The megabase-scale crossover landscape is independent of sequence divergence

Qichao Lian¹, Victor Solier¹, Birgit Walkemeier¹, Bruno Huettel³, Korbinian Schneeberger^{1,2*},
Raphael Mercier^{1*}

¹Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

²Faculty of Biology, LMU Munich, 82152, Planegg-Martinsried, Germany

³Max Planck-Genome-center Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Corresponding author: Raphael Mercier (mercier@mpipz.mpg.de), Korbinian Schneeberger
(schneeberger@mpipz.mpg.de)

Keywords: meiotic crossover, inbred line, genomic polymorphisms, chromatin accessibility,
Arabidopsis thaliana

Abstract

Meiotic recombination frequency varies along chromosomes and strongly correlates with sequence divergence. However, the causality underlying this correlation is unclear. To untangle the relationship between recombination landscapes and polymorphisms, we characterized the genome-wide recombination landscape in the absence of polymorphisms, using *Arabidopsis thaliana* homozygous inbred lines in which a few hundred genetic markers were introduced through mutagenesis. We found that megabase-scale recombination landscapes in inbred lines are strikingly similar to the recombination landscapes in hybrids, with the sole exception of heterozygous large rearrangements where recombination is prevented locally. In addition, we found that the megabase-scale recombination landscape can be accurately predicted by chromatin features. Our results show that polymorphisms are not causal for the shape of the megabase-scale recombination landscape, rather, favor alternative models in which recombination and chromatin shape sequence divergence across the genome.

Introduction

Meiotic recombination is initiated by the formation of numerous DNA double-strand breaks, a minority of which are repaired as crossovers (COs), resulting in reshuffling of the genetic material between generations, and is thus crucial for diversity, adaptation, evolution, and breeding¹⁻⁴. Two pathways have been described for CO formation (class I and II)^{1,3,4}. Class I COs represent the vast majority of COs and are subject to interference, the propensity of COs to be widely spaced along chromosomes⁵.

COs are not homogeneously distributed and recombination frequencies vary along chromosomes^{6,7}. Many different features are correlated with the recombination landscape. One consistent pattern across monocentric species is the suppression of COs at and next to centromeres^{3,8,9}. The landscape can also differ between the two sexes of the same species, a phenomenon called heterochiasmy¹⁰⁻¹³. Polymorphism between homologs can negatively affect crossovers, as observed very locally at crossover hotspots or even completely suppress crossovers in cases of large polymorphisms, like megabase-scale inversions^{7,14-21}. In contrast, however, heterozygous regions in *Arabidopsis thaliana* showed increased recombination rates when juxtaposed with homozygous regions suggesting that the density of small-scale sequence divergence can increase recombination rates²². In addition, increasing single nucleotide polymorphism (SNP) density in hybrids associates positively with COs, and the pericentromeric regions that are dense in polymorphisms are also elevated in COs, potentially due to a positive feedback of mismatch recognition during CO formation²¹. A positive correlation between polymorphisms and recombination landscapes can be also observed in natural populations: in many species, historical recombination landscapes as deduced from linkage disequilibrium are positively correlated with SNP densities²³⁻²⁷.

To better understand the relationship between polymorphisms and meiotic recombination, we here compared CO distribution along chromosomes in the absence (inbred lines) and presence (hybrids) of polymorphisms. In hybrids, the numerous DNA polymorphisms can be used to precisely map COs^{6,28-37}, while this is not an option in homozygous inbred lines. Instead, CO frequency could be estimated by cytological techniques, but this poses several challenges, such as the difficulty in identifying individual chromosomes, low resolution, and the limited number of cells that can be analysed. Alternatively, fluorescence-tagged lines (FTLs) could be used to measure recombination in intervals flanked by markers conferring fluorescence in seeds or pollen grains but are not suitable for mapping the genome-wide CO landscape³⁸⁻⁴⁰.

In this study, we developed a method to analyse genome-wide recombination landscapes in inbred lines based on the introduction of a few markers by chemical mutagenesis. We characterized the crossover landscapes of two *Arabidopsis* inbred lines (Columbia (Col) and Landsberg (*Ler*)) and compared them to the CO landscape of the Col/*Ler* F1 hybrid as well as with the historical recombination pattern in this species. The CO landscapes of the inbred lines were remarkably similar to each other as well as to the historical recombination landscape and that of the hybrid with the exception of local suppression due to large heterozygous rearrangements. This shows that polymorphism density, with the exception of large structural variations, is not a major determinant of the CO landscape.

It has been shown before that COs co-localize with gene promoters and with regions of open chromatin and low levels of DNA methylation^{3,30,41,42}. We now show that chromatin accessibility, gene density, and DNA methylation are sufficient to explain 85% of the megabase-scale recombination landscape in *Arabidopsis*.

Results

To investigate the landscape of meiotic recombination in *Arabidopsis thaliana* inbred lines, we applied moderate EMS mutagenesis to introduce genetic markers into the genomes of *A. thaliana* Col-0 and *Ler*. Independent M₂ mutants were crossed to generate F₁*s, and independent F₁*s were reciprocally crossed to generate F₁ populations, which were used to analyse recombination independently in female and male meiosis. (Figure 1, Supplemental Figure S1, Supplemental Table S1, Materials and methods). Through Illumina short-read genome sequencing of F₁*s and F₁s, we identified 838–955 high-confidence mutations segregating in the Col populations and 471–539 in the *Ler* populations (Supplemental Table S1), which is a negligible level compared to the natural divergence between these two accessions^{14,37,43}. The markers were pretty uniformly distributed across the chromosomes, which allowed the identification of meiotic CO events (Supplemental Figure S2). We analysed four independent pairs of populations of both accessions, with a total of 309 and 309 progeny derived from male and female meiosis in Col and 253 and 251 progenies derived from male and female meiosis in *Ler*. Overall, we identified 3,155 COs in Col (examples shown in Supplemental Figure S3 and S4; median resolution 522 kb; Supplemental Table S3) and 2,004 (median resolution 855 kb, Supplemental Table S4) COs in *Ler*. We observed a consistent CO frequency between the replicate populations (Figure 2A and 2B), arguing against the unlikely possibility that an EMS mutation dominantly affected CO numbers in the F₁*. CO numbers were correlated neither with sequence depth nor with the number of markers, suggesting an absence of bias in CO detection (Supplemental Figure S5). In Col male meiotic cells, CO number was previously estimated to be ~10.5 with MLH1-HEI10 foci counts^{44,45}. However, MLH1 and HEI10 are exclusive markers of class I COs, and do not mark class II COs, which are estimated to represent ~15% of the total COs^{1,46}, suggesting a total number of ~12.1 COs per male meiocytes. Here, we detected an average of 6.13 COs per Col male gamete, which corresponds to 12.26 COs per male meiocyte, fitting very well with the expected number. Altogether, this shows that our method robustly detects COs in inbred lines.

To compare the recombination landscapes of Col and *Ler* with the corresponding F₁ hybrid, we sequenced reciprocal back-crosses of Col/*Ler* F₁ hybrids with Col to identify COs in 428 and 294 progenies derived from female and male meiosis, respectively. We identified 1,192 COs (median resolution 739 bp) and 1,587 COs (median resolution 1,019 bp) in female and male hybrids,

respectively (Supplemental Figure S6, Supplemental Table S5). The female and male high-resolution CO distribution that we obtained is consistent with a previous dataset that described female/male CO landscapes with lower resolution¹¹ and CO distribution in the same hybrid in F2s that does not distinguish female and male COs⁷ (Supplemental Figure S7–8, Supplemental Table S6). Comparison of the genomic compartments where CO occurred did not reveal differences between females and males, with COs notably enriched in promotor regions in both sexes. This suggests that the factors driving fine-scale CO placement are similar in female and male meiosis (Supplemental Figure S7D–E).

In all three types of populations, Col, *Ler*, and hybrid, we observed heterochiasmy, i.e., significantly more COs in male than female meiosis (Mann–Whitney test, $p < 2.2 \times 10^{-16}$, Figure 2C). In male meiosis, both the highest (Col, 6.13) and the lowest (*Ler*, 4.62) numbers of COs are observed in inbred lines, with the hybrid exhibiting an intermediate number of COs (5.4), suggesting that CO number in males is mainly genetically controlled in trans and not by sequence polymorphism. In females, the highest CO number is also observed in Col (4.08), with less COs in *Ler* (3.34, $p = 1.2 \times 10^{-7}$), indicating that the same trans mechanism also influences CO frequency in females. However, an even lower level of COs is observed in the hybrid (2.79, $p = 0.0002$), suggesting that an additional phenomenon is responsible for the reduced CO frequency specifically in female hybrids. In all contexts, CO interference is more pronounced in females than males, with the strongest interference observed in female hybrids (Supplemental Figure S9–10). This anti-correlation between CO interference and CO numbers suggests that modulation of CO interference is an important driver of CO numbers. In both sexes of the three backgrounds, CO number is positively correlated with chromosome length, except for the female hybrid where the curve is almost flat at just above 0.5 CO per chromosome per gamete, corresponding to one CO per bivalent and a very strong CO interference (Supplemental Figure S11).

Along chromosomes, a strikingly similar pattern is observed in the three genetic backgrounds. COs are markedly suppressed at the centromeric regions and tend to be frequent at the edge of pericentromeres in both female and male meiosis. In all three backgrounds, the female and male recombination landscapes tend to diverge with decreasing distance from telomeres, with distal regions being among the highest recombination intervals in males and the lowest in females (Figure 2D–F, Supplemental Figure S12). Strikingly, CO distributions are more closely correlated

between the same sexes across the three different backgrounds than between the two sexes in the same background (Figure 2, Supplemental Figure S14). For example, female hybrids are more similar to female Col and *Ler* (Spearman's correlation $r_s = 0.62$ and 0.64) than to male hybrids ($r_s = 0.26$). Thus, sequence divergence appears to have a far lesser impact on the CO landscape than the sex of the meiocyte.

In addition, we wanted to compare the contemporary CO landscapes with the historical landscape. We reconstructed a historical recombination map using a set of non-singleton SNPs generated from 2,029 accessions (Figure 3A and Supplemental Figure S13)^{47,48}. Confirming previous findings, the historical CO landscape is strongly correlated with the sequence diversity (Figure 3A). The historical landscape is the result of combined female and male recombination, and we thus compared it to the merged female and male dataset for the inbred lines and to the previously described large Col/*Ler* F2 dataset⁷. (Figure 3D–F and Supplemental Figure S14). To facilitate the comparison of the landscapes independently of total CO numbers, we show both the observed CO density (cM/Mb, Figure 3B) and the corresponding normalized distribution (Figure 3A and 3C). Strikingly, the CO landscape in the two inbred lines and hybrid all appear remarkably similar to each other. The distributions of recombination in the inbred and the hybrid genomes closely match each other (Figure 3B), with co-localization of many peaks and valleys, including large peaks on both sides of the centromeric regions, but also in the middle of the arms (Figure 3C). This similarity is confirmed by genome-wide correlation analyses (non-linear correlation $r_n = 0.73$ – 0.78 , Figure 3D), which are even stronger when chromosome arms and peri-centromeres are considered separately (Figure 3E–F, $r_n = 0.78$ – 0.89). The historical recombination landscape is also very similar to the three contemporary landscapes, with most peaks being conserved ($r_n = 0.6$ – 0.7). This shows that the global CO landscape is largely independent of the presence (hybrid and historical) or absence (Col and *Ler*) of polymorphisms between the two chromosomes that recombined.

One notable exception is observed at position ~2 Mb of chromosome 4, with suppression of CO in the hybrid that is not observed in the inbred lines and the historical landscape (Figure 3A–B, 2D–F). This region corresponds to a large ~1.2 Mb genomic inversion, which suppresses recombination in the Col/*Ler* hybrid where it is heterozygous^{7,43,49,50}. We then asked if the smaller rearranged regions are also depleted for COs. We explored the overlap of COs with the non-syntenic regions by employing permutation tests in F2 hybrids and observed a strong depletion of CO in non-syntenic regions (Supplemental Figure S15B, $p = 0.0002$) and CO enrichment in the

adjacent regions (Supplemental Figure S15C, $p = 0.0002$), confirming that structural arrangements are correlated with inhibition of CO formation in hybrids. The CO resolution obtained for inbred lines did not allow us to test if these regions recombine normally in Col and *Ler*, as is the case for the unique large inversion.

Consistent with previous analyses²⁷, we found that the historical recombination rate is highly correlated with sequence diversity along chromosomes (Figure 3A and 3D–F, Supplemental Figure S14), and contemporary COs in the Col/*Ler* hybrid are correlated with SNP density between Col and *Ler*^{41,51}. As shown above, all the CO landscapes show high similarity. Consequently, the CO landscapes in Col and *Ler* are correlated with Col/*Ler* SNPs ($r_n = 0.46–0.57$) and sequence diversity ($r_n = 0.46–0.64$), whereas these polymorphisms were absent in the Col and *Ler* inbred lines where these COs were produced. This strongly argues against the possibility that the polymorphisms shape the CO landscape, as the CO landscape is largely unchanged when polymorphisms are absent (with the notable exception of large rearrangements).

To decipher the contributions of genomic and epigenomic features to shaping the CO landscape, we analysed the recombination distribution in Col with a total of 17 different features obtained in the same accession, including GC content; gene and transposable element densities; origins of DNA replication (BrdU-seq)⁵²; meiotic DSBs (SPO11-1-oligonucleotides)⁵³; chromatin accessibility in flowers (ATAC-seq and DNase-seq)^{54–57}; euchromatic (H3K4me1, H3K4me2 and H3K4me3, ChIP-seq)^{53,58} and heterochromatin histone modification marks in flower buds (H3K9me2 and H3K27me1, ChIP-seq)⁵⁸; DNA methylation in male meiocytes (CG, CHG and CHH contexts, BS-seq)⁵⁹; nucleosome occupancy in buds (micrococcal nuclease sequencing, MNase-seq)⁵³; and the meiotic cohesin REC8 occupancy (ChIP-seq, Figure 4A, Supplemental Table S2)⁵⁸.

Genome-wide, CO distribution is correlated with many genetic and epigenetic features, notably positively with open chromatin (ATAC) ($r_n = 0.71$), H3K4me1 ($r_n = 0.65$), gene density ($r_n = 0.64$) and CHG methylation ($r_n = 0.55$) (Supplemental Figure S16A). These correlations are at least partially driven by the centromeric regions, at which COs are abolished and where these features are strongly depleted (Supplemental Figure S16B and S17). However, considering only the chromosome arms, the correlations are almost the same between COs and ATAC ($r_n = 0.71$), H3K4me1 ($r_n = 0.64$), CHG methylation ($r_n = 0.59$) gene density ($r_n = 0.57$), and other features

(Figure 4A), suggesting a relationship between CO density and chromatin features beyond the centromere.

We next used a machine-learning algorithm (random forest) to analyse the contribution of 17 genomic features on chromosome arms in Col. We first developed a model to predict the frequency of meiotic recombination for a given interval with all the given features together and analysed how the model learned to perform the prediction. ~96.3% of variation can be explained by the random forest predictive model, as the predicted CO landscape is highly correlated with the observation (Spearman's $r_s = 0.997$, Figure 4D, Supplemental Figure S18A). As shown in Figure 4B and 4C, the most important feature was open chromatin (ATAC), which explained ~28.7% of the variation. In order to further investigate the performance of the random forest model, we generated a training set by simply sampling 70% of the intervals, with the remaining 30% serving as the testing set. The model trained with the training set performed well with the test set; it could explain ~94.5% of the variation in the training set ($r_s = 0.996$), resulting in a significant correlation ($r_s = 0.977$) between the prediction and the observations from the test set (Supplemental Figure S18B, S18E and S18F). Furthermore, we observed that the top three features (open chromatin, gene density, and CHH DNA methylation) can explain ~85% of the variation (Supplemental Figure S18C). Strikingly, when we trained the model with only these top three features, we observed a high correlation between the predicted and observed CO distribution ($r_s = 0.985$, training set $r_s = 0.982$, and testing set $r_s = 0.92$) (Figure 4E, Supplemental Figure S18D, S18G–H). Altogether, these results show that chromatin accessibility, gene content, and DNA methylation are the major predictors of the mega-scale distribution of meiotic recombination in *Arabidopsis thaliana*.

Discussion

In this work, we developed a method to analyse the genome-wide recombination landscape in inbred lines and applied it to the *Arabidopsis* accessions Col and *Ler*. This method is based on the introduction of a limited number of markers and allows robust detection of COs. The strategy can be applied to any species for which homozygous lines and mutagenesis are available. We expect this method to be particularly useful for exploring the natural variation of recombination landscapes in species that are inbred lines in the wild (e.g., *Arabidopsis*) and for exploring CO distribution in species where inter-strain crosses are problematic (e.g., in the fission yeast *Schizosaccharomyces pombe* because of killer meiotic drivers)⁶⁰.

Meiotic recombination frequency has previously been studied in hybrids in many species and varies along chromosomes and positively correlates with the distribution of polymorphisms^{3,7,12,21,23-27,51}. One of the possible causes for these correlations is that heterozygosity may favour the formation of COs^{21,22}. We showed here that the megabase-scale recombination landscape in inbred lines is strikingly similar to those of hybrids as well as to historical patterns. The observation that the CO landscape is maintained in the absence of polymorphism leads to the conclusion that polymorphisms are not a major determinant of the megabase-scale CO distribution. Polymorphisms, including SNPs and small rearrangements, influence the local recombination pattern¹⁸⁻²¹, but this effect is not manifest at the megabase-scale; at this range, the landscape appears to be unaffected by polymorphism density. An important exception is large-scale genomic rearrangements, such as the ~1.2 Mb inversion (between Col and *Ler*) on chromosome 4, where COs are abolished in hybrids, while the corresponding regions are CO-proficient in isogenic lines. Smaller structural variations are also associated with CO depletion in the hybrid and are presumably CO-prone in the inbred lines, though we cannot confirm this because of the relatively low resolution in CO position.

In many species, COs tend to colocalize with nucleosome- and methylation-depleted gene promoters^{3,7,30,34,41,61,62}, consistent with our observation in Col/*Ler* hybrids. Moreover, in this study, we found that among a total of 17 genomic and epigenomic features, open chromatin (ATAC), gene density, and DNA methylation in the CHH context are the most potent factors for predicting the distribution of COs along chromosomes in inbred Col. Interestingly, these three features were enough to predict ~85% of the CO variation along chromosome arms. We do not claim that these

three features alone directly control CO positions, but these results show that the chromatin context, which can be largely captured using these three features, efficiently predicts mega-base CO landscapes.

Our results suggest that the large-scale CO landscape is not driven by the polymorphism density. Thus, two possibilities may explain the correlation between polymorphism density and recombination observed in hybrid and historical landscapes. First, the recombination landscape could gradually shape the polymorphism density. Indeed, meiotic recombination is mutagenic, which might be an important driver of genetic diversity and genome evolution^{25,28,62-66}. In addition, selection tends to reduce polymorphisms in regions with low recombination rates: both the spread of beneficial mutations and the removal of deleterious mutations by selection reduce polymorphism levels and this effect is larger if recombination is low⁶⁷. A second, not mutually exclusive hypothesis, is that local differences in chromatin features not only influence the distribution of recombination, but that chromatin, independently of recombination, contributes to genomic diversity by shaping differences in local mutation rates along the genome.

While the recombination landscape is largely conserved between the inbred lines and the hybrid, they differ in the total CO number. Globally, there are more COs in Col than in *Ler*, with the hybrid having an intermediate number. This is consistent with previous observations in a few crossover reporter intervals and is probably largely driven by an allelic difference in the pro-CO factor HEI10; the Col allele was shown to increase the number of COs compared to the *Ler* allele in a co-dominant manner⁶⁸.

When female and male recombination are analysed separately, CO rates are always highest in Col and lowest in *Ler* and always higher in males than females. The hybrid exhibits an intermediate recombination rate of CO formation in males, but in contrast the hybrid female has less COs than the two inbred lines. This suggests that some mechanism specifically reduces CO frequency in female hybrids compared to the female inbred lines. One possibility is that class II COs, which represent a minority of COs, are inhibited in the presence of polymorphism and thus reduced in hybrids^{49,69}. This would have a proportionally larger effect on female meiosis where class I COs are less numerous than in males, and thus account for the very low level of COs in hybrid females. As class II COs are non-interfering, this would also explain why CO interference is stronger in

hybrid females than in inbred females and could especially account for the absence of very close double-COs (Supplemental Figure S10)⁷⁰. Interestingly, in female hybrids, the number of COs observed was very close to the obligate one crossover per bivalent (0.5 crossover per chromatid), suggesting that the CO landscape in female hybrids corresponds to the distribution of the obligate CO, which thus occurs highly preferentially in the proximal regions. The most striking contrast between females and males was the pronounced difference in the distal regions, where males tend to recombine more than females in both pure lines and hybrids. This further confirms that the megabase-scale recombination landscape is largely independent of polymorphisms, and instead suggests that the cellular environment plays a much more critical role, notably by controlling chromosome organization^{70,71}.

An improved understanding of the control of meiotic recombination along the chromosome opens the possibility of manipulating COs and increasing recombination rates globally^{49,69,72} and in reluctant regions. This would facilitate the reshuffling of genomic material, breaking of the linkage between beneficial and deleterious alleles, and allow the combination of favourable alleles in elite varieties.

Materials and methods

Isogenic population construction and sequencing

Plants were grown in greenhouses or growth chambers (16-h day/8-h night, 20 °C). Wild-type Col-0 and *Ler*-1 are 186AV1B4 and 213AV1B1 from the Versailles *Arabidopsis thaliana* stock center (<http://publiclines.versailles.inra.fr/>). For each accession, seeds were subjected to EMS mutagenesis as described in⁷³, and four independent M2 plants were crossed to produce two independent F1*, which were consequently heterozygous for a set of EMS mutations (Figure 1). Then, the two F1* plants were reciprocally crossed to generate two F1 populations. To test the robustness of the results and detect the unlikely possibility that a dominant modifier of recombination was caused by an EMS-induced mutation, two independent replicates of the entire process were performed for each accession. These F1* and F1 plants were then used for CO analysis by whole-genome sequencing (Figure 1, Supplemental Table S1). Leaf samples from the populations were used for DNA purification and library preparation for 2x150bp HiSeq 3000 Illumina sequencing⁷⁴. To detect the markers, we sequenced genomic DNA from the F1* (~59 x and ~16x, in Col and *Ler* respectively) and F1s (~4.8x and ~5.0x)

Identifying and genotyping EMS-induced mutations

For each individually sequenced F1* and F1 plant of *Arabidopsis* Col and *Ler* accessions, the whole-genome resequencing reads were aligned against the Col-0 TAIR10 reference genome^{75,76} and *Ler* assembled genome⁴³ by BWA v0.7.15-r1140⁷⁷ with default parameters, and variant calling in F1* populations was performed using inGAP-family³⁷, separately. To obtain high-quality mutation marker lists, we first removed non-allelic markers using inGAP-family with input from the tandem replicates and structural variants predicted using Tandem Repeats Finder v4.09⁷⁸ and inGAP-family, respectively, and further filtered variations that not meet the following criteria: (i) heterozygous genotype with alternative allele frequency from 0.4 to 0.6, (ii) specific to each of the F1*, and (iii) GA to CT substitution. Then, the read count and genotype map of mutation markers of each F1* was generated from their F1 progenies by inGAP-family, which was subsequently used for mutation phasing and CO identification. In order to properly compare CO landscapes in isogenic and hybrid lines, we transferred the coordinates of mutations in the *Ler* population to Col-0 by using syntenic alignments identified by SyRI v1.2⁷⁹.

Two additional replicates in Col (C and D) were discarded, because the marker analysis showed that one of the F1* was resulting from an accidental selfing and not from a cross. Two additional replicates in *Ler* were also discarded (C and H), because the number of detected mutation markers (<350) was insufficient for good genome coverage.

Phasing mutations and CO identification in inbred lines

To phase the EMS-induced mutation markers, we employed a hierarchical clustering-based sliding window method, with a window size of ten mutation markers and step size of one mutation marker (Supplemental Figure S1). For each window, the genotype map of the mutation markers was constructed and used as input for clustering, resulting in two groups: one consisting of wild-type samples and one comprising mutant samples. The genotype and phase of mutation was evaluated by the vote of covered sliding windows. During this process, uncovered and low-covered markers were imputed and corrected. The CO events were defined as consistent switches of phase of mutation markers along chromosome arms, and the border was further refined by examining the wild-type allele of the mutation. For the termini of chromosomes, COs were validated as switches with one well-supported mutant allele or more than ten reads supported by the consecutive wild-type allele of the variant marker. The CO interference was analyzed using MADpattern^{80,81}.

CO analysis in hybrid population

The Col/*Ler* and wild-type Col plants were reciprocally crossed to construct female and male populations (428 and 294 plants, respectively). Leaf samples of the backcross populations were collected for DNA purification, library preparation, and Illumina sequencing⁷⁴. In addition, the raw reads of the Col/*Ler* F2 population were downloaded from ArrayExpress with the accession number E-MTAB-8165⁷.

The quality of the raw sequencing datasets was checked using FastQC v0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and then adapters and low-quality bases were trimmed using Trimmomatic v0.38⁸², with parameters “LEADING:5 TRAILING:5 SLIDINGWINDOW:5:20 MINLEN:50”. The mapping of sequencing reads, generation of high-confidence SNP markers between Col and *Ler*, meiotic CO prediction, and filtering of the poorly covered and potentially contaminated samples were performed using a previously described

method⁴⁴. To avoid false genotyping, we selected 0.95 as the threshold allelic ratio for the determination of homozygosity in F1 hybrids. Identified COs were manually checked at random using inGAP-family.

SPO11-1-oligo, BrdU-seq, ChIP-seq, MNase-seq, DNase-seq, and ATAC-seq data analysis

The short reads from public datasets (Supplemental Table S2) were quality-checked with FastQC. The specific 3' adapter and 5' ends sequences were trimmed before alignment by Cutadapt v1.9.1⁸³ as described⁵³. For BrdU-seq and ATAC-seq datasets, the reads were processed with Trimmomatic to remove potential adapter sequences and low-quality bases, with "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Duplicated reads were removed using BBMap (<https://github.com/BioInfoTools/BBMap>). Then, clean reads were aligned to the TAIR10 reference genome using Bowtie2 v2.2.8⁸⁴ with settings "--very-sensitive -k 10" for single-end datasets and further settings "--no-discordant --no-mixed" for paired-end datasets. The uniquely mapped reads were kept for subsequent analysis, which were processed by Samtools v1.9⁸⁵ and Sambamba v0.6.8⁸⁶. For all sequencing data, coverage across the genome was evaluated and normalized with bins per million mapped reads (BPM) in bedGraph format using bamCoverage v3.4.3⁸⁷.

Bisulfite sequencing data analysis

The quality and adapter sequencing of raw reads were examined by FastQC. The sequencing reads were mapped to the TAIR10 reference genome with Bismark v0.22.0⁸⁸, with the following setting: -q -bowtie2 -N 1 -L 24. Reads that mapped to multiple positions and duplicated alignments were removed. Methylated cytosines in the CG, CHG, and CHH contexts, and the level of methylation were extracted for subsequent association analysis.

Genome-wide CO distribution correlation analysis

The chromosomal profiles of COs, genomic, and epigenomic features were estimated in 50kb windows along chromosomes. For a given window, the recombination frequency was normalized with the total CO number within the corresponding chromosome. Then, all of the COs, genomic and epigenomic data were smoothed with 40 nearby windows using the filter function and then

normalized using the scale function in the R environment. The non-linear correlation matrices were calculated using the nlcor package (<https://github.com/ProcessMiner/nlcor>) in R, at the genome, chromosome arms, and pericentromeric scale respectively. The constitution (pericentromeres, centromeres, and arms) of the TAIR10 reference genome was adopted from Underwood et al.⁶¹. Here, all the random forest models were trained using randomForest v4.6-14 package in R, with the setting of “mtry=3, importance=TRUE, na.action=na.omit, ntree=2000”.

Estimating nucleotide diversity and historical recombination rate

For the sequence polymorphism data of 2,029 Arabidopsis accessions from the 1,001 Genomes Project⁴⁷ and the RegMap population⁴⁸, we first selected diallelic SNP positions with <20% missing data and >5% minimum allele frequency using VCFtools v0.1.16⁸⁹. Then, we masked SNPs located in (i) tandem repeat regions (Tandem Repeats Finder output), (ii) repetitive elements and low-complexity regions (extracted from the masked TAIR10 reference genome), (iii) transposable elements (TAIR10 annotation) and (iv) centromeric regions (definition adopted from Underwood et al.⁶¹). Finally, we obtained a collection of 905,613 SNPs from 2,029 accessions for CO frequency analysis. FastEPRR v2.0⁹⁰ was employed for estimating population recombination rates ($\rho = 4Ner$, where N_e is the effective population size and r is the recombination rate of the window), with 50 kb non-overlapping window size. The nucleotide diversity of each 50 kb non-overlapping window along chromosomes was calculated using VCFtools.

Data availability

The raw sequencing data of individuals of Col, *Ler* inbred lines and the Col/*Ler* F1 hybrid can be accessed in ArrayExpress under the accession numbers E-MTAB-11248, E-MTAB-11249, E-MTAB-11250, E-MTAB-11251, E-MTAB-11254, respectively. The list of COs identified in Col, *Ler*, F1 hybrid (female and male), and F2 hybrid can be found in Supplemental Table S3-S6.

Acknowledgements

We would like to thank the Max Planck Genome center for DNA extraction, library preparation and sequencing, Hequan Sun and Wen-Biao Jiao for helpful discussion, Charles Underwood, Ian

Henderson and Andrew Tock for the help with SPO11-oligos analysis. We thank Wayne Crismani and Andrew Lloyd for critical reading of the manuscript.

This work was support by core funding from the Max Planck Society, and an Alexander von Humboldt Fellowship to Q.L.

Author contributions

Q.L., K.S. and R.M. designed the research. Q.L., K.S. and R.M. analyzed the data. V.S. and B.W. generated plant materials. B.H. supervised the whole-genome sequencing work. Q.L. and R.M. wrote the article with input from K.S.

References

1. Mercier R, Mezard C, Jenczewski E, Macaisne N, Grelon M. The Molecular Biology of Meiosis in Plants. *Annual Review of Plant Biology*, Vol 66 **66**, 297-327 (2015).
2. Wang Y, Copenhaver GP. Meiotic Recombination: Mixing It Up in Plants. *Annu Rev Plant Biol* **69**, 577-609 (2018).
3. Zelkowski M, Olson MA, Wang M, Pawlowski W. Diversity and Determinants of Meiotic Recombination Landscapes. *Trends Genet* **35**, 359-370 (2019).
4. Gray S, Cohen PE. Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annu Rev Genet* **50**, 175-210 (2016).
5. von Diezmann L, Rog O. Let's get physical - mechanisms of crossover interference. *J Cell Sci* **134**, (2021).
6. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479-485 (2008).
7. Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D. An Ultra High-Density Arabidopsis thaliana Crossover Map That Refines the Influences of Structural Variation and Epigenetic Features. *Genetics* **213**, 771-787 (2019).
8. Fernandes JB, Wlodzimierz P, Henderson IR. Meiotic recombination within plant centromeres. *Curr Opin Plant Biol* **48**, 26-35 (2019).
9. Nambiar M, Smith GR. Repression of harmful meiotic recombination in centromeric regions. *Semin Cell Dev Biol* **54**, 188-197 (2016).
10. Lenormand T, Dutheil J. Recombination difference between sexes: a role for haploid selection. *PLoS Biol* **3**, e63 (2005).
11. Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mezard C. Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. *PLoS Genet* **7**, e1002354 (2011).
12. Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci* **372**, (2017).
13. Sardell JM, Kirkpatrick M. Sex Differences in the Recombination Landscape. *Am Nat* **195**, 361-379 (2020).
14. Zapata L, et al. Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E4052-E4060 (2016).
15. Zetka MC, Rose AM. The meiotic behavior of an inversion in Caenorhabditis elegans. *Genetics* **131**, 321-332 (1992).
16. Jaarola M, Martin RH, Ashley T. Direct evidence for suppression of recombination within two pericentric inversions in humans: a new sperm-FISH technique. *Am J Hum Genet* **63**, 218-224 (1998).
17. Demirci S, Peters SA, de Ridder D, van Dijk ADJ. DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *Plant J* **95**, 686-699 (2018).
18. Borts RH, Haber JE. Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* **237**, 1459-1465 (1987).
19. Baudat F, de Massy B. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res* **15**, 565-577 (2007).
20. Serra H, Choi K, Zhao XH, Blackwell AR, Kim J, Henderson IR. Interhomolog polymorphism shapes meiotic crossover within the Arabidopsis RAC1 and RPP13 disease resistance genes. *Plos Genetics* **14**, e1007843 (2018).
21. Blackwell AR, et al. MSH2 shapes the meiotic crossover landscape in relation to interhomolog polymorphism in Arabidopsis. *Embo J* **39**, e104858 (2020).
22. Ziolkowski PA, et al. Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. *Elife* **4**, (2015).
23. Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature* **356**, 519-520 (1992).

24. Nordborg M, *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**, e196 (2005).
25. Spencer CC, *et al.* The influence of recombination on human genetic diversity. *PLoS Genet* **2**, e148 (2006).
26. Gore MA, *et al.* A First-Generation Haplotype Map of Maize. *Science* **326**, 1115-1117 (2009).
27. Kim S, *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* **39**, 1151-1155 (2007).
28. Qi J, Wijeratne AJ, Tomsho LP, Hu Y, Schuster SC, Ma H. Characterization of meiotic crossovers and gene conversion by whole-genome sequencing in *Saccharomyces cerevisiae*. *Bmc Genomics* **10**, 475 (2009).
29. Lu PL, *et al.* Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing *Landsberg erecta* and all four products of a single meiosis. *Genome Research* **22**, 508-518 (2012).
30. Wijnker E, *et al.* The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. *Elife* **2**, e01426 (2013).
31. Qi J, Chen YM, Copenhaver GP, Ma H. Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10007-10012 (2014).
32. Liu H, *et al.* Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biol* **16**, 15 (2015).
33. Si W, *et al.* Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F2 plants. *New Phytol* **206**, 1491-1502 (2015).
34. Kianian PMA, *et al.* High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. *Nature Communications* **9**, 2370 (2018).
35. Luo C, Li X, Zhang QH, Yan JB. Single gametophyte sequencing reveals that crossover events differ between sexes in maize. *Nature Communications* **10**, 785 (2019).
36. Dreissig S, *et al.* Natural variation in meiotic recombination rate shapes introgression patterns in intraspecific hybrids between wild and domesticated barley. *New Phytol* **228**, 1852-1863 (2020).
37. Lian Q, Chen Y, Chang F, Fu Y, Qi J. inGAP-family: Accurate Detection of Meiotic Recombination Loci and Causal Mutations by Filtering Out Artificial Variants due to Genome Complexities. *Genomics Proteomics Bioinformatics*, (2021).
38. Emmanuel E, Yehuda E, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA. The role of AtMSH2 in homologous recombination in *Arabidopsis thaliana*. *Embo Reports* **7**, 100-105 (2006).
39. Berchowitz LE, Copenhaver GP. Fluorescent *Arabidopsis* tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. *Nat Protoc* **3**, 41-50 (2008).
40. Wu G, Rossidivito G, Hu TQ, Berlyand Y, Poethig RS. Traffic Lines: New Tools for Genetic Analysis in *Arabidopsis thaliana*. *Genetics* **200**, 35-U53 (2015).
41. Choi KH, *et al.* *Arabidopsis* meiotic crossover hot spots overlap with H2A. Z nucleosomes at gene promoters. *Nat Genet* **45**, 1327-+ (2013).
42. Yelina NE, Lambing C, Hardcastle TJ, Zhao XH, Santos B, Henderson IR. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. *Gene Dev* **29**, 2183-2202 (2015).
43. Jiao WB, Schneeberger K. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nature Communications* **11**, 989 (2020).
44. Capilla-Perez L, *et al.* The synaptonemal complex imposes crossover interference and heterochiasmy in *Arabidopsis*. *Proc Natl Acad Sci U S A* **118**, e2023613118 (2021).
45. Morgan C, Fozard JA, Hartley M, Henderson IR, Bomblies K, Howard M. Diffusion-mediated HEI10 coarsening can explain meiotic crossover positioning in *Arabidopsis*. *Nat Commun* **12**, 4674 (2021).
46. Higgins JD, Vignard J, Mercier R, Pugh AG, Franklin FC, Jones GH. AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana*, but is not required for synapsis. *Plant J* **55**, 28-39 (2008).
47. Alonso-Blanco C, *et al.* 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481-491 (2016).

48. Arouisse B, Korte A, van Eeuwijk F, Kruijer W. Imputation of 3 million SNPs in the Arabidopsis regional mapping population. *Plant J* **102**, 872-882 (2020).
49. Serra H, *et al.* Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis. *Proc Natl Acad Sci U S A* **115**, 2437-2442 (2018).
50. Schmidt C, *et al.* Changing local recombination patterns in Arabidopsis by CRISPR/Cas mediated chromosome engineering. *Nat Commun* **11**, 4418 (2020).
51. Horton MW, *et al.* Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nat Genet* **44**, 212-216 (2012).
52. Costas C, *et al.* Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. *Nature Structural & Molecular Biology* **18**, 395-U190 (2011).
53. Choi K, *et al.* Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. *Genome Res* **28**, 532-546 (2018).
54. Zhang WL, Zhang T, Wu YF, Jiang JM. Genome-Wide Identification of Regulatory DNA Elements and Protein-Binding Footprints Using Signatures of Open Chromatin in Arabidopsis. *Plant Cell* **24**, 2719-2731 (2012).
55. Maher KA, *et al.* Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell* **30**, 15-36 (2018).
56. Alvarez JM, *et al.* Local Changes in Chromatin Accessibility and Transcriptional Networks Underlying the Nitrate Response in Arabidopsis Roots. *Mol Plant* **12**, 1545-1560 (2019).
57. Zhong Z, *et al.* DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. *Proc Natl Acad Sci U S A* **118**, e2023347118 (2021).
58. Lambing C, *et al.* Interacting Genomic Landscapes of REC8-Cohesin, Chromatin, and Meiotic Recombination in Arabidopsis. *Plant Cell* **32**, 1218-1239 (2020).
59. Kawakatsu T, *et al.* Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell* **166**, 492-505 (2016).
60. Bravo Nunez MA, Nuckolls NL, Zanders SE. Genetic Villains: Killer Meiotic Drivers. *Trends Genet* **34**, 424-433 (2018).
61. Underwood CJ, *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).
62. Halldorsson BV, *et al.* Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, eaau1043 (2019).
63. Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2109-2114 (2015).
64. Rattray A, Santoyo G, Shafer B, Strathern JN. Elevated Mutation Rate during Meiosis in *Saccharomyces cerevisiae*. *Plos Genetics* **11**, e1004910 (2015).
65. Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* **4**, e1000071 (2008).
66. Hellmann I, Prüfer K, Ji H, Zody MC, Pääbo S, Ptak SE. Why do human diversity levels vary at a megabase scale? *Genome Research* **15**, 1222-1231 (2005).
67. Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* **14**, 262-274 (2013).
68. Ziolkowski PA, *et al.* Natural variation and dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination. *Genes Dev* **31**, 306-317 (2017).
69. Fernandes JB, Seguela-Arnaud M, Larcheveque C, Lloyd AH, Mercier R. Unleashing meiotic crossovers in hybrid plants. *Proc Natl Acad Sci U S A* **115**, 2431-2436 (2018).
70. Lloyd A, Jenczewski E. Modelling Sex-Specific Crossover Patterning in Arabidopsis. *Genetics* **211**, 847-859 (2019).
71. Kleckner N, Storlazzi A, Zickler D. Coordinate variation in meiotic pachytene SC length and total crossover/chiasma frequency under conditions of constant DNA length. *Trends Genet* **19**, 623-628 (2003).
72. Mieulet D, *et al.* Unleashing meiotic crossovers in crops. *Nat Plants* **4**, 1010-1016 (2018).
73. Capilla-Perez L, *et al.* The HEM Lines: A New Library of Homozygous Arabidopsis thaliana EMS Mutants and its Potential to Detect Meiotic Phenotypes. *Front Plant Sci* **9**, 1339 (2018).

74. Rowan BA, Patel V, Weigel D, Schneeberger K. Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3 (Bethesda)* **5**, 385-398 (2015).
75. Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).
76. Lamesch P, *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research* **40**, D1202-1210 (2012).
77. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
78. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**, 573-580 (1999).
79. Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* **20**, 277 (2019).
80. Zhang L, Liang Z, Hutchinson J, Kleckner N. Crossover patterning by the beam-film model: analysis and implications. *PLoS Genet* **10**, e1004042 (2014).
81. White MA, Wang S, Zhang L, Kleckner N. Quantitative Modeling and Automated Analysis of Meiotic Recombination. *Methods Mol Biol* **1471**, 305-323 (2017).
82. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
83. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**, 3 (2011).
84. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
85. Danecek P, *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
86. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034 (2015).
87. Ramirez F, *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-165 (2016).
88. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
89. Danecek P, *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
90. Gao F, Ming C, Hu WJ, Li HP. New Software for the Fast Estimation of Population Recombination Rates (FastEPRR) in the Genomic Era. *G3-Genes Genomes Genetics* **6**, 1563-1571 (2016).

Figure legends

Figure 1. Experimental design for CO identification in Arabidopsis inbred lines.

M2 plants were derived from the selfing of independent EMS-treated seeds. Pairs of M2s were crossed to produce F1*s, which are then heterozygous for a set of unique mutations defining two phases, indicated by coloured ticks. Two F1*s were then reciprocally crossed to generate F1 populations. The DNA of leaves of F1* and F1 plants were sequenced using Illumina. The color-coded ticks indicate EMS-induced mutation markers. The red triangles represent COs, detected by phase switching. This design allows the detection of COs that occurred in the male and female meocytes of the F1* plants.

Figure 2. Analysis of female and male COs in Col, Ler, and F1 hybrid populations.

(A–B) The number of COs per gamete in each replicate population of Col and Ler, respectively. Each letter corresponds to one F1* plant (Figure 1). The Mann-Whitney test was used to evaluate the differences in CO numbers between female and male meiosis. (C) Distribution of CO number per gamete in female and male meiosis of Col, Ler, and F1 hybrids. The mean number of COs are color-coded and indicated by dashed lines. The sample sizes are indicated in parentheses. (D–F) The chromosomal distribution (sliding window-based, window size 2 Mb, step size 50 kb) of COs in female and male meiosis of Col, Ler, and F1 hybrids, with 90% confidence interval. The genome-wide mean level of recombination is shown with dashed lines. The pericentromeric and centromeric regions are indicated by grey and blue shading separately. The ~1.2 Mb inversion between Col and Ler on chromosome 4 is indicated by a red bar.

Figure 3. Comparison of the genome-wide CO landscape in Col, Ler, and F2 hybrids with genetic polymorphisms.

(A) The normalized landscapes (sliding window-based, window size 2 Mb and step size 50 kb) of historical recombination rate (4Ner per kb, red) and sequence diversity (π , blue) within 2,029 Arabidopsis accessions across genomes. (B) The chromosomal distribution of COs, with 90% confidence interval in Col, Ler (merged female and male), and Col/Ler F2 hybrids. The genome-wide mean level recombination is shown with dashed lines. (C) The normalized distribution of COs in Col, Ler, and F2 hybrids and the density of SNPs between Col and Ler along chromosomes. The peaks and valleys detected in the hybrid landscape are represented by black points prolonged

by vertical dashed lines. The pericentromeric and centromeric regions are indicated by grey and blue shading separately. The ~1.2 Mb inversion between Col and Ler on chromosome 4 is represented by a red bar. (D–F) The non-linear correlation coefficient matrices among distributions, considering the (D) whole genome, (E) chromosome arms, and (F) pericentromeric regions. Col_COs, Ler_COs, Hybrid_COs and pop_COs represent CO landscapes in Col, Ler, F2 hybrids and the population of 2,029 Arabidopsis accessions (historical recombination rate); SNPs represents SNP density between Col and Ler, and Seq_div represents sequence diversity in the population of 2,029 Arabidopsis accessions.

Figure 4. Association and prediction of CO distribution with genomic and epigenomic features.

(A) The non-linear correlation matrices show the comparison of pairwise features along chromosome arms, with differences in colour and size according to the correlation scale. Col_COs, Ler_COs and Hybrid_COs (CO landscapes in Col, Ler, and F2 hybrids), SNPs (SNPs density between Col and Ler), Seq_div (sequence diversity in the population of 2,029 Arabidopsis accessions), INV_TRANS (inversions and translocations between Col and Ler), BrdU_labelled (origins of DNA replication), SPO11 (SPO11-1-oligos), Genes, TEs and GC (gene, TE and GC density), ATAC and DNase (chromatin accessibility, ATAC-seq and Dnase-seq), H3K4me1/2/3, H3K9me2, H3K27me1 (euchromatin, heterochromatin and Polycomb histone marks, ChIP-seq), REC8 (cohesin, ChIP-seq), mCG, mCHG and mCHH (DNA methylation in CG, CHG and CHH contexts), MNase (nucleosome occupancy, MNase-seq). (B) The importance of each of the 17 features for the prediction of CO distribution. The size of points corresponds to the importance. (C) The cumulated proportion of variation that can be explained with the accumulation of features. The top three most important features are colored. (D–E) Spearman's correlation between the prediction (with the 17 and top three features) and observation of CO distribution, respectively.

Supplemental Figure legends

Figure S1. Mutation marker phasing and CO identification.

The EMS-induced mutation markers were first genotyped in each F1 population. Then, in each sliding window of markers, F1 samples were clustered into two groups, corresponding to the two different phases (P1 and P2, alleles coloured by blue and orange). After phasing, the count table was generated, indicating the number of supported windows of each phase on each marker position. With this, uncovered and mis-genotyped markers were imputed and corrected. Finally, COs can be identified as consistent switches of phase. “G” and “A” represent markers that genotyped as reference and mutant alleles, “n” represent markers not covered by sequencing reads.

Figure S2. The chromosomal distribution of phased EMS-induced mutation markers in Col and Ler.

Figure S3. The phased genotype map of mutation markers on chromosome 1 in Col A population.

Each line is a F1 individual, each column a marker. As the F1 results from reciprocal crosses between independent F1*, a marker in a given F1* can be either heterozygous or wild-type. Dark purple and dark red indicate the detection of mutant reads (phase 1 and phase 2, respectively), unambiguously scored as a heterozygote. When only wild-type reads are detected, the number of reads is indicated with a colour scale. (phase 1 and phase 2 are colored blue and orange, respectively. Maximum was set as 5 manually). The identified COs are marked with pairs of black arrows.

Figure S4. The genome-wide phase map of mutation markers in two samples of Col A population.

Figure S5. Correlation analysis of CO numbers, sequencing depths and number of covered and phased mutation markers in each replicate population of Col and Ler.

Figure S6. The distribution of interval length of COs and correlation analysis with sequencing depths in Col, Ler, and hybrids.

Figure S7. The chromosomal distribution of COs in Col/Ler F1 hybrids and relationship to genome features.

(A) Distribution of COs in female meiosis of Col/Ler F1 hybrids in this study (pink) compared with Giraut et al (black). (B) Distribution of COs in male meiosis of Col/Ler F1 hybrids in this study (blue) compared with Giraut et al (black). The exactly same window size is adopted from Giraut et al for comparison. (C) Sliding window-based distribution (window size 500 kb, step size 50 kb) of COs in female and male meiosis of Col/Ler F1 hybrids (this study). (D) Distribution of distance of COs from nearest TSS. (E) Permutation tests (5,000 shuffling) were performed to evaluate the overlap between promoter regions of proteins encoding genes and COs in Col/Ler F1 hybrids.

Figure S8. The distribution of CO frequency across genomes in Col/Ler detected in backcrosses (F1, this study) and in selfing (F2, Rowan et al) populations.

Comparison was done using a window size of 500 kb and step size of 50 kb (A) or a window size of 2 Mb and step size of 50 kb (B). Female and male data (this study, blue) were merged to be comparable with the selfing data (Rowan et al., orange).

Figure S9. Comparison of the CoC curve between female and male meiosis in Col, Ler, and F1 hybrids.

Chromosomes were divided into 15 intervals, and the mean coefficient of coincidence (CoC = observed number of coincident occurrence of COs in both intervals/expected number according to CO frequency in each interval) was calculated for pairs of intervals separated by a certain distance (proportion of chromosome length). A CoC lower than 1 indicates the presence of crossover interference.

Figure S10. Comparison of the CoC curve between Col, Ler, and F1 hybrids in female and male meiosis.

Chromosomes were divided into 15 intervals, and the mean coefficient of coincidence (CoC = observed number of coincident occurrence of COs in both intervals/expected number according to

CO frequency in each interval) was calculated for pairs of intervals separated by a certain distance (proportion of chromosome length). A CoC lower than 1 indicates the presence of crossover interference.

Figure S11. CO number distribution and correlation analysis with chromosome length in female and male Col, Ler, and hybrid individuals.

(A) Distribution of CO numbers per chromosome and per gamete in female and male meiosis of Col, Ler, and F1 hybrids. (B) Mean number of COs per chromosome per gamete versus chromosome size in Mb, in female and male Col, Ler, and F1 hybrids. Pearson's correlation analysis of relationship between CO number and chromosome length: Col female $r = 0.96$, $P = 0.008$, Col male $r = 0.96$, $P = 0.011$, Ler female $r = 0.97$, $P = 0.007$, Ler male $r = 0.94$, $P = 0.017$, Hybrid female $r = 0.71$, $P = 0.18$, Hybrid male $r = 0.97$, $P = 0.007$. (C) Pearson's correlation analysis of relationship between CO number and chromosome length in Col, Ler pseudo F2, and hybrids F2 (Rowan et al) (Col $r = 0.96$, $P = 0.01$, Ler $r = 0.98$, $P = 0.003$, Hybrid $r = 0.98$, $P = 0.004$).

Figure S12. Comparison of the chromosomal distribution of COs between Col, Ler, and F1 hybrids in female and male meiosis.

The sliding window-based distribution (window size 2 Mb, step size 50 kb) of COs in female (A) and male (B) meiosis of Col, Ler, and hybrids individually.

Figure S13. Historical recombination rate in the *Arabidopsis thaliana* genome.

(A) The geographical distribution of the 2,029 Arabidopsis accessions (the 1001 Genome Project and the RegMap lines, the latitude and longitude for eight accessions were not included). (B) The distribution of historical recombination rates (4Ner per kb) along chromosomes at 50kb window scale within 2,029 Arabidopsis accessions. Mean values are indicated by horizontal lines. The pericentromeric and centromere regions are indicated by grey and blue shading separately.

Figure S14. Correlation analysis of CO distribution with polymorphisms.

Spearman's correlation between each pair of features is shown in the upper triangle panel.

Figure S15. Permutation analysis of the overlap between non-syntenic and adjacent syntenic regions and COs in Col/Ler F2 hybrids.

(A) The percentages of non-syntenic and syntenic regions in Col and *Ler* and pericentromeric regions. (B-C) Permutation tests were performed to evaluate the overlap between non-syntenic and adjacent syntenic regions and COs in Col/Ler F2 hybrids, respectively. The adjacent syntenic regions were defined as the up- and downstream regions (same size) of non-syntenic regions. The vertical black dashed lines indicate the expected mean number of overlaps in 5,000 shuffled permutations. The observed number of overlaps is represented by vertical red dashed lines.

Figure S16. Non-linear correlation coefficient matrices for the distribution of COs, genomic and epigenomic features at genome and pericentromeric scales.

Col_COs, Ler_COs and Hybrid_COs (CO landscapes in Col, *Ler*, and F2 hybrids), SNPs (SNPs density between Col and *Ler*), Seq_div (sequence diversity in the population of 2,029 Arabidopsis accessions), INV_TRANS (inversions and translocations between Col and *Ler*), BrdU_labelled (origins of DNA replication), SPO11 (SPO11-1-oligos), Genes, TEs and GC (gene, TE and GC content density), ATAC and DNase (chromatin accessibility, ATAC-seq and Dnase-seq), H3K4me1/2/3, H3K9me2, H3K27me1 (euchromatin, heterochromatin and Polycomb histone marks, ChIP-seq), REC8 (cohesin, ChIP-seq), mCG, mCHG and mCHH (DNA methylation in CG, CHG and CHH contexts), MNase (nucleosome occupancy, MNase-seq).

Figure S17. Genomic landscape of Col COs, genomic and epigenomic features.

Sliding window-based distributions (window size 2 Mb, step size 50 kb) were normalized to the same scale. The pericentromeric and centromere regions are indicated by grey and blue shading separately. Col (CO landscapes in Col), BrdU (origins of DNA replication), SPO11 (SPO11-1-oligos), Genes and TEs (gene and TE density), ATAC and DNase (chromatin accessibility, ATAC-seq and Dnase-seq), H3K4me1 and H3K9me2 (euchromatin and heterochromatin histone marks, ChIP-seq), REC8 (cohesin, ChIP-seq), mCG, mCHG and mCHH (DNA methylation in CG, CHG and CHH contexts), MNase (nucleosome occupancy, MNase-seq).

Figure S18. The performance of the random forest model for predicting CO distribution in Col.

(A-B) The distribution of observed and predicted CO frequency by modelling with 17 features in the whole, training (70%) and testing (30%) datasets, respectively. (C-D) The distribution of observed and predicted CO frequency by modelling with the top three most important features in the whole, training and testing datasets, respectively. The pericentromeric and centromere regions are indicated by grey and blue shading separately. (E-H) Spearman's correlation between the predicted and observed CO distribution predicted by models constructed with 17 features and the top three most important features in the training and testing datasets, respectively.

Table S1. Summary of EMS-induced mutations in Arabidopsis inbred lines

Table S2. The public dataset used in this study

Table S3. CO positions in inbred Col

Table S4. CO positions in inbred Ler

Table S5. CO positions in Col x Ler backcross population

Table S6. CO positions in Col x Ler F2 population

Figure 1. Experimental design for CO identification in Arabidopsis inbred lines.

M2 plants were derived from the selfing of independent EMS-treated seeds. Pairs of M2s were crossed to produce F1*s, which are then heterozygous for a set of unique mutations defining two phases, indicated by coloured ticks. Two F1*s were then reciprocally crossed to generate F1 populations. The DNA of leaves of F1* and F1 plants were sequenced using Illumina. The color-coded ticks indicate EMS-induced mutation markers. The red triangles represent COs, detected by phase switching. This design allows the detection of COs that occurred in the male and female meiocytes of the F1* plants.

Figure 2. Analysis of female and male COs in Col, Ler, and F1 hybrid populations.

(A–B) The number of COs per gamete in each replicate population of Col and Ler, respectively. Each letter corresponds to one F1* plant (Figure 1). The Mann-Whitney test was used to evaluate the differences in CO numbers between female and male meiosis. (C) Distribution of CO number per gamete in female and male meiosis of Col, Ler, and F1 hybrids. The mean number of COs are color-coded and indicated by dashed lines. The sample sizes are indicated in parentheses. (D–F) The chromosomal distribution (sliding window-based, window size 2 Mb, step size 50 kb) of COs in female and male meiosis of Col, Ler, and F1 hybrids, with 90% confidence interval. The genome-wide mean level of recombination is shown with dashed lines. The pericentromeric and centromeric regions are indicated by grey and blue shading separately. The ~1.2 Mb inversion between Col and Ler on chromosome 4 is indicated by a red bar.

Figure 3. Comparison of the genome-wide CO landscape in Col, Ler, and F2 hybrids with genetic polymorphisms.

(A) The normalized landscapes (sliding window-based, window size 2 Mb and step size 50 kb) of historical recombination rate (4Ner per kb, red) and sequence diversity (π , blue) within 2,029 Arabidopsis accessions across genomes. (B) The chromosomal distribution of COs, with 90% confidence interval in Col, Ler (merged female and male), and Col/Ler F2 hybrids. The genome-wide mean level recombination is shown with dashed lines. (C) The normalized distribution of COs in Col, Ler, and F2 hybrids and the density of SNPs between Col and Ler along chromosomes. The peaks and valleys detected in the hybrid landscape are represented by black points prolonged by vertical dashed lines. The pericentromeric and centromeric regions are indicated by grey and blue shading separately. The ~1.2 Mb inversion between Col and Ler on chromosome 4 is represented by a red bar. (D–F) The non-linear correlation coefficient matrices among distributions, considering the (D) whole genome, (E) chromosome arms, and (F) pericentromeric regions. Col_COs, Ler_COs, Hybrid_COs and pop_COs represent CO landscapes in Col, Ler, F2 hybrids and the population of 2,029 Arabidopsis accessions (historical recombination rate); SNPs represents SNP density between Col and Ler, and Seq_div represents sequence diversity in the population of 2,029 Arabidopsis accessions.

Figure 4. Association and prediction of CO distribution with genomic and epigenomic features.

(A) The non-linear correlation matrices show the comparison of pairwise features along chromosome arms, with differences in colour and size according to the correlation scale. Col_COs, Ler_COs and Hybrid_COs (CO landscapes in Col, Ler, and F2 hybrids), SNPs (SNPs density between Col and Ler), Seq_div (sequence diversity in the population of 2,029 Arabidopsis accessions), INV_TRANS (inversions and translocations between Col and Ler), BrdU_labeled (origins of DNA replication), SPO11 (SPO11-1-oligos), Genes, TEs and GC (gene, TE and GC density), ATAC and DNase (chromatin accessibility, ATAC-seq and Dnase-seq), H3K4me1/2/3, H3K9me2, H3K27me1 (euchromatin, heterochromatin and Polycomb histone marks, ChIP-seq), REC8 (cohesin, ChIP-seq), mCG, mCHG and mCHH (DNA methylation in CG, CHG and CHH contexts), MNase (nucleosome occupancy, MNase-seq). (B) The importance of each of the 17 features for the prediction of CO distribution. The size of points corresponds to the importance. (C) The cumulated proportion of variation that can be explained with the accumulation of features. The top three most important features are colored. (D–E) Spearman's correlation between the prediction (with the 17 and top three features) and observation of CO distribution, respectively.