Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*

Kieran Samuk[1], Brenda Manzano-Winkler[1], Kathryn R. Ritz[1], Mohamed A.F. Noor[1]*

[1]Department of Biology, Duke University. Durham, NC, USA, 27708.
*Lead Contact, Correspondence: ksamuk@gmail.com

Summary

While recombination is widely recognized to be a key modulator of numerous evolutionary phenomena, we have a poor understanding of how recombination rate itself varies and evolves within a species. Here, we performed a comprehensive study of recombination rate (rate of meiotic crossing over) in two natural populations of *Drosophila pseudoobscura* from Utah and Arizona, USA.

We used an amplicon sequencing approach to obtain high-quality genotypes in approximately 8000 individual backcrossed offspring (17 mapping populations with roughly 530 individuals each), for which we then quantified crossovers. Interestingly, variation in recombination rate within and between populations largely manifested as differences in genome-wide recombination rate rather than remodeling of the local recombination landscape. Comparing populations, we discovered individuals from the Utah population displayed on average 8% higher crossover rates than the Arizona population, a statistically significant difference. Using a Q_{ST}-F_{ST} analysis, we found that this difference in crossover rate was dramatically higher than expected under neutrality, indicating that this difference may have been driven by natural selection. Finally, using a combination of short and long read whole-genome sequencing, we found no significant association between crossover rate and structural variation at the 200-400kb scale. Our results demonstrate that (1) there is abundant variation in genome-wide crossover rate in natural populations, (2) at the 200-400kb scale, recombination rate appears to vary largely genome wide, rather than in specific intervals and (3) interpopulation differences in recombination rate may be the result of local adaptation.

Keywords: recombination rate, evolution, meiosis, quantitative genetics, structural variation, genetic map, Q_{ST}-F_{ST}, *Drosophila pseudoobscura*

Introduction

Meiotic recombination is the exchange of genetic material between homologous chromosomes that occurs during meiosis. This exchange has two major forms, crossing over and non-crossover gene conversion, both of which are initiated by the formation of a double-strand break during meiosis. Recombination, particularly crossing over, is a key mediator of chromosome pairing during meiosis, with most species exhibiting an average of one crossover per chromosome arm [1,2].

While physical constraints often set a lower bound on rates of recombination, the evolution of recombination rate and particularly the rate of crossing over (i.e., number of crossovers per generation in a genomic interval) can have far-reaching effects on nearly every evolutionary process [2–4]. For example, recombination rates can modulate processes as diverse as adaptation to a new environment, the evolution of reproductive isolation, and the dynamics of introgression between populations [5–8]. More generally, recombination rate determines the degree to which an individual's parental chromosomes are mixed in their gametes – i.e., how often novel allelic combinations are generated in their gametes. Increases or decreases in this rate can be favored

under different evolutionary or ecological conditions. For example, increasing the rate of recombination can facilitate adaptation by increasing the probability that adaptive and maladaptive alleles will be decoupled or that adaptive alleles will be brought together in the same genotype (i.e., overcome Hill-Robertson interference [9]). Increased rates of recombination are similarly favored when fitness optima change rapidly between generations, e.g., under fluctuating selection [10]. On the other hand, lower recombination rates can be favored under scenarios in which adaptive combinations of alleles are at risk of being broken apart, such as under maladaptive gene flow [11]. Reduction/suppression also appears to have important consequences for the evolution of reproductive isolation [11,12] and patterns of introgression and divergence in the genome [8,13,14].

While there is a rich theoretical literature focused on the evolution of recombination rate, empirical studies have lagged somewhat behind. One reason for this may be that recombination rate is difficult to quantify directly – it generally requires the construction of a linkage map from a genetic cross and/or cytological visualization of recombination-associated proteins [2,15,16]. Recently, many studies have attempted to overcome this difficulty by instead estimating a population genetic quantity known as ρ, the *population scaled recombination rate [17]*. This quantity is the product of four times the effective population size and realized recombination rate (sometimes denoted "c") [18]. The general approach to estimating ρ is to perform coalescent simulations and fit a simulated value of ρ to observed patterns of linkage disequilibrium (LD) [19–21]. While this approach has proven successful at recapitulating many of the general features of the recombination landscape in many species, it is not able to disentangle changes in LD *per se* (e.g. as a result of selection or demography) from changes in recombination rate (either locally or genome-wide) [21,22]. Further, these methods are highly sensitive to increases in LD that occur as a result of gene flow between populations [22–24]. As such, LD-based methods are likely to be less appropriate for the study of the evolution of recombination rate than direct estimates of recombination rate.

In spite of methodological difficulties, there has been a recent resurgence of interest in the empirical study of the evolutionary causes and consequences of recombination rate [2,4,25]. One key contributor to this resurgence has been the democratization of high throughput genotyping, which has increased the tractability of creating high density linkage maps in non-model species (e.g. using pedigreed populations or gametic sequencing, [26,27]). The increased availability of such linkage maps has in turn led to a growing appreciation of the enormous diversity in recombination

rate that exists between taxa [25]. This variation can manifest globally, i.e. genome-wide, or locally, i.e. along a specific tract of a chromosome [25,28].

Studies using direct estimates of recombination rate have largely focused on describing differences in recombination between species or sexes [25,29,30]. However, there are surprisingly few studies focused on directly testing *evolutionary* hypotheses concerning variation in recombination rate. For example, a key question that emerges from the theoretical literature is: is variation in recombination rate shaped by natural selection [5,10,31]? While a tempting research direction, the difficulty in measuring and manipulating recombination rate makes testing adaptive hypothesis a non-trivial enterprise [2,4]. One approach may be experimental evolution, in which the proposed selective agent that favors/disfavors changes in recombination rate is experimentally varied, evolved differences in recombination rate are quantified, and these differences are then compared to a null (non-adaptive) expectation [32]. This approach is powerful but highly laborious and difficult to apply to natural systems. A more broadly applicable method for detecting the influence of natural selection on a quantitative trait is perhaps the Q_{ST}-F_{ST} approach [33]. Originating in the quantitative genetics literature, this powerful method is designed to answer the question: are the observed differences between populations in a quantitative trait greater than expected on the basis of drift alone [33,34]? This question is formalized as a statistical hypothesis test that compares variation in a quantitative trait (Q_{ST}) within and between populations to a null distribution of variation in neutral genetic markers (F_{ST}) within and between populations [34,35]. While the Q_{ST}-F_{ST} is subject to many of the same limitations and assumptions as other methods for studying natural selection in the wild it is also has a number of advantages, including the ability to detect very recent natural selection and robustness to a variety of common demographic perturbations (e.g. changes in population size or levels of migration). While the Q_{ST}-F_{ST} method has enjoyed great success in the quantitative and evolutionary genetics literature, it has not yet been applied to testing the role of selection in shaping recombination rate. Given its flexibility and applicability to any quantitative trait, we see Q_{ST}-F_{ST}as an ideal approach to this problem.

Along with quantifying intraspecific variation and the role of natural selection, we also have a poor understanding of the genetic basis of differences in recombination rate between populations and species. As is the case for other traits, identifying the genetic architecture of evolutionary changes in recombination rate allows for a more complete explanation for how and why recombination rate evolves [36]. One specific question is the degree to which variation in recombination rate manifests as a local vs. global phenomenon. Local variation in recombination can arise due to structural

variants that suppress recombination such as inversions and large deletions [37–39]. In contrast, global variation can arise from mutations in the genes involved in meiosis and/or double-strand break repair pathways [40]. Modifiers of both global and local rates of recombination have been identified in laboratory and/or interspecific crosses, but their occurrence in natural populations of individual species is only just beginning to be explored [26,29,40–42].

Here, we performed a comprehensive study of recombination rate (meiotic rates of crossover) in two natural populations of *Drosophila pseudoobscura* from Utah and Arizona, USA. We made use of modern sequencing and genetic map construction methods, along with the Q_{ST}-F_{ST} approach. We first constructed individual-level genetic maps and discovered ample quantitative genetic variation for recombination rate within and between populations of *D. pseudoobscura*. Interestingly, we found that this variation largely manifested as differences in genome-wide recombination rate rather than remodeling of the local recombination landscape. Interindividual differences in local genome structure (e.g. structural variation) did not appear to influence recombination rate at the scale of measurement, again suggesting that variation in recombination rate is largely governed by global modifiers. Finally, using the Q_{ST}-F_{ST} approach, we discovered that between-population differences in recombination rate are much greater than expected under a pure-drift model, suggesting that natural selection may have shaped recombination rate variation in *D. pseudoobscura*. Together, these results provide direct evidence for genetic variation in global modifiers of recombination and support the hypothesis that natural selection can and does act to shape recombination rate in natural populations.

Results

Genome wide recombination rate varies within and between populations

Genome-wide recombination rate varied significantly within and between the *D. pseudoobscura* populations we studied. Within lines, there was a range of 4.27–5.86 crossovers per genome, corresponding to 0.85–1.00 crossovers per chromosome arm on average (Figure 1A). This between-line variation was statistically significant ($p < 2.2\times10^{-16}$, Likelihood Ratio Test Statistic= 141.13, df=1, comparison via dropping random effect of inbred line). At the population level, lines from American Fork Canyon, UT had 5.20 ± 0.17 crossovers per genome on average, while lines from Madera Canyon, AZ had 4.82 ± 0.21 crossovers per genome on average, a significant difference

in genome-wide crossover rate (Figure 1B, Type II Wald Test, p=0.018, df=1; Likelihood Ratio Test, χ^2_1=4.794, p=0.028).

That said, despite genome-wide differences, the local rates of recombination were extremely similar among individuals and populations (Figure 1C & Figure 2A, R^2 = 0.96, correlation test t = 68.866, df = 207, < 2.2×10^{-16}). Indeed, in contrast to the aggregate genome wide difference we observed in Figure 1B, only 19 of the 209 recombination intervals we assayed displayed significant population-specific differences at the α = 0.05 level, and none were significant after FDR correction (Figure S4). That said, some recombination intervals did show a significant effect of inbred line identity (Figure S5) suggesting that there may be genetic variation for local recombination rates at the 200-400kb scale. Finally, we found that chromosome-scale recombination rates were highly correlated *within* lines, such that there was a strong trend that lines with high recombination rate on one chromosome tended to also have high recombination on other chromosomes (Figure 2B; average R^2 = 0.78, all correlations significant via correlation tests, p <0.0001). In sum, these results suggest that phenotypic variation in recombination rate within and between populations largely manifests at the genome-wide scale. That said, our marker density prevents us from ruling out finer-scale population-level differences in the recombination landscape (i.e. at the <200kb scale).

Population differences in recombination rate are greater than expected under neutrality

As expected from previous studies, genetic divergence between Madera Canyon, AZ and American Fork Canyon, UT was very low: genome-wide Weir and Cockerham's F_{ST} was approximately 0.0039 (Figure 3A, mean F_{ST} of 6 591 high quality SNPs, MAF > 0.1, LD > 0.2; F_{ST} computed using WGS from inbred lines was highly similar). Examining variation in recombination rate, we estimated a within population (between line) variance component of 0.066 and a between population variance component of 0.018, yielding an observed Q_{ST} of 0.212 (Figure 3A, dashed arrow). Our parametric bootstrap simulations of Q^n_{ST} suggest that this value of $Q_{ST}$is highly unlikely to be observed under neutrality (0 of 10,000 Q^n_{ST} replicates were greater than the observed value of Q_{ST}, thus p < 1.0×10^{-6}). Similarly, the parametric bootstrap estimates of Q_{ST}-F_{ST} under neutrality do not overlap with the parametric bootstrap observed values of Q_{ST}-F_{ST} , even when taking into account sampling variance (Figure 3B). Together, these results indicate that while the observed phenotypic difference in recombination rate between MC and AFC is modest, it greatly exceeds its expected value under neutrality. This result is consistent with the hypothesis that natural selection has driven the observed difference in recombination rates between populations.

Nonsynonymous differences in meiosis genes are correlated with recombination rate

Of the 46 candidate genes examined, 33 had at least one non-synonymous polymorphism. Of these 33 genes, there were a total of 357 codons (out of a total of 29 964) with at least one non-synonymous polymorphism. After controlling for multiple comparisons three of these sites in two genes (*asp* and *mei-41*) were significantly associated with crossover rate (FDR adjusted p-value < 0.05, Figure 4A). Both *asp* and *mei-41*play key roles in meiosis and recombination: *asp* is involved in spindle pole formation during cell division (both mitotic and meiotic) whereas *mei-41* (also known as ATR) is an important regulator of double strand break repair and meiosis checkpoint activation [43,44]. Homozygous, nonsynonymous polymorphisms in these genes were associated with a 5%-7% difference in recombination rate between lines (Figure 4B). There was, however, strong LD (r^2 > 0.8) between these alleles (e.g. lines with the lowest averaged crossover rates shared genotypic states for all three genes), and thus disentangling their independent effects on recombination rate was not possible. We also note that the small number of lines examined here precluded more powerful association methods (e.g. full GWAS) and further work will be required to experimentally validate the contribution of these genes to variation in recombination rate.

Structural variation does not explain differences in recombination rate

Both short and long-read sequencing revealed extensive structural variation between inbred lines of *D. pseudoobscura*. As expected, the three strategies we used to detect structural variation (GATK INDELs, PacBio SV and LUMPY/Smoove) varied in the number and relative proportions of the various classes of structural variant they identified (Figure S6). That said, all three methods suggested that the most common form of structural variation are small to mid-sized (10-100bp) INDELs, with larger deletions, insertions, and duplications being much rarer (Figure S6). Consistent with the observation that AFC and MC are highly similar in their chromosomal arrangements, our structural variant analysis found no evidence of large-scale chromosomal inversions differentiating any of the lines.

Structural variation between lines did not co-vary with recombination rate (Figure 5). First, there was no relationship between recombination rate and the estimated percent sequence homology between the tester and inbred lines (Figure 5B, likelihood ratio test comparison of GLMMs , df = 3, p = 0.3989). Second, there was no relationship between recombination rate and the count of differences in structural alleles between each inbred line and the tester line (Figure 5A, likelihood ratio test comparison of GLMMs , df = 3, p = 0.7617). This result was consistent across all methods

used to identify structural variation (likelihood ratio tests, comparison of GLMMs with and without method by count/homology interaction effects, all $p > 0.3$). As such, at the 300kb scale, there is no evidence that the local differences in recombination rate among inbred lines are a result of differences in homology or local genome structure.

Discussion

Recombination rate is a key modulator of many evolutionary processes, yet we have a poor understanding of how recombination rate itself evolves. Here, we studied how recombination rate varies using strains from two natural populations of *D. pseudoobscura* from Madera Canyon, AZ and American Fork Canyon, UT. We directly measured recombination rate in a total of 17 inbred lines from these populations and found substantial variation for recombination rate both within and between populations. Interestingly, the population from Madera Canyon, AZ exhibited an ~8% lower recombination rate on average than the population from American Fork Canyon, UT. Within and between-population variation in recombination rate manifested largely as differences in genome-wide recombination rate, rather than changes in the local recombination landscape. This finding is supported by a general pattern of covariation in recombination rate among chromosomes within lines. That said, our choice to assay greater numbers of individuals in fewer genomic intervals prevents us from ruling out the possibility of finer-scale differences in the recombination landscape between populations and lines. While overall differences in recombination rates between populations were modest in absolute terms (~8% depending on the interval), a Q_{ST}-F_{ST} analysis revealed that this difference vastly exceeds the amount of phenotypic divergence expected under neutral drift. This result is consistent with the hypothesis that local adaptation has driven differences in recombination rate between these populations.

We explored two possible mechanisms underlying recombination rate differences between lines. First, we found evidence that some differences in recombination rate between lines may involve non-synonymous coding changes in meiosis-related genes. Secondly, we found that local variation in recombination rate between lines does not correlate with local structural variation at the 300kb scale. These findings suggest that the differences in recombination we observed were driven by alleles resulting in genome-wide changes in recombination rate rather than local remodeling of the recombination landscape. Below, we discuss the relevance of our findings for the study of the evolution of recombination rate and relationships to previous work.

Recombination rate variation in natural populations

Previous work has shown that recombination can vary between individuals, or between populations/species [25,26,45–47]. These studies have ranged from early work on chiasma frequency in snails [48] to more recent work leveraging modern human population genomic data [49,50]. The bulk of this work has focused on describing variation in recombination and its potential molecular correlates. Further, most studies of natural populations have measured recombination in uncontrolled environments (e.g. in the wild, [25]. Our study contributes to this literature directly examining genetic variation for recombination rate both within and between natural populations of a single species and performing one of the first tests that this variation is shaped by natural selection. Together with previous work, our study contributes to a growing body of evidence that there is ample genetic variation for recombination rate in natural populations, and that recombination rate is actively evolving on observable timescales.

Secondly, we found that recombination rate varies primarily at the genome-wide scale rather than via variation in specific genomic regions. Our candidate gene analysis suggests that this variation in genome-wide recombination rate may be the result of allelic variation in meiosis-related genes (i.e. *asp* and *mei-41*). This is in line with previous work connecting genetic variation in genes regulating meiosis and/or crossover formation to variation in variation in genome-wide recombination rate [29,47,49,51,52]. The emerging evidence for natural variation in gene-wide modifiers of recombination is particularly intriguing given that many theoretical models of recombination evolution make use of abstract "modifier" alleles that alter genome wide rates of recombination [5,6]. Further characterization of such modifiers in natural populations may eventually allow direct tests of theoretical models of recombination evolution [2].

Local adaptation of recombination rate

Our Q_{ST}-F_{ST} analysis suggests that differences in recombination rate between *Drosophila pseudoobscura* populations from AZ and UT may have been driven by natural selection. To our knowledge, this is the first application of the Q_{ST}-F_{ST} method to the study of recombination, and among the first evidence for the role of selection acting on genome-wide recombination rate in natural populations [40]. However, while our results suggest a role for natural selection, the agent of selection underlying this change remains unknown. There are a wide variety of possible explanations for this difference [2]. For example, differences in recombination between the

populations may be directly favored, or other phenotypic differences may be divergently selected between the populations that incidentally affect recombination rate (via linkage or pleiotropy). One intriguing possibility is local differences in climate: recombination rate in *Drosophila* is known to be plastic with respect to ambient temperature [53]. Madera Canyon, Arizona has a mean annual temperature of approximately eleven degrees Celsius higher than American Fork Canyon, Utah (10.5°C vs 21.6°C, [54]). Assuming that the temperature reaction norm is similar in both populations, this higher temperature could, for example, cause an increase in realized recombination rate in the Madera Canyon population in the wild. We speculate that the difference in recombination rate we observed under constant conditions may be a compensatory response to an environmentally-induced increase in recombination rate in order to return genome-wide recombination rate to some optimum value (i.e., a response to maladaptive plasticity, [55]). Further work will naturally be needed to connect variation in recombination rates to specific agents of selection. One obvious extension of our approach would be a greater number of populations, perhaps existing over a climatic gradient (or paired populations in differing environments). We hope that our demonstration of the efficacy of the Q_{ST}-F_{ST} method inspires the undertaking of such eco-evolutionary studies of recombination rate.

One caveat regarding our application of the Q_{ST}-F_{ST} method is that our estimates of recombination come from F1s, and we were thus only able to observe genetic variation underlain by dominant and co-dominant effects. This is not ideal, as it potentially alters the distribution of Q_{ST} relative to F_{ST}, which could bias the outcome of the Q_{ST}-F_{ST} test [35]. A dedicated simulation study aimed at understanding the direction and magnitude of this bias would be of great utility for future work on recombination using inbred lines.

Structural variation as a modulator of recombination rate

We found no association between among-line variation in recombination rate and among-line variation in the abundance or size of structural variants. An important consideration here is that this analysis was not intended to test whether *average* recombination rate (across all lines) is associated with structural variation – this association is extremely well documented and is unquestionably present in our data [56–58]. Instead, our goal was to test if among-line variation in recombination rate in each genomic interval was explained by among-line structural differences, using normalized metrics of both recombination rate and structural variation within each genomic interval (as *Z*-scores, i.e. statistical controlling for average recombination rate).

Why was there no detectable association between structural variation and local rates of recombination? For one, our F1 cross design is not able to detect recessive-acting effects of structural variation (e.g. those that only affect recombination in homozygous form). Secondly, a key consideration in interpreting these results is the *scale* of our recombination estimates: much of the previous work describing the effects of heterozygous structural variation on crossing-over was performed at much finer scale, e.g. <1kb in *Arabidopsis* [59]. It may be that changes in recombination resulting from structural variation are restricted to finer genomic scales (i.e. <300kb) and that other types of regulators (e.g. variation in meiosis genes or the chromatin landscape) modulate recombination at larger scale [40]. A notable exception to this is large scale chromosomal inversions (notably absent in our lines), which are well known to affect recombination at scales much larger than 300kb – upwards of 10Mb in many cases [60,61]. However, inversions likely have outsized recombination suppressing effects compared to other forms of non-homology because of the loop structures they form during chromosome pairing [60,62]. Further work will be required to disentangle the relative contribution of structural and global/trans modifiers of recombination rate at different genomic scales.

Amplicon sequencing as a tool for genetic maps

Our ability to economically sequence hundreds of markers in thousands of individuals was made possible by the GT-seq amplicon sequencing approach [63]. This technique is highly scalable, and in our case, we likely could have sequenced many more markers (and/or individuals) while maintaining a very high depth per amplicon. This method is an alternative to the increasingly popular bulk-sequencing approaches, in which sample DNA is pooled prior to sequencing [64]. GT-seq avoids some of the complexity of these approaches. For one, because it is a PCR-based method, GT-seq does not require performing extraction, quantification and manual normalization of sample DNA. This is a non-trivial consideration when individual sample sizes are in the thousands. Further, unlike bulk-sequencing, amplicon sequencing provides individual-level genotypes. As such, the occurrence of double/triple/etc. crossovers can be directly resolved, and problematic individuals identified and removed during analyses. To our knowledge, these are both not currently possible with bulk sequencing (unless barcodes are employed, limiting the total number of individuals in the pool). The main drawbacks of amplicon sequencing are a decrease in resolution (number of markers), and the need to pre-identify mapping informative markers. That said, we believe GT-seq and amplicon sequencing more generally will be a useful tool for future studies of variation in

recombination rate and can be readily paired with other approaches depending on the goals of the study.

Conclusion

Recombination rate plays an important modulatory role in many evolutionary processes, but little is known about how recombination rate itself evolves. Here, we studied natural variation in recombination rate within and between two populations of *Drosophila pseudoobscura*. We found extensive genetic variation for recombination rate within and between populations, with the majority of variation detected manifesting as differences in overall genome-wide recombination rate. This suggests that the differences in recombination we detected between lines may be the result of genetic variation in trans-acting global regulators of recombination, an idea supported by a significant association between non-synonymous variation in meiosis-associated genes and recombination rate. We also found no evidence that among-line differences in local recombination rate at the 300kb scale were correlated with structural variation within the lines. Finally, we discovered that the magnitude of phenotypic difference in recombination rate between the two populations was far greater than expected under a model of neutral trait evolution, suggesting that the differences may have been driven by natural selection. Our study provides key insights in the quantitative genetics of recombination rate and lays the groundwork for future research focused on studying the recombination rate in natural populations.

Acknowledgements

Funding for this project was provided by National Science Foundation grants DEB-1545627, 1754022, and 1754439 to MAFN. KMS was additionally supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. The tester line (MV2-25) was generously provided by Dr. Steven Schaeffer. We thank Dr. Noah Whiteman and Dr. John Chaston for assistance during fieldwork. Dr. Katharine Korunes, Dr. Jenn Coughlan, members of the Noor Lab and three anonymous reviewers provided valuable feedback on the manuscript. Dr. Andrew MacDonald provided guidance on statistical analyses. Dr. Armin Töpfer provided advice on the use of the pbsv software.

Author Contributions

Conceptualization, MAFN and KS; Methodology KS, BMW, KRR and MAFN; Software KS, Formal Analysis KS; Investigation KS, BMW and KRR; Resources MAFN and BMW; Data Curation, KS; Writing – Original Draft, KS. Writing – Review & Editing, KS, MAFN, BMW, KRR. Visualization, KS; Supervision, MAFN; Project Administration, KS and BMW; Funding Acquisition, MAFN and KS.

Declaration of Interests

The authors declare no competing interests.

Supplemental Information

Document S1; Supplemental Figures S1-6.
Table S1 - Per-chromosome counts of crossover events for all lines.
Table S2 - Fine scale estimates of crossover location for all lines.
Table S3 - List of *D. melanogaster* candidate genes and *D. pseudoobscura* homologues.
Table S4 - List of primers used in GT-seq library preparation.

Figure Legends (Main Text)

Figure 1 | Recombination rate varies within and between populations of *D. pseudoobscura*. (A) Variation in genome-wide crossing over frequency for 17 inbred lines. Lines are colored according to their population of origin (Green, MC: Madera Canyon, AZ, Red, AFC: American Fork Canyon, UT.). Points depict the mean crossover frequency for each line with vertical lines representing 95% confidence intervals (n = 384 per line). (B) Differences in crossover frequency between AFC and MC. Jittered points are individual line means (from A), and larger points are marginal means derived from mixed model regression coefficients along with 95% confidence intervals (error bars). (C) Variation in recombination rate across the genome. Each panel depicts recombination rate along a single chromosome arm (columns) in one of two populations (rows). Thick lines depict population average recombination rates, with lighter lines depicting rates for individual inbred lines. Note that in *D. pseudoobscura* the X chromosome takes the place of a chromosome "1". See Figure S3 for an example of GLMM model fit diagnostics for this and other statistical comparisons.

Figure 2 | Recombination rate varies primarily at the genome-wide scale. (A) The correlation between recombination rate measured in genomic windows (~300kb in size) in the MC and AFC populations. Each dot depicts a single genomic window (all chromosomes combined). (B) The correlation between chromosome-wide mean recombination rate between all pairs of chromosomes. Each point represents the recombination rate on two chromosomes for a single inbred line. Points and lines are colored to indicate the particular pair of chromosomes being

compared. Positive trends indicate that recombination rates are consistent across chromosomes within lines (i.e. they vary genome-wide, and not idiosyncratically across chromosomes). See Figure S2 for comparisons of marker orders and recombination fractions between line-specific genetic maps, and S4 and S5 for more detailed analyses of local variation in recombination rate.

Figure 3 | Recombination rate Q_{ST}-F_{ST} exceeds neutral expectations. (A) Weir and Cockerham's F_{ST} from 6591 RADseq-derived SNPs (mean F_{ST} = 0.0039). The observed value of Q_{ST} for recombination rate (0.212) is indicated with an arrow. (B) Comparisons of the sampling distribution of Q_{ST}-F_{ST} expected under neutrality (green histogram) and the observed value (yellow histogram). Both distributions were simulated via a parametric bootstrap (see text). Black points with error bars indicate the mean and 95% confidence interval of the sampling distributions.

Figure 4 | Non-synonymous substitutions associated with variation in recombination rate. (A) Regression coefficients from linear models (y-axis) comparing genotype and crossover rate for sites (points) bearing non-synonymous, non-reference polymorphisms in a collection of meiosis-related candidate genes (x-axis). Red points indicate associations that were significant after adjustment via FDR correction (adjusted p-value < 0.05). (B) Mean recombination rates (crossovers per chromosome arm) for sites with significant associations (red points in A). Each panel depicts the mean and 95% confidence interval for crossover rates for each genotypic class (either homozygous reference or homozygous non-synonymous derived). cs.

Figure 5 | Structural variation is not correlated with recombination rate at the 300kb scale. (A) The relationship between normalized recombination rate and the normalized count of structural differences between each inbred line and the tester line. Each point represents a single recombination interval (all approximately 300kb in length) from one inbred line. Lines on each plot represent smoothed conditional means and are accompanied by 95% confidence intervals. Each column depicts the relationship using each of the three methods used to assay structural variation. (B) The relationship between normalized recombination rate and the difference in total sequence length between each inbred line and the tester line. See Figure S6 for a detailed summary of the frequency and size of different classes of structural variation.

Figure 6 | Schematic of the crossing design and one method of interfering crossovers. (a) Isolines from MC and AFC were individually crossed to tester lines to generate F1s, which were subsequently crossed to a "donor line" sharing the same genotype as all isolines, but a different genotype than the tester line at all marker loci. Further, all markers were selected such that only two alleles were found in all lines, with the tester line having one allele ("1") and all other lines including the donor line having the other ("0"). This allows for the scoring of crossovers as changes in heterozygosity, as shown in (b). (c) Example genotypic data from one chromosome showing the number of inferred crossovers. White genotype states indicate missing data. See also Figure S1 for details on the performance of GT-seq .

STAR★Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and protocols should be directed to and will be fulfilled by the Lead Contact, Dr. Kieran Samuk (ksamuk@gmail.com). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We collected wild male and female *Drosophila pseudoobscura* from Madera Canyon, AZ, USA (31°42'48.9"N, 110°52'22.4"W) and American Fork Canyon, UT, USA (40°26'38.9"N, 111°42'08.5"W) in May and July of 2015 respectively using bucket traps [65]. These populations were chosen because they were known to share similar karyotypic configurations (e.g. inversions) but also differ in their ecological context (i.e. xeric vs. sub-alpine). We returned live individuals to the laboratory, isolated females, and created inbred lines from their offspring (one line per surviving female). These lines were created by successive crosses between virgin siblings for a minimum of 14 generations. The inbred lines (and all subsequent lines) were reared in 20C incubators with 65% relative humidity and photoperiods of 14D:10N. The inbreeding process resulted in a total of 7 inbred lines from Arizona and 12 from Utah.

METHOD DETAILS

RAD-seq libraries from wild samples

To generate a set of SNPs for estimating F_{ST} between the Utah and Arizona populations, we performed double-digest RAD-seq reduced representation sequencing. To begin, we extracted DNA from single wild-caught individuals (excluding the females used to initiate the inbred lines) via phenol-chloroform DNA extraction. We then performed a RAD-seq library preparation protocol after [66]. The resulting libraries were sequenced in a single lane on an Illumina HiSeq 4000 at the Duke Center for Genomic and Computational Biology sequencing facility.

Whole genome sequencing of inbred lines

We performed both short read and long read whole genome sequencing on all 17 inbred lines, as well as our testers line (MV2-25 and Flagstaff-14). The short read libraries were prepared by first performing phenol-chloroform DNA extractions from pools of 20-30 individual female flies. We quantified DNA purity and concentration via Nanodrop (Thermofisher Inc.) and Qubit (Qiagen Inc.). The DNA samples were then submitted for library preparation and sequencing via Illumina NovaSeq (300-400bp insert, 150bp paired end reads) at the Duke Center for Genomic and Computational Biology sequencing facility.

The long-read libraries were prepared by first performing high-molecular weight DNA extractions from pools of 20-30 female flies using Qiagen Midi/Mini Prep DNA extraction kits (Qiagen Inc.). These were then assessed for fragment size via standard gel electrophoresis and submitted for sequencing on a PacBio Sequel (4 SMRT cells, 4-5 samples multiplexed per cell) at the Duke Center for Genomic and Computational Biology sequencing facility.

Whole genome variant calling: short read WGS and RAD-seq data

We identified variants in the short read data (both isoline whole genome sequencing and wild population RAD-seq) using an analysis pipeline based on the GATK best practices [67,68]. The complete code for this pipeline is available as a Github repository at http://github.com/ksamuk/samuk_et_al_curr_biol_2020. All tools were run with default settings unless otherwise indicated. Briefly, we aligned the reads for each sample to the *D. pseudoobscura* reference genome (version 3.04 from FlyBase, ftp://ftp.flybase.net/genomes/Drosophila_pseudoobscura/) using bwa mem version 0.7.17 [69]. We marked adapters and duplicates using PicardTools [70], and performed individual-level genotyping for each set of marked reads using the HaplotypeCaller. We then performed joint genotyping on the resulting set of GVCFs via GenotypeGVCFs. We filtered SNPs in the resulting VCF using the GATK Best Practices hard filters (see scripts for details), working in R 3.4.1 [71] with the vcfR and tidyverse packages [72,73].

Creation of mapping populations

To estimate variation in crossover rate in our inbred lines, we created backcross-like mapping populations (crossing scheme shown in Figure 6). We crossed groups of 3-5 males from each isoline to single virgin females from the *D. pseudoobscura* reference genome isoline (MV2-25, provided by

Dr. Steve Schaeffer). We then allowed the F_1 offspring to develop and collected virgin females from the resulting offspring. Finally, we crossed these virgin F_1 females to males from a second fixed isoline, Flagstaff-14 (a highly inbred isoline from Flagstaff, AZ). This resulted in a backcross-like mapping population for each of the AZ and UT lines, in which all BC1 offspring had one maternal chromosome from their F_1 mother and one paternal Flagstaff-14 chromosome with a fixed, known genotype (Figure 6A). This design allows for straightforward mapping of recombination events that occurred in F_1 females. As such, our estimates are unable to detect any variation in recombination due to recessive-acting effects and may underestimate total recombination rates (e.g. from modifiers that act in an additive fashion) in the pure inbred lines. Critically, this potential underestimation is identical across all F_1 families, and thus cannot (in and of itself) generate systematic differences in recombination rate between lines or populations.

Genotyping of mapping populations

Because our goal was to quantify the number of crossovers per generation rather than their precise location, we performed low density, genome-wide SNP genotyping using an amplicon sequencing approach. To do this, we adapted the 'GT-seq' method outlined in [63]. A summary of the design and performance of this method is depicted in Figure S1. To begin, we identified SNPs genotyped in the whole genome dataset that were unique to the MV2-25 isoline (i.e. fixed for one allele in all 19 inbred lines and Flagstaff-14 and fixed for another allele in MV2-25). Genotyping these markers in BC1 individuals allows the recovery of genotypic phase simply by examining the genotype of the marker SNPs – regions with UT or AZ ancestry are represented as runs of heterozygous SNPs and regions with MV2-25 ancestry are represented as runs of homozygous SNPs (see diagram in Figure 6B). In total, we selected 500 of these SNPs evenly spaced at approximately 300kb intervals along each chromosome (Figure S1 A & B). Note that this choice of marker density is optimized to detect small differences in *genome wide* recombination rate and cannot completely resolve fine scale (i.e. <300kb) variation in the recombination landscape.

We designed primer pairs to generate ~200-300bp amplicons containing each of our target SNPs. These primer pairs were optimized to minimize primer-primer interactions during multiplex PCR (primer design service provided by GT-Seek Ltd., Idaho, USA). With these primers in hand, we performed two test library preps using the GT-seq protocol described in Campbell et al. (2015). We sequenced the first test library on a MiSeq (V3 flow cell, Illumina Corp., California, USA), and identified poorly performing amplicons using the criteria outlined in Campbell et al. (2015), i.e. high

503 dropout, low representation among individuals, evidence of amplicons mapping to duplicate
504 regions, etc. (service provided by GT-Seek LTD, Idaho, USA). We then prepared a second test library
505 with the primers for the poor-performing amplicons omitted and sequenced it as above. A final
506 screen for poor-performing amplicons resulted in a final set of 390 amplicons ranging from 200-
507 300bp, each containing at least one recombination-informative SNP.

508 After optimizing our panel of amplicons, we used GT-seq to genotype approximately 400 BC1
509 offspring from each mapping population (400 individuals from each of 19 lines, a total of
510 approximately 7600 individuals). We created two pools of 40 plates (individuals and plates are
511 individually barcoded as part of GT-seq library preparation) and submitted these for sequencing on
512 an Illumina NextSeq 500 (1st pool: High Output Reagent 150 PE Reagent Kit, 2nd Pool: Mid Output 150
513 PE Kit, Illumina Corp., California) at the Duke Center for Genomic and Computational Biology
514 sequencing facility. We called SNPs in our sequenced GT-seq amplicons using an identical approach
515 to our whole genome short read data. The final dataset contained 679 total variants across all
516 amplicons, sequenced to an average depth of ~200X (Figure S1 C). While there was some variability
517 in sequencing depth between amplicons (mean coefficient of variation for depth of amplicon
518 sequence was ~0.75), the overall high depth of sequencing resulted in the vast majority of
519 amplicons having >100X coverage (Figure S1 C). We performed further quality control on the
520 resulting SNPs in R using the vcfR and tidyverse packages [72,73]. First, we dropped any markers
521 that mapped to genomic locations outside our original targeted amplicons. Next, we dropped any
522 individuals that had an average depth below 10X (19/7600 individuals). Finally, we removed any
523 markers that displayed any evidence that they were in fact not unique to the tester line. This was
524 done by removing markers displaying: (1) any evidence of segregation distortion, (2) any evidence
525 that any of the isolines were in fact polymorphic for the marker or (3) high dropout (i.e.
526 represented in fewer than 75% of samples). In some cases, the source of marker dropout was
527 clearly an undetected INDEL polymorphism in the amplified regions, which, for consistency among
528 lines, we erred on the side of removing rather than recoding as them as markers for mapping. The
529 final set contained 344 mapping-informative SNPs. After filtering, we recoded all SNP genotypes as
530 '0' for the isoline/donor line state and '1' for the tester line state. Because of the backcross design,
531 the only possible genotypes were thus '0/0' and '0/1'.

Detection of recombination events

We identified crossovers in two steps: (1) ancestry assignment of chromosome segments and (2) crossover counting. To begin, we updated the genomic ordering of our markers using the genomic scaffold ordering from [74]. Note that this reordering results in movement and replacement of contigs between chromosomes, and as such overall physical lengths of the reordered chromosomes are different from that of the most current *D. pseudoobscura* reference (version 3.04). After markers had been reordered, we assigned the ancestry (isoline or tester) of chromosomal segments by identifying runs of 0/0s and 0/1s. In regions with a single ancestry assignment, we imputed (via parsimony) across gaps of missing markers (e.g. due to filtration or dropout) shorter than 2 markers (~400kb). After local ancestry was assigned, we counted crossovers by counting the number of ancestry changes (from 0/0 to 0/1) along each chromosome in each individual using the function countXO in R/qtl [52]. Following the recommendations in [75], we ignored double crossovers spanning less than 2 markers (~400kb) and/or individuals displaying more than four crossovers on a single chromosome: crossover interference should make close range double crossovers exceedingly rare, and thus these cases likely represent genotyping or marker-order errors. It is also worth noting that our method of crossover detecting relies on quantifying crossover events in live-born offspring. As such, any extreme changes in crossover patterns incompatible with proper chromosome segregation during meiosis will not be observed (i.e. because they are lethal or lead to gamete degradation).

This crossover counting method assumes that the order of markers on each chromosome is identical in each line. Differences in marker order could, for example, generate spurious double crossovers (although ignoring short double crossovers reduces this problem). To directly address the possibility of different marker orders among lines, we created separate genetic maps for each isoline using the R packages r/QTL and ASMap [75,76]. Following the general recommendations from the documentation, these two packages agnostically infer linkage group assignment, marker order, and genetic distances between markers. Overall, there was high concordance in marker order between all the individually-inferred maps (Figure S2). Individual recombination rate estimates within each line were highly similar when using the reference genome marker order or individually-inferred marker orders (Figure S2, Spearman rank correlation = 0.93, $p < 2.2\times10^{-16}$). We thus elected to use the reference genome marker order (reordered based on [74]) for all subsequent analyses. Individual estimates of crossover events are provided in Table S1 and Table S2.

Candidate genes associated with recombination differences

We explored the possibility that between-line variation in meiosis-related candidate genes may underlie between-line differences in recombination rate. We were specifically interested in the hypothesis that coding changes in meiosis genes underlie any differences in recombination rate between inbred lines (and act dominantly or additively in the F1s). To do this, we first assembled a list of candidate genes from Anderson et al. 2009 and Hunter et al. 2016 [29,77] (full list in Table S3). We then obtained the FASTA sequences for these genes in each line by intersecting the short read variant calls (including INDELs) with the *D. pseudoobscura* reference genome CDS for each candidate gene. To ensure proper alignment, we then performed multiple alignment of the line-level FASTA sequences and the reference CDS using MAFFT version 7.407 [78]. Once the sequences had been aligned, we identified non-synonymous, non-reference alleles in each line.

Association between local structural variation and recombination rate

Along with the candidate gene approach to examine associations with genome-wide recombination rate, we also investigated the possibility that small-scale differences in genomic structure between the inbred lines may explain differences in recombination rate. This may be of particular importance given that our design required measuring recombination rate in F_1 individuals (inbred line × tester line), and that structural heterozygosity has a well-known negative association with recombination rate [37,60,79].

To test if differences in genome structure underlie local differences in recombination rate in our inbred lines, we first identified structural variants (SVs) using two approaches. First, we used the SVtools pipeline [80] to identify SVs using paired-end short read data. This pipeline identifies structural variation using a variety of genomic signatures, particularly split reads (different parts of a single read mapping to multiple discrete locations) and discordant reads (paired end reads separated by a much greater genomic distance than expected on the basis of their insert size). SVtools can identify insertions, deletions, inversions, duplications, and other classes of rearrangements. The general procedure is to identify split/discordant reads using the tools *SAMBAMBA* and *SAMBLASTER*, which are then analyzed and annotated with the SVtools variant callers [81,82]. The resulting structural variant VCF was filtered via empirical cut offs using the guidelines in [80]. Along with SVtools, we separately identified structural variation in the PacBio long reads dataset using the PacBio structural variant pipeline and tools, *pbsv* (https://github.com/PacificBiosciences/pbsv, see also [83]). This involves aligning the long reads

with *minimap2* (accessed via the *pbmm2* wrapper), identifying individual signatures of structural variation using pbsv, and jointly calling structural variation from the combined set of signatures. This again results in a VCF containing structural variants, which we filtered using empirical cutoffs as before.

After identifying structural variants, we next quantified the total difference in sequence homology between each line and the tester line (MV2-25) for each genomic interval where recombination was measured (~300kb windows). To do this, we summed the total number of non-shared, non-reference base pairs between each line and the tester line. We included SNPs, inversions, insertions, deletions, and translocations in this calculation. This method collapses multiple classes of genomic variation into a single, consistent metric and avoids the ambiguity associated with identifying shared locations of breakpoints for the structural variants (e.g. needed for per-variant associations). Further, this method focuses on the most likely biological cause of structurally-mediated recombination suppression, i.e. differences in homology *per se*, which has been widely demonstrated in many species [84–86]. We also tabulated the total count of structural variant alleles (of any type) that differed between each isoline and the tester line for each recombination interval. We normalized all homology estimates and structural variant counts in each window using both the total number of genotyped base pairs in each window as well as the mean depth per isoline.

QUANTIFICATION AND STATISTICAL ANALYSIS

Population differences in recombination rate

We quantified differences in recombination rate between populations using a generalized linear mixed model fitted with the R package *lme4* [87]. This model had the form *crossover count ~ population + (1|inbred line)*, with a Poisson error distribution and a log link function (in order to accommodate the non-normal nature of crossover counts). We checked for violations of model fit for this and all subsequent models using a QQ-plot and a fitted vs. Pearson residuals plot (see Figure S3 for an example). To test the fixed effect of population, we performed a Type II Wald Test using the function "Anova" from the *car* package [88], as well as a Likelihood Ratio Test comparing models with and without the population term. Note that these two tests focus on a slightly different hypothesis (i.e. that the populations are significantly different in recombination rate, on the basis of phenotypic variance alone) than the Q_{ST}-F_{ST} analysis below.

QST-FST Analysis

To test the hypothesis that population-level differences in recombination rates are driven by natural selection, we performed a Q_{ST}-F_{ST} analysis [33,35,89]. We began by computing a point estimate of Q_{ST} for genome-wide recombination rate using lm4 by fitting a linear mixed effects model with the following form: *crossover count = intercept + (1|inbred line) + (1|population)*. We extracted the variance components for population and inbred line (nested in population) using the R function varcomp(). Following [55] we computed Q_{ST} using the following formula:

$$(1)\ Q_{ST} = \frac{\sigma^2_{BG}}{\sigma^2_{BG} + \sigma^2_{WG}}$$

Where σ^2_{BG} denotes the between-group (population) variance and σ^2_{WG} denotes within-group (inbred line nested in population) variance. Note that the within-group variance term in the denominator is not multiplied by two in the case of haploids or completely inbred lines [89].

We computed F_{ST} using SNPs genotyped via RAD-seq in wild AZ and UT individuals. To do this, we converted the GATK VCF to a SNP table using vcfR and the tidyverse package in R (see analysis scripts). We then converted the resulting SNP table for manipulation in the R package SNPRelate [90]. Using SNPRelate, we first performed LD pruning (default settings, r < 0.2) to reduce statistical non-independence between SNPs [91]. This resulted in a dataset composed of 16 individuals for AZ and 42 for UT, with a total of 6 591 high quality SNPs. We then computed per-SNP estimates of Weir and Cockerham's F_{ST} using SNPRelate, requiring filtered sites to have a minimum minor allele frequency of 0.1.

We assessed the statistical departure from neutrality for each value of Q_{ST} using the Null-QST method outlined in [35] and [55] with a modification to accommodate trait data from inbred lines. The general approach outlined in these two references is to simulate the expected distribution of for a *neutral* trait (denoted Q^n_{ST}, neutral Q_{ST}) via a parametric bootstrap, and use this distribution as the basis of a statistical test of the hypothesis $Q_{ST} > Q^n_{ST}$.

To simulate the distribution of Q^n_{ST} we first estimated the between-group (σ^2_{BG}) and within-group (σ^2_{WG}) variance components. We obtained these values via REML estimation by fitting mixed-effects linear models using the function lmer in the R package lme4 [87]. These models took the form *crossover count ~ intercept + (1|population) + (1|line)*. We extracted the variance components (standard deviations of the random effects) using the function VarComp from lme4.

We next generated 10 000 (nonparametric) bootstrap estimates of the mean value of Weir and Cockerham's F_{ST} by resampling the RADseq SNPs with replacement, and computing genome-wide mean F_{ST} using SNPRelate. We then generated 10 000 matching parametric bootstrap estimates of the σ^2_{WG} by multiplying the REML point estimate by a random draw from a χ^2 distribution with degrees of freedom equal to the number of inbred families (df = 17). Next, we generated parametric bootstrap estimates of the expected values of σ^2_{BG} for a neutrally evolving trait using the equation:

$$(2)\ boot\ (\sigma^2_{BG}) = \frac{boot(F_{ST})boot(\sigma^2_{WG})}{1-boot(F_{ST})} \times \chi^2\ (n = 1, df = 1)$$

With (2) above being modified from [35] to accommodate complete inbreeding. In equation (2), "boot" indicates individual bootstrap samples for each quantity, and the χ^2 term represents a draw from a χ^2 distribution with degrees of freedom equal to the number of populations minus one (one, in this case). This procedure results in 10 000 bootstrap samples for σ^2_{BG} *and* σ^2_{WG}, from which we computed 10 000 bootstrap samples *of* Q^n_{ST} using equation (1). We finally computed a p-value for the observed value of Q_{ST} by determining the number of Q^n_{ST} values that exceeded the observed value of Q_{ST} distribution. We also computed a confidence interval for $Q_{ST} - F_{ST}$ (the *difference* between Q_{ST} and F_{ST}, expected to be zero under the neutral model) by subtracting each value the $Q^n_{ST} - F_{ST}$ distribution from the observed value of $Q_{ST} - F_{ST}$ (after [55]). Note that while we computed the distribution of F_{ST} from the independently-sourced RAD-seq data, the distribution of F_{ST} was nearly identical when computed using SNPs derived from the short read whole genome sequencing of the inbred lines themselves.

Candidate genes associated with recombination differences

Using the candidate gene data from the inbred lines, we tested for associations between inbred line recombination estimates and genotype at each site where at least one non-synonymous change occurred in each gene. To this end, we fit linear models with recombination rate as the response and genotype (at all variable non-synonymous sites) as the predictor. This yielded a p-value for each genotype vs. recombination comparison. In order to control for the possibility of false positives, which we adjusted via the FDR approach [92], with FDR < 0.05 adjustments performed using the function p.adjust in R. A main caveat to this approach is that the small number of lines and large number of variable sites limits our power and ability to include controls for genetic background, genotype at "non-meiosis" genes, etc. As such, we consider the function of this analysis

to be mainly hypothesis-generating and to serve as a bridge between our results in previous molecular work.

Association between local structural variation and recombination rate

We tested the association between normalized sequence homology and recombination rate via a hierarchical linear model fit using the function glmer from the R package lme4 [87]. This model had the form: *recombination rate ~ method * homology + (1|window identity) + (1|inbred line)*, with Poisson-distributed errors and a log link function. Assigning window identity (i.e. genomic region in which recombination and homology were measured) as a random effect controls for mean local variation in recombination rate (i.e. normalizes the absolute recombination rates among windows). Similarly, modelling inbred line identity as a random effect controls for genome-wide differences in recombination rate, which are unrelated to local variation. We assessed the significance of the homology term in the model by comparing the full model to a model with only random effects via a likelihood ratio test in R. We finally repeated this model fitting procedure with the normalized count of differences in structural variant alleles as the predictor.

DATA AND CODE AVAILABILITY

All analysis code employed throughout the paper is available as a Git repository at: https://github.com/ksamuk/samuk_et_al_curr_biol_2020

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Flagstaff14	This study	N/A

MV2-25	Dr. Steven Schaeffer	N/A
American Fork Canyon Inbred Lines	This study	N/A
Madera Canyon Inbred Lines	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
Qiagen Plus Multiplex PCR Master Mix	Qiagen	Cat # 206145
Charm Biotech Just-A-Plate 96 PCR Purification and Normalization Kit	Charm Biotech	Cat # JN-120-10
AmpureXP beads	Beckman Coulter	Cat # A63881
Deposited Data		
PacBio Sequel Whole Genome Sequencing	This paper	SRA PRJNA610090
Illumina HiSeq Whole Genome Sequencing	This paper	SRA PRJNA610029

Illumina HiSeq ddRAD-seq Sequencing	This paper	SRA PRJNA610904
Illumina NovaSeq Amplicon Libraries (High Output)	This paper	Dryad Accession TBD
Illumina NovaSeq Amplicon Libraries (Mid Output)	This paper	Dryad Accession TBD
Analysis scripts	This paper	https://github.com/ksamuk/samuk_et_al_curr_biol_2020
Oligonucleotides		
GT-Seq Primers - See Table S4	This paper	N/A
Illumina Small RNA sequencing primer (CGACAGGTTCAGAGTTCTACAGTCCGACGATC)	Illumina	N/A
Software and Algorithms		
samtools	Li et al. (2009)	http://www.htslib.org/
Bwa2	Li (2013)	https://github.com/lh3/bwa

GATK	Poplin et al. (2017)	https://gatk.broadinstitute.org/hc/en-us
R for Statistical Programming	R Core Team (2019)	https://www.r-project.org/
RStudio	RStudio Team (2019)	https://rstudio.com/
tidyverse (R Package collection)	Wickham et al. (2019)	https://www.tidyverse.org/
lme4 (R Package)	Bates et al. (2015)	https://cran.r-project.org/web/packages/lme4/index.html
r/qtl (R Package)	Broman (2003)	https://rqtl.org/
ASmap (R Package)	Taylor et al. (2017)	https://cran.r-project.org/web/packages/ASMap/index.html
vcfR	Knaus & Grunwald (2017)	https://cran.r-project.org/web/packages/vcfR/index.html

patchwork	Pedersen (2019)	https://cran.r-project.org/web/packages/patchwork/index.html
SNPRelate	Zheng et al. (2012)	https://github.com/zhengxwen/SNPRelate
lmerTest	Kuznetsova et al. (2019)	https://cran.r-project.org/web/packages/lmerTest/index.html
car (R Package)	Fox & Weisberg (2019)	https://cran.r-project.org/web/packages/car/index.html

Figure Legends (Supplemental Tables)

Table S1 | Per-chromosome counts of crossover events for all lines. Each row lists the number of observed crossovers for each chromosome (Columns E-H) in the re-ordered reference genome for a single F1-Backcross individual (Column D). Population, inbred line and plate information are provided (AFC = American Fork Canyon, MC = Madera Canyon). Note that for purposes of comparing recombination rates, the unit of replication is inbred line and not individual.

Table S2 | Fine scale estimates of crossover location for all lines. Each row corresponds to the observed "crossover state" (whether a crossover was observed between the previous and current position) for each location on all chromosomes of the re-ordered reference genome. Note that the total length of this file is greater than the maximum rows displayed by some versions of Microsoft Excel (i.e. the file will be truncated if viewed in Excel).

Table S3 | *D. melanogaster* candidate genes and *D. pseudoobscura* homologues used for the candidate gene analysis. Each row lists a single *D. melanogaster* candidate gene (using Flybase Gene

IDs), along with the corresponding homologous gene in *D. pseudooobscura*, and its location in the re-ordered reference genome.

Table S4 | List of PCR primers used in GT-seq library preparation. Each row lists the sequence of a single primer pair (forward, Column D and reverse, Column E). The primer ID is formatted as: [Reference Genome Chromosome]_ [PrimerID]_ [Reference Genome Start Coordinate].

References

1. Hughes, S.E., Miller, D.E., Miller, A.L., and Hawley, R.S. (2018). Female Meiosis: Synapsis, Recombination, and Segregation in Drosophila melanogaster. Genetics *208*, 875–908.

2. Dapper, A.L., and Payseur, B.A. (2017). Connecting theory and data to understand recombination rate evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160469.

3. Charlesworth, B., Betancourt, A.J., Kaiser, V.B., and Gordo, I. (2009). Genetic recombination and molecular evolution. Cold Spring Harb. Symp. Quant. Biol. *74*, 177–186.

4. Ritz, K.R., Noor, M.A.F., and Singh, N.D. (2017). Variation in Recombination Rate: Adaptive or Not? Trends in Genetics *33*, 364–374.

5. Barton, N.H., and Otto, S.P. (1997). The evolution of recombination: Removing the limits to natural selection. Genetics *147*, 879–906.

6. Otto, S.P., and Lenormand, T. (2000). The evolution of recombination in a heterogeneous environment. Genetics *156*, 423–438.

7. Marais, G., and Charlesworth, B. (2003). Genome evolution: recombination speeds up adaptive evolution. Curr. Biol. *13*, R68–70.

8. Martin, S.H., Davey, J.W., Salazar, C., and Jiggins, C.D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. *17*, e2006288.

9. Barton, N.H. (2010). Mutation and the evolution of recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 1281–1294.

10. Charlesworth, B. (1976). Recombination modification in a fluctuating environment. Genetics *83*, 181–195.

11. Kirkpatrick, M., and Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. Genetics *173*, 419–434.

12. Noor, M.A.F., Grams, K.L., Bertucci, L.A., and Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences *98*, 12084–12088.

13. Samuk, K., Owens, G.L., Delmore, K.E., Miller, S.E., Rennison, D.J., and Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination.

748 Mol. Ecol. *26*, 4378–4390.

749 14. Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, J.C., Sankararaman,
750 S., Andolfatto, P., Rosenthal, G.G., *et al.* (2018). Natural selection interacts with recombination
751 to shape the evolution of hybrid genomes. Science *360*, 656–660.

752 15. Wang, R.J., and Payseur, B.A. (2017). Genetics of Genome-Wide Recombination Rate Evolution
753 in Mice from an Isolated Island. Genetics *206*, 1841–1852.

754 16. Kulathinal, R.J., Noor, M.A.F., Fitzpatrick, C.L., and Bennett, S.M. (2008). Fine-scale mapping of
755 recombination rate in Drosophila refines its correlation to diversity and divergence. Proc. Natl.
756 Acad. Sci. U. S. A. *105*, 10051–10056.

757 17. Stumpf, M.P.H., and McVean, G.A.T. (2003). Estimating recombination rates from population-
758 genetic data. Nat. Rev. Genet. *4*, 959–968.

759 18. Wall, J.D. (2000). A comparison of estimators of the population recombination rate. Mol. Biol.
760 Evol. *17*, 156–163.

761 19. Auton, A., and McVean, G. (2007). Recombination rate estimation in the presence of hotspots.
762 Genome Res. *17*, 1219–1227.

763 20. Gärtner, K., and Futschik, A. (2016). Improved Versions of Common Estimators of the
764 Recombination Rate. J. Comput. Biol. *23*, 756–768.

765 21. Adrion, J.R., Galloway, J.G., and Kern, A.D. (2019). Inferring the landscape of recombination
766 using recurrent neural networks. bioRxiv, 662247.

767 22. Nei, M., and Li, W.H. (1973). Linkage disequilibrium in subdivided populations. Genetics *75*,
768 213–219.

769 23. Cutter, A.D. (2019). Recombination and linkage disequilibrium in evolutionary signatures. A
770 Primer of Molecular Population Genetics, 113–128. Available at:
771 http://dx.doi.org/10.1093/oso/9780198838944.003.0006.

772 24. McVean, G. (2007). Linkage Disequilibrium, Recombination and Selection. In Handbook of
773 Statistical Genetics, D. J. Balding, M. Bishop, and C. Cannings, eds. (Chichester, UK: John Wiley &
774 Sons, Ltd), pp. 909–944.

775 25. Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017). Variation in
776 recombination frequency and distribution across eukaryotes: patterns and processes. Philos.
777 Trans. R. Soc. Lond. B Biol. Sci. *372*. Available at: http://dx.doi.org/10.1098/rstb.2016.0455.

778 26. Johnston, S.E., Bérénos, C., Slate, J., and Pemberton, J.M. (2016). Conserved Genetic Architecture
779 Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis
780 aries). Genetics *203*, 583–598.

781 27. Hellsten, U., Wright, K.M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S.R., Schmutz, J., Willis, J.H., and
782 Rokhsar, D.S. (2013). Fine-scale variation in meiotic recombination in Mimulus inferred from
783 population shotgun sequencing. Proc. Natl. Acad. Sci. U. S. A. *110*, 19478–19482.

784 28. Smukowski, C.S., and Noor, M.A.F. (2011). Recombination rate variation in closely related

785 species. Heredity *107*, 496–508.

786 29. Hunter, C.M., Huang, W., Mackay, T.F.C., and Singh, N.D. (2016). The Genetic Architecture of
787 Natural Variation in Recombination Rate in Drosophila melanogaster. PLoS Genet. *12*,
788 e1005951.

789 30. Dumont, B.L., and Payseur, B.A. (2008). Evolution of the genomic rate of recombination in
790 mammals. Evolution *62*, 276–294.

791 31. Otto, S.P., and Michalakis, Y. (1998). The evolution of recombination in changing environments.
792 Trends Ecol. Evol. *13*, 145–151.

793 32. Aggarwal, D.D., Rashkovetsky, E., Michalak, P., Cohen, I., Ronin, Y., Zhou, D., Haddad, G.G., and
794 Korol, A.B. (2015). Experimental evolution of recombination and crossover interference in
795 Drosophila caused by directional selection for stress-related traits. BMC Biol. *13*, 101.

796 33. Leinonen, T., Scott McCairns, R.J., O'Hara, R.B., and Merilä, J. (2013). Q(ST)-F(ST) comparisons:
797 evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics *14*,
798 179–190.

799 34. Whitlock, M.C. (2008). Evolutionary inference from QST. Mol. Ecol. *17*, 1885–1896.

800 35. Whitlock, M.C., and Guillaume, F. (2009). Testing for Spatially Divergent Selection: Comparing
801 QST to FST. Genetics *183*, 1055–1063.

802 36. Barrett, R.D.H., and Hoekstra, H.E. (2011). Molecular spandrels: tests of adaptation at the
803 genetic level. Nat. Rev. Genet. *12*, 767–780.

804 37. Morgan, A.P., Gatti, D.M., Najarian, M.L., Keane, T.M., Galante, R.J., Pack, A.I., Mott, R., Churchill,
805 G.A., and de Villena, F.P.-M. (2017). Structural Variation Shapes the Landscape of
806 Recombination in Mouse. Genetics *206*, 603–619.

807 38. Völker, M., Backström, N., Skinner, B.M., Langley, E.J., Bunzey, S.K., Ellegren, H., and Griffin, D.K.
808 (2010). Copy number variation, chromosome rearrangement, and their association with
809 recombination during avian evolution. Genome Res. *20*, 503–511.

810 39. Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J., and Anderson, L.K. (2007). Recombination: an
811 underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. *8*, 77–84.

812 40. Brand, C.L., Cattani, M.V., Kingan, S.B., Landeen, E.L., and Presgraves, D.C. (2018). Molecular
813 Evolution at a Meiosis Gene Mediates Species Differences in the Rate and Patterning of
814 Recombination. Curr. Biol. *28*, 1289–1295.e4.

815 41. Baker, Z., Schumer, M., Haba, Y., Bashkirova, L., Holland, C., Rosenthal, G.G., and Przeworski, M.
816 (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9
817 across vertebrates. Elife *6*. Available at: http://dx.doi.org/10.7554/eLife.24133.

818 42. Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., Moreno, C.R., and Servin, B. (2017).
819 Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. Genetics
820 *207*, 767–784.

821 43. Wakefield, J.G., Bonaccorsi, S., and Gatti, M. (2001). The drosophila protein asp is involved in

822 microtubule organization during spindle formation and cytokinesis. J. Cell Biol. *153*, 637–648.

823 44. Joyce, E.F., Pedersen, M., Tiong, S., White-Brown, S.K., Paul, A., Campbell, S.D., and McKim, K.S.
824 (2011). Drosophila ATM and ATR have distinct activities in the regulation of meiotic DNA
825 damage and repair. J. Cell Biol. *195*, 359–367.

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

829 46. Chowdhury, R., Bois, P.R.J., Feingold, E., Sherman, S.L., and Cheung, V.G. (2009). Genetic analysis
830 of variation in human meiotic recombination. PLoS Genet. *5*, e1000648.

831 47. Ziolkowski, P.A., Underwood, C.J., Lambing, C., Martinez-Garcia, M., Lawrence, E.J., Ziolkowska,
832 L., Griffin, C., Choi, K., Franklin, F.C.H., Martienssen, R.A., *et al.* (2017). Natural variation and
833 dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination. Genes
834 Dev. *31*, 306–317.

835 48. Price, D.J., and Bantock, C.R. (1975). MARGINAL POPULATIONS OF CEPAEA NEMORALIS (L.)
836 ON THE BRENDON HILLS, ENGLAND. II. VARIATION IN CHIASMA FREQUENCY. Evolution *29*,
837 278–286.

838 49. Kong, A., Thorleifsson, G., Stefansson, H., Masson, G., Helgason, A., Gudbjartsson, D.F., Jonsdottir,
839 G.M., Gudjonsson, S.A., Sverrisson, S., Thorlacius, T., *et al.* (2008). Sequence variants in the
840 RNF212 gene associate with genome-wide recombination rate. Science *319*, 1398–1401.

841 50. Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters,
842 G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K.T., *et al.* (2010). Fine-scale recombination rate
843 differences between sexes, populations and individuals. Nature *467*, 1099–1103.

844 51. Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic variants
845 in REC8, RNF212, and PRDM9 influence male recombination in cattle. PLoS Genet. *8*,
846 e1002854.

847 52. Wang, R.J., Dumont, B.L., Jing, P., and Payseur, B.A. (2019). A first genetic portrait of
848 synaptonemal complex variation. PLoS Genet. *15*, e1008337.

849 53. Jackson, S., Nielsen, D.M., and Singh, N.D. (2015). Increased exposure to acute thermal stress is
850 associated with a non-linear increase in recombination frequency and an independent linear
851 decrease in fitness in Drosophila. BMC Evol. Biol. *15*, 175.

852 54. Menne, M.J., Williams, C.N., Gleason, B.E., Rennie, J.J., and Lawrimore, J.H. (2018). The Global
853 Historical Climatology Network Monthly Temperature Dataset, Version 4. J. Clim. *31*, 9835–
854 9854.

855 55. King, J.G., and Hadfield, J.D. (2019). The evolution of phenotypic plasticity when environments
856 fluctuate in time and space. Evol Lett *3*, 15–27.

857 56. Mugal, C.F., Smeds, L., Ellegren, H., Burri, R., Suh, A., Kawakami, T., and Nater, A. (2017). Whole-
858 genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and
859 consequences of fine-scale recombination rate variation in birds. Mol. Ecol. *15*, 900.

57. Kent, T.V., Uzunović, J., and Wright, S.I. (2017). Coevolution between transposable elements and recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*. Available at: http://dx.doi.org/10.1098/rstb.2016.0458.

58. Ross, C.R., DeFelice, D.S., Hunt, G.J., Ihle, K.E., Amdam, G.V., and Rueppell, O. (2015). Genomic correlates of recombination rate and its variability across eight recombination maps in the western honey bee (Apis mellifera L.). BMC Genomics *16*, 107.

59. Opperman, R., Emmanuel, E., and Levy, A.A. (2004). The Effect of Sequence Divergence on Recombination Between Direct Repeats in Arabidopsis. Genetics *168*, 2207.

60. Crown, K.N., Miller, D.E., Sekelsky, J., and Hawley, R.S. (2018). Local Inversion Heterozygosity Alters Recombination throughout the Genome. Curr. Biol. *28*, 2984–2990.e3.

61. Said, I., Byrne, A., Serrano, V., Cardeno, C., Vollmers, C., and Corbett-Detig, R. (2018). Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. Proc. Natl. Acad. Sci. U. S. A. *115*, 5492–5497.

62. Ortiz-Barrientos, D., Chang, A.S., and Noor, M.A.F. (2006). A recombinational portrait of the Drosophila pseudoobscura genome. Genet. Res. *87*, 23–31.

63. Campbell, N.R., Harmon, S.A., and Narum, S.R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. *15*, 855–867.

64. Chakraborty, P., Pankajam, A.V., Dutta, A., and Nishant, K.T. (2018). Genome wide analysis of meiotic recombination in yeast: For a few SNPs more. IUBMB Life *70*, 743–752.

65. Ritz, K.R., and Noor, M.A.F. (2015). North American Southwest collection of obscura-group Drosophila in summer 2015. Drosophila Information Service *98*. Available at: http://www.ou.edu/journals/dis/DIS98/RitzNoor(pg23).pdf.

66. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One *7*, e37135.

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

68. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al.* (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 11.10.1–33.

69. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

70. Wysoker, A., Tibbetts, K., and Fennell, T. (2013). Picard tools version 1.90. http://picard. sourceforge. net (Accessed 14.

71. R Core Team (2018). R: A language and environment for statistical computing (R Foundation

for Statistical Computing, Vienna, Austria.) Available at: https://www.R-project.org/.

72. Knaus, B.J., and Grünwald, N.J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. *17*, 44–53.

73. Wickham, H. (2017). tidyverse: Easily Install and Load the "Tidyverse" Available at: https://CRAN.R-project.org/package=tidyverse.

74. Schaeffer, S.W., Bhutkar, A., McAllister, B.F., Matsuda, M., Matzkin, L.M., O'Grady, P.M., Rohde, C., Valente, V.L.S., Aguadé, M., Anderson, W.W., *et al.* (2008). Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics *179*, 1601–1655.

75. Broman, K.W., and Sen, S. (2009). A Guide to QTL Mapping with R/qtl (New York, NY: Springer New York).

76. Taylor, J., and Butler, D. (2017). R Package ASMap : Efficient Genetic Linkage Map Construction and Diagnosis. Journal of Statistical Software *79*, 1–29.

77. Anderson, J.A., Gilliland, W.D., and Langley, C.H. (2009). Molecular population genetics and evolution of Drosophila meiosis genes. Genetics *181*, 177–185.

78. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

79. Korunes, K.L., and Noor, M.A.F. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. Mol. Ecol. *28*, 1302–1315.

80. Larson, D.E., Abel, H.J., Chiang, C., Badve, A., Das, I., Eldred, J.M., Layer, R.M., and Hall, I.M. (2019). svtools: population-scale analysis of structural variation. Bioinformatics *btz492*. Available at: https://doi.org/10.1093/bioinformatics/btz492.

81. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.

82. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics *30*, 2503–2505.

83. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., *et al.* (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. Available at: https://www.ncbi.nlm.nih.gov/pubmed/31406327.

84. Greig, D., Travisano, M., Louis, E.J., and Borts, R.H. (2003). A role for the mismatch repair system during incipient speciation in Saccharomyces. J. Evol. Biol. *16*, 429–437.

85. Rogers, D.W., McConnell, E., Ono, J., and Greig, D. (2018). Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast. PLoS Biol. *16*, e2005066.

86. Salomé, P.A., Bomblies, K., Fitz, J., Laitinen, R.A.E., Warthmann, N., Yant, L., and Weigel, D. (2012). The recombination landscape in Arabidopsis thaliana F2 populations. Heredity *108*,

447–455.

87. Bates, D., Sarkar, D., Bates, M.D., and Matrix, L. (2007). The lme4 package. R package version *2*, 74.

88. Fox, J., and Weisberg, S. (2018). An R Companion to Applied Regression (SAGE Publications).

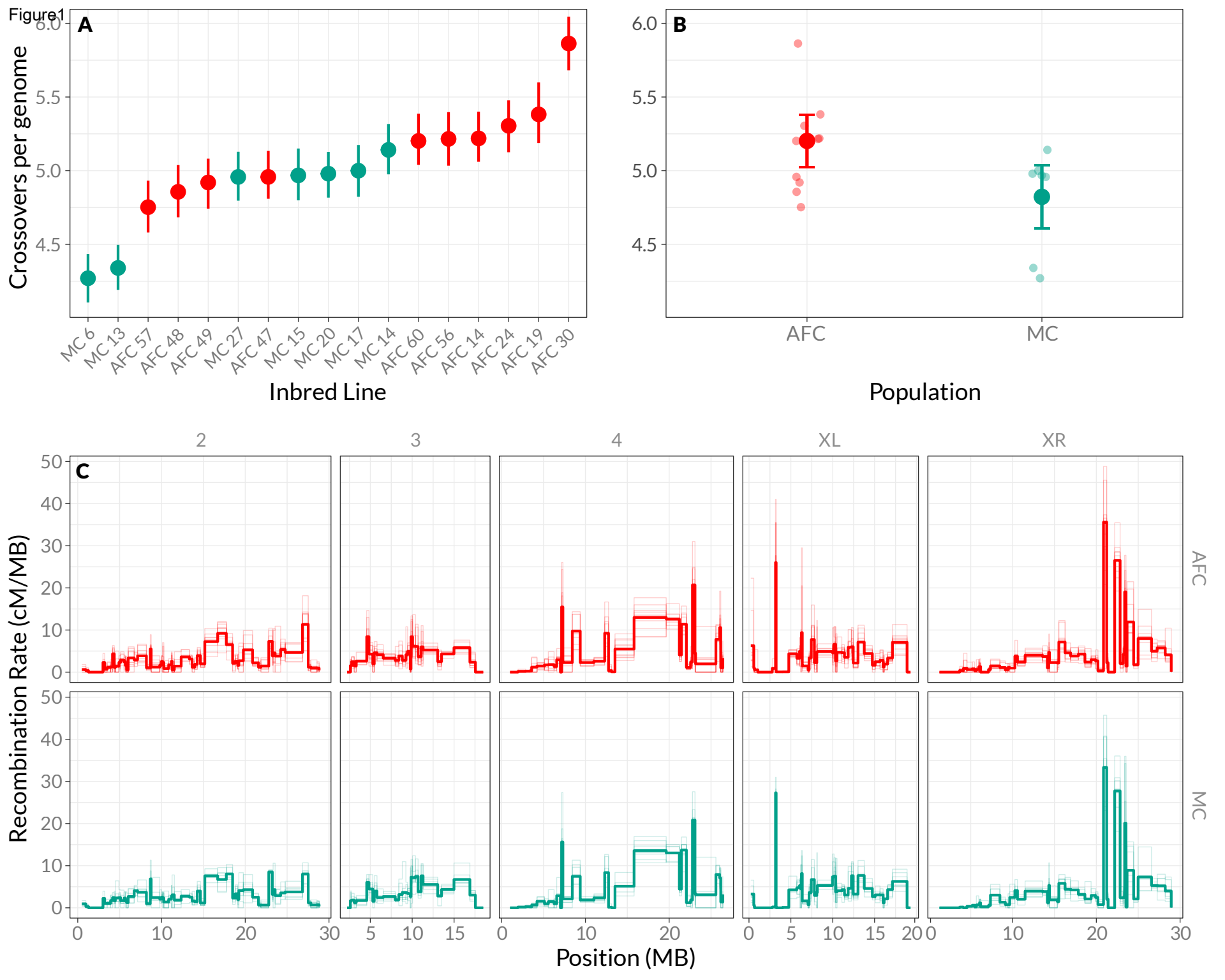
89. Walsh, B., and Lynch, M. (2018). Evolution and Selection of Quantitative Traits (Oxford University Press).

90. Shen, J., Levine, D., Weir, B.S., Laurie, C., Gogarten, S.M., and Zheng, X. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328.

91. Lotterhos, K.E. (2019). The Effect of Neutral Recombination Variation on Genome Scans for Selection. G3: Genes|Genomes|Genetics, g3.400088.2019.

92. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Figure1


A
Crossovers per genome
6.0
5.5
5.0
4.5
MC 6
MC 13
AFC 57
AFC 48
AFC 49
MC 27
AFC 47
MC 15
MC 20
MC 17
MC 14
AFC 60
AFC 56
AFC 14
AFC 24
AFC 19
AFC 30
Inbred Line
B
AFC
MC
Population
C
2
3
4
XL
XR
Recombination Rate (cM/MB)
AFC
MC
Position (MB)

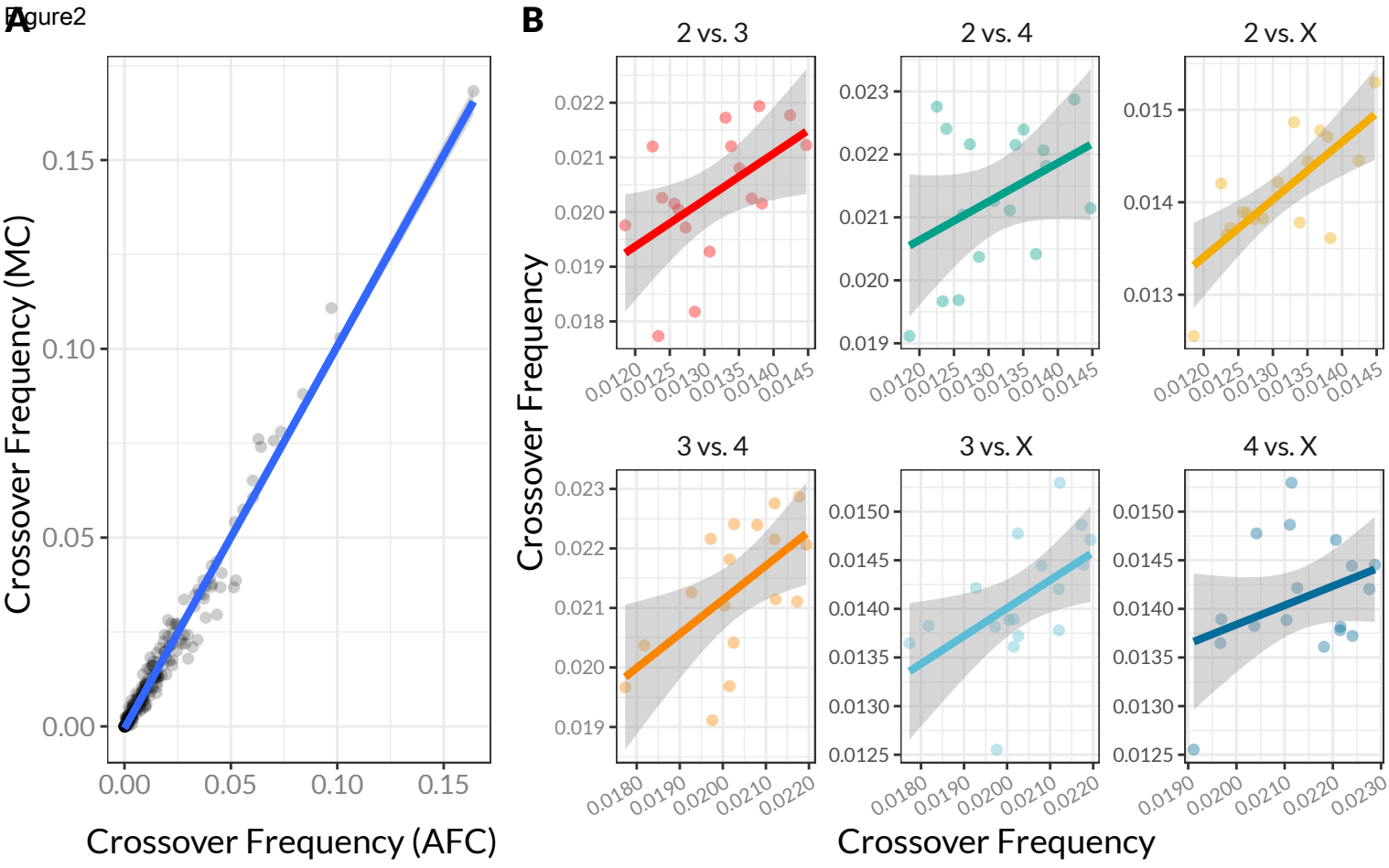

A
Crossover Frequency (MC)
0.15
0.10
0.05
0.00
0.00
0.05
0.10
0.15
Crossover Frequency (AFC)
B
2 vs. 3
2 vs. 4
2 vs. X
3 vs. 4
3 vs. X
4 vs. X
Crossover Frequency
Crossover Frequency

Figure3

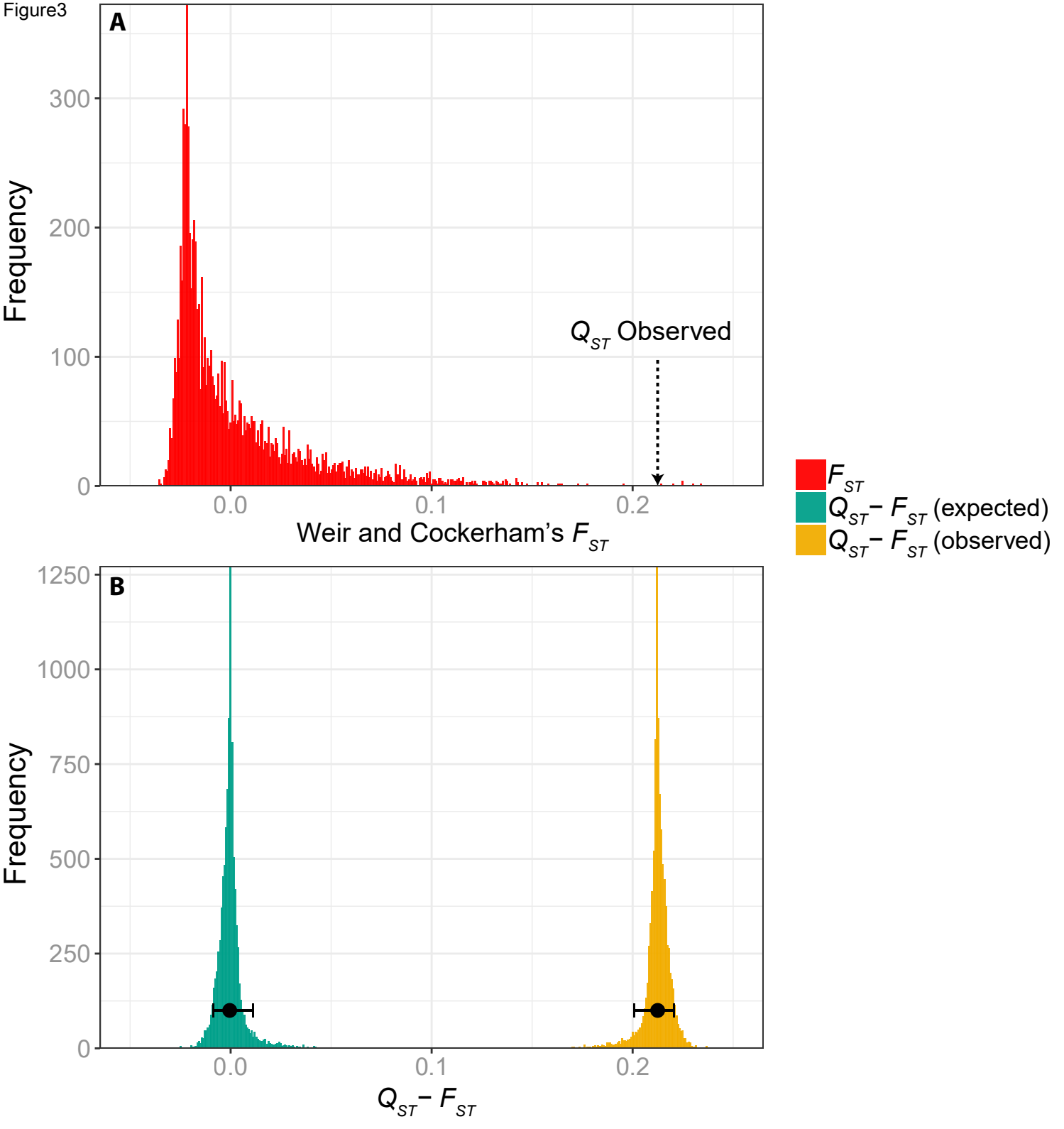

A
300
200
100
0
Frequency
Q_{ST} Observed
0.0
0.1
0.2
Weir and Cockerham's F_{ST}
F_{ST}
$Q_{ST}-F_{ST}$ (expected)
$Q_{ST}-F_{ST}$ (observed)
B
1250
1000
750
500
250
0
Frequency
0.0
0.1
0.2
$Q_{ST}-F_{ST}$

Figure 4

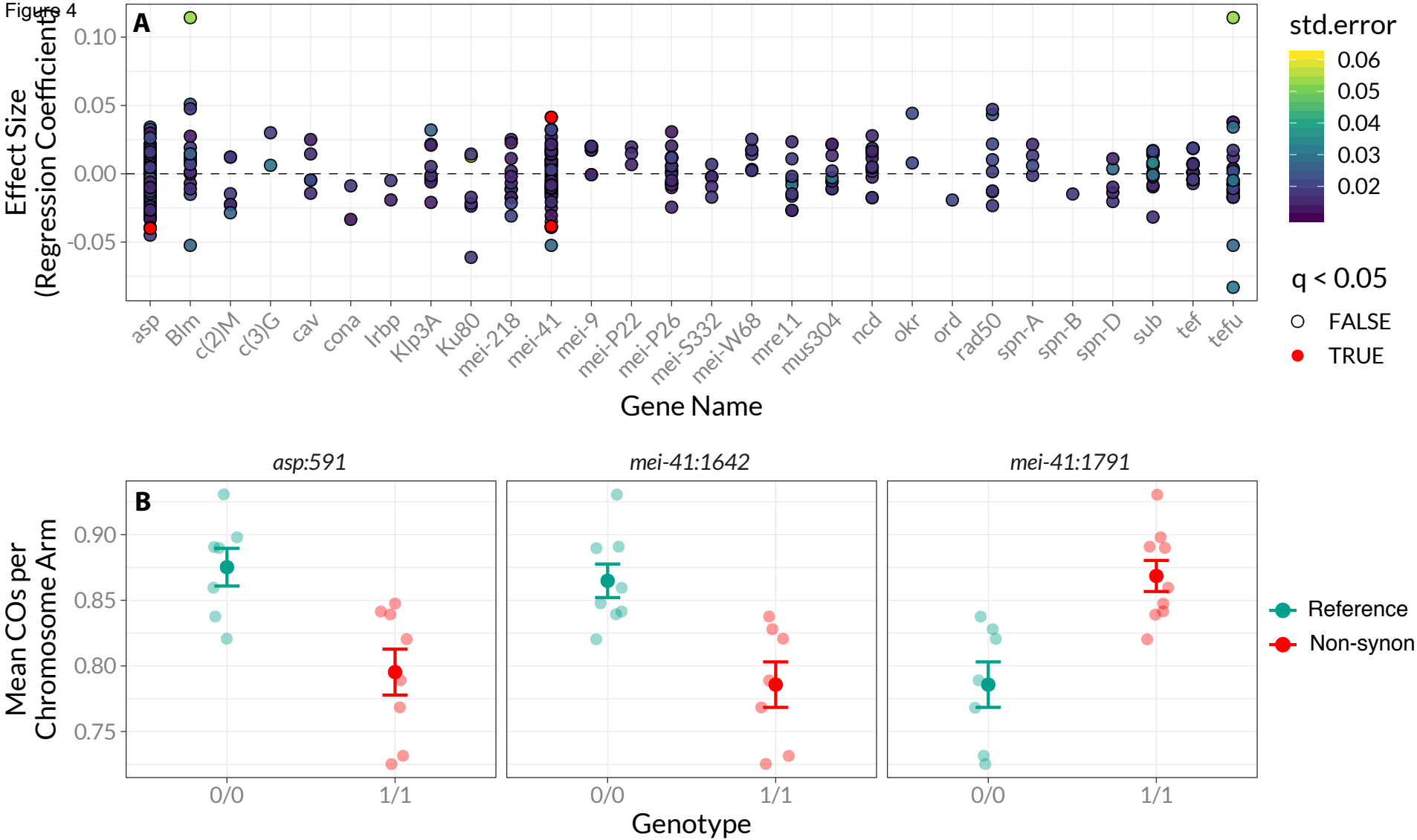

A
Effect Size
(Regression Coefficient)
0.10
0.05
0.00
-0.05
asp
Blm
c(2)M
c(3)G
cav
cona
Irbp
Klp3A
Ku80
mei-218
mei-41
mei-9
mei-P22
mei-P26
mei-S332
mei-W68
mre11
mus304
ncd
okr
ord
rad50
spn-A
spn-B
spn-D
sub
tef
tefu
Gene Name
std.error
0.06
0.05
0.04
0.03
0.02
q < 0.05
FALSE
TRUE
B
asp:591
mei-41:1642
mei-41:1791
Mean COs per
Chromosome Arm
0.90
0.85
0.80
0.75
0/0
1/1
Genotype
Reference
Non-synon

Figure5

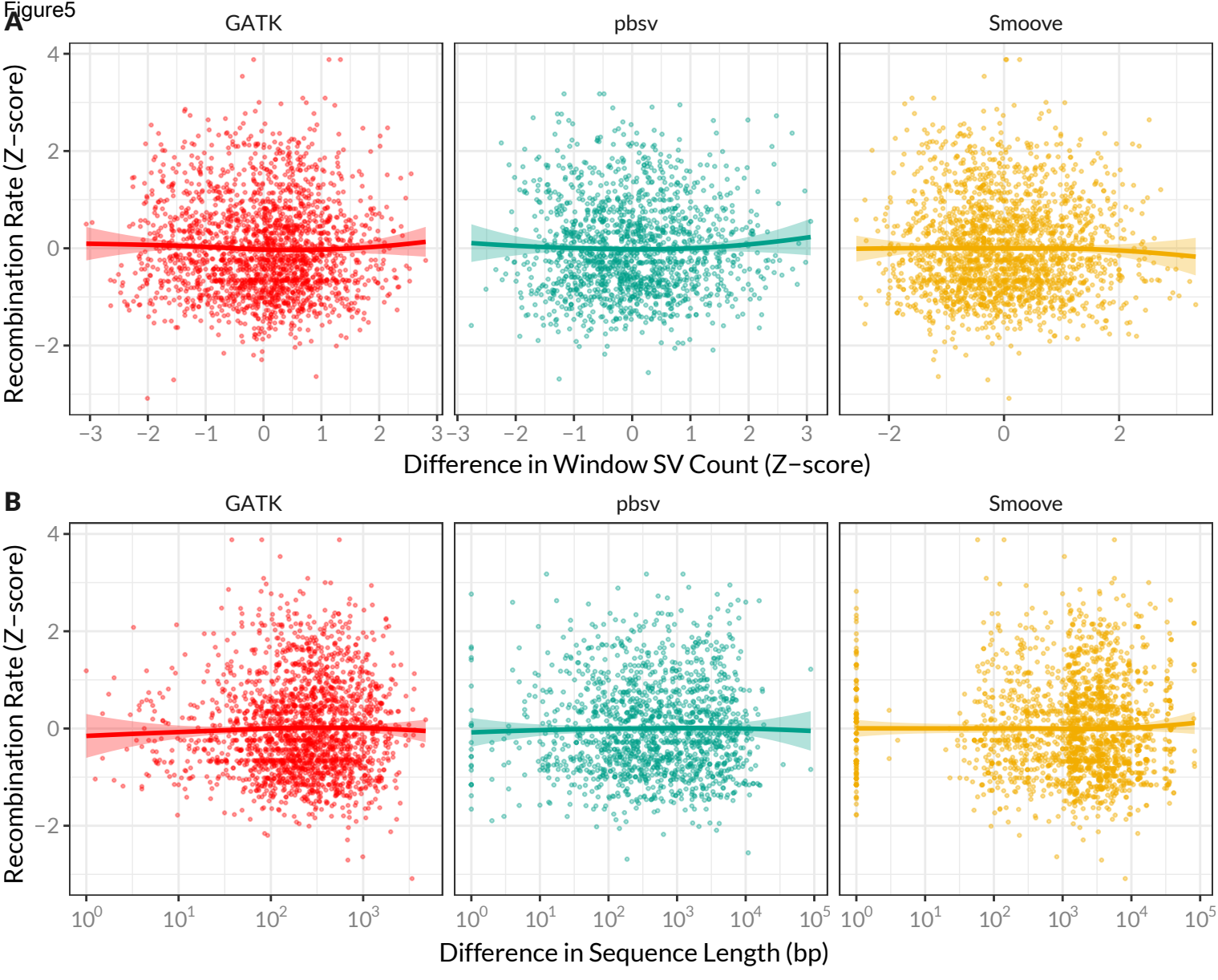
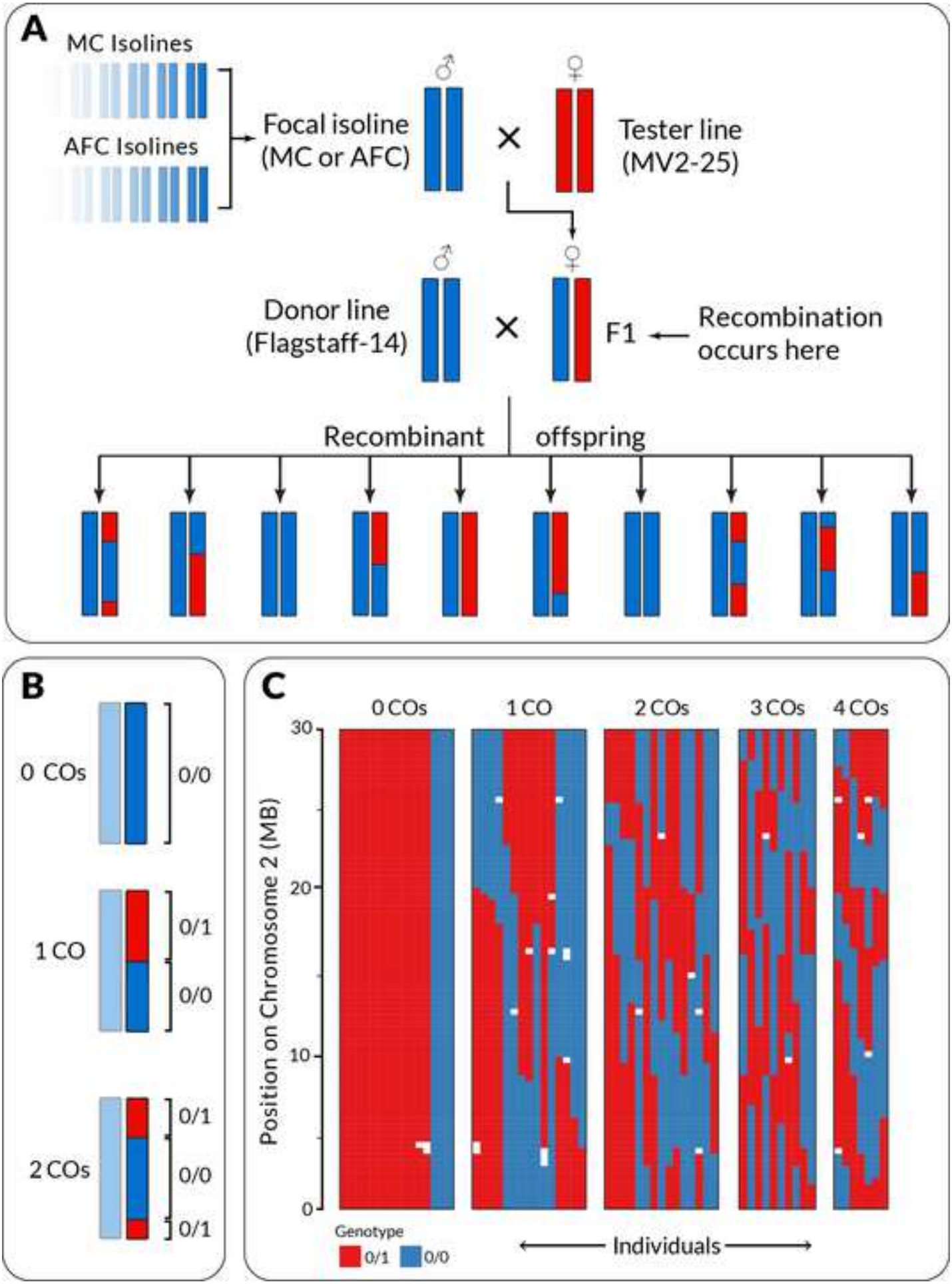

A
GATK
pbsv
Smoove
Recombination Rate (Z−score)
Difference in Window SV Count (Z−score)
B
GATK
pbsv
Smoove
Recombination Rate (Z−score)
Difference in Sequence Length (bp)

Figure6



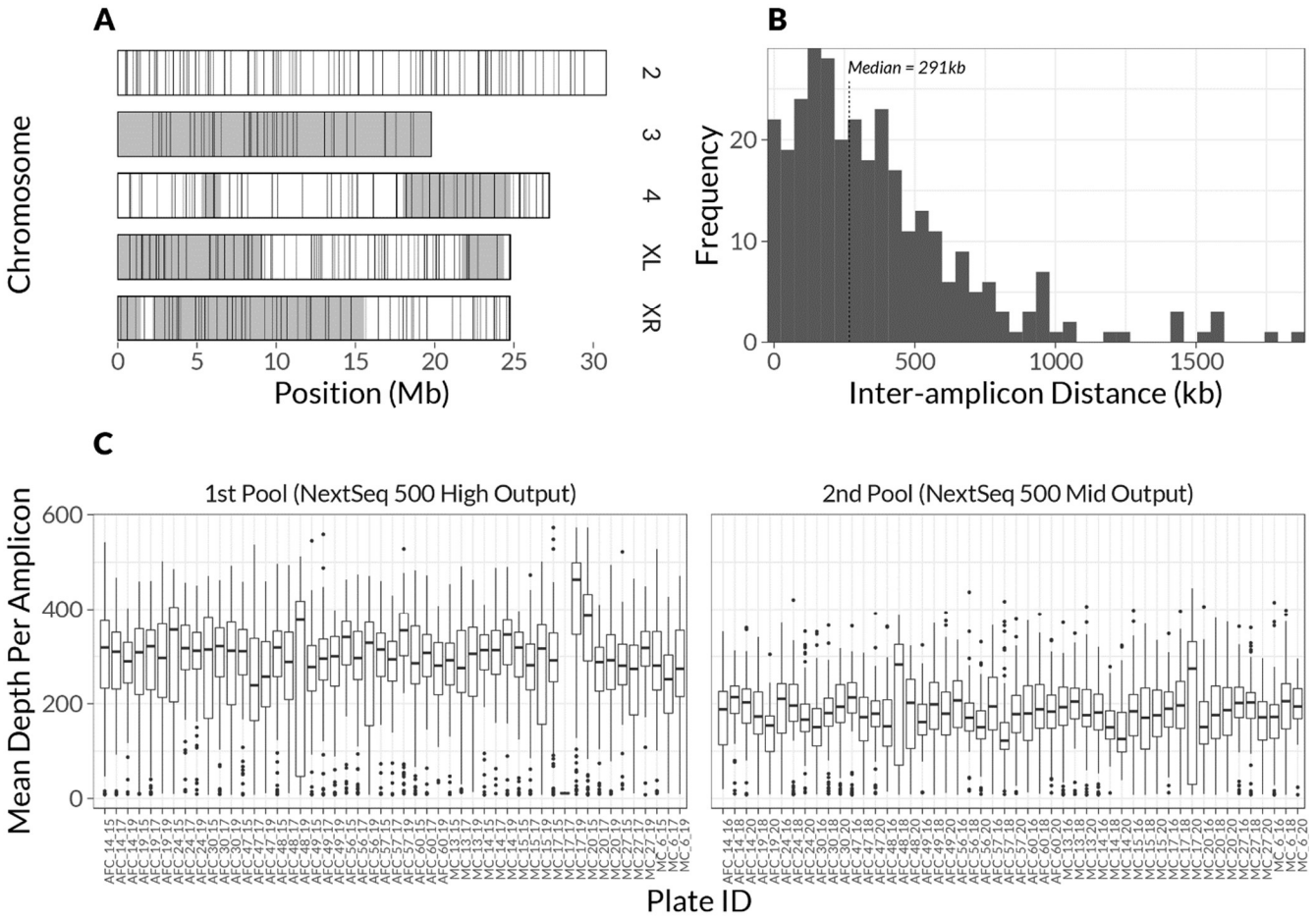


Figure S1| Summary of the design and performance of GT-seq amplicon sequencing. (A) Spacing of amplicons on the four chromosomes of *D. pseudoobscura* (the X chromosome is separated into its two arms, by convention). Each rectangle represents one chromosome in the *D. pseudoobscura* reference genome, with alternating grey and white regions indicating assembly contigs. Contigs are ordered based on Schaeffer et al. (2008). Vertical lines indicate the location of the mapping-informative amplicons we designed for use in our GT-seq protocol. (B) The distribution of distances between all amplicons in kilobases. The dotted line indicates the median value. (C) Mean sequenced depth per amplicon for each 96-well plate sequenced in the study. Each boxplot depicts the distribution of mean depths (in reads) per amplicon for all the individuals in a single plate after alignment to the reference genome. Two separate pools (left and right boxes) were sequenced using a high and mid output kit respectively.

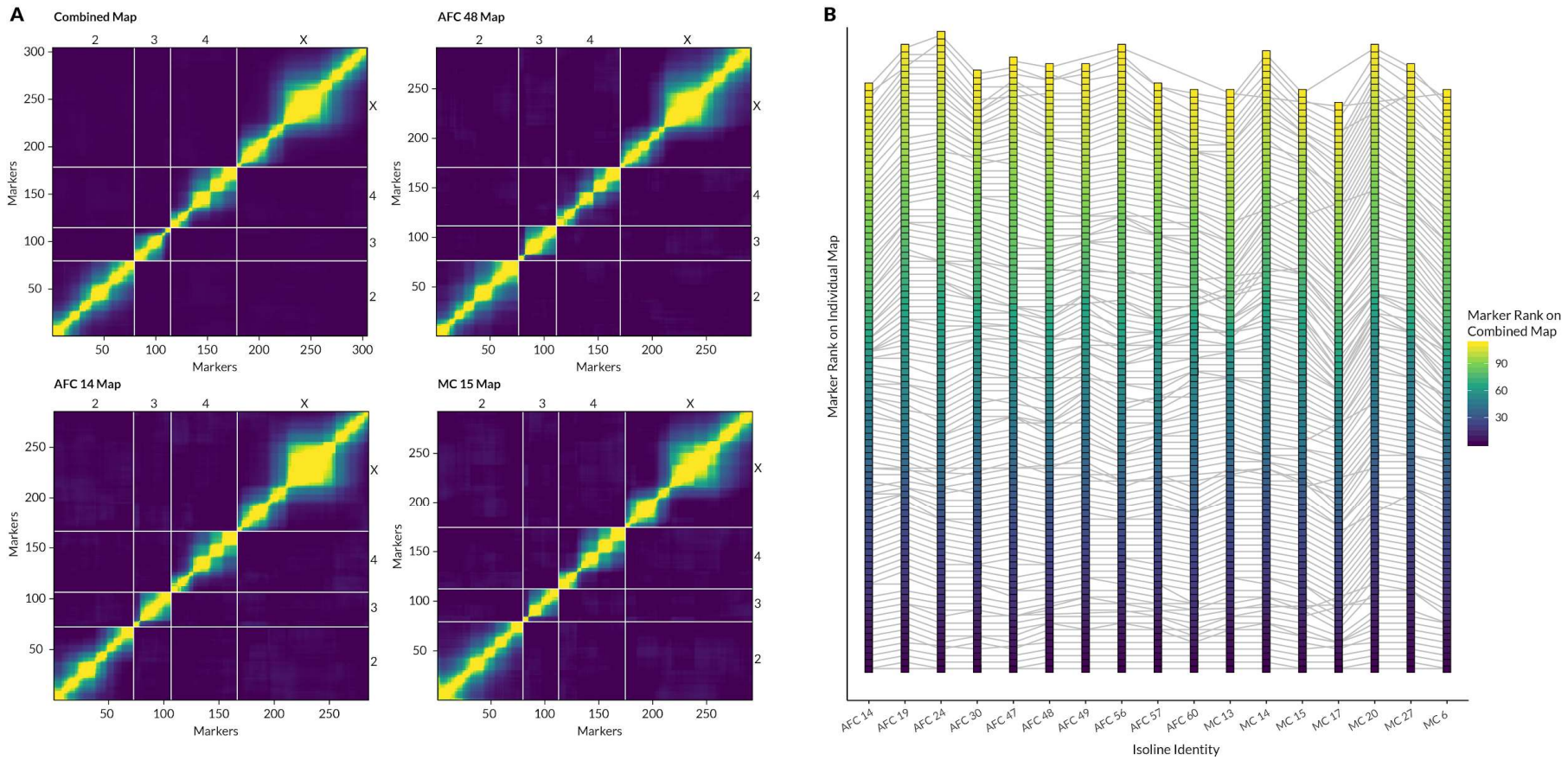


Figure S2 | Comparison of inferred marker orders in combined vs. individual datasets. (A) Recombination fraction heat maps for four representative maps. The combined map was built by pooling all samples before inferring marker order. Lighter colors correspond to larger recombination fractions between pairs of markers. (B) Inferred markers orders on Chromosome X for all isolines. Each square point is a single marker on chromosome X, and markers are ordered based on their rank (y-axis) in each isoline-level individual map (x-axis).horizontal lines between points show the position of the same marker in each individual map. The color scale indicates the rank inferred from the combined map: major changes in marker order between the combined and individual maps would manifest as jumbled colors in the gradient, of which there are none.

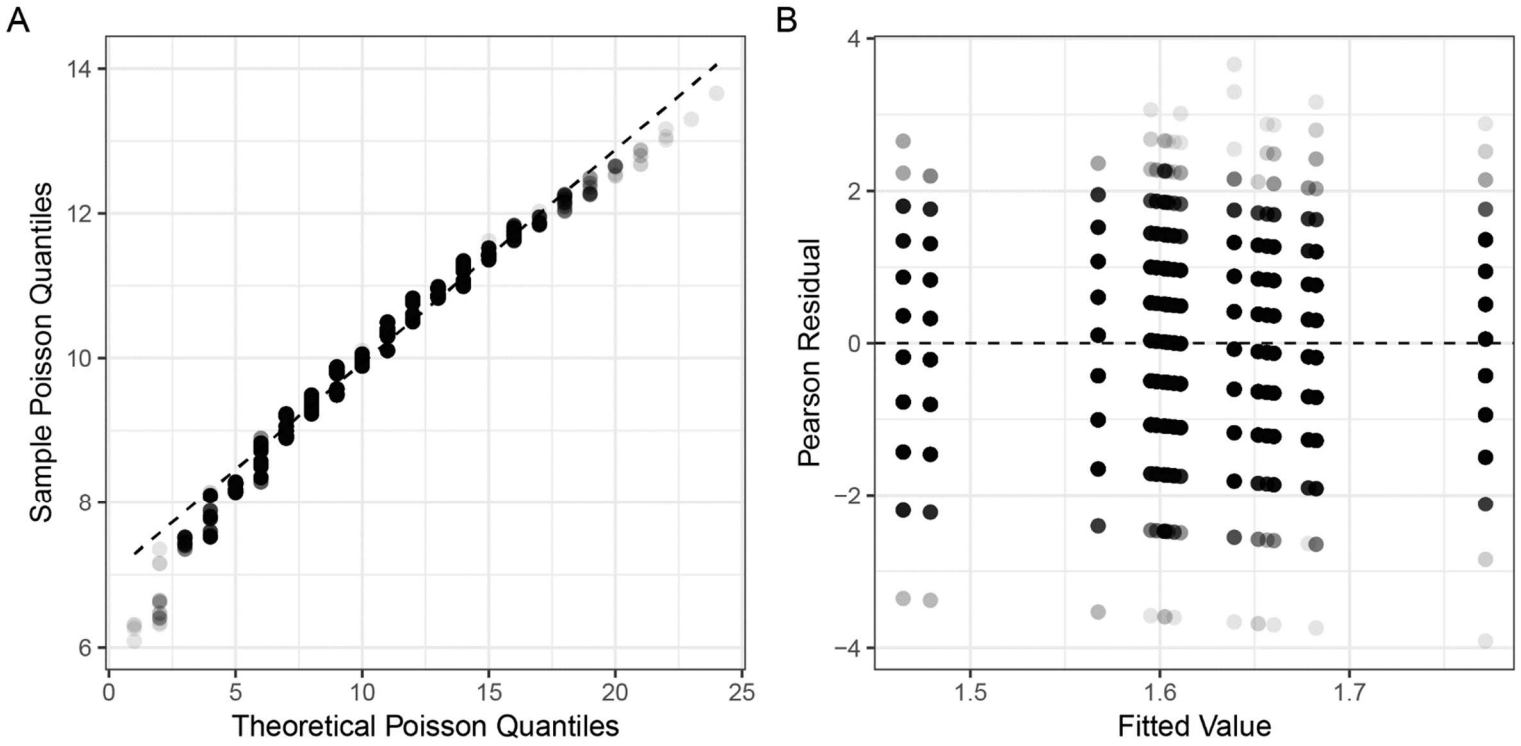


Figure S3 | Example of diagnostics of model fit for the generalized linear mixed models used throughout the paper. (A) and (B) above are derived from the model fitted in order to directly compare population differences in recombination rate. (A) A Q-Q plot showing the fit between sample Poisson quantiles and theoretical Poisson quantiles (the expectation is an approximate 1:1 fit, shown by the dashed line). To ensure positivity for plotting purposes, all samples values were transformed by adding a constant of +10 prior to plotting. (B) A plot of Pearson residuals vs. the fitted values of the model (the expectation is no trend in the mean or change in variance, shown by the dashed line).

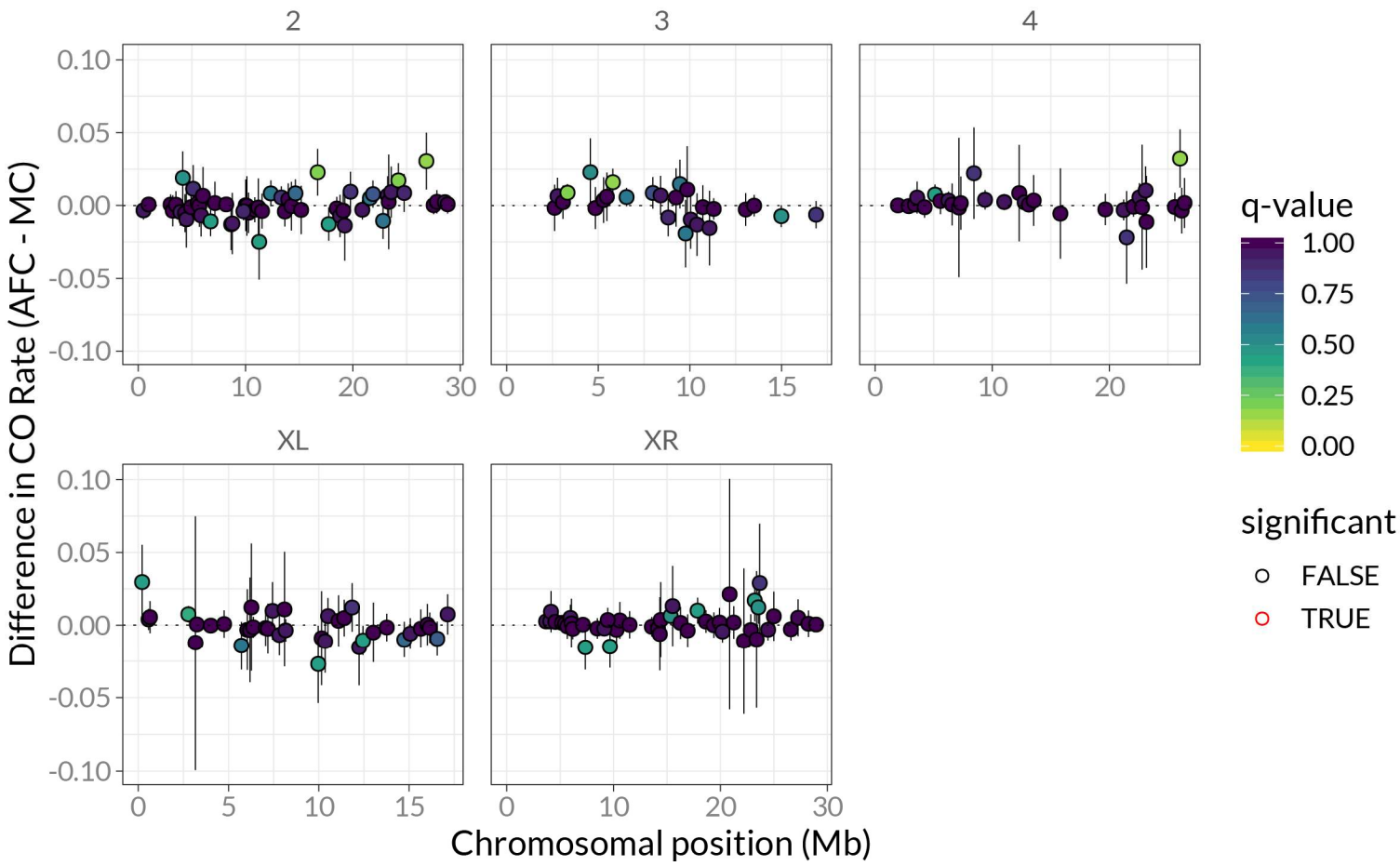


Figure S4 | Differences in recombination rate (AFC minus MC) for individual recombination intervals along the genome (position on x-axis). Each point represents the mean difference in rate for a single interval (all inbred lines considered together, n = 12 for all AFC intervals, n = 7 for all MC intervals), with vertical lines depicting 95% confidence intervals. Points are colored by their q-value (FDR corrected p-value) for a test comparing the to a null value of zero, with red outlines depicting intervals that exceeded intervals where $q < 0.05$ (of which there are none, in this case).

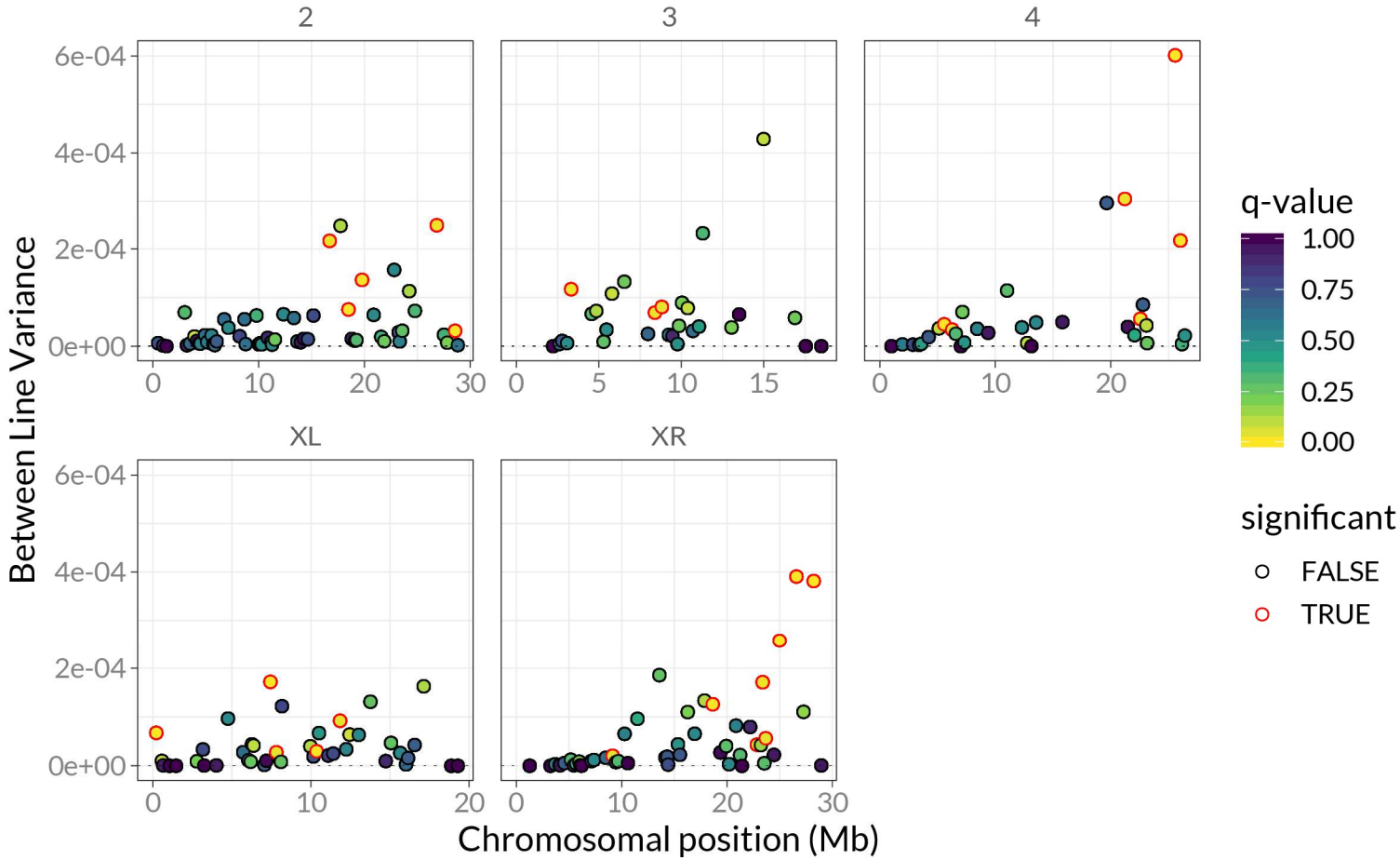


Figure S5 | Between line variance (variance in recombination rate for all inbred lines, y-axis) for individual recombination intervals along the genome (x-axis). Each point represents the interline variance for a single interval (all inbred lines considered together, n = 12 for all AFC intervals, n = 7 for all MC intervals). Points are colored by their q-value (FDR corrected p-value) derived from an ANOVA (F-test) with crossover rate as the response and inbred line as the explanatory variable, with red outlines depicting intervals where $q < 0.05$.

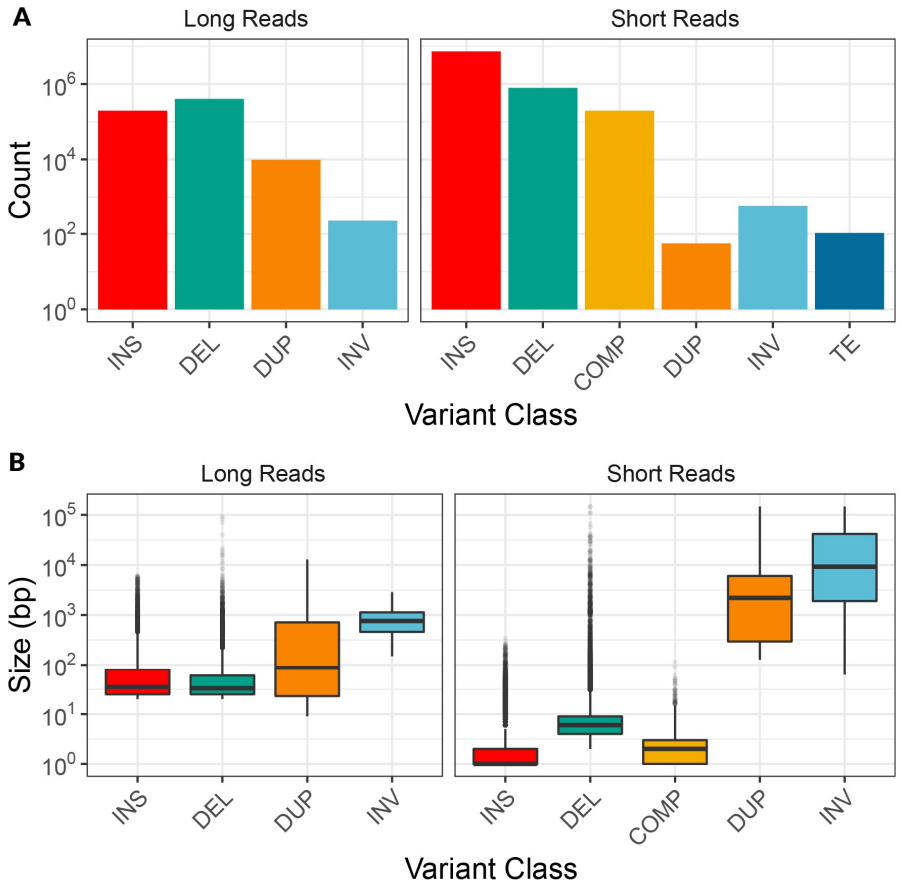


Figure S6| Summary of the count (A) and size in base pairs (B) of structural variants identified via short and long read sequencing. Structural variant classes are: insertions (INS), deletions (DEL), complex indels (COMP), duplications (DUP), inversions (INV), transposable elements (TE).