RBIs were encompassed only by the intergenic regions that comprise $> 85\%$ of the genome. The majority of genes (404/561) that contain an RBI (i.e., an apparent intragenic crossover) experienced a crossover in only a single RIL. But some genes appear to be "hotspots" such that crossovers were observed in at least two RILs (fig. 2B). Permutation testing (Materials and Methods) indicated that the number of "hotspot" genes was significantly more than expected (P -value < 0.001). These observations support both the previous finding that recombination is more likely to occur in genic than nongenic regions and that certain genes experience recombination events much more frequently than do others.

We also found that intragenic crossovers were $1.9\times$ more likely to occur within syntenic genes than within nonsyntenic genes (Fisher's Exact Test, P -value $= 5.9 \times 10^{-4}$). Syntenic genes are more likely to be expressed than are nonsyntenic genes and therefore intragenic crossovers in syntenic genes may be more likely to be detected using our RNA-seq based genotyping strategy. However, roughly equal proportions of syntenic and nonsyntenic genes contained at least two SNPs that passed all quality filtering parameters (49% and 48% respectively) and therefore had the potential for detectable intragenic crossover events. Hence, the enrichment of intragenic crossovers within syntenic genes is unlikely to have resulted from an ascertainment bias and instead provides additional support for the conclusion that syntenic genes experience more crossovers than do nonsyntenic genes.

A preference for crossovers to cluster in the 5'-ends of genes has been observed in several case studies of maize (e.g., Patterson et al. 1995; Xu et al. 1995) and later more globally in Arabidopsis (Choi et al. 2013; Shilo et al. 2015). To explore where in genes intragenic crossovers occur, each gene was divided into three equal parts: 5'-end, middle, and 3'-end (Materials and Methods). For 301 RBIs the entire interval is within one of these three parts: 141 (46.8%),

79 (26.3%), and 81 (26.9%) RBIs occur at the 5'-end, middle and the 3'-end, respectively. The overall distribution of SNPs identified by RNA-seq between the three segments is 25,486 (24%) 5'-end, 41,356 (38%) middle, 41,117 (38%) 3'-end, significantly different from the distribution of RBIs ($\chi^2=89.7$, $df=2$, P -value $< 2.2e-16$); thus, the observed distribution of intragenic RBIs is not an artifact of the distribution of SNPs.

Intragenic Crossovers Can Generate Novel Alleles with Transgressive Levels of Transcript Accumulation

Intergenic crossovers can produce novel haplotypes by shuffling existing alleles. Similarly, intragenic crossovers can shuffle parental polymorphisms to create novel alleles (Schnable et al. 1998). Novel alleles created by intragenic crossovers have the potential to exhibit novel functionality. To test this hypothesis, we compared the transcript accumulation levels, represented by RPKM (reads per thousand of genic region per million of total reads), of alleles derived from intragenic crossovers to the transcript accumulation levels of their parental alleles. The transcript accumulation levels of each of the 848 recombinant alleles generated by intragenic crossovers was compared with the transcript accumulation levels of the nonrecombinant (i.e., parental) alleles that were segregating in the other RILs, as well as in the parental inbreds, B73 and Mo17. This comparison identified a statistically significant number of recombinant alleles ($N=20$) that exhibited transgressive transcript accumulation in which the recombinant allele is expressed 5% higher or lower than the most extreme valued observed across the RIL set (and their parents) for nonrecombinant parental alleles (fig. 4, P -value = 0.004, see Materials and Methods for the permutation test).

Among these 20 recombinant alleles with transgressive transcript accumulation in 18 genes, 9 and 11 alleles exhibit high and low transgressive transcript accumulation, respectively. Five, two, and one RBIs were located at 5' ends, the middle, and 3' ends of the genes. We conducted a permutation test using randomly sampled genes to demonstrate that the probability that at least 20 of the 793 intragenic recombination alleles identified in this study would exhibit transgressive segregation is $\sim 1/250$ or 0.4%. Although this test does not provide certainty that any given recombinant allele exhibits transgressive segregation, it does provide strong evidence that as a group at least some of the identified recombinant alleles exhibit transgressive segregation. Because transgressive expression resulting from segregation of trans-regulatory factors should be equally common for not only alleles which experienced intragenic recombination and also for those which did not, it is unlikely that segregation of trans-regulatory factors explains the amount of transgressive expression we observed for alleles produced by intragenic recombination.

Two genes contain more than one recombinant allele with transgressive transcript accumulation. In the gene GRMZM2G074238, a putative alpha/beta-hydrolase, two intragenic recombinant alleles exhibited transgressive transcript accumulation lower than other alleles. Both alleles had the same haplotype and were the only two recombinant alleles

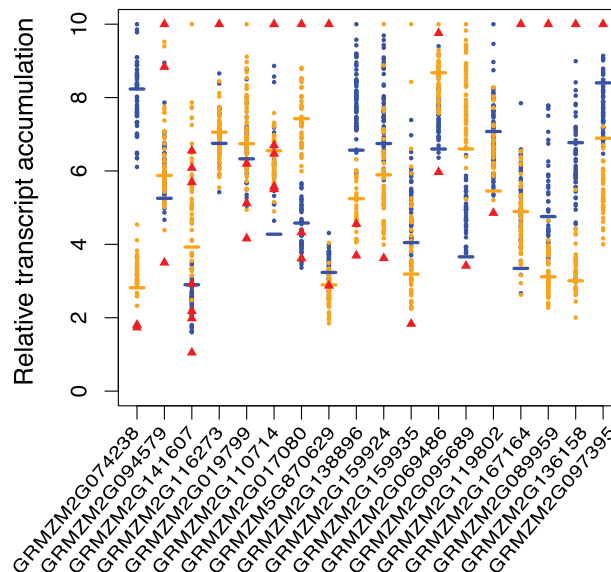


FIG. 4. Transgressive transcript accumulation of recombinant alleles. Shown are 18 genes with at least one recombinant allele (red triangles) that exhibit transgressive transcript accumulation. Blue dots indicate the transcript accumulation of RILs with a B73 allele, whereas orange dots indicate the expression of RILs with a Mo17 allele. The transcript accumulation of each gene is scaled such that the maximum transcript accumulation is 10 and no transcript accumulation is 0. The blue and orange bars are the transcript accumulation level of the B73 and Mo17 parents, respectively.

recovered in this gene. In the gene GRMZM2G094579, a putative glycosyltransferase, one allele with transgressive transcript accumulation exhibits higher transcript accumulation but the other exhibits lower transcript accumulation. Consistent with the hypothesis that transgressive expression of these alleles arises via an interaction of the alternative alleles at the 5' and 3' ends of the gene, in one of these alleles the 5' end of the gene was derived from B73 and in the other allele the 5' end was derived from Mo17. The genes for which we identified alleles with transgressive segregation are not enriched for those with low parental transcript accumulation; therefore, transgressive transcript accumulation is not simply a consequence of small amounts of variation in genes that are expressed at low levels (supplementary fig S5, Supplementary Material online). In summary, these findings demonstrate that crossover between polymorphic alleles can generate novel alleles with different patterns of gene regulation and, potentially, function.

GC Events

Few studies of GC, a form of meiotic recombination that involves a nonreciprocal exchange of genetic information have been conducted genome-wide in plants (Wijnker et al. 2013; Li et al. 2015). In yeast, each haploid spore of a tetrad can be separated and germinated to form a spore colony that provides sufficient material for sequencing to detect evidence of GC genome-wide (Fogel and Hurst 1967). In plants where tetrad analysis is technically challenging, putative GC events are typically detected primarily through indirect methods, such as assessing the segregation of markers

(Shi et al. 2010). Only a single tetrad study of maize microsporocytes has been conducted in maize to assess GCs, which was limited by sequencing coverage ($1.4\times$ for 41% of the maize genome) and the number of tetrads ($N = 24$ tetrads which resulted in the identification of 160 GC tracts, 150 which were >10 kb due to low marker density; Li et al. 2015). Two GC events were identified in the IBM RIL centromeres, and these events combined with an indirect estimate of GC based on haplotypes in a diversity panel were used to estimate the overall rate of GC in maize centromeres as $\sim 1 \times 10^{-5}$ per marker per generation (Shi et al. 2010).

To estimate the frequency of GCs in maize, we identified aGC among the IBM RILs. Evidence from *Arabidopsis* and maize suggests GC tracts are typically <2 kb (Yandeau-Nelson et al. 2005; Lu et al. 2012). SNPs that exhibited the opposite genotype from the chromosomal segment in which they were embedded were classified as aGCs (fig. 1E). After stringent filtering of RILs, SNPs, and aGCs to control for potential systemic errors described by Qi et al. (2014) (Materials and Methods), we were left with 97 RILs, 10,289,873 SNPs and 3,014 aGC SNPs covering 2,634 SNP sites, or an average of 31 aGC SNPs per RIL. 4.9×10^{-5} exchanges were estimated to occur per marker per generation, that is, an order of magnitude higher than the rate in maize centromeres (Shi et al. 2010).

Because we used RNA-seq data to discover aGCs, we considered the possibility that we were inadvertently detecting cases of RNA–DNA differences (Li et al. 2011). Although the frequency of RNA–DNA differences has been brought into question (Piskol et al. 2013), to rule out what we believed was a remote possibility, genomic sequence data from two RILs, M0022 and M0023 (Liu et al. 2012) were used to successfully cross-validate 5/6 SNPs associated with aGCs, indicating that at least most aGCs are not the result of RNA–DNA differences. Another potential explanation for the origin of aGCs would be single-nucleotide mutations. The probability of two such mutations resulting in a SNP that matches the nonparental SNP is much lower than the rate of aGC at $\sim 1.3 \times 10^{-11}$ (Materials and Methods). In addition, mutations would not be expected to cluster as aGCs do.

The aGCs are not randomly distributed across the genome. Indeed, 96% (2,518/2,634) of the aGCs cluster within 2 kb of another aGC SNP, and 636 aGC SNPs are located adjacent to other aGC SNPs within a RIL. Segmentation (Materials and Methods) of those aGC SNPs identified 143 genomic regions that harbor multiple aGC SNPs. Given the RNA-seq strategy used to detect aGCs it is not surprising that they cluster in genes and genes are not randomly distributed across the genome. But setting aside the distribution of aGCs in genic versus nongenic regions, some genes are enriched in aGCs. The 2,634 aGCs are located in only 946 genes, fewer genes than would be expected by chance (permutation test; $P < 0.001$). aGCs also exhibit little difference in frequencies in the 5', middle and 3' portions of genes (supplementary table S1, Supplementary Material online).

Discussion

By conducting RNA-seq on 105 RILs we identified 176,279 polymorphic SNPs, 7,574 RBIs, of which 848 are intragenic crossovers, and 3,014 aGCs.

Intragenic Crossover Generates Novel Alleles with Functional Differences

Mutations create genetic diversity upon which selection can act. However, even in the absence of mutation, meiotic recombination can create novel haplotypes that may prove to be adaptive, though few examples have been reported at the gene level. Previous investigations have identified intragenic recombination events (Brown and Sundaesan 1991; Civardi et al. 1994; Eggleston et al. 1995; Patterson et al. 1995; Xu et al. 1995; Dooner and Martínez-Férez 1997) and suggest that genes are recombination hotspots (Brown and Sundaesan 1991; Civardi et al. 1994; Fu et al. 2001; Yao et al. 2002; Rodgers-Melnick et al. 2015). Further, a handful of studies have shown for single genes across multiple species that novel alleles created via intragenic recombination can exhibit alterations in gene function or expression (Freeling 1976; Hagblom et al. 1985; McDowell et al. 1998; Ogasawara et al. 1998). Similarly, alleles derived via intragenic recombination at 31 loci were located within chromosomal intervals that have been shown via GWAS to exhibit novel functionality (Pan et al. 2016). The limitation of this approach for ascertaining functional novelty of intragenic recombinants is that GWAS typically does not provide gene level mapping resolution. In contrast, in the current study, we directly demonstrated that intragenic recombinants had novel levels of gene transcript accumulation. We used SNPs derived from RNA-seq data to map 848 intragenic crossovers to 561 nonredundant genes in a biparental population. A statistically significant number of these recombinant alleles exhibit transgressive levels of transcript accumulation relative to the parental alleles from which they were derived, thus providing direct evidence for their novel functionality.

In this study, transcript accumulation, as measured by RPKM was used as a proxy for gene expression. Transcript accumulation can be affected by sequences at both ends of genes, for example, promoters at the 5' end of the gene (Bilas et al. 2016) and 3' regions that affect RNA stability (Wang et al. 2009). Hence, an intragenic recombinant allele that combines a strong 5' promoter with a stabilizing 3' polymorphism could exhibit transgressive levels of transcript accumulation. Indeed, we identified one gene in which intragenic recombinant alleles that consisted of 5' regions derived from B73 and 3' regions derived from Mo17 had high transcript accumulation, whereas an intragenic recombinant allele with opposite composition (i.e., a 5' region derived from B73 and a 3' region derived from Mo17) had low transcript accumulation. Hence, our data demonstrate that intragenic crossovers can generate novel alleles with unique patterns of gene expression. While our data do not link transgressive transcript accumulation with a specific phenotype, variation in transcript accumulation have been shown to be associated with phenotypic variation (Lin et al. 2017).

The new alleles generated by intragenic recombination have the potential to create more extreme phenotypes which may be acted on by selection. Whereas the number of recombinant alleles identified with transgressive expression was small, over generations and across large populations, these rare recombination events can generate new alleles at a higher rate than does mutation.

Rate and Distribution of GC in Maize

Genome-wide discovery of GC events relies on the accuracy of identification of DNA polymorphisms. Qi et al. (2014) identified potential sources causing artifactual GC events, including misalignments of sequences from paralogs, regions exhibiting structural variation, and some tandem repeats, as well as random sequencing errors resulting in an apparent conversion from a homozygous genotype from one parent to a heterozygous genotype. In our RIL population, alignment issues are predicted to result in repeated observations of aGC at the same sites in multiple RILs. Here, we excluded aGC SNP sites observed in five or more independent RILs and regions that were >2 kb and that contained clustered aGC SNPs, which controlled for false GCs derived from misalignments. In our analysis, we only identified GCs that were fixed in the homozygous state. This is expected to substantially reduce the likelihood of false GCs caused by random sequencing errors that were discussed by Qi et al. (2014). Our ability to confirm 5/6 aGC via independent sequencing confirmed that the rate of false positive aGCs in this study was low.

On the basis of our analysis of ~3,000 SNPs associated with aGCs, we estimated the rate of GC in genic regions to be 4.9×10^{-5} per marker per generation, which is higher than the genome-wide rate (3.6×10^{-6}) reported for *Arabidopsis* (Wijnker et al. 2013) and the rate estimated for maize centromeres ($\sim 1 \times 10^{-5}$ per marker per generation; Shi et al. 2010), but within the range of estimates reported in other species (Korunes and Noor 2017). However, as no standard for estimating GC has been determined, the variation in reported GC rates might result from variation due to experimental design, error, or biology.

It is well established (and this study confirms) that crossovers exhibit nonrandom distributions within genes (i.e., 5' enrichment). The distribution of GCs within genes is less well established, in part because of the danger of ascertainment bias based on the distributions of markers within the limited numbers of genes that have been studied. However, Dooner and He (2014) reported that GC events occur more frequently at two ends (5' and 3') than in the middle of the maize *bz1* gene. In other species, other intragenic distributions have been observed (Schultes and Szostak 1990; Malone et al. 1994). Analysis of our collection of 3,014 aGCs affecting 946 genes provides no evidence to support a nonrandom distribution within genes (supplementary table S3, Supplementary Material online). The difference in distribution of crossovers and aGC is not surprising given our current understanding of meiotic recombination (Guillon et al. 2005; Yandea-Nelson et al. 2005), though the nonrandom intragenic distribution of aGC has not been previously reported.

Synteny and Crossover

As a consequence of the evolutionary history of the maize genome, its two subgenomes have undergone subsequent gene loss and rearrangement (Schnable et al. 2011). There is a growing body of evidence that genes that have remained syntenic and those that are not syntenic differ in multiple respects (Schnable 2015), including methylation (Eichten et al. 2011), transcript accumulation levels (Eichten et al. 2011) and functional importance as defined by mutant analysis and studies of natural phenotypic variation (Schnable and Freeling 2011; Renny-Byfield et al. 2017). Our results provide strong evidence that syntenic genes experience higher rates of crossover than do nonsyntenic genes and as such provide additional support for functional differences between those genes that have been retained at syntenic positions and those that have not.

Several studies have identified negative correlations between cytosine methylation and recombination in maize (Liu et al. 2009; Rodgers-Melnick et al. 2015) and later in *Arabidopsis* (Choi et al. 2013; Shilo et al. 2015). Hence, the differences in crossover frequency we observed between syntenic and nonsyntenic genes may be related to the observed differences in methylation between these classes of genes (Eichten et al. 2011). This could come about because following polyploidization, “genome-shock” can lead to the expression and translocation of transposons (Comai 2000), which in turn can disrupt synteny. Transposed genes that are no longer syntenic may undergo increases in cytosine methylation due to the proximity of transposons, which inhibits further movement and recombination (Eichten et al. 2012, 2013; Sehrish et al. 2014). These changes in methylation are thought to reduce transcript accumulation levels. Hence, it is intriguing to hypothesize that these changes in methylation may also be responsible for the reduced crossover frequency experienced by nonsyntenic genes (Eichten et al. 2011). It is, of course also possible that the reduced crossover frequency is a cause, rather than an effect, of altered patterns of methylation.

A stronger case can be made that the alterations in gene order and content that occur following polyploidization and subsequent fractionation may disrupt rates of crossover in nonsyntenic genes (Soltis et al. 2012). It is known, for example, that large structural rearrangements in the maize genome can reduce crossover frequency (Rhoades and Dempsey 1953; Rodgers-Melnick et al. 2015). Perhaps more relevant is the finding that crossovers are suppressed in genomic regions of mice that are polymorphic for copy number variation (Morgan et al. 2017). Furthermore, crossover frequency is suppressed in maize genes that are hemizygous for transposon insertions (Yao et al. 2002; Dooner and He 2008). Hence, it is reasonable to hypothesize that the reduced sequence conservation surrounding nonsyntenic genes directly contributes to their reduced crossover frequency.

Materials and Methods

Genetic Stocks and Illumina RNA-Seq

Tissue was collected from apices of 10 day old seedlings of B73, Mo17, and 105 RILs extracted from the Internated

B73 × Mo17 (IBM) Syn4 population (Lee et al. 2002 and subjected to single-end, 75 bp RNA-sequencing, Li et al. 2013). Raw reads (between 18.8 and 46.8 million per RIL) were scanned for low quality bases with PHRED quality values of 15 (out of 40) which were removed from each end of each read. The remaining nucleotides were then scanned using overlapping windows of 10 bp and sequences beyond the last window with average quality value less than the specified threshold were truncated.

Alignment of Reads to the Reference Genome

Trimmed reads (96–98% of the raw reads per RIL) were aligned to the B73 reference genome version 2 (RefGen2) using GSNAP (Wu and Nacu 2010), which allows for gapped alignments, including intron-spanning alignments. Reads were retained if they mapped uniquely in the genome, allowing two or less mismatches every 36 bp and fewer than five bases for every 75 bp in read length as unaligned “tails” (69–80% of the trimmed reads per RIL). The read depth of each gene was computed based on the coordinates of mapped reads and annotated transcript start and stop locations of genes in the reference genome (ZmB73_5b_FGS).

SNP Discovery

The coordinates of unique alignments were used for SNP discovery. Polymorphisms at each potential SNP site were examined and putative homozygous SNPs were identified after ignoring the first and last three aligned bases of each read. A polymorphic base was required to have a PHRED base quality value of at least 20, and at least five unique reads supported the SNP call.

Filtering to Extract Segregating SNPs

The IBM RILs are expected to be segregating 1:1 for each allele if both alleles are expressed. To obtain a set of confident SNPs that were expressed in the majority of RILs and were segregating in the population, we required that at least 20 RILs exhibited the B73 genotype and at least 20 RILs exhibited the Mo17 genotype. To remove SNPs showing extreme segregation distortion, we also required that the ratio of the number of RILs with each of the two alleles not exceed 2.5. We further filtered any segregating SNPs that were not discovered in the Mo17 RNA-seq. After filtering, 176,279 SNPs remained.

Segmentation to Identify Chimeric Chromosomal Structure of Each RIL

The scores of each filtered SNP marker were converted to 1 (B73 genotype) or 0 (Mo17 genotype). The converted binary data were subjected to the segmentation with the R package DNACopy (Olshen et al. 2004) using the following parameters: $\alpha = 0.01$, $nperm = 10000$, $p.method = "perm"$, $\eta = 0.01$, $min.width = 3$. These parameters required each generated segment to contain at least three SNP markers and a median absolute deviation equal to 0. We also required each segment to be at least 200 kb and the mean of each segment is ≥ 0.9 or ≤ 0.1 for the B73 segment and Mo17 segment, respectively. During segmentation, any segments > 2 Mb that were adjacent to a segment of the

same genotype were merged into a single segment. This generated 8,624 segments summed across the 105 RILs. Segments from all RILs were merged, and unique marker positions across the RILs were used to define the beginning and end of 7,867 segmental markers. A merged segment contained only a single genotype in each RIL: B73, Mo17, or recombinant. Adjacent recombinant segments are termed recombinant breakpoint intervals (RBI).

Constructing the Genetic Map

The software MSTmap (<http://alumni.cs.ucr.edu/~yonghui/mstmap.html>, last accessed September 11, 2018) was used to build a genetic map with the 7,867 segmental markers with B73 and Mo17 genotypes (recombinant genotypes were coded as missing) with the following parameters: `population_type DH`, `distance function kosambi`, `cut_off_p_value 0.000001`, `no_map_dist 15.0`, `no_map_size 2`, `missing_threshold 0.6`, `estimation_before_clustering yes`, `detect_bad_data yes`, and `objective_function COUNT`. A total of 13 markers had large disagreements in the physical and genetic positions and were removed from the segmental marker file. We then reran MSTmap with the filtered 7,854 segmental markers with the same parameters. All 7,854 segmental markers remained in the constructed genetic map, which contained 11 linkage groups. Two separated groups in which markers were all from chromosome 5 were concatenated.

Exploration of the Relationship between Crossover and Syntenic Retention

Genes within syntenic blocks (syntenic genes) between maize and sorghum have previously been determined (Schnable et al. 2011). Each chromosome of the B73 reference genome was divided into nonoverlapping 1 Mb bins. The total syntenic and nonsyntenic gene space in each bin was determined (ZmB73_5b_FGS). The amount of recombination in each bin representing primarily crossovers was determined by inferring the genetic start and end positions of each using a GAM function (Liu et al. 2009). The genetic distance between the start point and the end point was used to represent the crossover frequency of each bin.

Identification of Recombinant Alleles

The segmentation results from each individual RIL were used to identify RBIs. 7,574 RBIs were identified across the RILs. The number of RBIs is lower than the number of segmental markers because an RBI may contain multiple segments due to missing data. The sizes of RBIs vary depending on the surrounding informative SNP markers. Intragenic crossovers were defined as RBIs that occurred within a gene (ZmB73_5b_FGS). In this case, genes were defined as including an additional 1 kb of sequence upstream and downstream to include regulatory elements. Recombinant alleles were defined as alleles of these genes that include both B73 and Mo17 sequences. Most RBIs overlap with a gene (7,535/7,574) and a substantial fraction map to within a gene (848/7,574). Some RBIs occur within more than one overlapping gene. These RBIs were only counted once.

Enrichment of RBIs in hotspot genes (genes with two or more intragenic recombination events, $N = 13,298$) was tested by permutation. Only genes with two or more SNPs were included as potential genes with intragenic crossover. The probability of a gene having an RBI was determined in two ways. First, the probability was calculated as the distance between the first and last SNP in a gene divided by the sum of this distance for all genes. This method assumes crossovers are distributed equally across a gene. However, we show that crossovers are enriched at the 5' end of genes. Therefore, the second probability was modified in that any base pairs in the 5' end of a gene were counted twice, and the sum of the distance between all SNPs was adjusted accordingly. Random genes from this filtered set were selected based on the probability of each gene having an RBI without replacement for each RIL based on the number of intragenic RBIs observed in that RIL. The number of random genes with an RBI event and the number of random hotspot genes were recorded. This process was repeated 1,000 times. The P -value of the proportion of hotspot genes was determined as the proportion of random hotspot genes from all random genes with intragenic crossovers greater than the proportion of observed hotspot genes from all observed genes with intragenic crossover.

Transgressive Transcript Accumulation of Recombinant Alleles

The transcript accumulation of each recombinant allele in a particular RIL was compared with the transcript accumulations of the nonrecombinant alleles contained in other RILs and the two parental inbreds, B73 and Mo17. Transgressive transcript accumulation of recombinant alleles was defined as recombinant alleles that exhibit $>$ or $<5\%$ of the transcript accumulation of the most extreme nonrecombinant alleles. To determine if transgressive transcript accumulation occurs by chance, permutations were performed 1,000 times. For all genes with one or more recombinant alleles, the transcript accumulation of an equal number of alleles randomly selected from all RILs was tested for transgressive transcript accumulation. The number of random alleles with transgressive transcript accumulation for each gene was recorded and used as the null distribution. The observed number of genes with transgressive transcript accumulation from the real data was compared with this null distribution to obtain a P -value. The same process was performed to test for enrichment of up or down transgressive transcript accumulation separately.

Discovery of GC Events

To discover a set of possible GC events, the nonreciprocal exchange of genetic information, the SNPs with $<5\%$ missing data were further filtered. Identification of aGC events is highly dependent on missing data in a particular RIL; thus, only the 97 RILs with $>100k$ genotyped SNPs were included in this analysis. The genotype of each SNP was compared with the genotype of the segment at which the SNP was located. A SNP showing the alternate genotype to the corresponding segment is a SNP affected by an aGC. aGCs that were observed in >5 RILs were removed as they may be due to

systematic genotyping or sequencing errors. To obtain a consensus genotype for aGCs, the aGCs of each RIL were subjected to segmentation using DNACopy software with the criteria ($\alpha = 0.01$, $nperm = 10000$, $p.method = "perm"$, $\eta = 0.01$, $min.width = 2$). The input data for DNACopy are 1 and 0, representing aGC and not aGC of each of the SNP markers, respectively. Clusters of aGCs occurring in blocks >2 kb were excluded as these likely represent either double crossovers or regions incorrectly placed within the current maize pseudomolecules. The rate of GC per generation was calculated as the number of markers with an aGC divided by the total number of successfully genotyped markers divided by the number of generations. As RILs become more inbred, aGCs are less likely to be detected. Thus the number of generations in the RILs was modified as the number of rounds of intermating ($N = 4$) plus a modified value for the number of generations of selfing calculated as $1/2^i$ where i is the number of selfing generations as described by Shi et al. (2010) ($N = 2$) for a total of six generations.

The permutation tests to check for enrichment of GC events in certain genes was performed on RILs with $<5\%$ missing data and a filtered SNP set that only contains SNPs with a Mo17 genotype and a nearby SNP within 2 kb. Genes with SNPs that met these criteria were selected as genes with the potential for aGC. Permutations were performed by selecting a random set of genes equal to the observed number of genes with aGC per RIL. The number of unique genes randomly selected across RILs was determined and compared with the actual number of unique genes identified with GC in all RILs. This process was repeated 1,000 times.

The mutation rate of the population was estimated using a set of 97 RILs having low missing data (see below). Only SNP sites with $<5\%$ missing data and for which the Mo17 and B73 inbreds carry the same allele were used for this estimate. RILs in which a nonB73 allele was identified at a site and was present in ≤ 5 RILs were putative mutations. The mutation rate was calculated as the number of putative mutations divided by the total number of genotyped sites with enough reads to make a SNP call ($32,105/2,974,824,916 = 1.08 \times 10^{-5}$). For mutation to cause a GC-like event at least two sites must be mutated (1.08×10^{-5} squared) and each site must be mutated to the same allele (multiplied by $1/2$ squared). This estimate assumes mutations occur independently, which may not be accurate, and also does not take into account the fact that the two mutations must be adjacent for a GC event to be called and should thus be treated as a rough estimate of the improbability of mutations causing an aGC event in this population.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work is supported by National Science Foundation grants (IOS-0820610, IOS-1238142). Any opinions, findings, and conclusions or recommendations expressed in this

material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. 2007. SNP discovery via 454 transcriptome sequencing. *Plant J.* 51(5):910–918.
- Biłan R, Szafran K, Hnatuszko-Konka K, Kononowicz AK. 2016. Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tissue Organ Cult.* 127(2):269–287.
- Brinkman M, Frey KJ. 1977. Yield-component analysis of oat isolines that produce different grain yields. *Crop Sci.* 17(1):165–168.
- Brown J, Sundaresan V. 1991. A recombination hotspot in the maize A1 intragenic region. *Theor Appl Genet.* 81(2):185–188.
- Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet.* 45(11):1327–1336.
- Civardi L, Xia YJ, Edwards KJ, Schnable PS, Nikolau BJ. 1994. The relationship between genetic and physical distances in the cloned A1-Sh2 interval of the *Zea-mays* L genome. *Proc Natl Acad Sci USA.* 91(17):8268–8272.
- Comai L. 2000. Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol Biol.* 43(2–3):387–399.
- Dooner HK, He LM. 2008. Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. *Plant Cell.* 20(2):249–258.
- Dooner HK, He LM. 2014. Polarized gene conversion at the bz locus of maize. *Proc Natl Acad Sci USA.* 111(38):13918–13923.
- Dooner HK, Martínez-Férez IM. 1997. Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *Plant Cell.* 9(9):1633–1646.
- Eggleston WB, Alleman M, Kermicle JL. 1995. Molecular organization and germinal instability of R-Stippled maize. *Genetics* 141(1):347–360.
- Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, et al. 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell.* 25(8):2783–2797.
- Eichten SR, Ellis NA, Makarevitch I, Yeh C-T, Gent JI, Guo L, McGinnis KM, Zhang X, Schnable PS, Vaughn MW, et al. 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. *PLoS Genet.* 8(12):e1003127.
- Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu SZ, Yeh CT, Jia Y, Gendler K, Freeling M, et al. 2011. Heritable epigenetic variation among maize inbreds. *PLoS Genet.* 7(11):e1002372.
- 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.
- Fogel S, Hurst D. 1967. Meiotic gene conversion in yeast tetrads and the theory of recombination. *Genetics* 57(2):455.
- Freeling M. 1976. Intragenic recombination in maize: pollen analysis methods and the effect of parental adh1 isoalleles. *Genetics* 83(4):701–717.
- Fu HH, Park WK, Yan XH, Zheng ZW, Shen BZ, Dooner HK. 2001. The highly recombinogenic bz locus lies in an unusually gene-rich region of the maize genome. *Proc Natl Acad Sci USA.* 98(15):8903–8908.
- Fu Y, Wen TJ, Ronin YI, Chen HD, Guo L, Mester DI, Yang YJ, Lee M, Korol AB, Ashlock DA, et al. 2006. Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. *Genetics* 174(3):1671–1683.
- Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. 2007. Opinion – recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet.* 8(1):77–84.
- Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mezard C. 2011. Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. *PLoS Genet.* 7(11):e1002354.
- Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, et al. 2009. A first-generation haplotype map of maize. *Science* 326(5956):1115–1117.
- Guillon H, Baudat F, Grey C, Liskay RM, de Massy B. 2005. Crossover and noncrossover pathways in mouse meiosis. *Mol Cell.* 20(4):563–573.
- Hagblom P, Segal E, Billyard E, So M. 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria-Gonorrhoeae*. *Nature* 315(6015):156–158.
- He Y, Wang M, Dukowic-Schulze S, Zhou A, Tiang CL, Shilo S, Sidhu GK, Eichten S, Bradbury P, Springer NM, et al. 2017. Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. *Proc Natl Acad Sci USA.* 114(46):12231–12236.
- Jiao YN, Wickett NJ, Ayyampalayam S, Chanderali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–U113.
- Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467(7319):1099–1103.
- Korunes KL, Noor MAF. 2017. Gene conversion and linkage: effects on genome evolution and speciation. *Mol Ecol.* 26(1):351–364.
- Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A. 2002. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Mol Biol.* 48(5–6):453–461.
- Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, et al. 2013. Mendelian and non-Mendelian regulation of gene expression in maize. *PLoS Genet.* 9(1):e1003202.
- Li MY, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG. 2011. Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333(6038):53–58.
- Li X, Li L, Yan JB. 2015. Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. *Nat Commun.* 6(1):6648.
- Lin H-y, Liu Q, Li X, Yang J, Liu S, Huang Y, Scanlon MJ, Nettleton D, Schnable PS. (Lin2017 co-authors). 2017. Substantial contribution of genetic variation in the expression of transcription factors to phenotypic variation revealed by eRD-GWAS. *Genome Biol.* 18(1):192.
- Liu S, Ying K, Yeh CT, Yang J, Swanson-Wagner R, Wu W, Richmond T, Gerhardt DJ, Lai J, Springer N, et al. 2012. Changes in genome content generated via segregation of non-allelic homologs. *Plant J.* 72(3):390–399.
- Liu S, Yeh CT, Ji TM, Ying K, Wu HY, Tang HM, Fu Y, Nettleton D, Schnable PS. 2009. Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.* 5(11):e1000733.
- Lu PL, Han XW, Qi J, Yang JG, Wijeratne AJ, Li T, Ma H. 2012. Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing *Landsberg erecta* and all four products of a single meiosis. *Genome Res.* 22(3):508–518.
- Malone RE, Kim S, Bullard SA, Lundquist S, Hutchings-Crow L, Cramton S, Lutfiyya L, Lee J. 1994. Analysis of a recombination hotspot for gene conversion occurring at the HIS2 gene of *Saccharomyces cerevisiae*. *Genetics* 137(1):5–18.
- McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangi JL. 1998. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of arabidopsis. *Plant Cell.* 10(11):1861–1874.
- Morgan AP, Gatti DM, Najarian ML, Keane TM, Galante RJ, Pack AI, Mott R, Churchill GA, de Villena FPM. 2017. Structural variation shapes the landscape of recombination in mouse. *Genetics* 206(2):603–619.
- Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat Res.* 1(1):2–9.
- 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(5746):321–324.

- Ogasawara K, Yabe R, Uchikawa M, Bannai M, Nakata K, Takenaka M, Takahashi Y, Juji T, Tokunaga K. 1998. Different alleles cause an imbalance in A2 and A2B phenotypes of the ABO blood group. *Vox Sang*. 74(4):242–247.
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4):557–572.
- Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144(5):719–731.
- Pan QC, Li L, Yang XH, Tong H, Xu ST, Li ZG, Li WY, Muehlbauer GJ, Li JS, Yan JB. 2016. Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytol*. 210(3):1083–1094.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberler G, Hellsten U, Mitros T, Poliakov A, et al. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556.
- Patterson GI, Kubo KM, Shroyer T, Chandler VL. 1995. Sequences required for paramutation of the maize B-gene map to a region containing the promoter and upstream sequences. *Genetics* 140(4):1389–1406.
- Piskol R, Peng ZY, Wang J, Li JB. 2013. Lack of evidence for existence of noncanonical RNA editing. *Nat Biotechnol*. 31(1):19–20.
- Qi J, Chen Y, Copenhaver GP, Ma H. 2014. Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proc Natl Acad Sci USA*. 111(27):10007–10012.
- Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J. 2017. Gene fractionation and function in the ancient subgenomes of maize. *Mol Biol Evol*. 34(8):1825–1832.
- Rhoades MM, Dempsey E. 1953. Cytogenetic studies of deficient-duplicate chromosomes derived from inversion heterozygotes in maize. *Am J Bot*. 40(6):405–424.
- 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. *Proc Natl Acad Sci USA*. 112(12):3823–3828.
- Salome PA, Bombliès K, Fitz J, Laitinen RA, Warthmann N, Yant L, Weigel D. 2012. The recombination landscape in Arabidopsis thaliana F2 populations. *Heredity (Edinb)* 108(4):447–455.
- Schnable JC. 2015. Genome evolution in maize: from genomes back to genes. *Annu Rev Plant Biol*. 66:329–343.
- Schnable JC, Freeling M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE*. 6(3):e17855.
- Schnable JC, Freeling M, Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol*. 4(3):265–277.
- Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci USA*. 108(10):4069–4074.
- Schnable PS, Hsia AP, Nikolau BJ. 1998. Genetic recombination in plants. *Curr Opin Plant Biol*. 1(2):123–129.
- Schnable PS, Springer NM. 2013. Progress toward understanding heterosis in crop plants. *Annu Rev Plant Biol*. 64:71–88.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956):1112–1115.
- Schultes NP, Szostak JW. 1990. Decreasing gradients of gene conversion on both sides of the initiation site for meiotic recombination at the Arg4 locus in yeast. *Genetics* 126(4):813–822.
- Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2014. Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. *BMC Genomics*. 15(1):701.
- Shi JH, Wolf SE, Burke JM, Presting GG, Ross-Ibarra J, Dawe RK. 2010. Widespread gene conversion in centromere cores. *PLoS Biol*. 8(3):e1000327.
- Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. 2015. DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in Arabidopsis. *Plant Cell*. 27(9):2427–2436.
- Soltis DE, Buggs RJ, Barbazuk WB, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS. 2012. The early stages of polyploidy: rapid and repeated evolution in *Tragopogon*. In: Polyploidy and genome evolution. Berlin, Heidelberg: Springer. p. 271–292.
- Swigonova Z, Lai JS, Ma JX, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. 2004. Close split of sorghum and maize genome progenitors. *Genome Res*. 14(10a):1916–1923.
- Wang ZH, Treder K, Miller WA. 2009. Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. *J Biol Chem*. 284(21):14189–14202.
- Wijnker E, James GV, Ding J, Becker F, Klasen JR, Rawat V, Rowan BA, de Jong DF, de Snoo CB, Zapata L, et al. 2013. The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* 2:e01426.
- Wijnker E, van Dun K, de Snoo CB, Lelivelt CLC, Keurentjes JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R. 2012. Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. *Nat Genet*. 44(4):467–U156.
- Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol*. 8(6):e1000409.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.
- Xu XJ, 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(12):2151–2161.
- Yandeau-Nelson MD, Zhou Q, Yao H, Xu XJ, Nikolau BJ, Schnable PS. 2005. MuDR transposase increases the frequency of meiotic crossovers in the vicinity of a Mu insertion in the maize a1 gene. *Genetics* 169(2):917–929.
- 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. *Proc Natl Acad Sci USA*. 99(9):6157–6162.