PNAS RESEARCH ARTICLE | GENETICS OPEN ACCESS

Polymorphic transposable elements contribute to variation in recombination landscapes

Yuheng Huang[a], Zita Y. Gao[a], Kayla Ly[a], Leila Lin[a], Jan-Paul Lambooij[b], Elizabeth G. King[c], Aniek Janssen[b], Kevin H.-C. Wei[d], and Yuh Chwen G. Lee[a,1]

Affiliations are included on p. 10.

Edited by Douglas Koshland, University of California, Berkeley, CA; received January 2, 2025; accepted February 5, 2025

Meiotic recombination is a prominent force shaping genome evolution, and understanding why recombination rates vary within and between species has remained a central, though challenging, question. Variation in recombination is widely thought to influence the efficacy of selection in purging transposable elements (TEs), prevalent selfish genetic elements, leading to widely observed negative correlations between TE abundance and recombination rates across taxa. However, accumulating evidence suggests that TEs could instead be the cause rather than the consequence of this relationship. To test this prediction, we formally investigated the influence of polymorphic, putatively active TEs on recombination rates. We developed and benchmarked an approach that uses PacBio long-read sequencing to efficiently, accurately, and cost-effectively identify crossovers (COs), a key recombination product, among large numbers of pooled recombinant individuals. By applying this approach to Drosophila strains with distinct TE insertion profiles, we found that polymorphic TEs, especially RNA-based TEs and TEs with local enrichment of repressive marks, reduce the occurrence of COs. Such an effect leads to different CO frequencies between homologous sequences with and without TEs, contributing to varying CO maps between individuals. The suppressive effect of TEs on CO is further supported by two orthogonal approaches–analyzing the distributions of COs in panels of recombinant inbred lines in relation to TE polymorphism and applying marker-assisted estimations of CO frequencies to isogenic strains with and without transgenically inserted TEs. Our investigations reveal how the constantly changing TE landscape can actively modify recombination, shaping genome evolution within and between species.

transposable elements | recombination | epigenetic silencing | polymorphism | long-read sequencing

Significance

This study explores the cross-talk between two major drivers of genome evolution—meiotic recombination and transposable elements (TEs), selfish genetic elements that move within genomes. In many species, genomic regions with low recombination rates often contain more TEs. Previously, researchers thought this pattern existed because these regions are more permissive for TE accumulation. Here, we challenge this view by testing the converse hypothesis that TEs directly repress recombination. We developed a long-read pool-sequencing method to efficiently, accurately, and cost-effectively identify recombination events. Combining this with two other orthogonal approaches, we found that TEs suppress recombination and contribute to varying recombination landscapes between individuals. Our findings reveal that TEs can actively modify recombination, which influences genome evolution within and between species.

Homologous recombination, the exchange of genetic information between homologous chromosomes in sexually reproducing organisms, is a prominent force shaping genome evolution. During meiosis, recombination is initiated by programmed double-strand breaks (DSBs). When DSBs are repaired using homologous chromosomes and resolved into crossovers (COs), large blocks of homologous chromosomes are reciprocally exchanged, shuffling neighboring alleles into new combinations (1). This process allows for the purging of deleterious mutations (2, 3), facilitates adaptive evolution by bringing together beneficial variants (4, 5), and influences the efficacy of natural selection (6, 7). Given the critical evolutionary role of recombination, researchers have been studying its rate and distribution within genomes since the early days of genetics (8). However, robust estimation of these patterns of recombination is challenging and requires genotyping of large numbers of recombinant individuals (9). Despite this challenge, after decades of study, clear patterns have emerged: both the rates of recombination and the distribution of recombination events within a genome vary significantly across taxa (10), species (11), populations (12), and even among individuals of the same species (13). Understanding the mechanistic basis of this variability is essential for understanding how and why recombination landscapes evolve (9, 10, 14) and the consequential impact on genome evolution.

Variation in recombination is widely thought to influence the abundance and distribution of transposable elements (TEs) (15–17), prevalent selfish genetic elements found in nearly all eukaryotic genomes that can profoundly affect the function, fitness, and evolution of organisms (18, 19). In particular, the negative correlation between TE abundance and recombination rate within genomes is one of the few general, ubiquitously observed patterns of genome evolution across eukaryotes (reviewed in ref. 20). This prevalent association has largely been interpreted as a consequence of selection

Author contributions: Y.H. and Y.C.G.L. designed research; Y.H., Z.Y.G., K.L., L.L., J.-P.L., E.G.K., A.J., K.H.-C.W., and Y.C.G.L. performed research; Y.H., K.H.-C.W., and Y.C.G.L. contributed new reagents/analytic tools; Y.H., K.H.-C.W., and Y.C.G.L. analyzed data; and Y.H. and Y.C.G.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: grylee@uci.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2427312122/-/DCSupplemental.

Published March 18, 2025.

being unable to effectively purge TEs from genomic regions with low recombination rates (20–22). However, accumulating observations suggest that the causality might be reversed, and TE accumulation might cause, instead of result from, low recombination rates (20, 23). The potential suppressive impact of TEs on recombination is hinted by the widely observed low recombination rates in TE-infested pericentromeric heterochromatin (24–27). Within chromosome arms, clusters of TEs or high TE density are also less likely to co-occur with COs (28–30) or DSBs (31–33).

Several mechanisms could mediate the potential suppressive effects of TEs on COs. Polymorphic TEs are essentially insertions or deletions of several kilobases, and thus may have similar suppressive effects on COs as other structural DNA variants (34–36). In addition, as a consequence of host-directed epigenetic silencing of TEs, TEs within chromosome arms are oftentimes enriched with repressive epigenetic marks, particularly DNA methylation and H3K9me2/3 histone modifications (37). These repressive marks enriched at TEs could "spread" in *cis*, making TEs appear as islands of heterochromatin within the euchromatic genome (reviewed in ref. 23). In pericentromeric heterochromatin, the very same repressive marks have been demonstrated to suppress the formation of DSBs (33, 38) and the resolution of DSBs into COs (27, 39, 40). While the repressive marks associated with TEs in the euchromatic genome are less extensive than those in pericentromeric heterochromatin (several kilobases vs. several megabases), they may influence recombination through a similar mechanism. Consistent with this prediction, in mice, TEs lose the ability to suppress the formation of DSBs when their repressive epigenetic marks are removed in mutants (41). In humans, TEs with motifs of CO hotspots are converted to coldspots upon acquiring repressive epigenetic marks (42). While strongly suggestive, these studies infer the effects of TE on recombination through comparisons of different genomic regions or by examining mutant backgrounds. Such approaches make it challenging to exclude potential confounding influences from local sequence or chromatin context and to accurately assess the impact of TEs in natural populations. Moreover, many TEs identified as recombination modifiers are fragmented and likely inactive (43), and thus unlikely to contribute to natural variation and ongoing evolution of recombination landscapes.

Here, we directly test whether active, polymorphic TEs influence CO occurrence and drive variation in recombination landscapes between individuals of the same species. To this end, we developed an approach using long-read PacBio sequencing on pools of recombinant offspring to efficiently, accurately, and cost-effectively identify COs in wild-type *Drosophila melanogaster* strains. Using this method, we identified nearly 3,000 COs with a few DNA extractions and PacBio SMRT cell sequencing. We then studied the distributions of identified COs between homologous sequences with and without TE insertions in two *D. melanogaster* strains with distinct TE insertion profiles. Because of the near identity of homologous sequences, any observed differences in CO distributions could be attributed to the presence of TEs rather than other local sequence contexts. In addition, we employed two other approaches—examining the distributions of COs in panels of recombinant inbred lines [RILs; (44)] in relation to TE polymorphism and applying marker-assisted estimations of CO frequencies (45) to isogenic strains differing in transgenically inserted TEs. These three orthogonal approaches collectively support the suppressive impact of TEs on CO occurrence, thereby contributing to the varying recombination landscapes observed between wild-type individuals of the same species.

Results

Proposed Long-Read Pool-Sequencing Approach Accurately Identifies Crossovers. A typical approach to identify recombination events starts with crossing two inbred strains, followed by backcrossing the F1 to one of the parents and profiling genetic markers in the backcross offspring (Fig. 1*A*). The haplotype information of F2 offspring is critical for inferring CO events, but is labor-intensive and cost-prohibitive to acquire, requiring separate DNA extraction and genotyping for every F2 individual. To circumvent this limitation, we proposed to sequence F2 offspring in pools using PacBio long-read sequencing, which enables efficient assaying of large numbers of individuals while preserving the local haplotype information (Fig. 1*A*).

To investigate the feasibility of our approach, we generated a "benchmark" dataset (Fig. 1*A*). We performed crosses between two highly inbred *Drosophila* synthetic population resource (DSPR) founder strains (strains A4 and A6, ref. 46), which have been fully sequenced with PacBio long-read sequencing (47). We sequenced the DNA of 192 F2 female individuals both individually using Illumina short-read sequencing (short-read approach; with an average 11.5× depth of each individual) and as a pool using PacBio continuous long read (CLR) sequencing (pool-seq approach; with a read depth of 1,156×; *Materials and Methods*). COs were inferred by the transitions of SNP alleles originating from one parent to the other in the assembly (short-read approach) or within a read (pool-seq approach; *Materials and Methods*). The short distance between SNPs (an average of 204 ± 5 bp) allows us to narrow down crossover events to within 1 kb (*SI Appendix*, Fig. S1).

We identified 538 CO events among 191 F2 individuals using their short-read-based genome assemblies (one individual failed Illumina sequencing despite multiple attempts). From the same pool of 192 F2 individuals, our PacBio pool-seq approach detected 736 read containing CO events. Because the sequencing depth per haplotype in the pool-seq method (3.01x) allowed for potential multiple-read capture of a single CO event, we merged COs within 100 bp, yielding 363 potentially unique CO events. Assuming that the short-read approach exhaustively and accurately identified all COs in the F2 individuals, we estimated the false positive rate of the pool-seq approach to be 6.5% *per read* and 8.3% *per event.* Some of these false positive events could result from COs missed by the short-read approach.

Conversely, 203 of the 538 COs identified by the short-read approach were also detected using the pool-seq method. The COs not detected by our pool-seq approach (false negatives) were randomly distributed across the genome (*SI Appendix*, Fig. S2), indicating that such an approach shows no bias for or against particular genomic regions in CO detection. The relatively low recall rate of the long-read-based pool-seq approach (37.7%) is likely due to our stringent CO calling criteria, which include removing reads with potential gene conversion events and requiring identified COs to be at least 2 kb away from alignment boundaries (*Materials and Methods*). Relaxing these criteria, however, significantly increases not only the recall rate, but also the false positive rate (*SI Appendix*, Table S2). Therefore, for subsequent analyses, we maintained stringent criteria for calling COs in F2 pools to ensure high accuracy.

Proposed Long-Read Pool-Sequencing Approach Efficiently Identifies Large Numbers of Crossovers. To test the predicted suppressive effects of TEs on recombination, we aimed to compare the distribution of COs between strains with different TE insertion profiles. We chose three DSPR founder strains that have short average distances between SNPs (204 bp between

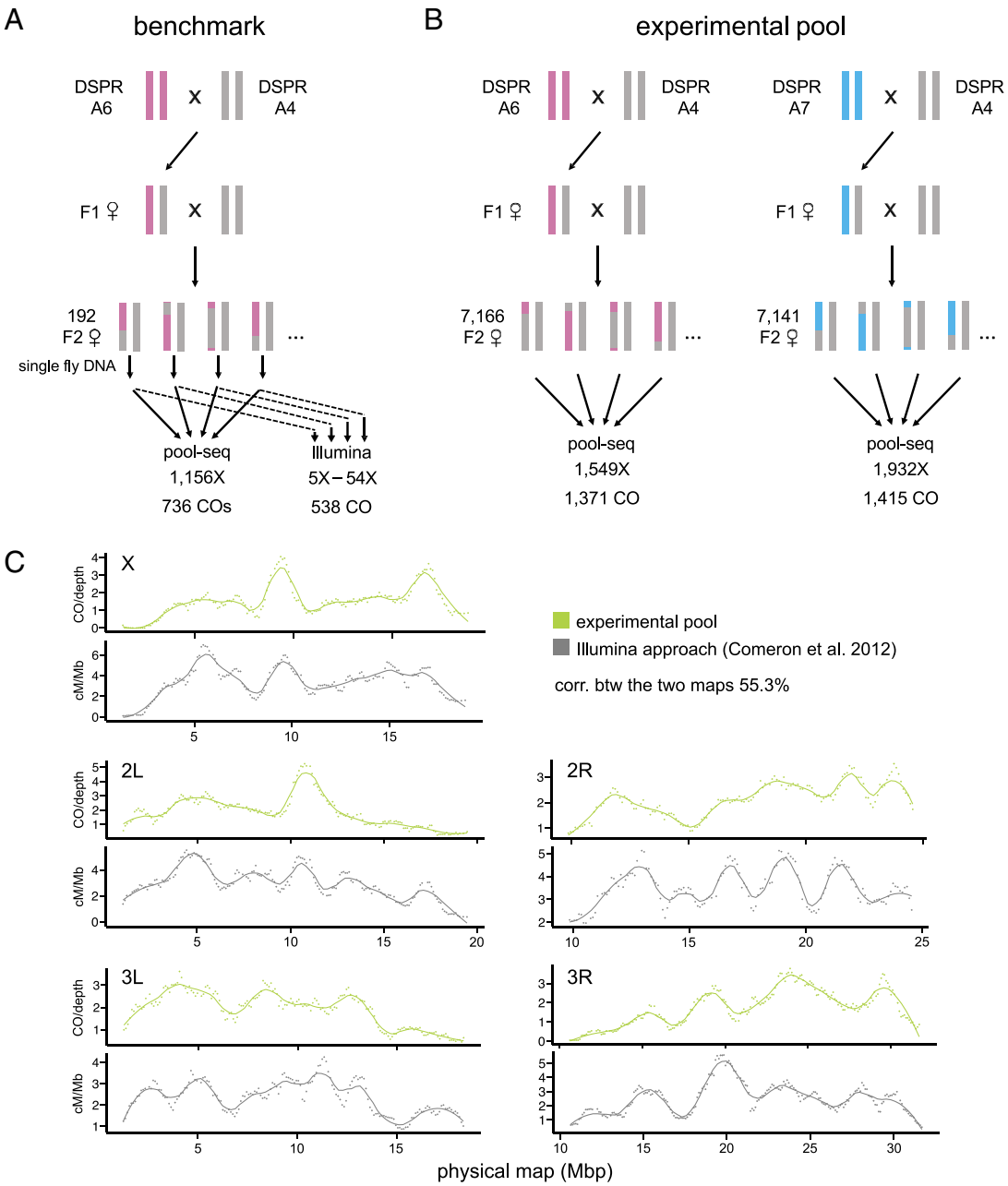

Fig. 1. Experimental design and CO maps inferred by the proposed pool-seq approach. (*A*) For the benchmark data, DNA was extracted individually from 192 F2 female backcrossed offspring of A6 × A4 crosses. Half of the DNA from each individual was sequenced by Illumina short-read sequencing, while the remaining DNA from all individuals was pooled in equal molar ratios and sequenced with PacBio CLR sequencing. (*B*) For the experimental pools, we set up two backcross experiments (A6 × A4 and A7 × A4). For each cross, over 7,000 F2 female backcrossed offspring were collected, pooled, DNA extracted en masse, and sequenced with PacBio CLR. (*C*) The read-depth normalized CO numbers identified in our experimental pools (in every 100 windows and combining the two experimental crosses) for each chromosome is shown on the top, and a previously generated recombination map based on a short-read Illumina sequencing approach (replotting from ref. 13) is showing on the bottom. The distribution of COs was smoothed by a sliding window of 1 Mb with a step size of 100 kb and LOESS with a 20% span. See *SI Appendix*, Fig. S3 for the distribution of CO numbers of individual strains.

A4 vs. A6 and 217 bp between A4 vs. A7) and conducted two backcross experiments (Fig. 1*B*). For each cross, we pooled more than 7,000 F2 females and sequenced them using PacBio CLR, with a final read depth of over 1,500× after filtering reads based on mapping quality and read length (Fig. 1*B*; *Materials and Methods*). To estimate the recall rate for our proposed method, we calculated the maximum number of gametes we could have detected with these sequencing runs, which equals half of the sequencing depth because the remaining reads originated from the backcross parents. Based on the average number of 2.5 CO events per female meiosis in *Drosophila* (48), our approach should on average detect 2.5 CO events per sequenced recombinant gamete under perfect sensitivity, or 1,936 COs at 1,549X sequencing depth (A6) and 2,415 COs at 1,932X sequencing depth (A7). This calculation indicates a ~65% recall rate for our proposed method (71% for A6 cross and 59% for A7 cross), much higher than that observed in the benchmark data (37.7%, also see below).

We compared the distribution of COs identified here to the most detailed currently available CO map, which was generated by using Illumina short-read sequencing to genotype F2 individuals separately (13). We found that the correlation between our maps and the published ones [*Spearman rank* ρ: 53.7% (A6 cross), 50.6% (A7 cross), and 55.3% (averaged between the two crosses),

$P < 10^{-16}$, Fig. 1*C* and *SI Appendix*, Fig. S3] is comparable in strength to previously reported correlations for different *Drosophila* strains [~45%, (13)]. This observation suggests that our proposed approach effectively and efficiently recapitulates the genome-wide variation in CO locations.

Best Practice for Detecting Crossovers Using the Proposed Long-Read Pool-Sequencing Approach. To ensure that any observed association between CO occurrence and TE presence is not due to CO detection biases, we conducted simulations to investigate potential confounding factors (see *SI Appendix*, *Supporting Text* for details about the simulations). Our analysis found no falsely identified COs when reads were simulated from nonrecombinant parental genomes, indicating the high precision of our approach. Using simulations with CO events, our approach achieved an 81.4% recall rate per read for reads that are properly mapped, overlap with CO events, and pass our stringent filtering criteria. Relaxation of these filtering criteria further increased the recall rate (*SI Appendix*, Table S2). Interestingly, we found that simulated reads that span TEs identify COs at a significantly higher rate than reads aligned to genomic regions without TEs (see *SI Appendix*, *Supporting Text* for details), which should make any observed negative association between the CO occurrence and TE presence conservative. Finally, while read length showed limited impact on CO detection, we observed that both higher sequencing depth and SNP density significantly increase CO detection rates (see *SI Appendix*, *Supporting Text* for details), and these two factors should be accounted for in the downstream analysis.

The substantial difference in recall rates between the benchmark data (37.7%) and experimental pool (65%) motivated us to investigate the underlying cause and cost-effective approaches for implementing our proposed method. One major difference between the two is the DNA pooling methods. In the benchmark data, DNA of F2 individuals was pooled post hoc after individual extraction. In contrast, DNA in the experimental pool was extracted en masse from a large number of F2 individuals. The former is likely to result in an unequal contribution of F2 DNA to the final sequencing pool, which is evidenced by many COs being detected by more than the expected number of reads in the benchmark pool (*SI Appendix*, Fig. S4 *A* and *B*), but not in the experimental pool (*SI Appendix*, Fig. S4*C*). Another major difference is the ratio of sequencing depth per recombinant F2 haplotype (referred to as depth-per-haplotype ratio hereafter; 3× for the benchmark data vs. 0.12× for the experimental pool). Our additional simulations (see *SI Appendix*, *Supporting Text* for details) showed that as the depth-per-haplotype ratio increases, not only does the recall rate increase (*SI Appendix*, Fig. S5*A*), but also the number of simulated reads recovering the same CO event increases (*SI Appendix*, Fig. S5*B*). Such an observation suggests that higher depth-per-haplotype ratios lead to inefficient use of sequencing reads, as they repeatedly identify the same COs (*SI Appendix*, Fig. S5*C*). Given the random distribution of false negative events across the genome (*SI Appendix*, Fig. S2), we argue that the implementation of our method should prioritize maximizing CO detections per unit of sequencing depth rather than complete CO recovery. Accordingly, we recommend extracting DNA from many individuals and gradually increasing sequencing depth while maintaining low depth-per-haplotype ratios until reaching the desired CO count. This strategy balances sequencing costs and CO detection efficiency. See *SI Appendix*, *Supporting Text* for detailed discussions.

CO Occurrence Is Suppressed Around Euchromatic TEs When Compared to Other Euchromatic Regions. We first investigated the predicted negative impact of TEs on CO occurrence by comparing euchromatic windows with and without TEs within the individual strain. We annotated TEs using RepeatModeler2 (49) and identified 620 (A6) and 501 (A7) TEs in the euchromatic regions after filtering TEs with short lengths, in large clusters, or shared between strains (*Materials and Methods*). We estimated the number of COs detected in the F2 recombinant pools in 5 kb windows upstream and downstream to TEs (TE flanking windows). After filtering windows that overlap or contain fewer than five SNPs, we analyzed 509 and 376 TE-flanking windows in A6 and A7 crosses respectively, representing 7.7% and 5.7% of the euchromatic sequences. Because of the large number of F2 recombinant individuals (~7,000) and high sequencing depth (>1,500×) in each pool, more than one CO event could be detected within a given window (Fig. 2*A*). We used the number of COs in 10 kb windows that are at least 30 kb away from TEs as controls (control windows, Fig. 2*A*). TE flanking windows tend to have lower sequencing depth and fewer SNPs than the control windows (*Mann–Whitney U test,* $P < 0.05$ for all comparisons; *SI Appendix*, Fig. S6). These differences could potentially confound our analysis because, similar to observations with simulated data (see above), the number of COs identified across non-TE windows are positively correlated with sequencing depth [*Spearman rank* $\rho = 0.018$, $P = 0.12$ (A6) and $\rho = 0.055$, $P < 10^{-6}$ (A7)] and SNP numbers [*Spearman rank* $\rho = 0.095$, $P = 6.0 \times 10^{-10}$ (A6) and $\rho = 0.10$, $P = 6.7 \times 10^{-12}$]. Accordingly, we included sequencing depth and the number of SNPs as covariables when assaying the impacts of TEs on the distributions of COs. On the other hand, read length would not have confounded our analysis because we found no associations between that and the number of COs detected [*Spearman rank correlation tests,* $P = 0.15$ (A6) and 0.14 (A7)].

For both strains, we found that TE flanking windows have significantly lower numbers of CO than control windows [*Mann–Whitney U test,* $P = 0.017$ (A6) and 1.5×10^{-4} (A7); Fig. 2*B*], a result that is robust when controlling for the confounding effects of sequencing depth and number of SNPs using generalized linear models [regression coefficients of TE effect: −0.20, $P = 0.24$ (A6) and −0.71, $P = 0.00083$ (A7); *Materials and Methods*]. We also compared the physical distance from the focal site to the nearest CO events, which should be longer when CO occurrence is suppressed around the focal site. Consistent with analysis comparing CO numbers, the distance from the TE insertion site to the nearest CO is significantly longer than that of the control site in both strains [*Mann–Whitney U test,* $P = 8.6 \times 10^{-7}$ (A6) and $P = 6.9 \times 10^{-7}$ (A7); Regression coefficient for TE effect: 1.52 (A6) and 1.29 (A7), $P < 10^{-16}$ for both; Fig. 2*C*]. These observations support the prediction that TEs suppress the occurrence of COs.

We next investigated what biological attributes make a TE exert stronger suppressive effects on COs. The two strains have a slightly different relative abundance of different classes and types of TEs, with A6 having many more DNA-based TIR than other RNA-based TE families (A6: 255 TIR, 228 LTR, and 137 non-LTR; A7: 95 TIR, 230 LTR, and 176 non-LTR). While we did not find differences in the numbers of COs near different classes of TEs (RNA vs. DNA, and LTR vs. non-LTR vs. TIR, regression coefficient $P > 0.05$ for all pairwise comparisons; *SI Appendix*, Table S3), we did find significantly longer distances to the nearest CO around RNA-based TEs when compared to DNA-based TEs in one strain [A6; Regression coefficient: 0.506 (RNA vs. DNA), 0.508 (LTR vs. DNA), and 0.505 (non-LTR vs. DNA), $P < 0.01$; *SI Appendix*, Table S3]. We then tested whether the enrichment of H3K9me3 may act as a driver for the suppressive effects of TEs on COs. Two measures for the enrichment of H3K9me3 were used: the average H3K9me3 enrichment level in TE flanking windows (adjacent

H3K9me3 enrichment) and accumulated H3K9me3 across the entire extent of H3K9me3 spread from TEs (H3K9me3 mass; *SI Appendix*, Fig. S7). We used the former measure for analysis using CO numbers and the latter for analyzing distance to the nearest CO, because the extent of the suppressive effect of H3K9me3 is unknown. We found significant negative associations between adjacent H3K9me3 enrichment and the number of COs identified in one strain (A6: *Spearman rank* $\rho = -0.097$, $P = 0.028$; regression coefficient of H3K9me3 enrichment −0.42, $P = 0.047$; Fig. 2*D*), while associations using H3K9me3 mass were not significant (Fig. 2*E* and *SI Appendix*, Table S3).

Because previous studies found that RNA-based TEs have stronger enrichment of H3K9me3 (50, 51), which we also observed (*SI Appendix*, Table S4), we sought to differentiate whether the observed negative association between H3K9me3 enrichment and CO occurrence is due to TE class or the enrichment of H3K9me3 itself. We repeated the above analysis within TE class, and still found significantly negative correlations between the enrichment of H3K9me3 and CO number for RNA-based TEs [A6: *Spearman rank* $\rho = -0.18$, $P = 0.0014$ (RNA-based TE), $\rho = -0.18$, $P = 0.015$ (LTR), $\rho = -0.18$, $P = 0.057$ (non-LTR); regression coefficient of H3K9me3 enrichment −0.59, $P = 0.057$ (RNA-based TE), −0.46, $P = 0.12$ (LTR), and −1.90, $P = 0.053$ (non-LTR); all other comparisons, including analysis using distance to CO, have $P > 0.05$; see *SI Appendix*, Table S3 for details]. Such a result confirms the impact of TE-induced enrichment of repressive marks on CO occurrence.

Interestingly, even for TEs without H3K9me3 enrichment, they are also associated with fewer CO numbers [regression coefficient of TE presence −0.086, $P = 0.68$ (A6) and −0.684, $P = 0.031$ (A7)] and show longer distance to the nearest CO [regression coefficient of TE presence 1.38 (A6) and 1.44 (A7), $P < 10^{-16}$ for both] than the control windows. These TEs may impact CO occurrence similar to other SVs, an effect expected to be stronger for longer TEs. We tested this possibility by focusing on TEs without the enrichment of H3K9me3, but found no significant associations between TE length and the number of COs detected nor distance to the nearest COs (Spearman rank correlation analysis and regression analysis, $P > 0.05$ for all comparisons, *SI Appendix*, Table S3). In summary, our observations suggest that TEs, especially RNA-based TEs and TEs with high local enrichment of repressive marks, suppress the occurrence of COs in their vicinity.

CO Occurrence Is Suppressed in the Presence of Euchromatic TEs When Compared to Homologous Sequences Without TEs. While our above analysis found a reduced occurrence of COs around TEs when compared to TE-free genomic regions within a strain, similar patterns could also be driven by other local features influencing CO occurrence, as well as TE insertion preference or differential selection against TEs (reviewed in ref. 20). To investigate the suppressive impacts of TEs on CO while accounting for these potential confounding effects, we compared the distributions of COs around homologous alleles with and without TEs. Given that homologous sequences are nearly identical, observed differences in CO occurrence can be primarily attributed to the major distinguishing feature between alleles: the presence/absence of TEs. Specifically, we tested the association of TE presence with the differences in two indexes, the number of COs within windows and the distance to the nearest CO (Fig. 2*F*). Because of the pronounced impacts of sequencing depth on the detection of COs (see above), we inferred indexes of CO distribution after downsampling sequencing reads to ensure equal sequencing coverage between homologous alleles with and without TEs (TE set). These estimates were compared to those from control sites at least 30 kb away from TEs (control set) to account for the difference in recall rate and, potentially, genetic map length between the two strains (*Materials and Methods*). In addition, our analysis controlled for SNP number (*Materials and Methods*) because the ratio of the number of SNPs between homologous alleles is marginally significantly smaller for the TE set than the control set [*Mann–Whitney U test*, $P = 0.06$ (A6) and 0.08 (A7)], Significance levels of differences in CO distribution between homologous alleles with and without TEs were determined by two approaches: 1) the *median Mann–Whitney U test P-values* for the comparisons between the TE set and control set across 1,000 downsampled data and 2) comparing the observed difference in CO indexes of the TE set to the null distribution generated by bootstrapping control sets (*Materials and Methods*).

We first tested whether the presence of TEs negatively impacts CO occurrence, which should be reflected as smaller values for the difference in CO number but larger values for the ratios of distance to the nearest CO for the TE set than those of the control set (Fig. 2*F*). Consistently, compared to the control set, the TE set has a significantly lower difference in CO number [*Mann–Whitney U test, median p:* 0.24 (A6) and 0.00052 (A7); *bootstrapping P:* 0.25 (A6) and <0.001 (A7), Fig. 2*G* and *SI Appendix*, Fig. S8] and larger ratios of distance to CO [*Mann–Whitney U test, median P:* 0.17 (A6) and 0.0071 (A7); *bootstrapping P:* 0.501 (A6) and 0.246 (A7), Fig. 2*H* and *SI Appendix*, Fig. S8]. We found limited differences in the suppressive impact of TEs of different classes and types, with non-LTR showing significantly larger ratios in the distance to COs than TIRs in one of the two strains (*Mann–Whitney U test, median* $P = 0.037$, all other comparisons $P > 0.05$; see *SI Appendix*, Table S5).

To investigate whether the enrichment of H3K9me3 around TEs contributes to varying suppressive effects of TEs on COs, we compared the differences in CO numbers between TEs with and without H3K9me3 enrichment in their 5 kb flanking window (adjacent enrichment level > 1). Notably, non-TE induced enrichment of repressive marks does occur in the euchromatic genomic region (e.g., ref. 52), and, to mitigate the consequential confounding effect, we excluded windows where alleles without TEs similarly show enrichment of H3K9me3. TEs with H3K9me3 enrichment tend to have lower differences in CO number than TEs without, although the difference is not statistically significant [*Mann–Whitney U test, median P:* 0.11 (A6) and 0.43 (A7); *bootstrapping P:* 0.18 (A6) and 0.45 (A7), Fig. 2*I* and *SI Appendix*, Fig. S9]. Alternatively, we correlated the level of H3K9me3 enrichment to the difference in CO number and found a significant negative association for one of the two strains [*median Spearman correlation coefficient*: −0.12 (A6, *median* $P = 0.043$) and −0.03 (A7, *median* $P = 0.36$), *SI Appendix*, Fig. S10]. Removing four *Roo* elements near COs, despite being highly enriched with H3K9me3, in the insignificant strain leads to a stronger correlation (*median Spearman correlation coefficient*: −0.10, *median* $P = 0.11$ in A7). We also performed a similar analysis using ratios of the distance to the nearest CO and used the total H3K9me3 mass to categorize TEs (high and low H3K9me3 enrichment), excluding those with enrichment for H3K9me3 in the TE-free homologous alleles (*Materials and Methods*). TEs with high H3K9me3 mass show significantly larger ratios of the distance to the nearest CO, especially in one of the two strains [*Mann–Whitney U test, median P:* 0.015 (A6) and 0.438 (A7); *bootstrapping P:* 0.005 (A6) and 0.078 (A7), Fig. 2*J* and *SI Appendix*, Fig. S9]. In the same strain, we also found a significant positive correlation between H3K9me3 mass and the ratios of the distance to the nearest CO [*median Spearman correlation coefficient*: 0.11, $P = 0.041$ (A6); *SI Appendix*, Fig. S10]. Importantly, significant negative associations between TE-induced enrichment of H3K9me3 and difference in

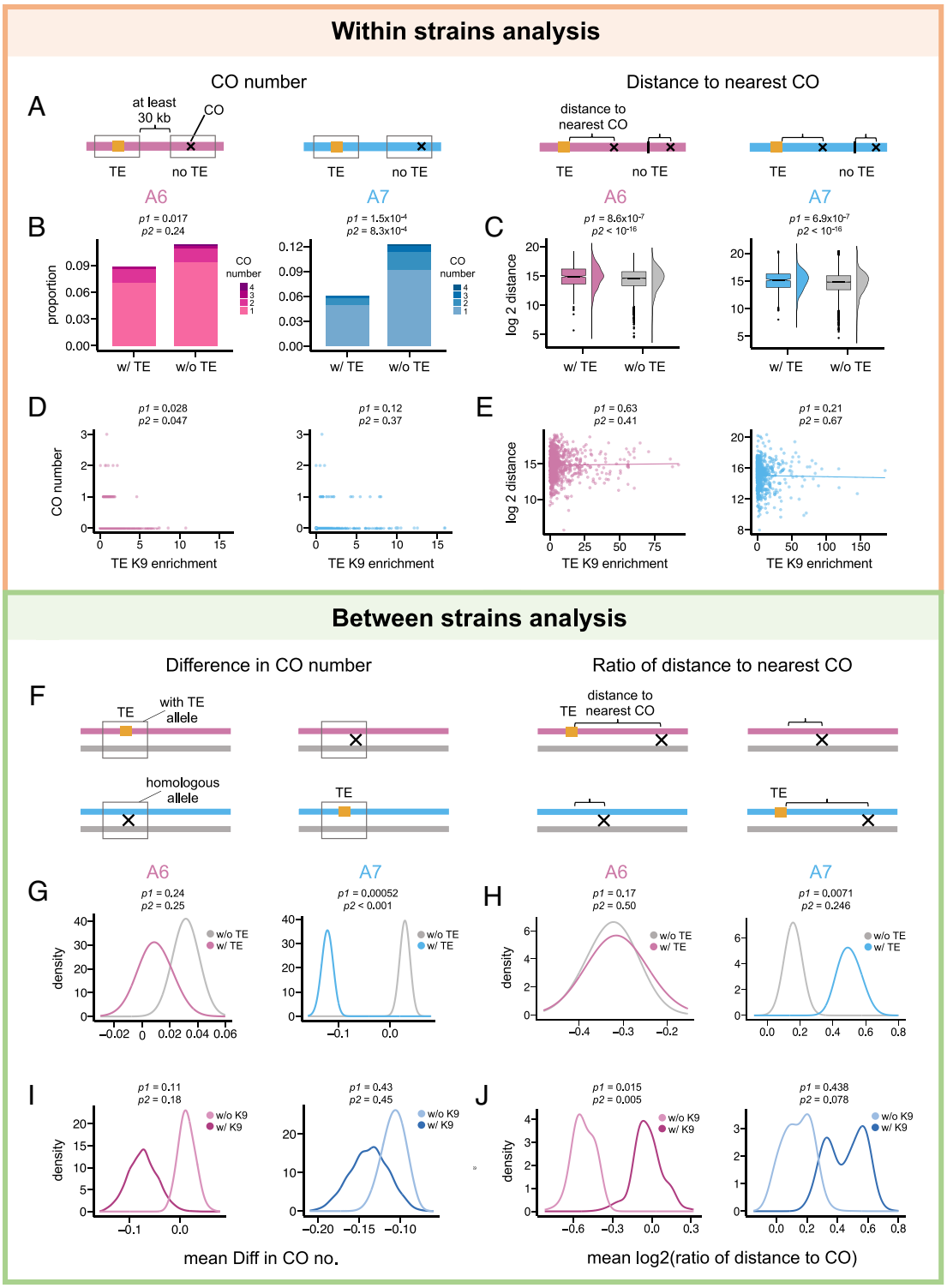

Fig. 2. Euchromatic TEs suppress CO occurrence. (*A–E*) Within-strain analysis to investigate the impact of TEs on the occurrence of COs. (*A*) Cartoon illustrating measures for the occurrence of COs: CO numbers of 5 kb windows flanking focal sites and the physical distance from the focal site to the nearest CO. (*B*) Within-strain comparisons of the number of COs detected in the F2 recombinant pools for TE windows (509 and 376 windows for A6 and A7, respectively) and control windows without TEs (4,185 and 4,720 windows for A6 and A7 respectively). The stacked bar indicates the proportion of windows with a certain CO number. (*C*) Within-strain comparisons of the distance to the nearest COs TEs (620 and 501 TEs for A6 and A7, respectively) and controls (4,256 and 4,924 control sites for A6 and A7, respectively). A significant level was obtained via both the *Mann–Whitney U* test (*p*1) and the regression analysis (*p*2). (*D* and *E*) Associations between TE-induced H3K9me3 enrichment and CO numbers (*D*) and the distance to the nearest CO (*E*). A significant level was obtained via both the *Spearman* correlation (*p*1) and the regression analysis (*p*2). (*F–J*) Between strain analysis to test the impact of TEs on the occurrence of COs. (*F*) Cartoon illustrating measures for the impacts of TEs on CO occurrence: difference in CO numbers in the TE flanking windows (*Left*) and log2 ratio of the distance to nearest CO (*Right*) between homologous alleles with and without TEs. (*G*) Shown distributions compare the mean differences in CO numbers for 1,000 downsampled sets of TE pairs that have equal sequencing coverage between homologous alleles with and without TEs (345 and 232 TEs for A6 and A7, respectively) and the same number of downsampled control window pairs as TEs (bootstrapped from 1,434 and 1,507 control windows in A6 and A7, respectively). Because most windows overlap with no COs, the distribution of the median difference in CO number is less visually informative than the mean presented here. (*H*) Shown distributions compare the mean differences in distance to the nearest COs for 1,000 downsampled sets of TE pairs (489 and 360 TEs in A6 and A7, respectively) and the same number of control windows as TEs (bootstrapped from 1,831 and 2,041 control windows in A6 and A7, respectively). (*I*) Comparisons for the difference in CO numbers between TE pairs with (70 and 61 TEs in A6 and A7, respectively) and without H3K9me3 enrichment (135 and 90 TEs in A6 and A7, respectively). (*J*) Comparisons for the ratios of the distance to nearest CO between TE pairs with high and low H3K9me3 mass (123 and 93 TEs in either category for A6 and A7, respectively). A significant level was obtained via both the *Mann–Whitney U* test (*p*1) and the bootstrapping test (*p*2) for between-strain analysis (see main text). The numbers of TEs in (*B* and *G*) differ from those of (*C* and *H*) because, for CO number, we required at least five SNPs within both 5 kb flanking windows (*B* and *G*), while for analysis using distance to nearest CO (*C* and *H*), this criterion was not required. The smaller number of TEs studied in (*I* and *J*) analysis when compared to (*G* and *H*) is due to the stringent filtering of TE pairs whose without TE alleles also show some enrichment of H3K9me3. w/ TE: TE windows/sites; w/o TE: control windows/sites without TEs; w/ K9: with H3K9me3 enrichment; w/o K9: without H3K9me3 enrichment.

CO numbers remain when restricting the analysis to RNA-based TEs [*median Spearman correlation coefficient*: −0.24, P = 0.003 (RNA-based TEs), −0.26, P = 0.005 (LTR), −0.24, P = 0.1 (non-LTR); see *SI Appendix*, Table S5 for all other comparisons], suggesting such an association is not due to difference in H3K9me3 enrichment between TE classes (*SI Appendix*, Table S4). Finally,

similar to within-strain analysis, we found that even TEs without H3K9me3 show suppressive effects on COs in one of the strains [CO number: *Mann–Whitney U test median* $P = 0.030$ and bootstrapping $P = 0.016$ (A7); *SI Appendix*, Table S5], but found no correlations between TE length and difference in CO distribution for TE insertions where alleles with and without TEs are both depleted of H3K9me3 enrichment ($P > 0.05$ for all comparisons, *SI Appendix*, Fig. S11 and Table S5). Overall, our between-strain analyses that compare the distribution of COs between homologous alleles with and without TEs corroborate our findings observed within strains: TEs, especially those that lead to local enrichment of repressive marks, suppress the occurrence of COs in the euchromatic genome.

TE-Mediated Suppressive Effects Contribute to Between-Strain Differences in CO Landscapes. Having established that the presence of TEs significantly reduces the occurrence of COs, we aim to quantify these TE-mediated effects. We first estimated how much reduction in CO occurrence could be driven by TEs when compared to the homologous TE-free alleles. To control for factors that could potentially confound the detection of COs, we selected TE-free control windows with similar between-homolog SNP ratios to those of TE windows and performed downsampling to acquire comparable sequencing depth between alleles. These control windows allow the estimation of normalizing factors that correct for the varying total CO map length while accounting for the above-mentioned potential confounding factors, enabling direct comparisons of COs between homologous alleles with and without TEs (*SI Appendix*, *Supporting Text* and Fig. S12*A*). We found that, compared to homologous TE-free alleles, the presence of TEs led to an average 36.2% reduction in COs (95% bootstrapped CI: −58.2 to −10.4%; *SI Appendix*, Fig. S12*B*), with TEs showing enrichment of H3K9me3 leading to an even stronger effect (−49.5%, 95% bootstrapped CI: −78.3 to −0.6%; *SI Appendix*, Fig. S12*B*). These estimates further corroborate that TE-mediated suppressive effects on COs could be substantial.

To quantify how much between-strain variation in CO landscapes may be attributed to the suppressive effects of TEs, we analyzed TE prevalence in euchromatic windows that show CO differences between the two studied strains. After normalizing for potential confounding factors (see *SI Appendix*, *Supporting Text* for details), we identified 2,082 10 kb windows with between-strain differences in CO numbers, representing 22.0% of all euchromatic windows. Among windows with CO differences, TEs are found in 12.2% (n = 254) of them, and within these TE-containing windows, TE-mediated suppressive effects accounted for 61.3% of the CO differences (95% bootstrapped CI: 55.3 to 67.2%; *Fisher's Exact Test*, *P-value* $< 10^{-5}$, odds ratio = 3.5). Collectively, TEs explained 7.5% (95% bootstrapped CI: 6.4 to 8.6%) of the total between-strain CO variation by reducing CO occurrence in the strain where they were present. These estimations point to TEs as an important contributor to varying CO landscapes between individuals.

Suppressive Effect of TEs on Recombination Is Supported by Orthogonal Approaches. To further corroborate our observed suppressive effects of TEs on COs in the euchromatic genome, we used two orthogonal approaches to independently test this finding. In the first approach, we investigated the distributions of CO breakpoints in the RIL panel of DSPRs (44) with respect to the presence/absence of TE insertions. We inferred the CO breakpoints in the subset of Illumina-sequenced RILs, estimated the distance between an euchromatic TE insertion site to the nearest recombination breakpoint, and compared such distances between homologous alleles with and without TEs (*Materials and Methods*). Consistently, alleles with TEs have a significantly longer physical distance to the nearest CO breakpoints than homologous alleles without TEs (Fig. 3 *A* and *B*; *paired Mann–Whitney U test,* $P < 10^{-16}$), with the presence of TEs adding a median of 207,894 bp to the nearest breakpoints. Analysis focusing on insertion sites where the median distance to breakpoint is within 100 kb reaches the same conclusion (presence of TEs add 230,454 bp, *paired Mann–Whitney U test,* $P < 10^{-16}$). Among all TE insertion sites studied, 66.82% of them have longer physical distances to the nearest CO breakpoint when the TE is present than when the TE is absent, a proportion that is significantly larger than randomly sampled genomic locations with matching chromosomal distributions and numbers (Fig. 3*C*, permutation test $P < 0.001$). However, we did not find differences in the suppressive effects of TEs of different classes (*Mann–Whitney U test,* $P > 0.05$ for all comparisons).

In the other approach to test the causal impacts of TEs on recombination, we compared the distribution of COs around alleles with and without transgenically introduced TEs in a controlled genetic background. Such an approach mitigates the potential difference in genetic background, which is hard to ascertain in the above genome-wide approaches. Specifically, we inserted mCherry containing transgenic constructs with different TE sequences into the same euchromatic location (*Materials and Methods*). To assess local changes in recombination rate in the presence of TE sequences, we employed a recently developed technique that infers chromosome-wide recombination rates from bulk sequencing of marker-selected pools (45). We chose to integrate two RNA-based TEs that showed the strongest *cis* spreading of repressive marks in our previous study [*roo* and *copia,* (50, 51)], and a DNA-based TE identified to exert *cis* epigenetic effects [*1360,* (53)]. In an F2 recombinant backcross (Fig. 3*D*), we selected F2 homozygotes with the visible *black* marker, which is ~7 Mb from the transgenic insertion site. In the selected F2 pool, the allele frequency of the inserted chromosome (blue in Fig. 3*D*) will increase as a function of increased genetic distance from *black* (Fig. 3*E*), and we extrapolated the local recombination rate as the rate of change in genetic distance after accounting for paternal contribution and secondary fitness effects. Within the 500 kb flanking windows of different transgenic constructs, there is significant variation in the recombination rate (ANOVA F = 7.2, $P = 0.017$), with RNA-based TEs having significantly lower recombination rate than the control (TukeyHSD test, adjusted $P = 0.02$; Fig. 3*F*). It is worth noting that the transgenic background coincidentally shares several large haplotype blocks with the *black* mutant strain just downstream of the transgenic insertion site, leading to a sharp decrease in informative SNP sites, and, therefore, suboptimal estimation of allele frequency. Nevertheless, both approaches provide orthogonal support for the suppressive effects of TEs on local recombination rate.

Discussion

By developing a long-read-based method to identify COs among pools of recombinant individuals, we compared the distribution of ~2,800 COs with respect to TE insertions. Our analysis revealed that the presence of TEs, especially RNA-based TEs and TEs with enrichment of repressive epigenetic marks, significantly reduces the occurrence of COs. These TE-mediated suppressive effects cause substantive reductions in CO occurrence (36.2%) and contribute to 7.5% of all the between-strain variation in CO distribution, despite TEs occupying only 12.2% of the analyzed euchromatic windows. Our two other approaches that compare

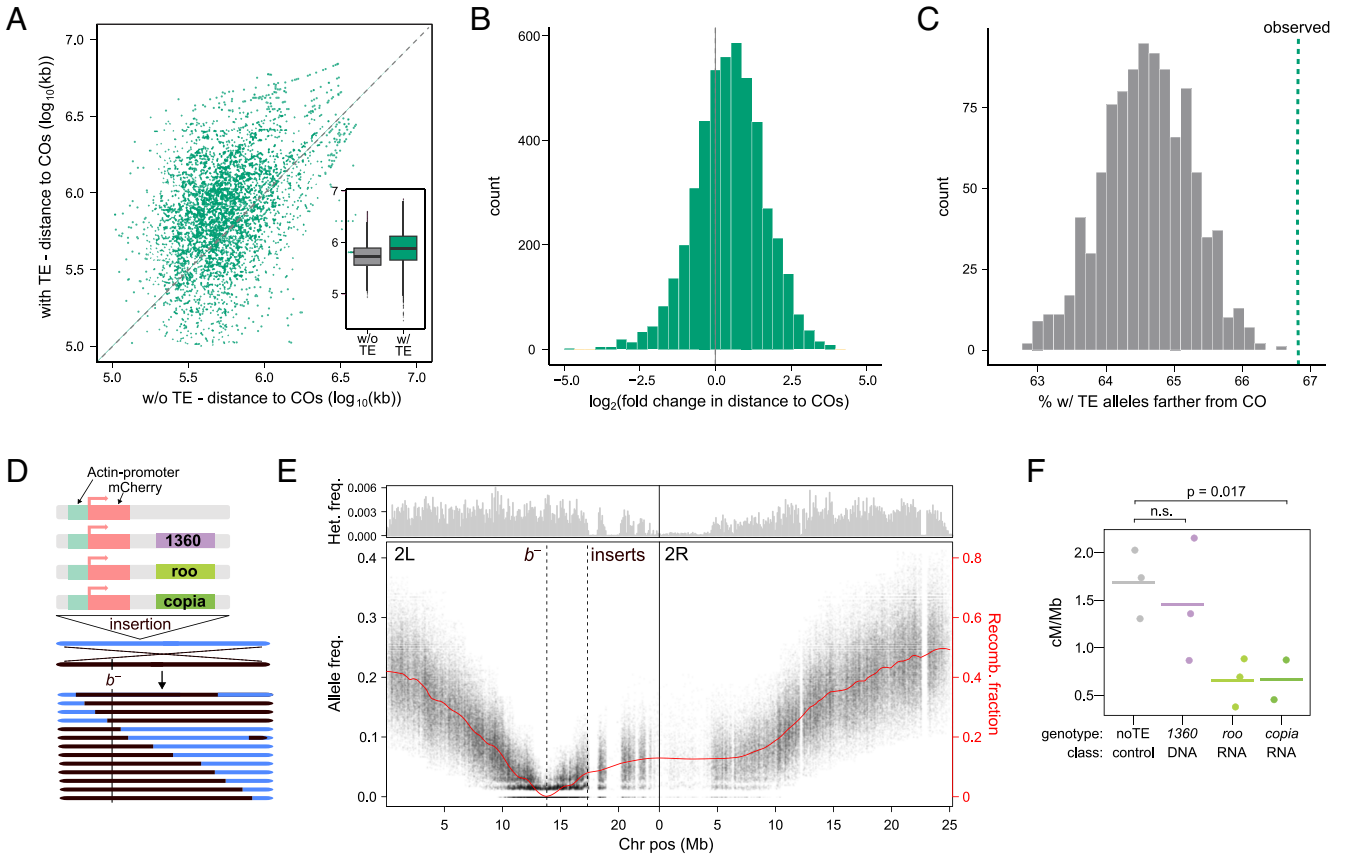

Fig. 3. Orthogonal approaches support the conclusion that TEs suppress CO occurrence. (*A–C*) Comparisons of the distance to nearest CO between homologous alleles with and without TEs using RILs of DSPR. (*A*) log 10 distance to the nearest CO breakpoints from TE insertion sites between alleles without (w/o, *x* axis) and with (*y* axis) TEs, which is biased toward larger distance for with TE alleles. Insert: boxplot showcasing the differences in distance to the nearest breakpoint between homologous alleles with and without TEs. (*B*) The distribution of fold change in the distance to the nearest CO breakpoints from TE insertion sites between homologous alleles with and without TEs. (*C*) Comparisons of the observed proportion of TE insertions with (w/) TE alleles having a farther distance to the nearest CO breakpoints than the homologous alleles (dashed green line) to the distribution of that of randomly sampled genome locations with matching properties (gray). (*D*) Cartoon representation of the transgenic strains used and how genetic distance is captured in allele frequency of marker selected pool. (*E*) After sequencing, allele frequency is estimated at every diagnostic SNP site where the parental strains are fixed for different alleles. The density of such sites is displayed in the *Top* panel (in 100 kb windows). Decay of allele frequency around *black* is then used to estimate the recombinant fraction, which approximates the genetic distance (red). The rate of change (slope) then approximates the recombination rate. (*F*) Estimated recombination rate using SNPs ±500 kb of the selected locus *black* in crosses with different transgenically introduced TEs. Significance differences between different TE insertions were estimated using a linear model: COrates ~ class + genotype. Because the genotype effect was not significant ($P = 0.98$), only the class effect was analyzed using the ANOVA and TukeyHSD tests.

CO distributions in panels of RILs or isogenic strains differing in transgenically inserted TEs reached the same conclusion, further corroborating the suppressive effects of TEs on recombination. These three orthogonal approaches compare nearly, if not completely, identical homologous sequences with and without TEs, effectively excluding the confounding influence of sequence context in driving differences in CO occurrence. Importantly, all the TEs included in our study are polymorphic and likely still active. This is in contrast to TEs suggested to be recombination hotspots in plants and mammals, where most of those TEs are no longer active and oftentimes share similar genetic features (e.g., DNA motifs) with other non-TE recombination hotspots (42, 43, 54).

Our method of applying PacBio long-read sequencing on pooled recombinant F2 is similar in concept to previous work using linked reads to detect COs in gametic pools (55, 56) but with higher accuracy (false positive rate 6.5% vs. 14.9%), easier analysis, and greater cost-effectiveness. Our algorithm identifies COs by detecting the transition of SNPs from one parent to the other within a single PacBio read, eliminating the need to reconstruct sequencing molecules and filter shared barcodes as required with linked-read sequencing (e.g., refs. 55 and 57). For individual *Drosophila* strains, our approach can detect COs of over five hundred meiosis events with just one DNA extraction, one PacBio library preparation, and two PacBio SMRT cell sequencing. The entire process cost roughly $4,000 several years ago and is likely much cheaper now. In contrast, identifying a similar number of COs using the short-read sequencing approach would require over five hundred separate DNA extractions and library preparations, more than doubling the costs in reagents and sequencing alone while also demanding significantly more labor and time.

Importantly, our benchmark data and in silico simulations not only allow us to estimate the accuracy of the approach, but also to identify potential sources of false-negative events and inform practices to maximize the efficiency of our approach. Our analyses reveal that, while a high sequencing depth per F2 recombinant haplotype leads to improved CO recall, it also results in redundant capture of the same COs, an inefficient use of sequencing efforts (*SI Appendix*, Fig. S6 *A–C*). In addition, such a high depth-to-haplotype ratio is susceptible to unequal sampling of F2 individuals. Considering the randomly distributed false negative events in the benchmark data (*SI Appendix*, Fig. S2) and the nearly linear relationship between sequencing depth and the number of unique COs detected in the experimental pools (*SI Appendix*, Fig. S6*E*), we recommend extracting DNA en masse from thousands of F2 recombinant individuals and sequencing the DNA in batches until the desired number of COs is detected. Crucially, a

low sequencing depth-to-haplotype ratio should be maintained to help maximize the number of COs detected per unit of sequencing effort. It is worth noting that we were unable to infer gene conversion events accurately due to the relatively high error rate of PacBio CLR [~10%, (58)], the distance between SNPs in the parental strains, and the typically short track length of gene conversion events [~500 bp, (13, 40)]. Recent applications of the much more accurate, but shorter, PacBio HiFi sequencing demonstrate the possibility of identifying other recombination products from gametic pools (59, 60). Further benchmarking, like those developed in this study, will help assess the specificity and sensitivity of using PacBio HiFi sequencing to detect various recombination products.

Our findings reveal that RNA-based TEs are more likely to suppress CO occurrence. This finding aligns with previously reported heterogeneity in the effects of different TE classes on recombination, particularly the more pronounced negative associations of RNA-based TEs with the formation of DSBs (32, 33) and the occurrence of CO (30, 54). However, RNA-based TEs tend to result in stronger local enrichment of repressive marks in *Drosophila* (51). Their negative effects on COs might be mediated by the known suppressive impacts of these marks on the various steps of COs (33, 38, 39), which is supported by our observed negative associations between the H3K9me3 enrichment around TEs and CO occurrence within TEs classes.

Interestingly, TEs without the enrichment of H3K9me3 are also significantly associated with reduced CO occurrence in one of the two strains, indicating that other mechanisms also mediate the suppressive effects. One possibility is that, similar to the reported effects of other structural variants (34, 61), sequence heterogeneity introduced by TEs interferes with homolog pairing, leading to the resolution of DSBs into non-CO products. Alternatively, because our approach can only detect COs in surviving adults, COs that reduce an individual's survival are likely underrepresented in the F2 pools. This bias is particularly likely when nonhomologous TE sequences serve as templates for DSB repair. The resulting "ectopic recombination" between nonallelic TEs can generate chromosomal rearrangements, causing significant fitness reductions in affected individuals (62, 63) and, consequently, the underrepresentation of detected COs around TEs. Longer TEs may present a more significant challenge to the homology required for DSB repair (64) or be more likely to undergo ectopic recombination (65). Yet, our analyses found no associations between CO occurrence and the length of TEs without enrichment of H3K9me3, suggesting that other mechanisms may mediate the CO-suppressing effects of TEs. It is important to note that these potential mechanisms can only arise because we estimated the CO-suppressing impacts of heterozygous TEs, which likely resemble the situation in natural populations given the predominantly low population frequencies of TEs (66, 67).

The observed suppressive effect of TEs on recombination may have a multifold impact on the evolution of both TEs and host genomes. Such an effect could influence TE evolutionary dynamics by limiting their potential for ectopic recombination (20, 23) and, in turn, shape the evolution of host-mediated epigenetic silencing targeting TEs (68). Given the prevalence of TEs in the gene-rich euchromatic genome, their suppressive effects on recombination could fundamentally influence genic evolution, hindering the fixation of adaptive variants or the removal of deleterious alleles due to selective interference (6, 7). Considering the abundance of TEs in *D. melanogaster* (67), the estimated 36.5% TE-mediated reduction in CO occurrence would translate to a roughly 4% decrease in population recombination rate (4Nr or ρ) for the entire euchromatic genome. While this rate may seem modest, the distribution of TEs across the euchromatic genome is not uniform, and some regions may be especially susceptible to such TE-mediated suppressive effects. For example, in genomic regions where TEs preferentially insert (69), ρ may be sufficiently reduced to have a substantial evolutionary impact. In other genomic regions, the influences of TEs on ρ would depend jointly on their suppressive effects on recombination and population frequencies. TEs with moderate enrichment of repressive marks that have minimum fitness impacts (e.g., in intergenic sequences) should have higher population frequencies and, thus, a greater overall impact on ρ. In either case, the consequentially reduced selection efficacy against TEs due to suppressed recombination should increase TE population frequencies and drive the accumulation of new TE insertions. This process can further extend and exacerbate the suppression of recombination, eventually resulting in blocks of genomes with diminished recombination rates. These effects may be especially important for generating microhaplotypes that help maintain coadapted alleles (70) or the evolution of chromosomes with nearly complete suppression of recombination (20, 71). It is worth noting that higher TE density (72) and stronger TE-mediated spreading of repressive marks (51) have been observed in other *Drosophila* species, where the suppressive effects of TEs on population recombination rate could be even more pronounced.

The ability of TEs to self-replicate and move between genomic locations generates substantial variation in the abundance, composition, and location of TE insertions between individuals (66, 73, 74) and species (18), making them a significant force shaping the eukaryotic genome. Beyond their previously known role in contributing to varying epigenome (50, 74) and transcriptome (75), our findings reveal that dynamically changing TEs can actively alter recombination landscapes. This effect contributes to varying recombination within genome, between individuals, and likely across time and even between species, underscoring their critical role in driving genome evolution through diverse mechanisms.

Materials and Methods

Generation of Benchmark Data and Experimental Pool. We chose three DSPR founder strains (76) that have the maximum differences in SNPs and set up two crosses: A4 × A6 and A4 × A7. These three strains were originally collected from Zimbabwe, Africa (A4), Georgia, USA (A6), and Kenting, Taiwan (A7). To generate F2 recombinants, F1 female offspring from each cross were backcrossed to A4. 192 F2 backcross females were collected for the benchmark experiment, while 7,166 and 7,141 F2 backcross female offspring were collected for A4 × A6 and A4 × A7 experimental pools, respectively.

For the benchmark data, high–molecular weight (HMW) DNA was extracted individually using Qiagen MagAttract HMW DNA Kit. The DNA of each individual was divided equally for standard Illumina short–read sequencing (using the Illumina Nextera DNA Flex Library Prep Kit and Illumina NovaSeq 6000) and pooled with equal molar ratios for PacBio CLR sequencing. For the experimental pool, HMW DNA was extracted en masse (~7,000 F2 females) for each cross using Qiagen Blood & Cell Culture DNA Midi Kit and sequenced with PacBio CLR.

Annotation of TEs in the Assembled Genomes of Parental Strains. For all sequence analysis, we used the following genome assemblies from ref. 77: NCBI GCA_003401745.1 (A4), GCA_003401885.1 (A6), and GCA_003401915.1 (A7). To annotate TEs, we ran Repeatmodeler2 [version 2.0.3; (49)] and used blastn (78) to assign candidate TEs to the family level. We excluded TEs shorter than 200 bp, of INE-1, which is mostly fixed in *D. melanogaster* (67), in clusters larger than 10 kb, and shared between strains [inferred using minimap2 (v. 2.24 (79)) to identify homologous alleles of TE insertions].

Identification of COs with Short-Read and Long-Read Sequencing Data. We first filtered heterozygous sites in each parental strain by resequencing these strains with Illumina NovaSeq 6000 and used minimap2 to call SNPs between two parental genome assemblies. To identify COs in each F2 individual in the

benchmark data, we mapped Illumina short reads to A6 genome, used 50-SNP sliding windows to identify the transition in parental origin (i.e., switch between homozygous A4 origin and heterozygous A6 origin), called a CO event when the parental tracks are consistent for at least 200 SNPs, and visually inspected each called event. To identify CO in PacBio pools, we mapped CLR reads to both parental genomes separately using minimap2. We used 10-SNP sliding windows along each read to identify transitions in parental origin that are characteristic of CO events. For a read to be considered containing a CO event, each parental haplotype/track needs to cover at least 5 SNPs, and the CO event is at least 2 kb from the edge of the alignment. Reads with more than one switch in parental origin were filtered due to them likely being caused by sequencing error, gene conversion, or, rarely, double crossovers.

CUT&Tag Experiment and Analysis. We followed the CUT&Tag@home protocol [V.3, (80, 81)] using H3K9me3 rabbit antibody (Abcam EPR16601) and anti-rabbit Secondary Antibody binding (EpiCypher 13-0047) to profile the distribution of H3K9me3 in A6 and A7 16 to 18 h embryos. The final CUT&Tag libraries were sequenced with Illumina NovaSeq 6000. We followed CUT&Tag Data Processing and Analysis Tutorial (81, 82) to analyze the generated data and estimated histone-modification magnitude (HM) for each position as the sequencing coverage. For H3K9me3 enrichment around TEs, we calculated the mean HM for 5 kb upstream and downstream flanking each TE normalized by the mean HM in 20 to 40 kb upstream and downstream windows of the respective TE, following ref. 50. Normalized HM above one is deemed enriched with H3K9me3. The H3K9me3 total mass is the sum of normalized HM across 1 kb windows where the normalized HM is above one.

Testing the Associations Between the Presence of TEs and CO Occurrence. For within-strain comparisons, we employed *Mann–Whitney U tests* and glm models to compare CO numbers and *Mann–Whitney U tests* and linear regression to compare the distance to the nearest CO between TE and control windows. Both regression analyses included type (whether a window is with or without TE), sequencing depth, and SNP density in a window as predictors.

For between-strain analysis, we selected control windows that were 1) >10 kb from TEs in both strains, 2) lacking H3K9me3 enrichment (<1) in both alleles, and 3) having similar SNP density ratios to TE windows. We compared CO differences (TE minus non-TE allele) between TE and control windows using two methods. First, we downsampled sequencing depth for TEs and control windows, performed one-tailed *Mann–Whitney U tests,* and repeated the process 1,000 times to get a median *P* for the *Mann–Whitney U tests.* Second, we bootstrapped control windows to match TE window counts per arm, downsampled a bootstrapped set, calculated mean CO differences for downsampled sets, and generated a null distribution of the median of the mean differences in CO number for 1,000 bootstrapped sets. One-tailed *P-values* were then estimated by comparing the median of mean differences in CO for downsampled TE windows. For analysis using distance to the nearest CO, if the nearest CO was unsampled during downsampling, we used the next nearest CO instead. All other analysis procedures are the same as those for analysis using CO numbers.

We classified TEs into those with and without H3K9me3 enrichment (adjacent H3K9me3 enrichment > 1 or not) or high/low H3K9me3 mass (use 50% quantile of H3K9me3 mass), depending on the analysis. TEs whose non-TE homologous alleles have adjacent H3K9me3 enrichment > 1 or H3K9me3 mass > 3 were excluded. We used the same two approaches as above to investigate whether TEs with H3K9me3 enrichment show significantly stronger suppression on CO occurrence, with the only difference being the null distribution was generated from bootstrapped non-H3K9me3 enriched TEs. For *Spearman correlation tests*, we used the median correlation coefficient and *P-values* of 1,000 downsampled sets.

Using DSPR RILs to Test the Effect of TEs on CO Occurrence. The DSPR RILs consist of inbred lines whose genomes are mosaics of the original eight parental lines. A subset of the DSPR RILs (28 RILs) have whole-genome Illumina sequencing data available [NCBI SRA PRJNA1161589; (44)], and we used previously developed hidden Markov model (76) to infer the underlying founder ancestry at each genomic position in each RIL. We inferred CO events at intervals where founder assignment probability transitioned from 95% for one founder to 95% for another in each RIL. We removed COs in heterochromatin and those present in multiple RILs to avoid double counting, yielding 1,801 euchromatic COs. We used SV identified in ref. 77 and the above approach to identify, annotate, and filter TEs in DSPR RILs, yielding 4,686 euchromatic TE insertions. We defined TE alleles as those where TEs are present in one of the two strains from which the CO was generated. Accordingly, an allele may be considered as a TE allele even if that particular TE is not present in the RIL. All other alleles were considered without TE alleles. We then estimated the median physical distance (in bp) between a TE insertion site and the nearest COs for alleles with and without TEs.

Estimation of Recombination Rates Around Transgenic TE Insertion. pAct5C promoter and the mCherry coding sequences were synthesized and cloned into pBS-KS-attB1_2 vector [obtained from DGRC, stock number 1322; (83)] by Genscript (Piscataway, NJ). Plasmid with TE sequences downstream of mCherry was obtained by cloning PCR amplified transposons into the above plasmid using Takara infusion technology. All plasmids (mCherry only, mCherry + 1360, mCherry + Roo, and mCherry + copia) were confirmed by sequencing and injected into MIMIC strain with docking site at 2L:17,325,574 (BDSC 48358) by BestGene (Chino Hills, CA). To estimate recombination rates around the construct insertion sites, we crossed each transgenic strain to *b/b* homozygotes (BDSC 227), collected ~1,000 mixed-sex F2 backcrossed offspring with *b/b* genotype, extracted DNA from selected F2 as a pool following (45), and sequenced it with Illumina NovaSeq 6000. Sequencing data were then analyzed using the analytical pipeline described in ref. 45. See *SI Appendix, Supporting Text* for detailed descriptions of materials and methods.

Data, Materials, and Software Availability. All the sequence data generated in this study was deposited to SRA under the Accession No. PRJNA1159377 (84). The scripts for performing the analyses are available at https://github.com/jumpingTE-LeeLab/TE_crossover (85) and https://github.com/weikevinhc/Marsupial (86).

ACKNOWLEDGMENTS. We appreciate Charles Langley for suggesting the use of PacBio sequencing, Serafin Colmenares for suggestions on the design of transgenic constructs, Jasmine Osei-Enin for laboratory assistant, Stephen Wright, Andrea Betancourt, Tony Long, J. J. Emerson, and Harsh Shukla for helpful discussion of the project, and Andrea Betancourt and Ching-Ho Chang for critically reading the manuscript. We thank the University of California High-Throughput Genomics Facility and High-Performance Cluster at University of California, Irvine for sequencing and computational resources. This work was supported by National Institute of Health (NIH) F32 GM099382 to E.G.K., the European Research Councilunder the European Union's Horizon 2020 research and innovation program, Grant Agreement No. 850405 and VIDI VI.Vidi.203.001 financed by the Dutch Research Council (NWO) to A.J., Natural Sciences and Engineering Research Council of CanadaDiscovery Grant (RGPIN-2023-05390) to K.H.-C.W., and NIH R35GM142494 to Y.C.G.L.

Author affiliations: [a]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; [b]Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands; [c]Division of Biological Sciences, University of Missouri, Columbia, MO 65211; and [d]Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

1. M. Zelkowski, M. A. Olson, M. Wang, W. Pawlowski, Diversity and determinants of meiotic recombination landscapes. *Trends Genet.* **35**, 359–370 (2019).
2. H. J. Muller, The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9 (1964).
3. A. S. Kondrashov, Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
4. R. A. Fisher, *The Genetical Theory of Natural Selection: A Complete. Variorum Edition* (OUP, Oxford, UK, 1930).
5. H. J. Muller, Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
6. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
7. J. Felsenstein, The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
8. A. H. Sturtevant, The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* **14**, 43–59 (1913).
9. S. E. Johnston, Understanding the genetic basis of variation in meiotic recombination: Past, present, and future. *Mol. Biol. Evol.* **41**, msae112 (2024).
10. J. Stapley, P. G. D. Feulner, S. E. Johnston, A. W. Santure, C. M. Smadja, Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philos. Trans. R. Soc. B: Biol. Sci.* **372**, 20160455 (2017).
11. C. S. Smukowski, M. A. F. Noor, Recombination rate variation in closely related species. *Heredity* **107**, 496–508 (2011).

12. K. Samuk, B. Manzano-Winkler, K. R. Ritz, M. A. F. Noor, Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. *Curr. Biol.* **30**, 1517–1528.e6 (2020).
13. J. M. Comeron, R. Ratnappan, S. Bailin, The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1002905 (2012).
14. K. R. Ritz, M. A. F. Noor, N. D. Singh, Variation in recombination rate: Adaptive or not? *Trends Genet.* **33**, 364–374 (2017).
15. B. Charlesworth, C. H. Langley, The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* **23**, 251–287 (1989).
16. M. G. Barrón, A.-S. Fiston-Lavier, D. A. Petrov, J. González, Population genomics of transposable elements in *Drosophila*. *Annu. Rev. Genet.* **48**, 561–581 (2014).
17. Y. C. G. Lee, C. H. Langley, Transposable elements in natural populations of *Drosophila melanogaster*. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* **365**, 1219–1228 (2010).
18. J. N. Wells, C. Feschotte, A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
19. G. Bourque *et al.*, Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
20. T. V. Kent, J. Uzunović, S. I. Wright, Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B* **372**, 20160458 (2017).
21. C. H. Langley, E. Montgomery, R. Hudson, N. Kaplan, B. Charlesworth, On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* **52**, 223–235 (1988).
22. E. S. Dolgin, B. Charlesworth, The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169–2177 (2008).
23. J. Y. Choi, Y. C. G. Lee, Double-edged sword: The evolutionary consequences of the epigenetic silencing of transposable elements. *PLoS Genet.* **16**, e1008872 (2020).
24. A. Kong *et al.*, A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
25. C. Ellermeier *et al.*, RNAi and heterochromatin repress centromeric meiotic recombination. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8701–8705 (2010).
26. C. J. Underwood *et al.*, Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Res.* **28**, 519–531 (2018).
27. J. B. Fernandes *et al.*, Structural variation and DNA methylation shape the centromere-proximal meiotic crossover landscape in *Arabidopsis*. *Genome Biol.* **25**, 30 (2024).
28. H. K. Dooner, L. He, Maize genome structure variation: Interplay between retrotransposon polymorphisms and genic recombination. *Plant Cell* **20**, 249–258 (2008).
29. L. He, H. K. Dooner, Haplotype structure strongly affects recombination in a maize genetic interval polymorphic for Helitron and retrotransposon insertions. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8410–8416 (2009).
30. B. Darrier *et al.*, High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. *Genetics* **206**, 1373–1388 (2017).
31. J. Pan *et al.*, A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
32. S. Yamada *et al.*, Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. *Cell Cycle* **16**, 1870–1884 (2017).
33. K. Choi *et al.*, Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Res.* **28**, 532–546 (2018).
34. A. P. Morgan *et al.*, Structural variation shapes the landscape of recombination in mouse. *Genetics* **206**, 603–619 (2017).
35. K. N. Crown, D. E. Miller, J. Sekelsky, R. S. Hawley, Local inversion heterozygosity alters recombination throughout the genome. *Curr. Biol.* **28**, 2984–2990.e3 (2018).
36. B. A. Rowan *et al.*, An ultra high-density *Arabidopsis thaliana* crossover map that refines the influences of structural variation and epigenetic features. *Genetics* **213**, 771–787 (2019).
37. R. K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285 (2007).
38. J. C. Peng, G. H. Karpen, Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet.* **5**, e1000435 (2009).
39. L. Maloisel, J.-L. Rossignol, Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev.* **12**, 1381–1389 (1998).
40. D. E. Miller *et al.*, A whole-chromosome analysis of meiotic recombination in *Drosophila melanogaster*. *G3* **2**, 249–260 (2012).
41. N. Zamudio *et al.*, DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* **29**, 1256–1270 (2015).
42. N. Altemose *et al.*, A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *eLife* **6**, e28383 (2017).
43. C. J. Underwood, K. Choi, Heterogeneous transposable elements as silencers, enhancers and targets of meiotic recombination. *Chromosoma* **128**, 279–296 (2019), 10.1007/s00412-019-00718-4.
44. E. G. King *et al.*, Genetic dissection of a model complex trait using the *Drosophila* synthetic population resource. *Genome Res.* **22**, 1558–1566 (2012).
45. K.H.-C. Wei, A. Mantha, D. Bachtrog, The theory and applications of measuring broad-range and chromosome-wide recombination rate from allele frequency decay around a selected locus. *Mol. Biol. Evol.* **37**, 3654–3671 (2020).
46. E. G. King *et al.*, Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res.* **22**, 1558–1566 (2012).
47. M. Chakraborty, J. J. Emerson, S. J. Macdonald, A. D. Long, Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.* **10**, 4872 (2019).
48. S. E. Hughes, D. E. Miller, A. L. Miller, R. S. Hawley, Female meiosis: Synapsis, recombination, and segregation in *Drosophila melanogaster*. *Genetics* **208**, 875–908 (2018).
49. J. M. Flynn *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020), 10.1073/pnas.1921046117.
50. Y. C. G. Lee, G. H. Karpen, Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. *eLife* **6**, e25762 (2017).
51. Y. Huang, H. Shukla, Y. C. G. Lee, Species-specific chromatin landscape determines how transposable elements shape genome evolution. *eLife* **11**, e81567 (2022).
52. N. C. Riddle *et al.*, Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.* **21**, 147–163 (2011).
53. M. F. Sentmanat, S. C. R. Elgin, Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14104–14109 (2012).
54. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
55. A. Dréau, V. Venu, E. Avdievich, L. Gaspar, F. C. Jones, Genome-wide recombination map construction from single individuals using linked-read sequencing. *Nat. Commun.* **10**, 4309 (2019).
56. H. Sun *et al.*, Linked-read sequencing of gametes allows efficient genome-wide analysis of meiotic recombination. *Nat. Commun.* **10**, 4310 (2019).
57. P. Xu, T. Kennell, M. Gao, R. P. Kimberly, Z. Chong, MRLR: Unraveling high-resolution meiotic recombination by linked reads. *Bioinformatics* **36**, 10–16 (2020).
58. G. A. Logsdon, M. R. Vollger, E. E. Eichler, Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **21**, 597–614 (2020).
59. P. S. Porsborg *et al.*, Insights into gene conversion and crossing-over processes from long-read sequencing of human, chimpanzee and gorilla testes and sperm. bioRxiv [Preprint] (2024). https://www.biorxiv.org/content/10.1101/2024.07.05.601967v1 (Accessed 13 August 2024).
60. R. Schweiger *et al.*, Insights into non-crossover recombination from long-read sperm sequencing. bioRxiv [Preprint] (2024). https://www.biorxiv.org/content/10.1101/2024.07.05.602249v1 (Accessed 13 August 2024).
61. R. H. Borts, J. E. Haber, Meiotic recombination in yeast: Alteration by multiple heterozygosities. *Science* **237**, 1459–1465 (1987).
62. M. Kupiec, T. D. Petes, Allelic and ectopic recombination between Ty elements in yeast. *Genetics* **119**, 549–559 (1988).
63. D. E. Symer *et al.*, Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* **110**, 327–338 (2002).
64. A. S. Waldman, Ensuring the fidelity of recombination in mammalian chromosomes. *BioEssays* **30**, 1163–1171 (2008).
65. D. A. Petrov, Y. T. Aminetzach, J. C. Davis, D. Bensasson, A. E. Hirsh, Size matters: Non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol.* **20**, 880–892 (2003).
66. J. M. Cridland, S. J. Macdonald, A. D. Long, K. R. Thornton, Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol. Biol. Evol.* **30**, 2311–2327 (2013).
67. G. E. Rech *et al.*, Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.* **13**, 1948 (2022).
68. Y. Huang, Y. C. G. Lee, Blessing or curse: How the epigenetic resolution of host-transposable element conflicts shapes their evolutionary dynamics. *Proc. Biol. Sci.* **291**, 20232775 (2024).
69. T. Sultana, A. Zamborlini, G. Cristofari, P. Lesage, Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* **18**, 292–308 (2017).
70. T. Connallon, C. Olito, Natural selection and the distribution of chromosomal inversion lengths. *Mol. Ecol.* **31**, 3627–3641 (2022).
71. K.H.-C. Wei, L. Gibilisco, D. Bachtrog, Epigenetic conflict on a degenerating Y chromosome increases mutational burden in *Drosophila* males. *Nat. Commun.* **11**, 5537 (2020).
72. B. Y. Kim *et al.*, Highly contiguous assemblies of 101 drosophilid genomes. *eLife* **10**, e66405 (2021).
73. P. H. Sudmant *et al.*, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
74. L. Quadrana *et al.*, The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5**, e15716 (2016).
75. X. Cao *et al.*, Polymorphic mobile element insertions contribute to gene expression and alternative splicing in human tissues. *Genome Biol.* **21**, 185 (2020).
76. E. G. King, S. J. Macdonald, A. D. Long, Properties and power of the *Drosophila* synthetic population resource for the routine dissection of complex traits. *Genetics* **191**, 935–949 (2012).
77. M. Chakraborty, J. J. Emerson, S. J. Macdonald, A. D. Long, Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.* **10**, 1–11 (2019).
78. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinf.* **10**, 421 (2009).
79. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
80. H. S. Kaya-Okur *et al.*, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
81. S. Henikoff, J. G. Henikoff, H. S. Kaya-Okur, K. Ahmad, Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife* **9**, e63274 (2020).
82. Y. Zheng, K. Ahmad, S. Henikoff, CUT&Tag data processing and analysis tutorial (2020). https://dx.doi.org/10.17504/protocols.io.bjk2kkye. Accessed 14 January 2023.
83. K. J. T. Venken *et al.*, MiMIC: A highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat. Methods* **8**, 737–743 (2011).
84. Y. Huang *et al.*, Polymorphic transposable elements contribute to variation in recombination landscapes. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1159377. Deposited 9 September 2024.
85. Y. Huang *et al.*, Polymorphic transposable elements contribute to variation in recombination landscapes. GitHub. https://github.com/jumpingTE-LeeLab/TE_crossover. Deposited 24 September 2024.
86. K.H.-C. Wei, A. Mantha, D. Bachtrog, The theory and applications of measuring broad-range and chromosome-wide recombination rate from allele frequency decay around a selected locus. GitHub. https://github.com/weikevinhc/Marsupial. Deposited 18 May 2020.