Lawrence Berkeley National Laboratory

Title

High-resolution mapping reveals hotspots and sex-biased recombination in Populus trichocarpa

Permalink

https://escholarship.org/uc/item/8jh4150r

Journal

G3: Genes, Genomes, Genetics, 13(1)

ISSN

2160-1836

Authors

Abeyratne, Chanaka Roshan
Macaya-Sanz, David
Zhou, Ran
et al.

Publication Date

2023-01-12

DOI

10.1093/g3journal/jkac269

Peer reviewed

High-resolution mapping reveals hotspots and sex-biased recombination in *Populus trichocarpa*

Chanaka Roshan Abeyratne [ID] ,[1] David Macaya-Sanz [ID] ,[2] Ran Zhou [ID] ,[3] Kerrie W. Barry [ID] ,[4] Christopher Daum [ID] ,[4] Kathy Haiby [ID] ,[5] Anna Lipzen [ID] ,[4] Brian Stanton [ID] ,[5] Yuko Yoshinaga [ID] ,[4] Matthew Zane [ID] ,[4] Gerald A. Tuskan [ID] ,[6] Stephen P. DiFazio [ID] [1],*

[1]Department of Biology, West Virginia University, Morgantown, WV 26506, USA,
[2]Department of Forest Ecology & Genetics, CIFOR-INIA, CSIC, Madrid 28040, Spain,
[3]Warnell School of Forestry and Natural Resources, Department of Genetics, and Department of Plant Biology, University of Georgia, Athens, GA 30602, USA,
[4]Department of Energy Joint Genome Institute, Berkeley, CA 94720, USA,
[5]Poplar Innovations, Camas, WA 98607, USA,
[6]Biosciences Division, Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

*Corresponding author: Department of Biology, West Virginia University, Morgantown, WV 26506, USA. Email: spdifazio@mail.wvu.edu

Abstract

Fine-scale meiotic recombination is fundamental to the outcome of natural and artificial selection. Here, dense genetic mapping and haplotype reconstruction were used to estimate recombination for a full factorial *Populus trichocarpa* cross of 7 males and 7 females. Genomes of the resulting 49 full-sib families ($N = 829$ offspring) were resequenced, and high-fidelity biallelic SNP/INDELs and pedigree information were used to ascertain allelic phase and impute progeny genotypes to recover gametic haplotypes. The 14 parental genetic maps contained 1,820 SNP/INDELs on average that covered 376.7 Mb of physical length across 19 chromosomes. Comparison of parental and progeny haplotypes allowed fine-scale demarcation of cross-over regions, where 38,846 cross-over events in 1,658 gametes were observed. Cross-over events were positively associated with gene density and negatively associated with GC content and long-terminal repeats. One of the most striking findings was higher rates of cross-overs in males in 8 out of 19 chromosomes. Regions with elevated male cross-over rates had lower gene density and GC content than windows showing no sex bias. High-resolution analysis identified 67 candidate cross-over hotspots spread throughout the genome. DNA sequence motifs enriched in these regions showed striking similarity to those of maize, *Arabidopsis*, and wheat. These findings, and recombination estimates, will be useful for ongoing efforts to accelerate domestication of this and other biomass feedstocks, as well as future studies investigating broader questions related to evolutionary history, perennial development, phenology, wood formation, vegetative propagation, and dioecy that cannot be studied using annual plant model systems.

Keywords: recombination; heterochiasmy; cross-over; sex-bias; *Populus*

Introduction

Meiotic recombination shuffles genetic variation, and may bring together beneficial alleles or purge deleterious alleles to create more fit haplotypes (Felsenstein 1974). In the complete absence of recombination, deleterious mutations would accumulate faster than selection can remove them (Muller 1964) and such a scenario would leave fitness consequences at the mercy of infrequent reverse mutations. Therefore, meiotic recombination is selectively favored and increases the efficiency of adaptive evolution in finite populations in concert with a backdrop of mutation, genetic drift and selection (Hill and Robertson 1966; Felsenstein 1974). In a breeding context, recombination serves to dampen gains made by artificial selection and yet it benefits breeding programs by reducing linkage drag. Accordingly, genome-wide local recombination rates in both natural and structured populations will be under selective pressure stemming from the need to maintain a tradeoff between reducing genetic load and reducing the immediate negative effects of

recombination load due to disruption of favorable combinations of alleles (Charlesworth and Barton 1996).

Meiotic recombination is a highly regulated process that occurs within germline cells of sexually reproducing organisms. It is initiated with programmed DNA double stranded breaks (DSBs) during meiotic prophase I (Choi and Henderson 2015). The DSB repair process resolves these strand breaks as either cross-over (CO) or non-CO (NCO) events leading to recombined chromosomes or gene conversions, respectively (Wang and Copenhaver 2018). In the latter case, haplotype fidelity is maintained on either side of the breakpoint, whereas CO events combine haplotypes from different chromosomes.

Recombination rates may differ inter or intraspecifically (Smukowski and Noor 2011; Bauer *et al.* 2013), between sexes (Lenormand and Dutheil 2005), and even across genomic regions within the same individual (Slavov *et al.* 2012; Rodgers-Melnick *et al.* 2015; Gion *et al.* 2016; Kianian *et al.* 2018). In certain species such as mouse, yeast and *Caenorhabditis elegans*, there is a stable upper limit for CO counts per chromosome (CO homeostasis).

However, such a trend has not yet been observed in plants where CO counts show a linear relationship with DSBs. As observed in maize and *Arabidopsis* a minimum of one CO per chromosome is obligatory and results in proper segregation of chromosomes during meiosis (Sidhu *et al.* 2015; Lambing *et al.* 2017). On a chromosomal scale, regions near centromeres and telomeres generally show a reduction of recombination, although such patterns are not clearly displayed in relatively shorter, acrocentric, or telocentric chromosomes (Haenel *et al.* 2018). Correspondingly, COs are not independently and uniformly distributed across a genome, and display nonrandom spatial heterogeneity among extensively studied species where fine-scale recombination hotspots are interspersed throughout the genome (Kauppi *et al.* 2004). Hotspots are generally 1–10 kb in size and have high probability of carrying a CO event, compared to background recombination or flanking cold regions suppressed for recombination. However, CO hotspot existence, size, and frequency within a genome may widely vary from species to species.

Hotspots are well established in humans and mice in which the most salient feature associated is enrichment of DNA-binding sequences for the histone methyl transferase PRDM9 (Baudat *et al.* 2010). PRDM9 orthologs are absent in nonmammalian taxa, including arthropods, birds, fungi, and plants. Instead, proteins involved in DSB repair, including SPO11, REC8, RAD51, and MUS81 have been associated with recombination hotpots (He *et al.* 2017; Lambing *et al.* 2020). In various plant species recombination frequency is reportedly associated with localized genomic features related to chromatin compaction such as histone remodeling, DNA-methylation, gene content, and repeat content, thus implying accessibility of DNA for DSBs is a major deciding factor (Saintenac *et al.* 2011; Yelina *et al.* 2012; Shilo *et al.* 2015; Rodgers-Melnick *et al.* 2016; Darrier *et al.* 2017; He *et al.* 2017; Choi *et al.* 2018). CO interference is an additional factor that may control spacing between adjacent CO events and is responsible for punctate distribution of COs across the genome. This process is driven by the interference sensitive (Type-I) DSB resolution pathway and is considered the dominant form displayed by most plant species (Choi and Henderson 2015; Wang and Copenhaver 2018).

Suppressed recombination is observed within sex chromosomes and sex-determination regions (SDRs) of nascent sex chromosomes which could perpetuate genetic-based differences between the sexes (Lenormand 2003; Charlesworth *et al.* 2005). Apart from this, some species of animals and plants show sex-based differences in rates of recombination observed in autosomes (Burt *et al.* 1991; Lenormand and Dutheil 2005; Kong *et al.* 2010). This bias may range from complete absence of recombination in one sex (achiasmy) as in the case of Drosophila, or differential rates (heterochiasmy, Lenormand 2003) observed between males and females where either sex can have the higher rate (Bherer *et al.* 2017; Kianian *et al.* 2018). The differences in recombination could also be extended to spatial localization and broader patterns across the genome in some species (Zelkowski *et al.* 2019; Sardell and Kirkpatrick 2020). Although sex-based differences in recombination were established early on, definitive local features that explain the finer scale variation still need clarification. Differential rates of recombination at various hierarchical scales are under selection (Dapper and Payseur 2017), and identification of associated conserved sequence motifs, molecular markers, and other localized genomic features could provide insights into mechanisms underlying adaptive trait variation (Penalba and Wolf 2020).

Much progress has been made toward revealing fine-scale genome-wide recombination patterns in plant species such as *Arabidopsis thaliana*, *Zea mays*, *Triticum aestivum*, and *Oryza sativa* largely due to the availability of large-structured mapping populations or multigeneration breeding programs for these species (Rodgers-Melnick *et al.* 2015; Wang and Copenhaver 2018; Rowan *et al.* 2019; Lambing *et al.* 2020). However, these species represent a biased subset of plants confounded by high levels of inbreeding and/or long histories of domestication. Conversely, there is a dearth of such studies in undomesticated, primarily outbreeding perennial species such as forest trees, which represent the majority of biomass in terrestrial ecosystems (Neale and Kremer 2011; Silva Junior and Grattapaglia 2015). Trees are subject to considerably different selection regimes that could profoundly shape their recombination landscapes in comparison to classical plant model systems such as *Arabidopsis* or maize (Petit and Hampe 2006).

Black cottonwood (*Populus trichocarpa*) is an undomesticated, outbreeding, pioneer riparian species with moderate life span that is widely used as a model for basic and applied research on many aspects of tree biology. *Populus* is also a promising renewable feedstock for bioenergy and bioproducts given its rich genomic resources, short-rotation cycle and desirable lignocellulosic characteristics (Bradshaw *et al.* 2000; Tuskan *et al.* 2006; Sannigrahi *et al.* 2010; Porth and El-Kassaby 2015). As such, both commercial and ecological success converge on similar adaptive traits (e.g. adventitious rooting of stem and branches). At this juncture, selection favoring trait combinations both in managed and natural populations depends heavily on the recombination landscape. *Populus* is a genetically tractable model system with high-quality male and female reference genomes, and a range of other molecular resources that make it a robust model system to use for studying recombination patterns.

Fine-scale genome-wide patterns of recombination can be studied using population-based linkage disequilibrium (LD) estimates, but these are subject to vagaries of population demography, drift, natural selection, and mutation rate, and they only provide information on sex averaged recombination rates (Auton and McVean 2007). Conversely, resolution of pedigree-based genetic maps is largely limited only by the size of the mapping population and the power of the genetic marker system to detect recombination breakpoints. Pedigrees can therefore provide insight into sex-based differences in recombination rates as well as more accurate estimates of single generation recombination rates. Most existing pedigree based genetic maps for *Populus* have been constructed using interspecific hybrids and large full-sib families and a limited number of molecular markers (Bradshaw *et al.* 1994; Yin *et al.* 2004; Gaudet *et al.* 2007). Recent advances in high-throughput whole-genome resequencing platforms, improved variant discovery methods and better reference genome assemblies facilitate the use of genome-wide SNP and INDEL markers at sufficient density to enable fine-scale mapping (Fang *et al.* 2018). However, mapping population size can still be a challenge, limiting the number of observed recombination events and map resolution.

Here, we use a large-scale full-factorial cross in *P. trichocarpa*, coupled with whole-genome resequencing to reveal the genome-wide recombination landscape and patterns of inheritance at a finer scale than has previously been possible in undomesticated plants. We produced dense genetic maps for 14 parents that contain a balanced representation of each sex. Data contributed from multiple individuals allowed us to conduct fine-scale analyses of sex dimorphism and intraspecific variation of genome-wide recombination patterns. We show how recombination rates

vary within and among genomes, and between the sexes, and elucidate key genomic features that may play a role in shaping recombination rates at a scale of approximately 960 kb.

Materials and methods
Populus trichocarpa mapping population
The *P. trichocarpa* mapping population used in the study consists of a total of 829 progeny from 49 full-sib families derived from a full-factorial cross between 7 males and 7 females (Harman-Ware *et al.* 2021) (Supplementary Table 1). The parents were originally collected as part of a population of 448 genotypes from natural riparian stands in WA and OR, USA (Supplementary Fig. 1; Wegrzyn *et al.* 2010). The parents were selected to represent the full range of natural variation in lignin composition observed in the population.

DNA isolation, whole-genome resequencing, and variant calling pipeline
Genomic DNA was extracted from foliage from all progeny and parents using the DNeasy 96 Plant DNA isolation kit (Cat. No. 69181; Qiagen, Valencia, CA, USA). The sample library preparation, quality control and whole-genome resequencing up to an expected coverage of 5× and 15× for all progeny and parental clones, respectively, was carried out using the Illumina HiSeq 2500 and NovaSeq 6000 sequencing platforms subject to established standard quality control and sequencing protocols (https://jgi.doe.gov/user-programs/pmo-overview/protocols-sample-preparation-information/) at the DOE Joint Genome Institute (JGI). Paired-end reads were aligned to a male *P. trichocarpa* Stettler-14 male reference assembly (Hofmeister *et al.* 2020), modified as described in Zhou *et al.* (2020). Mapping was accomplished with BWA v.0.7.10-r789 with default parameters. The MarkDuplicates tool in Picard v.1.131 was used to locate and flag duplicate reads in BAM files. We used an adaptation of the short variant discovery best practices pipeline for Genome Analysis Toolkit v.4.2.0 (DePristo *et al.* 2011) to identify genome-wide SNP and INDELs. GATK's HaplotypeCaller tool was used to call SNPs and INDELs simultaneously. A "truth dataset" containing 1,030,941 biallelic SNPs and 125,772 INDELs was derived from a starting set of 37,661,220 variants, by imposing strict biological constraints on Mendelian segregation and Mendelian violations (Supplementary Table 2). The truth set was subsequently used to train GATK's VQSR pipeline in order to identify 14,656,278 biallelic SNPs and INDELs enriched for true variants.

Parental genotypes: error correction and phasing
Parentage was verified and corrected as needed for all offspring based on degree of match to putative parents using 1,731,769 hard-filtered SNPs that represented a random subset of genome-wide variants (Supplementary Table 2). Parental genotypes were then imputed or corrected based on maximizing the likelihood of observing the validated half-sib family pedigree genotypes (Wang 2004). Given the moderate rate of missing genotypes in the offspring, estimations of joint-likelihood for predicted parental genotypes were based on at least a half-sib family size of 50 progeny.

A consensus of 2 independent methods was used to phase the parental chromosomes. Our first method assumed congruence in physical marker order between the focal parent and the reference genome, where pairwise LD estimates between neighboring markers were used to incrementally infer parental haplotype configuration (i.e. positive LD meaning markers are in coupling phase and negative meaning repulsion phase). A semi-

automated inspection was carried out to solve situations where the absolute value of LD between adjacent markers was low, meaning that the reference genome was inaccurate. As a precursor to our second independent method, corrected trio genotypes were used to identify the focal parent allele contribution in each offspring. Here, if all individuals in the trio were heterozygous or Mendelian violations were observed, the focal parent allele contribution was designated as missing and loci with more than 25% missing data were excluded. The two-point pairwise linkage analysis implemented in the Onemap R-package v2.1.1 (Margarido *et al.* 2007) was used to cluster markers into phased linkage groups using the following cutoffs: max.rf = 0.5 and LOD > 8. The resulting framework haplotypes were then used to infer the phases of all intervening markers segregating in the family. The final parental haplotypes resulted from the consensus of both adjacent marker LD and Onemap based clustering approaches and consisted of approximately 900,000 fully informative biallelic SNPs and INDEL markers for each parent.

Offspring genotypes: phasing, imputation, and demarcation of CO break-points
Initial offspring phasing was performed by identifying haplotypes derived from the common parent of each of the 14 half-sib families (Supplementary Fig. 2). These haplotypes were refined by imputing offspring haplotype configurations in bins using all available segregating markers, based on a sliding-window smoothing algorithm that assumes CO events are rare within an LG. Each window contained 20 heterozygous markers, with 10 markers overlapping with neighboring windows. This smoothing algorithm was implemented from both ends of each LG to obtain the consensus. However, inferred CO region (i.e. the region between 2 markers where a CO could be unequivocally ascertained) resolution was limited by the number of high-quality informative markers in the flanking regions. The median size of the inferred CO regions was 30.71 kb (Supplementary Fig. 3). The minimum haplotype blocks were defined as regions exceeding 500 kb in size (excluding the ends of chromosomes). With this filtering, we are confident that we detected the vast majority of the COs for this mapping population while eliminating most gene conversion events.

Construction of genetic linkage maps
A linkage map was constructed for the common parent of each half-sib family using the OneMap R-package. The marker datasets consisted of approximately 900,000 fully- informative biallelic SNPs and INDEL markers spread across the whole genome for a given half-sib family. Since the parental haplotype configurations were fully determined, we recoded the markers for the common parent as an F1-backcross. Markers with completely redundant genotype information across offspring were binned for computational efficiency. Using the criteria of max.rf. = 0.5 and using the "suggest_lod" function in onemap (which resulted in cutoffs greater than LOD 8 in all cases) yielded exactly 19 LGs per parent, corresponding to the base chromosome number of *Populus*. Marker order within LGs was determined using Onemap's "order_seq" function with the following adjustments of default parameters: n.init = 8, subset.search="twopt," twopt.alg="rec," touchdown=TRUE. The higher setting for n.init was used to obtain a more accurate and stable initial marker-ordering. For ordering markers, a two-point based algorithm which minimizes the total number of recombination events, implemented with the RECORD algorithm as per Van Os *et al.* (2005) was used. Using the "touchdown=TRUE" argument resulted in ordering markers that did not initially enter the

framework map using lower stringency [i.e. the "try_seq" function with a lower threshold (THRES = 2)]. Genetic distances between markers were estimated using the "Kosambi" mapping function allowing for possible CO interference. However, since a given marker order produced for each linkage group was not based on an exhaustive search, markers were rippled with a default window size of 4 using the "ripple_seq" function to identify and reposition misplaced markers.

A consensus genetic map was constructed using the phased haplotypes and recombination breakpoints for all 829 progeny. For each of the offspring, imputed parental haplotypes were divided into 10 kb-genomic-windows, resulting in 39,026 pseudomarkers spread across the genome. The consensus linkage map was constructed using the same method as described for half-sib families. For computational efficiency, LGs were subdivided into overlapping segments, and maps were merged using default settings of the LPmerge R-package (Endelman and Plomion 2014).

Statistical models for analyzing chromosome scale sex-based differences in CO counts

To test the role of sex in determining CO counts on a chromosome scale, we used a generalized linear mixed model (GLMM) from the lme4 R-package and a Poisson link function as detailed below.

$$Y_{ijk} \sim \text{Poisson}(\lambda_{ijk})$$

$$\lambda_{ijk} = \exp(\alpha + x_i\beta_i + x_j\beta_j + x_{ij}\beta_{ij} + x_k\beta_k + b_{ijk})$$

where Y represents the total CO count and i = gender of the focal parent in the half-sib family (either female or male); j is the chromosome identity (Chr01, Chr02, ..., Chr19); k is the identity of the half-sib family nested within ith sex; λ is the expected total CO count. β_i, β_j, β_{ij}, and β_k are fixed effect coefficients for sex (female or male), chromosome identity, sex-chromosome interaction effect, and half-sib family size, respectively, while x_i, x_j, x_{ij}, and x_k denote their respective indicator variables. The last term, b_{ijk} represents half-sib family identity as a random factor nested within sex. The selected model provided the lowest Akaike Information Criterion (AIC) out of all models considered (Supplementary Table 3). There were no discernible patterns in Pearson-residuals across any of the covariates used in this model or fitted values (Supplementary Fig. 4).

Comparing average pairwise marker LD to cumulative CO counts

In order to analyze historic recombination rates in a natural population of *P. trichocarpa*, we used a collection of 220 unrelated trees from the portion of the *P. trichocarpa* range that overlaps with the collection sites of the parents of the 7×7 trial (Supplementary Fig. 1). The rationale for restricting the collection to this area was to avoid genetic structure that could have distorted LD patterns. This subset of the population was identified from a larger collection that had been fully resequenced, as described in Chhetri *et al.* (2019). The resequenced reads for this natural population were aligned to the same Stettler-14 male reference genome that was used for the mapping population. An input variant dataset (SNPs) was derived using similar variant calling pipeline and parameters as described previously. The subpopulation was selected by evaluation of cluster admixture proportions using the software fastStructure (Raj *et al.* 2014) with the SNP dataset as inputs. Briefly, 10 iterations were run with cluster

number (K) from 1 to 7. The number of clusters was selected using the algorithm recommended by fastStructure developers and implemented in the script chooseK.py. Five clusters were the optimum to explain structure, one of them corresponding to the core population. Assignment to that cluster was determined when the mean admixture proportion (Q value) of an individual for that cluster was above 0.8. In order to estimate LD within nonoverlapping genomic windows, a subset of 13,989,405 SNPs was extracted for this subpopulation that intersected with the 7×7 SNP dataset, and this was used to estimate the squared correlation coefficient between genotypes (r^2) using the vcftools -geno-r2 function with ld-window-bp parameter set to 10,000. The average pairwise LD was then calculated for the same set of nonoverlapping genomic windows of 960 kb that were used for the CO counts.

Statistical model for analyzing finer-scale sex-based differences in cumulative CO counts

The finer scale analysis for heterochiasmy seeks to identify genomic regions that show differences in CO counts between sexes. As informed by the wavelet analysis (Supplementary methods), most of the variance in CO count signal for a given sex was contributed by lower dyadic scales (Supplementary Figs. 5 and 6). Yet, the dataset may suffer from poor resolution at finer scales such as 60 kb through 240 kb, largely due to the resolution of the CO region demarcation. At higher scales such as 3.84–7.68 Mb, the analysis may be redundant with the chromosome scale analysis described above. The wavelet analysis identified a nonoverlapping window size of 960 kb to be the most appropriate finer scale that unravels sex differences in CO rates. The response variable included cumulative CO counts for each sex at each nonoverlapping window across the whole genome. First, windows that overlap putative centromeres and telomeres were removed by imposing a chromosome-wide minimum average CO count after determining the global distribution of CO counts across the genome. Windows from the tails of the index of dispersion (IOD) distribution (≤5th and ≥95th percentile) were removed as outliers, resulting in a total of 280 windows to be tested. The effect of sex for each window was tested using a Poisson exact test implemented in the exactci R-package (Fay 2010), with mid-P-values (Heller and Gur 2011), and significance was determined using a false discovery procedure (FDR$_{BH}$) as implemented in the p.adjust function in R (Benjamini and Hochberg 1995).

Estimation of repeat content, gene content, and AT/GC composition within genomic windows

Repeats were identified using the RepeatModeler (v1.0.8) package, and used to mask the genome with Repeatmasker v4.0.3 (Smit *et al.* 2013–2015) following the same approach as Zhou *et al.* (2020). The AT/GC composition, long-terminal repeat (LTR) counts (i.e. *Gypsy*, *Copia*, and simple repeats) and gene content were estimated in nonoverlapping 960 kb windows with bedtools utilities v2.17.0 (Quinlan and Hall 2010). Given that LTRs are harbored near telomeric and centromeric regions, and cumulative CO counts tend to show extreme values near these regions, genomic windows overlapping putative centromeric and telomeric regions were removed as outliers using a genome-wide fixed cutoff (cumulative CO count per window <50).

Analyzing association of genomic correlates and male biased heterochiasmy

AT/GC composition, LTR repeat counts, and gene content were estimated within nonoverlapping 960-kb genomic windows as

explained above. A subset (140) of the total genomic windows of the same size (960 kb), were identified in 2 comparator groups as male biased (15) and background (125). "Male biased windows" included the 15 genomic windows that showed significant elevation of CO counts in males in comparison to females (described earlier in methods). A "background window" was defined as one which exhibits a nominal male bias (female:male ratio <1 at $0.4 \leq \text{FDR}_{\text{BH}} \geq 0.8$). The association between each of the genomic correlates and type of window categorized as either "male biased" or "background" were investigated using a 2-tailed Wilcoxon rank-sum test.

Localized CO pattern prediction using genomic correlates and sex

Genomic correlates were fitted to CO counts within genomic windows of size 960 kb, using backward stepwise linear regression with the "lm" function in R (R Core Team 2013). All covariates were centered and scaled for this analysis and the following models were selected based on lowest AIC and homoscedasticity of standardized residuals (Supplementary Table 4):

$$Y = \alpha + \beta_{\text{gc}} + \beta_{\text{gene}} + \beta_{\text{repeat}} + \beta_{\text{Copia}} + \beta_{\text{Gypsy}} + \beta_{\text{sex}} + e \quad \text{(Model-1)}$$
$$Y = \alpha + \beta_{\text{gc}} + \beta_{\text{gene}} + \beta_{\text{Copia}} + \beta_{\text{Gypsy}} + e \quad \text{(Model-2)}$$

$$e \sim N\left(0, \ \sigma_e^2\right)$$

where Y represents the total CO count for a half-sib family within a window; α, β_{gc}, β_{gene}, β_{repeat}, β_{Copia}, β_{Gypsy}, and β_{sex} are the intercept and fixed effects for % GC content, count of genes, count of simple repeats, count of *Copia*-LTR elements, count of *Gypsy*-LTR elements, and binary factor of sex (either male of female), respectively, within a 960-kb nonoverlapping genomic window; *e* represents the residual which is a random variable that is normally distributed with a mean of zero and variance σ_e^2. Model robustness of both models was assessed using 5-fold repeated cross-validation using the caret v 6.0-86 R-package (Kuhn 2008).

Genome-wide CO hotspot identification

The genome was partitioned into 30-kb nonoverlapping windows and each CO event was assigned to a single window based on the CO region mid-point. Window size was determined based on the median CO region (approximately 37 kb). CO hotspots were defined as windows showing significant deviation from the expected number of cumulative COs per window, with expectation under the null hypothesis that the probability of COs across the genome follows a Poisson distribution ($\lambda = 2.98$; $\text{FWER}_{\text{Bonferroni}} \leq 0.05$). Although a naïve approximation, stringent FWER cutoff enables detection of contiguous windows that far exceed the genome-wide expectation, highlighting candidate genomic regions enriched for CO hotspots. Due to structural rearrangements observed in parents GW-2393 and GW-6909, a stretch of windows was excluded from further analyses (more details in *Results*). Windows with <2 informative markers on average within 60 kb of window borders were removed to minimize the possibility of undetected COs in a neighboring window elevating the counts within a window. Also, windows where one or more parents presented anomalously high CO counts were removed as outliers. These outlier windows were flagged when IOD for a given window (calculated using cumulative CO counts in half-sib families) exceeded the top 90th percentile. Using similar methods, CO hotspots were also estimated separately for males and females ($\lambda_{\text{female}} = 1.38$; $\lambda_{\text{male}} = 1.54$; $\text{FWER}_{\text{Bonferroni}} \leq 0.05$).

Identification of DNA sequence motifs and genomic correlates associated with COs

Of the observed CO regions, 2,054 that were demarcated to a 10-kb genomic region or less were used for this analysis. DNA sequence information in FASTA format for these narrowly demarcated CO regions was extracted using an in-house developed Perl script. These sequences were provided as a single set of positive inputs to the STREME (Simple, Thorough, Rapid, Enriched Motif Elicitation) algorithm within the MEME v5.3.0 suite of software (Bailey *et al.* 2009), where differential enrichment to an automatically generated control set was investigated by setting the "–objfun" parameter to "de." The enriched DNA sequence motifs identified in this analysis were compared against a database of genomewide 15-mers estimated for the Stettler-14 reference genome using jellyfish 2.2.10 (Marcais and Kingsford 2011). Enriched DNA sequence motifs were aligned to the 15-mer database using the blastn algorithm with modifications to standard settings (-dust no; -task blastn-short) on BLAST 2.3.0 (Altschul *et al.* 1990). In order to assess the similarity of CO associated DNA sequence motifs identified in this study, with such motifs identified in selected set of previous studies (Supplementary Table 11), motifs were aligned to a second database consisting of DNA sequence motifs associated with recombination hotspots in *Arabidopsis*, maize, and wheat using the same methods as explained earlier.

In order to investigate whether there are DNA sequence motifs enriched for CO hotspots in comparison to regions of the genome that show average CO counts, a second set of 119 nonoverlapping 30-kb genomic regions previously identified as CO hotspots were provided as inputs to MEME. Similar parameters were used as before with the exception of now explicitly defining a negative control in the form of 3,214, 30-kb genomic regions that displayed average CO rates.

In order to investigate whether there are genomic features enriched for CO hotspots in comparison to genome-wide background levels, the same set of background genomic windows (3,124) were contrasted with 67 nonoverlapping 30-kb genomic regions previously identified as CO hotspots. The full set of genomic correlates, AT/GC composition, LTR repeat counts (*Gypsy*, *Copia*, and simple repeats) and gene content were all re-estimated for these 30-kb genomic windows and differential enrichment investigated using a 2-tailed Wilcoxon rank sum test.

Results
Construction of 14 genetic linkage maps using marker segregation within half-sib families

All 14 genetic-linkage maps (7 for each sex) converged on 19 linkage groups, representing the haploid chromosome count for *P. trichocarpa* (Supplementary Table 5). This study represents the first attempt at producing multiple linkage maps for each sex using half-sib family marker segregation in a single *P. trichocarpa* mapping population. Framework genetic maps contained 1,820 SNP/INDEL markers on average and consist exclusively of high-quality markers mapping to a single position with a conservative threshold. In addition, a consensus map constructed with all progeny contained a total of 5,029 binned pseudo-markers. Genetic maps span 376.7 Mb of physical length across 19 chromosomes in the Stettler-14 male reference assembly, and the consensus map was 2,322.9 cM in length (Supplementary Table 6), yielding a physical to genetic distance ratio of 162 kb/cM. Overall median physical distance between adjacent mapped markers

was approximately 126 kb and the median genetic distance was 1.03 cM. The genetic maps show high collinearity with the Stettler-14 male reference sequence assembly (Fig. 1). One notable exception is LG-XVII in parents GW-2393 and GW-6909, which show anomalously high map distances at the fringe of genomic region 9,044,429–10,304,428 bp. This is likely an artifact caused by large structural rearrangements in these 2 parent genomes. Also, LG-I of GW-6909 was trimmed due to a lack of informative markers beyond 29 Mb due to a preponderance of very high segregation distortion in this region. These 2 regions were excluded from further chromosome-scale analyses to avoid possible biases due to structural artifacts.

Although parental maps generally show high collinearity of marker order, there is considerable variation in recombination rate among individuals, as demonstrated by divergence of the scatter-plots representing genetic maps of each half-sib family
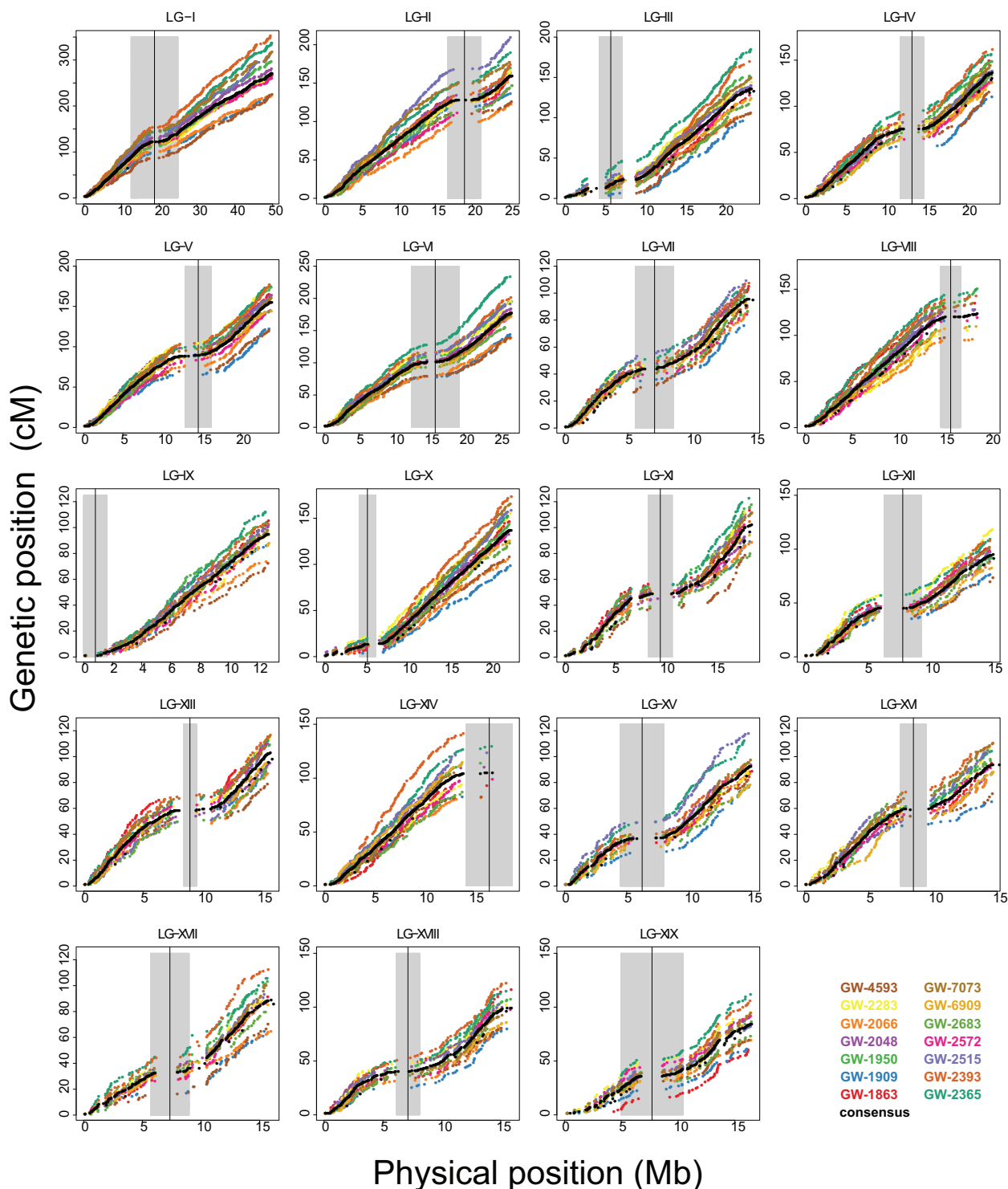

Fig. 1. Fourteen linkage maps constructed using phased markers for each half-sib family. Genetic maps are presented as scatter plots of genetic vs physical map positions for all chromosomes. Each half-sib family is represented in a different color for each LG. The consensus genetic map is shown in black. Gray shading indicates the putative centromere boundaries. The Y-axis has been scaled to the same height across all scatter plots in this image.

(Fig. 1). Nevertheless, a high correlation (Supplementary Fig. 7) was observed between genetic map averages produced in our study in comparison to similar maps from an independent study using SSR markers (Yin *et al.* 2004). This was true for both male and female median map sizes (median female maps: $r = 0.92$, $P = 1.709 \times 10^{-8}$; median male maps: $r = 0.93$, $P = 5.868 \times 10^{-9}$). Average recombination rate across LGs fluctuated within a narrow range and varied from 4.24 to 7.53 cM·Mb^{-1} for females and 5.61–7.75 cM·Mb^{-1} males, with the consensus map generally falling in between these. The lowest and highest recombination rate estimates for each sex occurred on LG-XIX and LG-IX, respectively (Supplementary Table 6).

Genome-wide recombination rate analyzed using CO counts

The accuracy of map-based estimates of recombination rate depends on informative marker density for the region under investigation. The median marker distance in our genetic maps was 126 kb. To improve resolution, we directly estimated CO rate by demarcating recombination breakpoints to a median resolution of 37 kb based on inferred haplotypes in the progeny. A total of 38,846 CO events were observed within 1,658 gametes (829 diploid offspring) averaging approximately 1.2 COs per chromosome. Of the total observed, 18,355 and 20,491 COs occurred within female and male groups, respectively, indicating a significant gross cumulative difference of 2,136 CO events between the sexes ($\chi^2_{df=1}$; $P = 2.2 \times 10^{-16}$).

Chromosome-scale differences in CO rate between sexes

As expected, LG identity was the most significant factor affecting CO count, and physical length of the LG explained much of the variance between the average number of COs within half-sib families ($r^2 = 0.77$, $P = 2.2 \times 10^{-16}$) (Supplementary Fig. 8). Nevertheless, there was still a considerable amount of unexplained residual variance among half-sib families for a given LG. We observed unequal cumulative CO counts among parents of the 2 sexes within LGs I, II, III, VI, X, XIV, XVI, and XIX ($\chi^2_{df=1}$; FWER ≤ 0.05) which hinted at possible sex-based recombination rate differences (Supplementary Table 7). Male parents showed higher CO rates (male-biased heterochiasmy) for 8 out of 19 LGs, but no LGs showed female-biased heterochiasmy. A GLMM was used to model cumulative CO count data across half-sib families, that included sex as a fixed effect among other covariates (Supplementary Table 8). This model yielded a significant positive coefficient for males, that translates to approximately 20% more COs in males in comparison to females. Furthermore, we observed that sex-based differences depended on the LG considered. The significant negative estimates for "LG × SEX" interaction terms for LGs IV, V, VII, VIII, IX, XI, XII, XIII, XV, and XVIII indicated reduced or absent male-biased heterochiasmy, which was consistent with lack of significant differences for these groups in the χ^2 analysis of sex effects (Supplementary Table 7).

Wavelet analysis for identifying the scale of sex differences in CO counts

A fine-scale comparison of CO counts between males and females requires choosing a window size that minimizes spatial autocorrelation, while still providing sufficient resolution to detect underlying factors that drive differences between them. We implemented a wavelet analysis, which is a signal processing technique that can decompose the total variance in CO counts across a chromosome for a wide range of scales. This revealed the optimum scale at which features such as CO counts and associated genomic features could be compared. CO count differences between males and females were most prominent at scales 480 kb to 1.9 Mb. At finer scales, differences remained indiscernible and higher scales are redundant with the chromosome scale analysis presented earlier (Fig. 2; Supplementary Fig. 6). The optimum window size was selected as 960 kb (Supplementary methods). All subsequent analyses related to feature contrasts between sexes are carried out at this scale.

Pairwise LD vs cumulative CO count

In order to evaluate the legitimacy of using CO counts as a proxy for recombination rate, cumulative CO count within 960-kb nonoverlapping genomic windows was compared to mean pairwise LD within the same set of windows in a subpopulation of 220 trees collected from the same region as the parents of the 7×7 cross (Supplementary Fig. 1). As expected, a strong negative correlation was observed between cumulative CO counts and mean pairwise LD ($r_{Spearman} = -0.69$; $P = 4.0 \times 10^{-4}$), consistently across all of the 19 linkage groups (Supplementary Fig. 9).

Prediction of localized CO pattern using genomic correlates

As is the general trend for plants, CO counts were lower in proximity to inferred centromere and telomere positions and higher at pericentromeric regions when chromosomes were metacentric or submetacentric. Genomic features such as repeat content (*Gypsy*-LTR, *Copia*-LTR, or simple repeats) GC-content and gene content were significantly associated with expected cumulative CO counts in 960-kb intervals (Model-1 cross-validation $R^2 = 0.52$, RMSE = 12.93). *Gypsy*-LTR, %-GC, and *Copia*-LTR content all had significant negative effects on cumulative CO counts while gene content and simple repeats were positively associated with CO counts (Supplementary Table 9). Dropping simple repeat content from the linear model increased the magnitude of the effects for other factors which may be explained by the multicollinearity exhibited between these variables (Supplementary Fig. 10). Dropping simple repeats from the prediction model also had minimum effect of the predictability of local CO counts within these genomic windows (Model-2 cross-validation $R^2 = 0.52$, RMSE = 13.03).

Fine-scale differences in CO counts between sexes

At the Chromosome-scale CO count per meiosis shows a dramatic variation among individuals (Fig. 3a), which is evident in both sexes. To further investigate the differences in CO counts between sexes (Fig. 3a), they were projected to a finer scale within chromosomes. Informed by the wavelet analysis, a nonoverlapping window size of 960 kb was used in the fine-scale analysis. Differences in cumulative CO counts between males and females showed similar trends across the genome ($r^2_{Spearman} = 0.78$; P-value $< 2.2 \times 10^{-16}$) and were not qualitatively different (Kolmogorov–Smirnov test, $P = 0.6093$, Fig. 2; Supplementary Fig. 6). Nevertheless, 16 genomic windows were identified (Exact-Poisson test; FDR$_{BH} \leq 0.25$) that show differences in CO counts between females and males, spread through the genome (Fig. 3b). Interestingly, only one genomic window showed higher CO counts for females (Fig. 3b). The overwhelming majority of windows showed higher CO rates for males. Furthermore, there were long consecutive chromosome segments for which male CO counts were consistently higher than females, such as at the beginning of LG-I, though the differences were not statistically
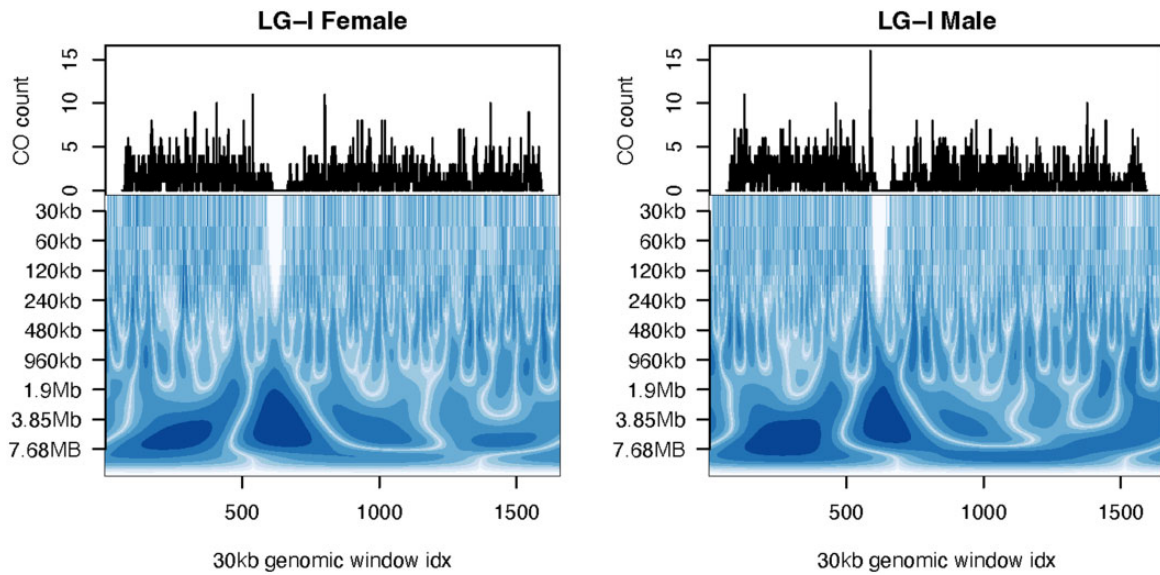
Fig. 2. Continuous wavelet transform (CWT) of female and male cumulative CO count for LG-I. Line graphs display cumulative CO counts for each sex (in columns) across 30 kb genomic windows. Wavelet coefficients at each scale using a CWT are displayed as a power spectrum at each scale.
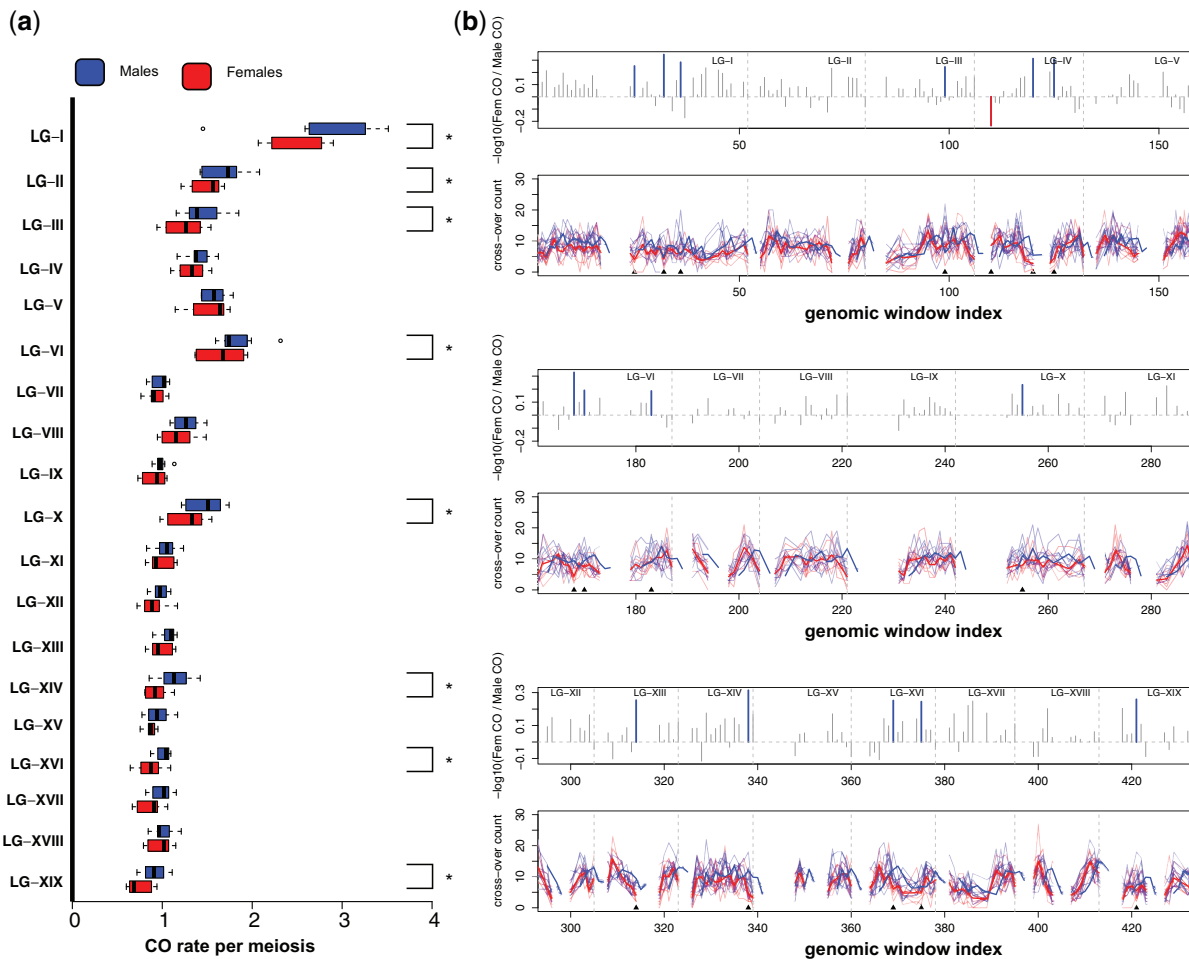

Fig. 3. Broad to fine scale CO count differences between sexes: a) CO rate per chromosome per meiosis in males (blue, top) vs females (red, bottom). LGs marked with an asterisk shows significant differences in cumulative CO counts between the sexes (χ^2 df = 1; FWER ≤ 0.05). b) Fine scale differences in recombination between sexes are shown. Log_{10} transformed male to female cumulative CO count ratios are shown in grey bars for each window and blue and red bars indicate male biased (positive values) and female biased (negative values) heterochiasmy, respectively. Only one window showed female bias. Mean CO count shown for male (thick blue line) and female (thick red line) parents. Lighter colored lines represent counts for individual parents of each sex. Gaps indicate centromeric or telomeric regions that were removed from our statistical analysis due to sparsity of counts.

significant for individual intervals (Fig. 3b). Windows with a higher number of COs in males had lower GC content (Wilcoxon Rank-sum test; $W = 655$; $P = 0.057$) and gene content (Wilcoxon Rank-sum test; $W = 623$; $P = 0.034$) compared to windows that did not show significant sex bias. None of the other variables related to repeat content (*Copia*, *Gypsy*, or simple repeats counts) had any significant difference (FPR ≤ 0.05) between these 2 groups (Fig. 4).

Identification of genome-wide CO hotspots

At yet a finer resolution of 30 kb, the observed distribution of CO counts showed an excess of windows with zero CO events, most likely due to windows that overlap centromeres and a longer than expected tail that may consist of candidate CO hotspots (Supplementary Fig. 11). CO hotspot regions were defined as 30-kb windows that showed significantly elevated cumulative CO counts in comparison to the genome-wide average. Initially, 119 candidate CO hotspot regions were identified ($\lambda = 2.98$, FWER$_{Bonferroni} \leq 0.05$), which were subsequently enriched for true-positives using 2 further criteria. The first set of outliers were identified as windows that do not represent a balanced contribution by all half-sib families, identified using inflated IOD of CO count within a 30-kb-genomic-window (10th percentile < IOD > 90th percentile of the global distribution of IOD across all windows). The second set of outliers were identified as windows that may have artificially inflated CO counts due to the lack of markers in adjacent genomic windows (windows containing at least one CO per window but average marker count in 2 adjacent

flanking windows is less than 1). Applying these criteria reduced the number of hotspot windows to 67. These robust CO hotspot candidates were spread across the genome in clusters of adjacent genomic regions (Fig. 5a). Genome-wide hotspots were evaluated separately for males and females with adjusted CO-counts per window ($\lambda_{female} = 1.38$; $\lambda_{male} = 1.54$) where 38 and 24 CO hotspot windows were identified, respectively (FWER$_{Bonferroni} \leq 0.05$). Of these, 21 and 11 windows overlapped with hotspots identified with females and males combined, respectively. Although female and male hotspots were within the general vicinity of each other, only a single CO hotspot window was shared between the sexes in our analysis (Supplementary Fig. 12). However, the lower number of CO hotspot windows detected for males in comparison to females could be due to the higher null expectation and variance of CO counts for males leading to a more stringent FWER threshold for meeting statistical significance in comparison to females.

The enrichment of genomic features in CO hotspots was evaluated through comparison with windows with background levels of CO counts using a 2-tailed Wilcoxon rank sum test (Supplementary Fig. 13). Simple repeat count was most strongly elevated in hotspots (Wilcoxon Rank-sum test; $W = 67,698$; $P = 1.795 \times 10^{-7}$), while GC content (Wilcoxon Rank-sum test; $W = 130,366$; $P = 0.003048$) and *Copia*-LTR content (Wilcoxon Rank-sum test; $W = 124,020$; $P = 0.0198$) were significantly lower. Neither *Gypsy*-LTR nor gene content showed significant differences (FPR ≥ 0.05).
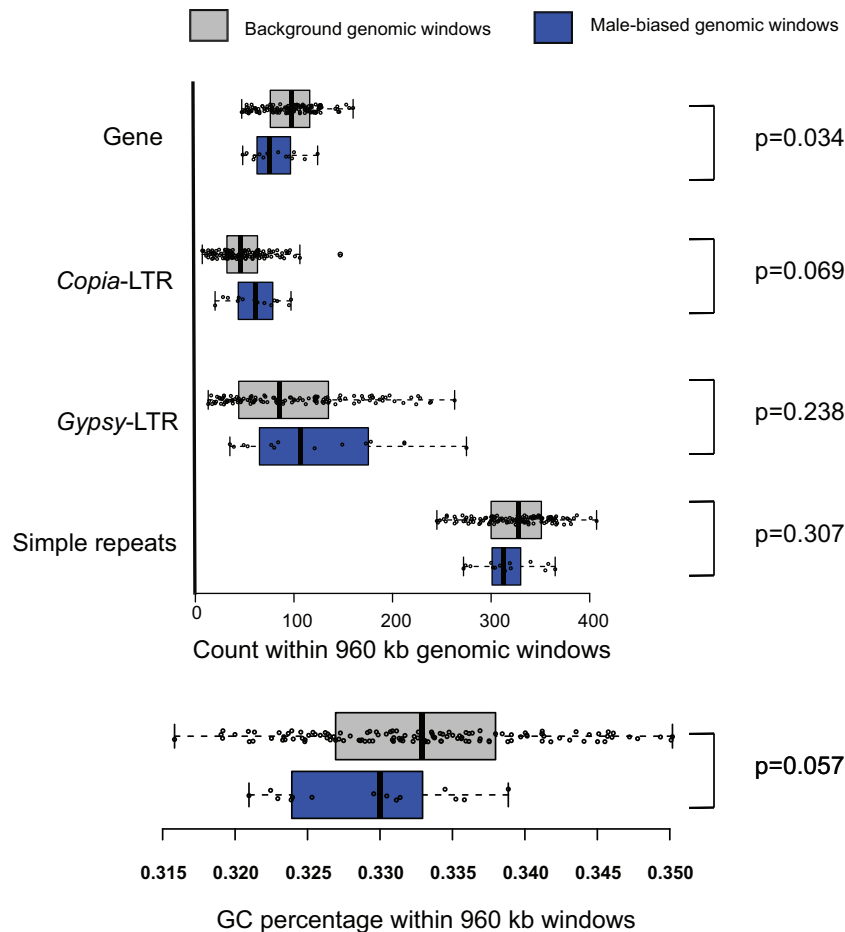

Fig. 4. Counts of genomic features in 960 kb windows across the genome, and their association with elevated CO counts in males. Enrichment of genomic features in windows that showed significantly elevated CO counts were compared with windows with background male/female ratios.
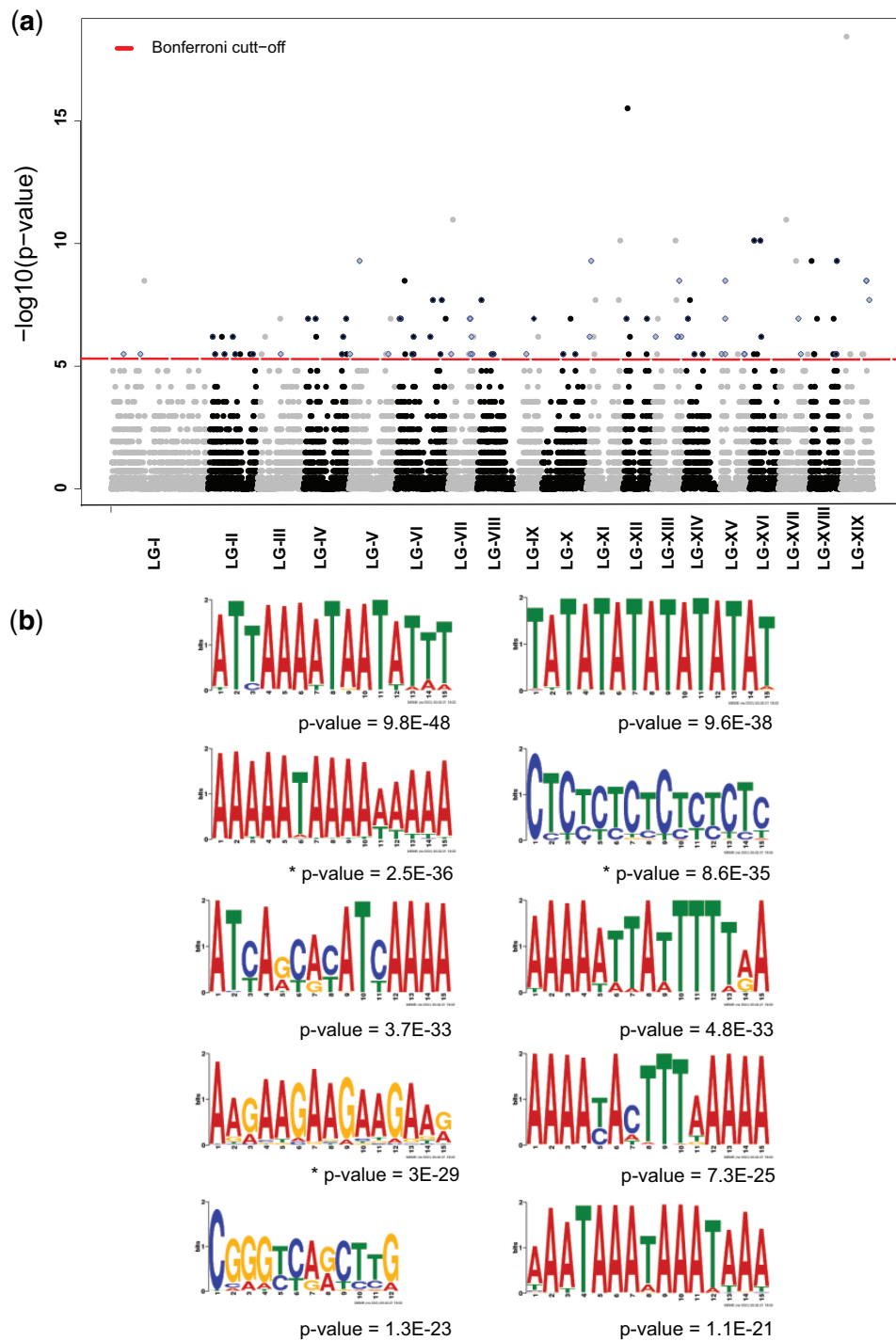
(a)

(b)

p-value = 9.8E-48

p-value = 9.6E-38

* p-value = 2.5E-36

* p-value = 8.6E-35

p-value = 3.7E-33

p-value = 4.8E-33

* p-value = 3E-29

p-value = 7.3E-25

p-value = 1.3E-23

p-value = 1.1E-21

Fig. 5. Candidate CO hotspot regions. a) Points marked as blue diamonds represent nonoverlapping windows of size 30 kb that represent CO count hotspots based on a probability model that describes genome-wide average CO per window by a Poisson null distribution ($\lambda = 3$; FWER ≤ 0.05). b) The top 10 DNA sequence motifs enriched in CO regions that were demarcated to a genomic region of 10 kb or less. *P*-values denoted with a "*" identifies DNA sequence motifs that have significant blast hits to recombination hotspot associated DNA sequence motifs previously identified for *Arabidopsis* and wheat.

DNA sequence motifs associated with COs

In order to investigate whether there are DNA sequence motifs associated with higher CO probability, the analysis was restricted to narrowly delimited CO regions. A total of 199 DNA sequence motifs putatively associated with COs are reported and of those, motifs that ranked within the first 3 were all A/T-rich repeat sequences (Fig. 5b). Interestingly, most of the enriched DNA sequence motifs identified within the top 10 showed considerable abundance across the whole genome. The full list of enriched DNA sequence motifs and matching or similar sequences across the genome is detailed in Supplementary Table 10. The enriched DNA sequence motifs also showed high similarity to such motifs associated with recombination hotspots in wheat and *Arabidopsis* as well (Fig. 5b; Supplementary Tables 11 and 12). However, our

analysis was not able to identify any DNA sequence motifs that are associated with CO hotspots when compared with regions with average COs at a maximum resolution of 30 kb.

Discussion
High-fidelity genetic maps
We reported 14 dense framework genetic maps each with 19 LGs, corresponding to the accurate haploid number of chromosomes in *P. trichocarpa*, all of which were derived from intraspecific crosses. Full-sib family sizes were too small for accurate genetic mapping, so we used half-sib families for map construction (Grattapaglia *et al.* 1996; Littrell *et al.* 2018). Given that each parent was crossed with the same set of individuals from the opposite sex in a full-factorial cross, genetic maps among individuals within a sex are highly comparable with each other. The concordance, collinearity, and the congruity of our genetic maps can be mainly attributed to key steps in our pipeline such as (1) pedigree error correction, (2) robust variant calling pipeline, (3) genotype error correction and imputation of parents, (4) phasing and identification of haplotype configuration in offspring, and (5) constructing the initial framework maps using a moderate number of intensively filtered markers to avoid map inflation. Although whole-genome resequencing provided a very high density of SNP and INDEL markers distributed across the genome, the resolution of our maps was largely limited by the number of meioses within each half-sib family, which resulted in a large number of markers being binned due to redundant genetic information (Da *et al.* 1998). Nevertheless, a consensus map making use of all 829 phased progeny takes full advantage of all observed recombinations. This represents the most dense genetic map available to date for *Populus*, with 5,029 nonredundant pseudo-markers spread across the genome. This should provide an invaluable resource for population genomics applications that require accurate estimates of recombination.

We reported comparable recombination rates across LGs and considerable variation in recombination rate among individuals. Average genome-wide recombination rate estimates across LGs observed in our analysis were in close agreement with a previous study in *P. trichocarpa* (Supplementary Fig. 7), yet markedly lower than reported in a similar study in *P. tremula* (Apuli *et al.* 2020). Interestingly, a comparative genomic study including these 2 species, also observed similar disparities in which population-scaled recombination rates in *P. trichocarpa* were found to be 4 times lower in comparison to *P. tremula* (Wang *et al.* 2016). Linkage-map-based recombination rate estimates are not immediately comparable between our study and Apuli *et al.* (2020), due to differences in method, number of markers, number of families and their respective size used to produce genetic maps. However, it will be interesting to investigate whether disparities discussed for these 2 species have roots in early evolutionary divergence of the genus *Populus* (Eckenwalder 1996).

Our approach simulates a structured tree breeding program and has high probability of identifying large haplotype blocks with a history of limited internal recombination (Jansen *et al.* 2003; Di Pierro *et al.* 2016). Such information is potentially useful to breeding programs because quantitative trait loci (QTL) falling within these regions can be targets for marker assisted selection (Wu *et al.* 1998; Muranty *et al.* 2014). Fine-scale estimates of recombination rates and exact recombination breakpoints identified in this study will be useful in QTL delineation for feedstock relevant phenotypes. These estimates provide an important foundation for accelerated domestication as well as basic investigations into the evolutionary biology and molecular genetics of this model forest tree (Grattapaglia and Resende 2011; Isik 2014).

Higher rates of CO in males
One of our most striking findings was the higher rates of CO in male parents of *P. trichocarpa* (male-biased heterochiasmy) where we had an equal number of gametes (829) for each sex. A higher rate of recombination in males had been previously reported for an interspecific cross of *Populus* (Yin *et al.* 2002). However, a revisit of these data in a metanalysis failed to confirm the finding (Lenormand and Dutheil 2005). A possible reason could be that the metanalysis only considered markers that were shared between the male and female maps, which also resulted in overall shorter cumulative map lengths. In our study, we compared cumulative CO counts in each LG accounting for individual variance and other covariates such as LG identity and half-sib family size, after which 8 out of 19 LGs showed male-biased heterochiasmy at the chromosomal scale.

We also observed higher CO counts in male gametogenesis in an overwhelming majority of the 960-kb windows analyzed genome-wide. Overall, male and female CO landscapes were highly correlated for a given LG and were not qualitatively different, unlike in *Arabidopsis* (Lloyd and Jenczewski 2019), human (Kong *et al.* 2010), or mouse (Brick *et al.* 2018) where patterns of CO distribution are dissimilar between sexes. This suggests a model where the CO landscape is spatially conserved across sexes in *P. trichocarpa*, at least at 960-kb resolution. Heterochiasmy is commonly observed across a wide range of organisms, including plants (Lenormand and Dutheil 2005; Sardell and Kirkpatrick 2020). It is a fast-evolving trait, as evidenced by the observation that phylogenetic inertia does not limit its emergence (Lenormand and Dutheil 2005). Five possible explanations have been proposed for the observed differences in CO counts between sexes: (1) cellular, molecular mechanisms; (2) pleiotropic effects of the sex chromosome or SDR; (3) sex dimorphism in haploid selection (Hunt and Hassold 2002; Petkov *et al.* 2007; Lloyd and Jenczewski 2019); (4) meiotic drive mechanisms (Brandvain and Coop 2012); and (5) differential external factors during meiosis (Phillips *et al.* 2015; Coulton *et al.* 2020). Evidence for each of these is discussed below.

Female and male meioses are fundamentally different in plants. Heterochiasmy can result from differences in the amount and distribution of DSBs, their resolution and maturation as either CO or NCO as well as the sensitivity to CO interference (Capilla-Pérez *et al.* 2021). The length of the synaptonemal complex (SC), which is a structure formed during meiosis and responsible for ensuring proper chiasma formation and maintenance, is shown to have a positive correlation to CO occurrence in *Arabidopsis* (Drouaud *et al.* 2007; Giraut *et al.* 2011), leading to higher rates of recombination in male gametogenesis. Furthermore, SCs are deemed essential in Arabidopsis for CO interference, and CO frequencies are shown to equalize in the absence of SCs (Capilla-Pérez *et al.* 2021). Heterochiasmy in *Zea mays* also reveals more COs in male gametogenesis and more hotspots. This varies by genotype and has been attributed to variation in CO maturation, which could also be related to differences in SC length (Luo *et al.* 2019). Differential methylation and chromatin structure have also been implicated for sex differences in recombination (Wang and Copenhaver 2018).

In *Populus balsamifera* which is a sister species to *P. trichocarpa* (Rood *et al.* 1986; Eckenwalder 1996; Wang *et al.* 2020) male biased expression in a subset of chromatin remodeling and DNA

methylation regulator genes have been observed (Cronk *et al.* 2020). Although we did not directly investigate chromatin modifications or DNA methylation in our study, we reported that observed male-biased heterochiasmy had close association with gene content and %GC content in 960-kb windows. These factors could themselves determine specific locales and amount of DNA methylation or chromatin modifications shaping the quantitative differences we observed. However, this remains an open question, and the underlying molecular mechanisms that are responsible for these differences have yet to be identified (Bergero *et al.* 2021).

The *Populus* SDR contains at least 2 genes involved in differential DNA methylation between sexes (Geraldes *et al.* 2015; Zhou *et al.* 2020; Kim *et al.* 2021). However, whether these genes could have pleiotropic effects that extend to other regions of the genome to affect DNA methylation and histone remodeling is still unclear (Sardell and Kirkpatrick 2020). Intriguingly, a study of DNA methylation across a range of tissues in *P. trichocarpa* found the highest rates of methylation in male floral tissue (Vining *et al.* 2012), though the significance of this for chromatin structure and CO occurrence is unclear. On the other hand, not all chromosomes or fine-scale genomic-windows in our study showed male biased heterochiasmy, which means that the pleiotropy-based explanation does not completely address our observations for windows in which female rates were higher or in which no significant bias was observed either way.

Another potential contributor to heterochiasmy in plants is differential opportunity for haploid selection (Lenormand and Dutheil 2005). In plants there is a substantial time period in which haploid spores, unmasked from effects of homologous partners, may express their genes that are then subject to selective pressures. In wind-pollinated species such as *P. trichocarpa* the opportunity for haploid selection and the distance of dispersal may favor mechanisms that promote genetic variation and thus increased shuffling of haploid genomes (Slavov *et al.* 2009). In *P. trichocarpa*, pollination distances can exceed dozens of kilometers (Slavov *et al.* 2009), and modeling studies suggest that male gametes typically are dispersed much further than female gametes (DiFazio *et al.* 2012). This mechanism could account in part for the observed male-biased heterochiasmy. However, increased recombination may not be favored when recombination breaks-up favorable epistatic interactions thus reducing the mean fitness of the resulting population in subsequent generations (Charlesworth and Barton 1996). In angiosperms, males are subjected to increased selective pressure due to pollen competition and should in theory have lower recombination rates that preserve favorable haplotype combinations minimizing recombination load (Lenormand and Dutheil 2005). However, the validity of this theory may not apply to *Populus*, which has several potentially mitigating characteristics, including dioecy, anemophily, and extensive clonal reproduction through sprouting and rooting of woody propagules.

Recombination rate modifiers that alter the efficiency of sex-specific meiotic drive mechanisms have also been implicated for sex-based differences in recombination rate. Although specific mechanistic processes are still unclear, a leading hypothesis based on population genetic modelling suggests that female meiotic drive could either enhance or suppress recombination rates, based on the interaction of whether the drive system functions during meiosis-I (MI) or meiosis-II (MII) and whether it is linked or unlinked to a recombination rate modifier (Brandvain and Coop 2012). As per this reasoning, relatively higher rates of recombination in males as we observed could be the result of an active suppression of recombination in females due to either an MI meiotic drive system in phase with a recombination suppressor or an MII drive system unlinked to a recombination suppressor during female gametogenesis. However, under this model, recombination rate differences between males and females should be more prominent closer to the centromeric region which was not observed in our study. Although an MII meiotic driver system affecting recombination has been characterized in maize (Hiatt and Dawe 2003), evidence for such systems are not reported for any *Populus* species. Female biased sex ratios are observed in a number of Salicaceae species, and even here the only mechanism identified does not implicate sex chromosome meiotic drive (Pucholt *et al.* 2017).

Higher recombination rates for male gametogenesis have been reported in interspecific crosses of *Populus* (Yin *et al.* 2002), as well as within male strobili of *Pinus* (Groover *et al.* 1995; Plomion and O'Malley 1996), and in the latter it may be attributed to the temporal differences in meiosis between male and female strobili where external temperature differences during gametogenesis are driving sex differences (Moran *et al.* 1983). Details or direct evidence of differential external or internal environmental effects during gametogenesis are not reported for *P. trichocarpa* other than possible ambient chemical composition differences within catkins.

Genome-wide CO hotspots

Conserved DNA sequence motifs associated with CO hotspots have been identified in domesticated plants such as *Zea mays*, *Triticum aestivum* (Darrier *et al.* 2017), and *Oryza sativa* (Mezard 2006), yet similar studies on undomesticated forest trees are rare (Slavov *et al.* 2012; Silva Junior and Grattapaglia 2015; Apuli *et al.* 2020). Although LD-based recombination rate estimates have been previously obtained for *P. trichocarpa* (Slavov *et al.* 2012), these estimates are subject to stochasticity of population demography and do not provide the flexibility to decompose sex-based differences. This study revealed the genome-wide recombination landscape and patterns of inheritance at a finer scale than has previously been possible in a single generation of *Populus*, and rarely achieved in undomesticated plants. Using high density phased and imputed SNPs, we were able to delineate recombination breakpoints within 30-kb windows. We identified 67 windows enriched for recombination hotspots and these sites displayed a genome-wide distribution yet clustered away from putative centromeres and telomeres. A role for absence of chromatin compaction and trimethylation of lysine 4 in histone H3 has been identified as strong markers of DSBs and recombination in yeast and mammals (Pan *et al.* 2011; Smagulova *et al.* 2011; Lange *et al.* 2016). However, such patterns cannot be generalized in plants such as *Zea mays* in which association between DSBs and recombination has been shown to be moderated by other factors such as the amount of repetitive DNA (Rodgers-Melnick *et al.* 2016; He *et al.* 2017). Molecular mechanisms that regulate DSBs and subsequent recombination are complex and involve the interaction of localized genomic features related to chromatin compaction such as histone remodeling, DNA-methylation, gene content, and repeat content in *Arabidopsis* (Yelina *et al.* 2012; Shilo *et al.* 2015).

Regions with high AT richness have been associated with elevated CO frequency in *A. thaliana* (Choi *et al.* 2018). Poly(dA:dT) tracts and their flanking regions are shown to be depleted of nucleosomes in vitro, in addition to blocking the spread of post-translational histone modifications that limit accessibility to recombination and transcriptional machinery (Segal and Widom

2009). Similarly, hotspots identified in our study were associated with higher AT content and were comparatively enriched for simple repeat elements. This observation is further supported by the association of AT-rich DNA sequence motifs at CO sites. Although not exact matches, we reported several DNA sequences that showed close similarity to such sequences that were observed in wheat (Darrier *et al.* 2017) and *Arabidopsis* (Shilo *et al.* 2015) that are associated with DNA methylation and histone remodeling.

We also presented a statistical model that accounts for 53% of the variation in CO counts in 960-kb nonoverlapping windows using localized genomic correlates. Consistent with other plant species (Mezard 2006; Gaut *et al.* 2007), we observed that COs tend to be higher in regions enriched for genes and simple repeats while negatively influenced by higher %GC content along with LTR-*Gypsy* and LTR-*Copia*. It is difficult to parse out the multicollinearity among these explanatory variables and by excluding the centromeric and telomeric regions we have tried to minimize this issue. Phylogenetically independent contrasts also show similar trends as we observed (Tiley and Burleigh 2015) and suggest fundamental genetic or evolutionary driving forces acting on CO control. It is hypothesized that transposable elements accumulate within regions with low recombination rates, which initiates a positive feedback loop spreading recombination suppression to adjacent regions in a model that describes coevolutionary interaction (Kent *et al.* 2017). We observed a positive correlation between gene density and CO counts that concur with the general tendency in plant species which provides a stark contrast to the absence of a conclusive relationship observed in animals and fungi at these broader scales (Mezard 2006; Gaut *et al.* 2007; Haenel *et al.* 2018). This is in line with the recombination modifier theory for the existence of CO hotspots where mutations of sequence variants that promote shuffling of flanking regions are favored, since such a situation makes selection and local adaptation more efficient especially in the case of perennial forest trees (Hey 2004).

Conclusions

Our analysis revealed the genome-wide recombination landscape at a finer scale than has previously been possible in most undomesticated woody plant species. In *P. trichocarpa*, this information will be used to delineate QTL for phenotypes of breeding relevance (height, diameter, bud set, and disease resistance) and along with recombination rate estimates to improve genomic prediction models. Given that *P. trichocarpa* is a promising renewable feedstock for bioenergy and bioproducts, we believe that our findings on recombination and CO rate estimates will be useful for ongoing efforts to accelerate domestication of this and other feedstocks, as well as future studies that investigate broader questions such as evolutionary history, perennial development related to phenology, wood formation, vegetative propagation, and dioecy that cannot be studied using conventional plant model systems such as *Arabidopsis*, rice, or maize.

Data availability

Raw sequence reads are publicly available at the JGI genome portal under proposal ID 502915, Award DOI: 10.46936/10.25585/60001012 and metadata related to sequence read archive (SRA) accessions used in this study are provided in Supplementary Table 13. Input data for genetic maps, relevant metadata, and analysis scripts used in this study are available at https://github.com/roshanabeyratne/DiFaziolab_Ptrichocarpa_recombination_7x7.

Supplemental material is available at G3 online.

Conflicts of interest

None declared.

Author contributions

CRA carried out genetic mapping and subsequent data analysis and wrote the manuscript. DMS provided advice on genetic mapping, data analyses, performed data analyses, wrote, and edited the manuscript. RZ provided methods and scripts for estimating repeats content, gene content, and AT/GC composition within genomic windows. KB led the resequencing of the genomes of individuals in this study at JGI. CD, MZ, YY, and AL carried out DNA sequence library preparation, sequencing and quality control. BS designed the 7×7 cross. KH generated the 7×7 crosses and managed the clonal progeny field trial. GAT provided funding and leadership for the project. SPD led the study and helped write and edit the manuscript. All authors read and approved the final manuscript.

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410.

Apuli R-P, Bernhardsson C, Schiffthaler B, Robinson KM, Jansson S, Street NR, Ingvarsson PK. Inferring the genomic landscape of recombination rate variation in European Aspen (*Populus tremula*). G3 (Bethesda). 2020;10(1):299–309.

Auton A, McVean G. Recombination rate estimation in the presence of hotspots. Genome Res. 2007;17(8):1219–1227.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server Issue): W202–W208.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science. 2010;327(5967): 836–840.

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W, *et al.* Intraspecific variation of recombination rate in maize. Genome Biol. 2013;14(9):R103.

Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B Stat Methodol. 1995;57(1):289–300.

Bergero R, Ellis P, Haerty W, Larcombe L, Macaulay I, Mehta T, Mogensen M, Murray D, Nash W, Neale MJ, *et al.* Meiosis and beyond—understanding the mechanistic and evolutionary processes shaping the germline genome. Biol Rev. 2021;96(3): 822–841.

Bherer C, Campbell CL, Auton A. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nat Commun. 2017;8:14994.

Bradshaw HD, Ceulemans R, Davis J, Stettler R. Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. J Plant Growth Regul. 2000;19(3):306–313.

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF. Molecular genetics of growth and development in Populus. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet. 1994;89(2–3):167–178.

Brandvain Y, Coop G. Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics. 2012;190(2): 709–723.

Brick K, Thibault-Sennett S, Smagulova F, Lam K-WG, Pu Y, Pratto F, Camerini-Otero RD, Petukhova GV. Extensive sex differences at the initiation of genetic recombination. Nature. 2018;561(7723): 338–342.

Burt A, Bell G, Harvey PH. Sex-differences in recombination. J Evol Biol. 1991;4(2):259–277.

Capilla-Pérez L, Durand S, Hurel A, Lian Q, Chambon A, Taochy C, Solier V, Grelon M, Mercier R. The synaptonemal complex imposes crossover interference and heterochiasmy in *Arabidopsis*. Proc Natl Acad Sci USA. 2021;118(12):11.

Charlesworth B, Barton NH. Recombination load associated with selection for increased recombination. Genet Res. 1996;67(1):27–41.

Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity. 2005;95(2):118–128.

Chhetri HB, Macaya-Sanz D, Kainer D, Biswal AK, Evans LM, Chen J-G, Collins C, Hunt K, Mohanty SS, Rosenstiel T, *et al.* Multitrait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. New Phytol. 2019;223(1):293–309.

Choi K, Henderson IR. Meiotic recombination hotspots - a comparative view. Plant J. 2015;83(1):52–61.

Choi K, Zhao X, Tock AJ, Lambing C, Underwood CJ, Hardcastle TJ, Serra H, Kim J, Cho HS, Kim J, *et al.* Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. Genome Res. 2018;28(4): 532–546.

Coulton A, Burridge AJ, Edwards KJ. Examining the effects of temperature on recombination in wheat. Front Plant Sci. 2020;11: 230–213.

Cronk Q, Soolanayakanahally R, Brautigam K. Gene expression trajectories during male and female reproductive development in balsam poplar (*Populus balsamifera* L.). Sci Rep. 2020;10(1):14.

Da Y, VanRaden PM, Li N, Beattie CW, Wu C, Schook LB. Designs of reference families for the construction of genetic linkage maps. Anim Biotechnol. 1998;9(3):205–228.

Dapper AL, Payseur BA. Connecting theory and data to understand recombination rate evolution. Philos Trans Roy Soc B Biol Sci. 2017;372:11.

Darrier B, Rimbert H, Balfourier F, Pingault L, Josselin A-A, Servin B, Navarro J, Choulet F, Paux E, Sourdille P, *et al.* High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. Genetics. 2017; 206(3):1373–1388.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–498.

Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst-van Putten HJ, Kruisselbrink JW, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, *et al.* A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hort Res. 2016;3(1):13.

DiFazio SP, Leonardi S, Slavov GT, Garman SL, Adams WT, Strauss SH. Gene flow and simulation of transgene dispersal from hybrid poplar plantations. New Phytol. 2012;193(4):903–915.

Drouaud J, Mercier R, Chelysheva L, Bérard A, Falque M, Martin O, Zanni V, Brunel D, Mézard C. Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. PLoS Genet. 2007;3(6):e106.

Eckenwalder JE. Systematics and evolution of *Populus*. In: RF Stettler, HD Bradshaw, PE Heilman, TM Hinckley, editors. Biology of Populus and Its Implications for Management and Conservation. Ottawa (Canada): NRC Research Press; 1996. p. 7–30.

Endelman JB, Plomion C. LPmerge: an R package for merging genetic maps by linear programming. Bioinformatics. 2014;30(11): 1623–1624.

Fang LC, Liu HL, Wei SY, Keefover-Ring K, Yin TM. High-density genetic map of *Populus deltoides* constructed by using specific length amplified fragment sequencing. Tree Genet Genomes. 2018;14(5):10.

Fay MP. Two-sided exact tests and matching confidence intervals for discrete data. R J. 2010;2(1):53–58.

Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78(2):737–756.

Gaudet M, Jorge V, Paolucci I, Beritognolo I, Mugnozza GS, Sabatti M. Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genet Genomes. 2007;4(1):25–36.

Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet. 2007;8(1):77–84.

Geraldes A, Hefer CA, Capron A, Kolosova N, Martinez-Nuñez F, Soolanayakanahally RY, Stanton B, Guy RD, Mansfield SD, Douglas CJ, *et al.* Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol Ecol. 2015;24(13): 3243–3256.

Gion J-M, Hudson CJ, Lesur I, Vaillancourt RE, Potts BM, Freeman JS. Genome-wide variation in recombination rate in *Eucalyptus*. BMC Genomics. 2016;17(1):12.

Giraut L, Falque M, Drouaud J, Pereira L, Martin OC, Mézard C. Genome-wide crossover distribution in *Arabidopsis thaliana* meiosis reveals sex-specific patterns along chromosomes. PLoS Genet. 2011;7(11):e1002354.

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR. Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genetics. 1996;144(3):1205–1214.

Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7(2):241–255.

Groover AT, Williams CG, Devey ME, Lee JM, Neale DB. Sex-related differences in meiotic recombination frequency in *Pinus taeda*. J Hered. 1995;86(2):157–158.

Haenel Q, Laurentino TG, Roesti M, Berner D. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol Ecol. 2018;27(11): 2477–2497.

Harman-Ware AE, Macaya-Sanz D, Abeyratne CR, Doeppke C, Haiby K, Tuskan GA, Stanton B, DiFazio SP, Davis MF. Accurate determination of genotypic variance of cell wall characteristics of a *Populus trichocarpa* pedigree using high-throughput pyrolysis-molecular beam mass spectrometry. Biotechnol Biofuels. 2021;14(1):15.

He Y, Wang M, Dukowic-Schulze S, Zhou A, Tiang C-L, Shilo S, Sidhu GK, Eichten S, Bradbury P, Springer NM, *et al.* Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. Proc Natl Acad Sci USA. 2017;114(46):12231–12236.

Heller R, Gur H. False discovery rate controlling procedures for discrete tests. arXiv preprint, arXiv:1112.4627, 2011; DOI: 10.1128/mSystems.00092-17.

Hey J. What's so hot about recombination hotspots? PLoS Biol. 2004; 2(6):e190.

Hiatt EN, Dawe RK. Four loci on abnormal chromosome 10 contribute to meiotic drive in maize. Genetics. 2003;164(2):699–709.

Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8(3):269–294.

Hofmeister BT, Denkena J, Colomé-Tatché M, Shahryary Y, Hazarika R, Grimwood J, Mamidi S, Jenkins J, Grabowski PP, Sreedasyam A, *et al.* A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial *Populus trichocarpa*. Genome Biol. 2020;21(1):27.

Hunt PA, Hassold TJ. Sex matters in meiosis. Science. 2002;296(5576): 2181–2183.

Isik F. Genomic selection in forest tree breeding: the concept and an outlook to the future. New Forests. 2014;45(3):379–401.

Jansen RC, Jannink JL, Beavis WD. Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci. 2003;43(3):829–834.

Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: recombination distributions in mammals. Nat Rev Genet. 2004;5(6): 413–424.

Kent TV, Uzunovic J, Wright SI. Coevolution between transposable elements and recombination. Philos Trans Roy Soc B Biol Sci. 2017;372:11.

Kianian PMA, Wang M, Simons K, Ghavami F, He Y, Dukowic-Schulze S, Sundararajan A, Sun Q, Pillardy J, Mudge J, *et al.* High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. Nat Commun. 2018;9(1):10.

Kim G, Montalvao APL, Kersten B, Fladung M, Muller NA. The genetic basis of sex determination in *Populus* provides molecular markers

across the genus and indicates convergent evolution. Silvae Genet. 2021;70(1):145–155.

Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir A, Gylfason A, Kristinsson KT, *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010;467(7319):1099–1103.

Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1–26.

Lambing C, Franklin FCH, Wang CJR. Understanding and manipulating meiotic recombination in plants. Plant Physiol. 2017;173(3): 1530–1542.

Lambing C, Tock AJ, Topp SD, Choi K, Kuo PC, Zhao X, Osman K, Higgins JD, Franklin FCH, Henderson IR, *et al.* Interacting genomic landscapes of REC8-Cohesin, chromatin, and meiotic recombination in *Arabidopsis*. Plant Cell. 2020;32(4):1218–1239.

Lange J, Yamada S, Tischfield SE, Pan J, Kim S, Zhu X, Socci ND, Jasin M, Keeney S. The landscape of mouse meiotic double-strand break formation, processing, and repair. Cell. 2016;167(3): 695–708.e16.

Lenormand T. The evolution of sex dimorphism in recombination. Genetics. 2003;163(2):811–822.

Lenormand T, Dutheil J. Recombination difference between sexes: a role for haploid selection. Plos Biol. 2005;3(3):e63.

Littrell J, Tsaih S-W, Baud A, Rastas P, Solberg-Woods L, Flister MJ. A high-resolution genetic map for the laboratory rat. G3 (Bethesda). 2018;8:2241–2248.

Lloyd A, Jenczewski E. Modelling sex-specific crossover patterning in *Arabidopsis*. Genetics. 2019;211(3):847–859.

Luo C, Li X, Zhang QH, Yan JB. Single gametophyte sequencing reveals that crossover events differ between sexes in maize. Nat Commun. 2019;10(1):8.

Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6): 764–770.

Margarido GRA, Souza AP, Garcia AAF. OneMap: software for genetic mapping in outcrossing species. Hereditas. 2007;144(3):78–79.

Mezard C. Meiotic recombination hotspots in plants. Biochem Soc Trans. 2006;34:531–534.

Moran GF, Bell JC, Hilliker AJ. Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. J Hered. 1983;74(1):62–62.

Muller HJ. The relation of recombination to mutational advance. Mut Res. 1964;1(1):2–9.

Muranty H, Jorge V, Bastien C, Lepoittevin C, Bouffier L, Sanchez L. Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. Tree Genet Genomes. 2014;10(6):1491–1510.

Neale DB, Kremer A. Forest tree genomics: growing resources and applications. Nat Rev Genet. 2011;12(2):111–122.

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell. 2011;144(5):719–731.

Penalba JV, Wolf JBW. From molecules to populations: appreciating and estimating recombination rate variation. Nat Rev Genet. 2020;21(8):476–492.

Petit RJ, Hampe A. Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst. 2006;37(1):187–214.

Petkov PM, Broman KW, Szatkiewicz JP, Paigen K. Crossover interference underlies sex differences in recombination rates. Trends Genet. 2007;23(11):539–542.

Phillips D, Jenkins G, Macaulay M, Nibau C, Wnetrzak J, Fallding D, Colas I, Oakey H, Waugh R, Ramsay L. The effect of temperature

on the male and female recombination landscape of barley. New Phytol. 2015;208(2):421–429.

Plomion C, O'Malley DM. Recombination rate differences for pollen parents and seed parents in *Pinus pinaster*. Heredity. 1996;77(4):341–350.

Porth I, El-Kassaby YA. Using *Populus* as a lignocellulosic feedstock for bioethanol. Biotechnol J. 2015;10(4):510–24524.

Pucholt P, Hallingback HR, Berlin S. Allelic incompatibility can explain female biased sex ratios in dioecious plants. BMC Genomics. 2017;18(1):12.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing; 2013.

Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573–89589.

Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, Li C, Li Y, Buckler ES. Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc Natl Acad Sci USA. 2015;112(12):3823–3828.

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. Proc Natl Acad Sci USA. 2016;113(22):E3177–E3184.

Rood SB, Campbell JS, Despins T. Natural poplar hybrids from southern Alberta. I. Continuous variation for foliar characteristics. Can J Bot. 1986;64(7):1382–1388.

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. 2019;213(3):771–787.

Saintenac C, Faure S, Remay A, Choulet F, Ravel C, Paux E, Balfourier F, Feuillet C, Sourdille P. Variation in crossover rates across a 3-Mb contig of bread wheat (*Triticum aestivum*) reveals the presence of a meiotic recombination hotspot. Chromosoma. 2011;120(2):185–198.

Sannigrahi P, Ragauskas AJ, Tuskan GA. Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuels, Bioprod Bioref. 2010;4(2):209–226.

Sardell JM, Kirkpatrick M. Sex differences in the recombination landscape. Am Nat. 2020;195(2):361–379.

Segal E, Widom J. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol. 2009;19(1):65–71.

Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. Plant Cell. 2015;27(9):2427–2436.

Sidhu GK, Fang C, Olson MA, Falque M, Martin OC, Pawlowski WP. Recombination patterns in maize reveal limits to crossover homeostasis. Proc Natl Acad Sci USA. 2015;112(52):15982–15987.

Silva Junior OB, Grattapaglia D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. New Phytol. 2015;208(3):830–845.

Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, *et al.* Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. New Phytol. 2012;196(3):713–725.

Slavov GT, Leonardi S, Burczyk J, Adams WT, Strauss SH, Difazio SP. Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. Mol Ecol. 2009;18(2):357–373.

Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. Nature. 2011;472(7343):375–378.

Smit AFA, Hubley R, Green P. *RepeatMasker Open-4.0*; 2013–2015. [accessed 2022 October 14]. http://www.repeatmasker.org.

Smukowski CS, Noor MAF. Recombination rate variation in closely related species. Heredity. 2011;107(6):496–508.

Tiley GP, Burleigh G. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol Biol. 2015;15(1):14.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313(5793):1596–1604.

Van Os H, Stam P, Visser RGF, Van Eck HJ. RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet. 2005;112(1):30–40.

Vining KJ, Pomraning KR, Wilhelm LJ, Priest HD, Pellegrini M, Mockler TC, Freitag M, Strauss SH. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. BMC Genomics. 2012;13:27.

Wang J, Street NR, Scofield DG, Ingvarsson PK. Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related *Populus* species. Genetics. 2016;202(3):1185–1200.

Wang JL. Sibship reconstruction from genetic data with typing errors. Genetics. 2004;166(4):1963–1979.

Wang M, Zhang L, Zhang Z, Li M, Wang D, Zhang X, Xi Z, Keefover-Ring K, Smart LB, DiFazio SP, *et al.* Phylogenomics of the genus Populus reveals extensive interspecific gene flow and balancing selection. New Phytol. 2020;225(3):1370–1382.

Wang YX, Copenhaver GP. Meiotic recombination: mixing it up in plants. Annu Rev Plant Biol. 2018;69:577–609.

Wegrzyn JL, Eckert AJ, Choi M, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai C-J, Neale DB. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol. 2010;188(2):515–532.

Wu R, Bradshaw HD, Stettler RF. Developmental quantitative genetics of growth in *Populus*. Theor Appl Genet. 1998;97:1110–1119.

Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, Miller N, Drouaud J, Grelon M, Copenhaver GP, *et al.* Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genet. 2012;8(8):e1002844.

Yin TM, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theor Appl Genet. 2004;109(3):451–463.

Yin T, Zhang X, Huang M, Wang M, Zhuge Q, Tu S, Zhu L-H, Wu R. Molecular linkage maps of the *Populus* genome. Genome. 2002;45(3):541–555.

Zelkowski M, Olson MA, Wang MH, Pawlowski W. Diversity and determinants of meiotic recombination landscapes. Trends Genet. 2019;35(5):359–370.

Zhou R, Macaya-Sanz D, Schmutz J, Jenkins JW, Tuskan GA, DiFazio SP. Sequencing and analysis of the sex determination region of *Populus trichocarpa*. Genes. 2020;11(8):843.

Communicating editor: A. Kern