Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize

Yan Heab,1, Minghui Wangac,1, Stefanie Dukowich-Schulzeab,1, Adelle Zhoua, Choon-Lin Tianga, Shay Shiloa, Gaganpreet K. Sidhuab,2, Steven Eichtenb, Peter Bradburyc, Nathan M. Springerc, Edward S. Bucklerab,9, Avraham A. Levyab, Qi Suna, Jaroslav Pillardy, Penny M. A. Kianian, Shahryar F. Kianian, Changbin Chenf, and Wojciech P. Pawlowskig,3

aSchool of Integrative Plant Science, Cornell University, Ithaca, NY 14853; bNational Maize Improvement Center, China Agricultural University, Beijing 100083, China; cBioinformatics Facility, Cornell University, Ithaca, NY 14853; dDepartment of Horticultural Science, University of Minnesota, St. Paul, MN 55108; ePlant and Environmental Sciences Department, The Weizmann Institute of Science, Rehovot 76100, Israel; fDepartment of Plant Biology, University of Minnesota, St. Paul, MN 55108; gUS Department of Agriculture–Agricultural Research Service, Ithaca, NY 14853; and hCereal Disease Laboratory, US Department of Agriculture–Agricultural Research Service, St. Paul, MN 55108

Edited by Joachim Messing, Rutgers University, Piscataway, NJ, and approved September 27, 2017 (received for review July 26, 2017)

Meiotic recombination is the most important source of genetic variation in higher eukaryotes. It is initiated by formation of double-strand breaks (DSBs) in chromosomal DNA in early meiotic prophase. The DSBs are subsequently repaired, resulting in crossovers (COs) and noncrossovers (NCOs). Recombination events are not distributed evenly along chromosomes but cluster at recombination hotspots. How specific sites become hotspots is poorly understood. Studies in yeast and mammals linked initiation of meiotic recombination to active chromatin features present upstream from genes, such as absence of nucleosomes and presence of trimethylation of lysine 4 in histone H3 (H3K4me3). Core recombination components are conserved among eukaryotes, but it is unclear whether this conservation results in universal characteristics of recombination landscapes shared by a wide range of species. To address this question, we mapped meiotic DSBs in maize, a higher eukaryote with a large genome that is rich in repetitive DNA. We found DSBs in maize to be frequent in all chromosome regions, including sites lacking COs, such as centromeres and pericentromeric regions. Furthermore, most DSBs are formed in repetitive DNA, predominantly Gypsy retrotransposons, and only one-quarter of DSB hotspots are near genes. Genic and nongenic hotspots differ in several characteristics, and only genic DSBs contribute to crossover formation. Maize hotspots overlap regions of low nucleosome occupancy but show only limited association with H3K4me3 sites. Overall, maize DSB hotspots exhibit distribution patterns and characteristics not reported previously in other species. Understanding recombination patterns in maize will shed light on mechanisms affecting dynamics of the plant genome.

Significance

Meiotic recombination is a process in plants, animals, and fungi during which chromosomes exchange their parts. It generates new genetic variation in the progeny and is one of the reasons why progeny are both similar to and different from their parents. Recombination is initiated by formation of breaks in chromosomal DNA. We generated a high-resolution map of sites where these breaks are formed in the genome of maize. Surprisingly, we found that DNA breaks are abundant in all genome regions, including sites where recombination was thought to be limited, such as repetitive DNA. The map will allow understanding of how recombination patterns shape the genome and aid development of more efficient breeding methods.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

Data deposition: RAD51 ChIP sequence data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=https%3A%2F%2Fhpf%3Dyssznqrcj&acc=GSE53701 (accession no. GSE53701); H3K4me3 ChIP and RNA sequence data have been deposited in the NCBI database, https://www.ncbi.nlm.nih.gov/biocad/biocad_sleri/link?name=bioprole&bio_sample_all&from_id=1685187 (accession nos. SAMN01884257-SAMN01884260) (H3K4me3 ChIP), and SAMN01889176 and SAMN01889179 (RNA); nucleosome occupancy sequence data have been deposited in the NCBI GEO database, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54968 (accession no. GSE54968).

1Y.H., M.W., and S.D.-S. contributed equally to this work.
2Present address: Institute for Cancer Genetics, Columbia University Medical Center, New York, NY 10032.
3To whom correspondence should be addressed. Email: wp45@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713225114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1713225114

PNAS | November 14, 2017 | vol. 114 | no. 46 | 12231–12236
intermediates but utilizes a synthesis-dependent strand annealing mechanism (14, 15). Some NCOs are also formed through alternative resolution of double Holliday junctions (16, 17). Repair of the majority of meiotic DSBs leads to NCOs. In maize male meiocytes, there are ~500 DSBs, yet their repair leads to fewer than ~20 COs (18). The distribution of DSB hotspots may not necessarily mirror the CO hotspot distribution. In budding yeast and mice, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28).

**Results**

**DSB Hotspots Are Ubiquitous in All Chromosome Regions.** To generate a map of DSB sites, we used a chromatin immunoprecipitation (ChIP) approach in which chromatin from flowers containing early stages of meiocytes was enriched in fragments associated with the RAD51 protein. RAD51 forms distinct foci on chromosomes at the sites of meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28). In maize, the CO hotspot distribution is PRDM9, a histone methyl transferase that initiates meiotic DSBs, facilitating their repair (12, 28).

DNA fragments recovered from ChIP were sequenced on the Illumina platform, and the sequence reads were mapped to the reference genome (26). To identify sites in the genome occupied by RAD51, we compared the distribution of RAD51 ChIP-seq reads to the following: (i) meiotic chromatin that was not subjected to ChIP (input chromatin); (ii) ChIP products generated from meiotic chromatin using preimmune IgG, instead of the RAD51 antibody; and (iii) ChIP products generated from leaf tissue chromatin using the anti-RAD51 antibody. As expected, the two latter treatments yielded very low amounts of DNA (~22% and ~5% of those recovered in RAD51 ChIP on meiotic flowers, respectively). Using a conservative false discovery rate of <0.01, we identified 3,126 meiotic RAD51 peaks across the maize genome (SI Appendix, Fig. S2).

We utilized a recombination-deficient mutant to validate the 3,126 peaks that we identified at meiotic DSBs. As a spo11 mutant in maize has not been described previously, we employed the phs1 mutant, which is deficient in the RAD51 homolog (9). The phs1 mutant lacks RAD51 foci; instead of the 500 foci observed in the wild type, phs1 meiocytes show only about three RAD51 foci per cell (29). Very little DNA was recovered in the phs1 ChIP experiments, ~2% of the amount obtained in the wild type. Furthermore, we found only about 60 RAD51 peaks in the ChIP-seq data in phs1, with none of them at the same location as in the wild type. Thus, we concluded that the 3,126 wild-type peaks indeed represent hotspots of meiotic DSB formation.

We found that DSB hotspots were, on average, 1.2 kb long (SI Appendix, Fig. S3A) and exhibited relatively random spacing (SI Appendix, Fig. S3B). The number of hotspots per chromosome was strongly correlated with chromosome length (r = 0.95; P < 2.2 e^-16) (SI Appendix, Fig. S3C). Hotspots were present in all chromosome regions, including centromeric and pericentromeric locations (Fig. 1C and SI Appendix, Fig. S4). Despite constituting a large fraction of the genome, most repetitive elements in maize are unique sequence-wise due to mutations as well as transposon insertions into repetitive DNA, which create unique sequence borders (30). Due to the substantial sequence diversity, maize centromeres are well assembled (31), even though they consist predominantly of tandemly arranged CentC repeats interspersed with retrotransposons (31).

Finding DSBs in centromeres/pericentromeres was interesting, as these regions lack COs (3, 27). To confirm the DSB presence
at centromeres, we conducted experiments combining the anti-RAD51 antibody with a fluorescence in situ hybridization (FISH) probe detecting the maize centromere repeat (29). These experiments demonstrated the presence of RAD51 foci on centromeres in midzygotene meiocytes (Fig. 1D).

Another type of highly repetitive region in which we found evidence of meiosis-specific DSBs was rRNA loci (Fig. 2 and SI Appendix, Figs. S5 and S6). This finding was also not anticipated as rRNA loci have been reported to be excluded from DSB formation in yeast (32).

**Most DSB Hotspots Are in Repetitive DNA.** To further dissect the distribution of DSB hotspots, we examined their genomic context. We found that the majority of DSB hotspots were located in repetitive DNA, primarily Gypsy retrotransposons (Fig. 2 and SI Appendix, Fig. S5). To determine if there were any specific regions within Gypsy in which DSB hotspots were more prevalent, we aligned the hotspots in Gypsy elements to sequences of the ∼220 Gypsy retrotransposon types identified in maize (33). About 77% of Gypsy-located hotspots primarily matched internal retrotransposon regions, while the remaining 23% primarily matched the long terminal repeats (LTRs). However, when the relative length of LTRs versus the internal region was taken into account, DSB hotspots appeared to be 2.4-fold more frequent in LTRs than in the internal region.

**DSB Hotspot and Chromatin State.** To search for factors controlling DSB location, we examined chromatin features of DSB hotspot regions. To do this analysis, we first investigated the positions of maize hotspots relative to open chromatin sites marked by low nucleosome occupancy. This experiment was conducted using microccoral nuclease digestion of chromatin extracted from meiotic anthers. We found a strong association between hotspot presence and nucleosome-free chromatin for both hotspots located in genic regions and those in repetitive DNA (Fig. 3A).

We also examined H3K4me3 as another marker of active chromatin. However, we discovered that only about 20% of hotspots overlapped with H3K4me3 sites detected using ChIP with an anti-H3K4me3 antibody on chromatin from isolated zygotene meiocytes (Fig. 3B). We confirmed this conclusion using cytological immunolocalization, which showed that only about 5% of RAD51 foci at midzygotene overlapped with H3K4me3 foci (Fig. 3C, n = 20). As immunolocalization detects a much smaller number of H3K4me3 sites than ChIP, these two results are essentially in agreement in showing limited colocalization of RAD51 and H3K4me3 sites. We then examined the association of ChIP H3K4me3 sites with DSB hotspots located in different genome contexts. Overlap with H3K4me3 was low (∼5%) for DSB hotspots located in repetitive DNA but higher for those in genic regions (∼55%) (Fig. 3B).

Since cytosine methylation has been shown to affect the CO landscape in plants (3, 23, 34–38), we examined the DNA methylation status of DSB hotspot sites. To do this, we analyzed total levels of C methylation in a 2-kb window around DSB hotspot centers. We found that DNA methylation was dramatically reduced at DSB hotspots (P = 0) (Fig. 3D). We also examined a set of 104 maize CO events mapped with a resolution of less than 2 kb (SI Appendix, SI Methods) and found a similarly significant reduction (P < 2.2e-16).

Finally, to examine the association of DSBs with heterochromatin, as suggested by the large fraction of DSB hotspots located in repetitive DNA, we conducted immunocolocalization experiments with RAD51 and a heterochromatin marker, H3 histone lysine 9 dimethylation (H3K9me2) (39). We found that, at zygotene, about 5% of RAD51 foci colocalized with H3K9me2-marked heterochromatin (Fig. 3E). Overall, our data indicate that DSBs in maize are formed at nucleosome-free and DNA-hypomethylated sites. However, the larger context of DSBs varies and only a minority of DSBs overlap with sites marked by H3K4me3. Even though a large fraction of DSB hotspots are located within transposons, they are present in nucleosome-free regions of transposon DNA rather than in condensed heterochromatin.

**Hotspot Presence in Genes Is Not Correlated with Transcription.** About 26% of maize DSB hotspots were in single-copy genic DNA. However, taking into account that genic DNA amounted to only about 14% of the Illumina reads that could be aligned to the maize genome, the relative frequency of hotspots was still higher in genic regions than in repetitive DNA (SI Appendix, Fig. S5). Within genes, there were two strong DSB formation peaks, located around the transcription start and transcription termination sites (Fig. 4A).

Based on the association between hotspots and active chromatin, we hypothesized that DSB formation in genic DNA may take place in or near highly expressed genes, as gene expression is generally associated with active chromatin. To test this hypothesis, we examined the transcriptome of isolated male
Maize Hotspot Sequence. Maize DSB hotspot regions showed a higher GC content than the average for the Illumina reads that could be aligned to the maize genome (SI Appendix, Fig. S7). Further investigations of DNA sequence of DSB hotspots identified a 20-bp-long GC-rich degenerate DNA sequence motif (named MHS, Maize Hotspot Sequence) present in about 72% of genic hotspots (Fig. 5A). In contrast, we have not been able to find a statistically significant motif in repetitive DNA hotspots.

By examining the dataset of 104 maize high-resolution-mapped CO events (SI Appendix, SI Methods), we detected a sequence motif associated with CO sites with similarity to MHS (Fig. 5B). The maize CO motif was also significantly similar (P = 2.25e-6) to a previously identified CO motif in Arabidopsis (38). MHS copies are distributed along the entire chromosome length but tend to be more frequent near telomeres (SI Appendix, Fig. S8). The RAD51 signal was enriched around MHS (SI Appendix, Fig. S9). However, presence of MHS alone is insufficient for DSB hotspot formation, as the B73 genome contains 27,371 sequences that match the MHS sequence consensus at 95% identity or more. To examine if there are differences in MHS copy numbers among maize inbreds, we generated whole-genome Illumina sequences of two other inbreds, Mo17 and CML228. The Mo17 genome contained 28,947 MHS copies, while the CML228 had 6,614 copies. Interestingly, CML228 exhibits much lower DSB and CO numbers than B73 or Mo17 (18), suggesting that the MHS copy number may be a determinant of recombination frequencies.

An intriguing question is what distinguishes MHS copies associated with hotspots (hotspot copies) from those that are not (nonhotspot copies). As MHS is cytosine-rich, we examined the DNA methylation status of the motif. We analyzed the three types of DNA methylation present in plants, CG, CHG, and CHH (where H is any nucleotide other than G), which are controlled by distinct genetic pathways (40). CG and CHG methylation levels were lower at hotspot copies compared with nonhotspot copies (Fig. 5C). The high-GC methylation levels of nonhotspot copies suggested that they were located in heterochromatin. Some DSB formation was present at MHS copies in methylated DNA, although it was not sufficient to result in formation of detectable hotspots (SI Appendix, Fig. S10). However, this observation suggests that the DNA methylation level of MHS might be a regulator of hotspot strength. The presence of DSB formation in regions of reduced DNA methylation implies sensitivity of the DSB machinery to methylation, perhaps involving SPO11, as it was recently found in mice (41). Altering DNA methylation levels could activate new DSB hotspots, which may provide a mechanistic explanation for the observations that reduced DNA methylation alters CO patterns in plants (34–37).

Only DSBs in Genic Regions Are Likely to Contribute to CO Formation. As the number of meiotic DSBs in maize is over 25-fold higher than the number of COs (28), most DSBs do not become CO sites. To examine how DSB distribution corresponds to CO distribution, we compared the DSB hotspot map to a CO map based on recombination data from the maize Nested Association Mapping (42). This comparison showed that CO distribution in maize does not follow the distribution of DSBs (r = 0.045; SI Appendix, Fig. S11). However, the correlation was much stronger (r = 0.41; SI Appendix, Fig. S11) and similar to the one reported in mouse (r = 0.46–0.64) (1) when only genic DSB hotspots were considered. CO site and MHS distributions were also strongly correlated (r = 0.72; SI Appendix, Fig. S11). These analyses imply that COs in maize are predominantly drawn from DSBs formed in genic regions while DSBs formed in repetitive DNA do not result in COs. Previous studies showing that COs in maize are generally formed in gene-rich regions (3, 27, 43) are consistent with this conclusion.

Of note, we found a strong DSB hotspot upstream from the bronzel (hcl) locus on chromosome 9 (SI Appendix, Fig. S12). hcl is the location of one of the few known maize CO hotspots, which exhibits CO frequencies over 100-fold higher than the genome average (44). Presence of this hotspot may be responsible for the high CO and gene conversion activities at this locus.

Discussion

DSB Hotspots in Maize Are Frequent in Repetitive DNA. We found a ubiquitous presence of DSB hotspots in centromeric and pericentromeric regions of maize chromosomes. As these regions are
The differing characteristics of DSBs in maize centromeric DNA (45, 46). Nevertheless, finding DSB hotspots at centromeres was surprising, as DSB formation in centromeric regions has not been reported in either yeast or mammals (1, 2, 21, 47). It is highly unlikely that these hotspots could be ChIP-seq mapping artifacts because (i) maize centromeres are well-assembled (31), which enables accurate mapping of ChIP-seq reads; and (ii) the ChIP-seq mapping data are corroborated by immuno-FISH experiments showing colocalization of RAD51 foci with centromeric repeats. Our data are consistent with previous reports of the presence of early recombination nodules in pericentromeric regions of maize chromosomes (48). Early nodules are cytological structures visible in transmission electron microscopy that correspond to RAD51/DMC1 protein complexes (48). Furthermore, signatures of recombination during centromere evolution have been found in rice, which is another species with well-assembled centromere region sequences (49).

We also discovered that the majority of DSBs were formed in retrotransposons in rice, which was unexpected as well. DSB formation in retrotransposons has been reported in budding yeast (50). Thus, the presence of a large fraction of DSBs in retrotransposons could be a reflection of the abundance of these elements in the maize genome (26). However, other maize transposons, such as Copia, LINE, and DNA transposons, exhibited fewer DSB hotspots relative to their ubiquity in the genome (SI Appendix, Fig. S5). Furthermore, yeast lack the highly condensed heterochromatin ubiquitous in maize (51), suggesting that DSB formation in retrotransposons in maize may differ at the mechanistic level from that in yeast.

**Two DSB Classes in Maize.** The differing characteristics of DSB hotspots located in genic vs. repetitive genome regions suggest that they represent two distinct classes of DSBs. DSB hotspots in genes were associated with the presence of MHS, while hotspots in repetitive DNA were not. Hotspots in genic regions were also more likely to overlap active chromatin sites marked with H3K4me3 than hotspots in repetitive DNA.

Differences in characteristics between genic and nongenic hotspots may reflect differences in how the two DSB classes are formed. They also likely affect the way in which DSBs are repaired, as mainly genic DSBs contribute to CO formation. It has been proposed that DSB fate may depend on the timing of DSB formation; DSBs made earlier are more likely to be repaired as COs than those made later (52, 53). DSBs in euchromatin regions may form earlier than those in heterochromatin and thus be more prone to result in COs.

**Mechanistic Underpinnings of DSB Hotspot Recognition in Maize.** Comparing recombination patterns in maize with those of yeast and mammals will be helpful for identifying universal recombination landscape features that are shared by a broad spectrum of species. Several general characteristics of maize DSB hotspot distribution patterns, such as hotspot width and spacing, are similar to those reported in mammals and yeast (1, 2, 21, 47). Association of DSB hotspots with nucleosome-free chromatin is also shared among species (2, 7), including maize hotspots located in genic regions and repetitive DNA. These commonalities suggest that there are conserved aspects of the DSB-targeting mechanism, which may have implications for genome evolution.

A striking difference in DSB distribution patterns between maize vs. yeast and mammals is the relatively low overlap of DSB hotspots with H3K4me3 sites in maize. In mice, the 93.9% of DSB-hotspot overlap with H3K4me3 (1) is a result of the action of PRDM9 (7). In PRDM9-lacking mouse mutants, DSBs are still produced but show altered distribution (54). These data suggest that PRDM9 functions primarily to sequester DSB formation away from 5′ regulatory regions of genes. In budding yeast, colocalization between DSB hotspots and H3K4me3 marks is also strong, but it may be a result of the compactness of the yeast genome rather than a causal relationship (24). Our data in maize suggest that the presence of H3K4me3 is not an absolute requirement for DSB formation.

**Methods**

**DSB Mapping Using RAD51 ChIP.** Polyclonal antibodies were raised in rabbits against a recombinant protein produced by expressing a full-length coding sequence of the zmRAD51A1 gene in Escherichia coli. Two rounds of immunoprecipitation were performed on chromatin extracted from male flowers at the zygotene stage of meiotic prophase I. Details of the experimental procedures are described in SI Appendix, SI Methods.

**H3K4me3 ChIP-Seq.** Male meiocytes in leptotene and zygotene were collected using the capillary collection of meiocytes method as previously described (58). Ten microliters of rabbit polyclonal anti-trimethyl-histone (Lys4) antibody (EMD Millipore) was used on chromatin extracted from ~30,000 meiocytes, following instructions of the MAGnify ChIP Kit (Invitrogen). Standard Illumina protocols were used for library construction and sequencing.

**Nucleosome Occupancy Mapping.** Nuclei were prepared from zygotene-stage anthers as described in the ChIP protocol except that EDTA was omitted from buffers. Details of the procedure are described in SI Appendix, SI Methods.

**Cytological Analyses.** Preparation of immunolocalization and immuno-FISH microscopic slides is described in detail in SI Appendix, SI Methods. The slides were examined using a DeltaVision imaging station (Applied Precision). Three-dimensional stacks of images were collected across the entire thickness of the specimen with optical sections 150 nm apart. The image stacks were subjected to constrained iterative deconvolution and analyzed with the softWoRx.
Software (Applied Precision). Colocalization analyses were conducted in 3D space to distinguish between actual colocalization and accidental overlap. For counting, protein foci were located automatically in the datasets by identifying local peak intensities in three dimensions.

**Computational Analyses.** A computational pipeline was developed to process ChIP-seq datasets. Details of the pipeline procedure, as well as protocols for peak calling, generating, and analysis of the DHS hotspot map, are described in SI Appendix, SI Methods. Standard procedures were used to identify DNA sequence motifs associated with the presence of DHS hotspots and COs.

**Acknowledgments.** We thank R. Kelly Dawe for advice on maize centromeres and Teresa Pawlickska and Wayne Crismani for comments on the manuscript. This research was supported by National Science Foundation Grants IOS-1025881 and IOS-1546792 (to W.P.P.) and by United States-Israel Binational Agricultural Research and Development Fund Grant US-4829-15 to W.P.P. and A.A.L.