

Recombining without Hotspots: A Comprehensive Evolutionary Portrait of Recombination in Two Closely Related Species of *Drosophila*

Caiti S. Smukowski Heil^{1,2,*}, Chris Ellison³, Matthew Dubin¹, and Mohamed A. F. Noor¹

Running title: Recombination rate evolution in two species of *Drosophila*

Author Affiliations:

¹Biology Department, Duke University; Box 90338, Durham, NC 27708

²Genome Sciences Department, University of Washington; Foege Building S-250, 3720 15th Ave NE, Seattle WA 98195

³ Department of Integrative Biology, University of California, Berkeley; 3060 VLSB, Berkeley, CA 94720-3140

*Corresponding author

Email: cssh@uw.edu

Phone: 206-616-2856

Abstract

Meiotic recombination rate varies across the genome within and between individuals, populations, and species in virtually all taxa studied. In almost every species, this variation takes the form of discrete recombination hotspots, determined in Metazoans by a protein called PRDM9. Hotspots and their determinants have a profound effect on the genomic landscape, and share certain features that extend across the tree of life. *Drosophila*, in contrast, are anomalous in their absence of hotspots, PRDM9, and other species-specific differences in the determination of recombination. To better understand the evolution of meiosis and general patterns of recombination across diverse taxa, we present what may be the most comprehensive portrait of recombination to date, combining contemporary recombination estimates from each of two sister species along with historic estimates of recombination using linkage-disequilibrium-based approaches derived from sequence data from both species. Using *Drosophila pseudoobscura* and *Drosophila miranda* as a model system, we compare recombination rate between species at multiple scales, and we replicate the pattern seen in human-chimpanzee that recombination rate is conserved at broad scales and more divergent at finer scales. We also find evidence of a species-wide recombination modifier, resulting in both a present and historic genome wide elevation of recombination rates in *D. miranda*, and identify broad scale effects on recombination from the presence of an inter-species inversion. Finally, we reveal an unprecedented view of the distribution of recombination in *D. pseudoobscura*, illustrating patterns of linked selection and where recombination is taking place. Overall, by combining these estimation approaches, we highlight key similarities and differences in recombination between *Drosophila* and other organisms.

Author Summary

Recombination, or crossing over, describes an essential exchange of genetic material that occurs during egg and sperm development and has consequences for the proper segregation of chromosomes, and for the evolution of genomes and genomic features. In our study, we compare genome wide recombination rate in two closely related species of the fruit fly *Drosophila* to understand if and how recombination changes over time. We find that recombination does indeed change, we observe globally increased recombination in one species, and differences in regional recombination likely reflecting the result of a chromosomal rearrangement in both species. Moreover, we show that the extent that recombination changes is dependent on the physical scale at which recombination is measured, likely reflecting selection pressures on recombination distribution and replicating a pattern seen in human-chimpanzee recombination. Apart from between-species differences, we note several ways in which the *Drosophila* recombination landscape has changed since *Drosophila* diverged from other organisms. In contrast to species of fungi, plants, and animals, *Drosophila* recombination is not concentrated in discrete regions known as hotspots, nor is it increased near the start of genes, suggesting that despite the importance of the recombination process, the determinants of recombination have been shifting over evolutionary time.

Introduction

Homologous meiotic recombination is a crucial mechanistic and influential evolutionary process. The creation of a physical link between homologous chromosomes during meiosis I promotes proper segregation of chromosomes, preventing aneuploidy and cell death. Evolutionarily, recombination breaks down linkage between loci, allowing sites to segregate independently, and thereby facilitating adaptation and the purging of deleterious mutations. Indeed, recombination is often touted as the fundamental benefit of sexual reproduction. Nonetheless, measuring recombination elicits many challenges. Researchers have used a myriad of ways to estimate recombination rate, which can broadly be grouped into two categories: assessing direct, current rates or capturing indirect, historical rates (1, 2). The most common approach for assessing direct, current rates entails genotyping individuals from known pedigrees or controlled crosses, which provides a coarse estimate of recombination rate in a single meiosis or small number of meioses. However, this method is costly, laborious, and very difficult to achieve fine scale resolution, often involving a trade-off between number of markers and number of individuals to survey. In contrast, with the rise of population sequencing data over the past decade, a statistical approach has been developed and now widely used that estimates the population recombination parameter, ρ , to computationally infer genome wide, population-averaged recombination rate (3-5). The approach can achieve very fine scale resolution, unmasking genomic features and leading to the discovery of determinants of recombination rate variation.

The development of recombination maps in a wide range of species through both direct and indirect methods has dramatically enhanced our understanding of where recombination takes place in the genome, what factors are responsible, and the selective pressures governing changes in recombination rate over time. For example, one of the most fundamental patterns observed is the existence of variation in recombination rate across the genome, and between populations and species in

every organism examined (6-9). Moreover, this variation in recombination rate is non-random, and dependent upon the physical scale at which it is examined. For instance, in many organisms, the broad scale recombination landscape is shaped by large scale structural features of the chromosomes, while the fine scale variation manifests as “hotspots”: discrete genomic regions, typically 1-2 kilobases (kb) in length in which recombination increases a magnitude fold above the background rate (2, 10-12).

With an increasing number of recombination maps developed for related species, it has become apparent that selective pressures governing recombination rate are also dependent on physical scale. Recombination rates observed at the megabase scale or greater exhibit conservation between closely related species (6, 8, 13-18), while recombination hotspots show nearly no overlap between closely related species (19-21). Some of the conservation at broad scales may be due to the necessity for at least one crossover per chromosome to ensure proper segregation of chromosomes during meiosis (8, 22-24). This mandates a minimum level of recombination, as a crossover rate that is too low is likely to be highly deleterious. In contrast, too much recombination breaks apart advantageous allele combinations, threatens genomic integrity, and may introduce mutations at hotspots (2, 12, 25). Therefore, on a broad scale, recombination is likely constrained between a lower bound dictated by chromosome segregation, and an upper bound determined by maintaining genomic integrity. At fine scales, the recombination landscape is shaped by a phenomenon known as the “hotspot paradox” (26-28), which describes the self-destructive nature of the resolution of meiotic crossover products. A method to counteract the loss of hotspots was unknown until the recent discovery of the zinc finger histone-methyltransferase, *Prdm9*, which controls the distribution of recombination in mice and humans (29, 30). Positive selection on *Prdm9*'s zinc fingers has created new DNA binding residues, thereby initiating new populations of hotspots (29, 31-36), and resulting in rapid turnover of hotspots and fine scale recombination rate divergence between individuals, populations, and species who have *Prdm9* (16, 21, 31-34, 36-41).

Furthermore, *Prdm9* may also determine the distribution of recombination events across the genome, possibly directing recombination machinery away from promoters and other functional genomic elements (33, 40, 42). In some of the organisms lacking *Prdm9* such as dogs, plants, and yeast, (35, 43, 44), hotspots are often found overlapping promoter regions/transcription start sites (TSS), presumably due to the open chromatin structure allowing recombination machinery access to the DNA (11, 45-50). These observations set up a model in which organisms without *Prdm9* are opportunistic: they form recombination hotspots in “windows of opportunity,” most often nucleosome depleted regions around promoters/TSS. On the other hand, in organisms with *Prdm9*, *Prdm9* most likely functions to introduce meiotic specific H3K4me3 modifications and directs recombination machinery to initiate at these locations away from genomic elements (51, 52).

Drosophila stands out from these systems in its unique recombination landscape. Like yeast and plants, *Drosophila* are missing *Prdm9* (53), but unlike these species, *Drosophila* lack highly localized recombination hotspots, the only system known to be missing such hotspots besides *C. elegans* (54-59). Moreover, *Drosophila* recombination is sex-specific (males don't experience crossing over), and chromosome pairing and synapsis proceed normally in its absence (60-62). *Drosophila* recombination poses an intriguing system to further explore such questions as: How is recombination distributed in a species lacking hotspots? Are recombination rates conserved or divergent at varying scales? Are there additional differences between *Drosophila* recombination and recombination in other taxa?

To address these questions, we have developed both empirical and LD-based estimates of recombination in each of two closely related species, *Drosophila pseudoobscura* and *Drosophila miranda*, to better understand the recombination landscape and how it changes over evolutionary time. Empirical estimates of recombination rate are available for both species from McGaugh et al. (63), and here, we present genome wide LD-based estimates of recombination (via LDhelmet (64)). We show that LD-based recombination maps generally resemble current recombination rates in these species, and

demonstrate general conservation between species over three million years divergence. We find that *D. miranda* recombination rates have been and still are higher than those of *D. pseudoobscura*, suggestive of a global recombination modifier. We uncover potential historic effects of an inversion on broad-scale recombination rate, and similar to mice and humans, recombination rates are more conserved between species at larger genomic scales (>100kb) than finer genomic scales. Our fine scale recombination rate estimates illuminate where recombination takes place in the *Drosophila* genome, and confirm the lack of highly localized recombination hotspots in this taxon. In total, this work highlights key similarities and differences of *Drosophila* recombination compared to other organisms, adding a more thorough understanding to the evolution of meiosis and recombination across disparate taxa.

Results and Discussion

How well do Linkage Disequilibrium-based linkage maps reflect present-day measures of recombination?

The use of statistical programs on population sequencing data has drastically accelerated the study of recombination rate variation, particularly at fine scales. However, while Linkage Disequilibrium (LD)-based inference of recombination rate is a powerful tool, this method has its drawbacks too: linkage disequilibrium is influenced by many other factors apart from recombination, including demography, selection, and genetic drift (5). Additionally, linkage disequilibrium represents a historical population averaged recombination rate while empirical methodologies typically survey only one population for one generation of meiosis. Thus, LD-based maps can potentially deviate from direct, single-generation empirical estimates of recombination for several reasons (65-68). Therefore, to first clarify how well our LD-based maps replicate empirical recombination rates, we created genome wide LD-based recombination maps for *D. pseudoobscura* and *D. miranda*. The LD-based recombination maps were made with the program LDhelmet (64), similar to the widely used program LDhat (4), but explicitly designed for *Drosophila* to account for biological and technical differences in the datasets, such as a

much higher background recombination rate than human. We confirmed that estimates of recombination from LDhelmet correlated strongly with a simple direct measure of linkage disequilibrium decay in our dataset (**S1 Fig.**). We then compared the LD-based recombination estimates to those from the empirical maps generated by McGaugh *et al.* (63). The empirical maps were generated by genotyping SNP markers at 150-200 kb intervals for several of the largest chromosome groups for two populations of *D. pseudoobscura*, and one of *D. miranda*, and several selected regions of 20kb and 5 kb of chromosome 2 for one population of *D. pseudoobscura* (55).

Recombination rates obtained from these two strategies are significantly correlated at a broad scale (~200kb, **Table 1, Fig. 1, S2-S4 Figs**), although empirical-LD-based correlations are generally stronger in *D. pseudoobscura*. A possible explanation for this difference is that LD-based rates typically don't capture extreme values seen in the empirical landscape (**Fig. 1, S2-S4 Figs**), which may explain why the correlation is weaker in the elevated recombination map of *D. miranda*. The particular strains used in the *D. miranda* empirical study may also have an atypical recombination distribution relative to the historical population, or that there could be greater among-strain variation across *D. miranda*. *D. miranda* XLgroup1a in particular stands as an outlier ($r=0.227$), although the empirical map in this region is broken into three pieces due to failure of markers, potentially introducing more error into these estimates. We also have empirical estimates for a small portion of *D. pseudoobscura* at finer scales of 20kb and 5kb; the correlation between empirical and LD-based maps persists at 20 kb ($r=0.68$, **S5 Fig.**), but there is too little data at the 5kb scale to be conclusive (**S6 Fig.**).

Fig. 1. A comparison of empirical and LD-based estimations of recombination rate. Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) across chromosome 2 for *D. pseudoobscura* Flagstaff (top) and *D. miranda* (bottom). All recombination rates

are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (black, solid).

Table 1. Empirical vs. LD-based comparisons of recombination rate.

	Interval size Average, Median (Mb)	Number of intervals	Average empirical recombination rate (cM/Mb)	Average LD- based recombination rate (cM/Mb)	Correlation (Pearson's r, p-value)
2					
<i>D. pseudoobscura</i> Flagstaff	0.215, 0.142	140	3.55	3.71	0.624, p<0.0001
<i>D. pseudoobscura</i> Pikes Peak	0.194, 0.149	158	3.47	3.75	0.658, p<0.0001
<i>D. miranda</i>	0.189, 0.147	154	4.85	4.87	0.639, p<0.0001
XLgroup1a					
<i>D. pseudoobscura</i> Flagstaff	0.215, 0.148	7	3.76	5.17	-
<i>D. pseudoobscura</i> Pikes Peak	0.211, 0.198	24	3.32	3.47	0.657, p=0.0005
<i>D. miranda</i>	0.247, 0.202	18	4.95	5.82	0.227, p=0.1806
XRgroup6					
<i>D. pseudoobscura</i> Flagstaff	0.164, 0.143	68	3.94	3.66	0.710, p<0.0001
<i>D. pseudoobscura</i> Pikes Peak	0.166, 0.146	71	4.44	4.55	0.719, p<0.0001
<i>D. miranda</i>	0.163, 0.149	74	5.40	5.67	0.506, p<0.0001
XRgroup8					
<i>D. pseudoobscura</i> Flagstaff	0.269, 0.172	32	4.52	4.03	0.750, p<0.0001
<i>D. pseudoobscura</i> Pikes Peak	0.248, 0.209	36	4.72	4.47	0.682, p<0.0001
<i>D. miranda</i>	0.200, 0.187	34	6.12	6.06	0.403, p=0.0182

Empirical recombination rates were obtained from McGaugh *et al* 2012 . LD-based recombination rates were averaged over empirical windows and converted to cM/Mb (see Methods). A dash indicates not enough data to assess the correlation.

Few studies have compared empirical and LD-based estimates, but similar results were obtained from LDhelmet in *D. melanogaster* (Raleigh, NC population; ~200kb: 2L $r=0.73$, 2R $r=0.75$, 3L $r=0.61$, 3R $r=0.73$, X $r=0.71$) (64). Chan *et al.* used data from Singh *et al.* (57) and FlyBase (69), at resolutions of approximately 100-200kb, so we present the first data obtained from LDhelmet demonstrating correspondence in LD-based and empirical rates at scales under 100kb. One of the only other comparisons is from human, where correlations between LD-based and empirical recombination rates are nearly replicates at the megabase scale ($r=0.98$)(42). Our results and those of Chan *et al.* clearly exhibit a lower correlation than observed in human, however, our comparisons are done at the ~200 kb scale where there is more variation in recombination rate (see below, **Table 3**). Furthermore, empirical-empirical correlations between different populations of *D. pseudoobscura* are only moderately higher than the LD-based-empirical correlations (**Table 2**), suggesting that the LD-based maps are capturing much of the recombination variation in these species.

Table 2. Empirical recombination rate comparisons.

	Interval size Average, Median (Mb)	Number of intervals	Correlation (Pearson's r)
2			
<i>D. pseudoobscura</i> Flagstaff – Pikes Peak	0.262, 0.170	115	0.718 p<0.0001
<i>D. pseudoobscura</i> Flagstaff - <i>D. miranda</i>	0.247, 0.152	115	0.621 p<0.0001
<i>D. pseudoobscura</i> Pikes Peak- <i>D. miranda</i>	0.235, 0.150	123	0.669 p<0.0001
XRgroup6			
<i>D. pseudoobscura</i> Flagstaff – Pikes Peak	0.232, 0.202	47	0.916 p<0.0001
<i>D. pseudoobscura</i> Flagstaff - <i>D. miranda</i>	0.228, 0.194	48	0.656 p<0.0001
<i>D. pseudoobscura</i> Pikes Peak- <i>D. miranda</i>	0.197, 0.164	60	0.617 p<0.0001
XRgroup8			
<i>D. pseudoobscura</i> Flagstaff – Pikes Peak	0.388, 0.280	21	0.873 p<0.0001
<i>D. pseudoobscura</i> Flagstaff - <i>D. miranda</i>	0.297, 0.257	20	0.776 p<0.0001
<i>D. pseudoobscura</i> Pikes Peak- <i>D. miranda</i>	0.296, 0.249	23	0.612 p<0.0001

Empirical recombination rates were obtained from McGaugh *et al* 2012. Intervals were condensed between each pairing to correctly assess recombination rates as per McGaugh *et al* 2012.

Table 3. LD-based comparison of recombination rates between *D. pseudoobscura* and *D. miranda*.

Chromosome	Size (bp)	Scale			
		50kb	100kb	250kb	500kb
2	30794191	0.479 p<0.0001	0.499 p<0.0001	0.535 p<0.0001	0.510 p<0.0001
4		0.617 p<0.0001	0.668 p<0.0001	0.710 p<0.0001	0.786 p<0.0001
group1	5278962	0.365 p<0.0001	0.409 p=0.0024	0.403 p=0.0697	-
group2	1235211	0.424 p=0.0347	-	-	-
group3	11692073	0.660 p<0.0001	0.720 p<0.0001	0.774 p<0.0001	0.842 p<0.0001
group4	6587037	0.593 p<0.0001	0.643 p<0.0001	0.762 p<0.0001	-
XL		0.462 p<0.0001	0.485 p<0.0001	0.557 p<0.0001	0.669 p<0.0001
group1a	9151740	0.592 p<0.0001	0.608 p<0.0001	0.702 p<0.0001	0.930 p<0.0001
group1e	12523135	0.168 p=0.0077	0.182 p=0.0418	0.173 p=0.2303	0.151 p=0.4705
group3a	2690836	0.730 p<0.0001	0.790 p<0.0001	0.822 p=0.0019	-
group3b	386183	-	-	-	-
XR		0.529 p<0.0001	0.617 p<0.0001	0.654 p<0.0001	0.652 p<0.0001
group3a	1468912	0.557 p=0.0014	0.673 p=0.0059	-	-
group5	740492	0.039 p=0.8898	-	-	-
group6	13314419	0.532 p<0.0001	0.638 p<0.0001	0.693 p<0.0001	0.717 p<0.0001
group8	9212921	0.614 p<0.0001	0.720 p<0.0001	0.756 p<0.0001	0.716 p<0.0001

The Pearson correlation coefficient and p-value for LD-based recombination rates of *D. pseudoobscura* and *D. miranda*. Recombination rates are averaged over a given interval size, first for the whole chromosome (all chromosome groups concatenated), then for each chromosome group individually. Dashes indicate not enough data to assess the correlation.

Recombination rates between species

Broad scale recombination rate patterns

Present-day empirically assayed recombination rate was previously shown to be conserved at broad scales (~200kb) between *D. pseudoobscura* and *D. miranda* (**Table 2**) (63), and with our LD-based recombination maps, we recapitulate and extend these results. We detect and confirm genome wide broad scale conservation of recombination rates between species (**Table 3, Fig. 2, S8-S9 Figs**), but with several notable exceptions. First, *D. miranda* rates remain elevated relative to *D. pseudoobscura*, supporting the conclusion that the putative global recombination modifier identified in empirical work persists species-wide in *D. miranda* (63). Estimates of the effect of the modifier are quite similar between empirical and LD-based data, recombination rate is 1.26x higher on chromosome 2 in *D. miranda* (compared to 1.28-1.32 higher from empirical work), and 1.71x higher on chromosome XR (empirical: 1.4-1.47). Moreover, this difference persists despite a difference in effective population sizes: *D. pseudoobscura* N_e is thought to be several times that of *D. miranda* (70-72).

Fig. 2: A comparison of LD-based estimations of recombination rate between two species. LD-based *D. pseudoobscura* recombination rates (red) are plotted with LD-based *D. miranda* recombination rates (blue) across the genome. Average rates were calculated for 100kb window sizes. Chromosomes 4, XL, and XR are split in to multiple groups, each labeled on the x-axis, according to the reference assembly for *D. pseudoobscura* (denoted with alternating background color). All recombination rates are reported in ρ /bp. Error bars are not shown for clarity.

Another pattern apparent in the broad scale recombination landscape is that some chromosome segments exhibit a weaker correlation of recombination rate between species, potentially indicative of regional differences. For example, chromosome arm XL group1e is a clear outlier in the comparison

between *D. pseudoobscura* and *D. miranda* (Table 3, Fig. 2, S8-S9 Figs). Interestingly, the species differ by two inversions on the XL chromosome arm: a large 12 Mb inversion that corresponds almost precisely with group1e and a small, ~4Mb inversion that took place following the large inversion at the distal end of group1e (73, 74). This is suggestive that an inversion caused historic broad scale changes in recombination rate.

Fine scale recombination rate patterns

While recombination rates are generally conserved at broad scales, they are divergent at fine scales in the few organisms in which they have been examined. Analysis of fine scale recombination in *Drosophila* has been limited to isolated genomic regions in *D. melanogaster* (56-58, 75) and *D. pseudoobscura* (55, 63, 76), and has not been compared between species. With LD-based maps, we are able to characterize the genome-wide fine-scale recombination landscape in *D. pseudoobscura* and *D. miranda* for the first time, and to test whether turnover in fine scale recombination rates occurs outside of mammals.

By comparing recombination maps across different scales, variation and amplitude in recombination rate appears to increase at finer scales: for every chromosome assembly but one, the correlation coefficient between the species was weaker in 50kb windows than in every larger window size (100kb, 250kb, or 500kb: see Table 3). This phenomenon was previously noted in several *Drosophila* recombination studies (58, 63). In McGaugh *et al.* (63), the 20kb recombination map revealed recombination rates ranging from 3.5–21.2 cM/Mb, while the ~200kb map measured rates of 5.6 and 4.4 cM/Mb covering the same region(17.5–17.7 Mb of chromosome 2). If we examine this same region in the LD-based maps (S7 Fig.), at position 17.5 Mb with resolution of 50kb, the recombination rate in *D. pseudoobscura* is 0.32 ρ /bp, while in the 500kb resolution map, the recombination rate is 0.16 ρ /bp, mirroring the empirical results. More generally, recombination rates peak at 0.32 ρ /bp in the 50kb

chromosome 2 map for both species, and top out at 0.20 ρ /bp (*D. miranda*) and 0.14 ρ /bp (*D. pseudoobscura*) in the 1 Mb chromosome 2 map. It follows that correlations of recombination rate between two species are significantly higher at 500kb scale than at 50 kb scale ($p=0.0094$), illustrating stronger conservation at broad scales (**Table 3, S7-S9 Figs**). It appears that even in the absence of *Prdm9* and hotspots, there is less constraint at finer scales, although the mechanism behind this in *Drosophila* remains unclear.

While our study achieves the finest scale genome wide recombination map of both *D. pseudoobscura* and *D. miranda* to date, our resolution is limited biologically due to a lower SNP density in *D. miranda*, unfortunately restricting the level at which we can compare rates between species. The effective population size of *D. miranda* is estimated to be up to five times smaller than *D. pseudoobscura*, and number of segregating sites almost half of *D. pseudoobscura* ($\theta_{pse} = 0.01$, $\theta_{mir} = 0.0058$). Because statistical programs like LDhelmet depend on SNP density to estimate recombination, we are prohibited from examining recombination rates at a scale finer than 50kb in *D. miranda*.

However, we were still able to examine recombination rates at finer scales in *D. pseudoobscura*, and identified several putative hotspots in a similar manner to Chan *et al.* (64). Chan *et al.* detected 21 total hotspots in two different populations of *D. melanogaster* and were able to independently verify 10. In *D. pseudoobscura*, we discovered 19 regions between 500-5000bp where the recombination rate was more than ten times greater than the chromosomal recombination average (**S10 Fig.**). However, we did not verify these *D. pseudoobscura* hotspots independently with alternative means, so their presence should be seen as provisional, and these isolated hotspots remain both far less in total number and in magnitude than those seen in other organisms. This conclusion remains consistent with previous data,

confirming that the recombination landscape of *Drosophila* possesses fine scale variation, but does not exhibit recombination hotspots like those found in humans, yeast, and others.

The distribution of recombination across the genome

Looking at the proportion of recombination that occurs in a proportion of sequence reveals the degree to which recombination events cluster across the genome. In humans, recombination is highly punctate, with about 80% of recombination happening in less than 20% of the sequence (77), or a Gini coefficient of approximately 0.8 (59, 78). As previously discussed, *Drosophila* lack hotspots characteristic of other organisms (55, 57, 58), although fine scale variation has been identified (see above), and Chan *et al.* even detected a handful of more typical hotspots (64). This is reflected in *D. pseudoobscura*'s intermediate Gini coefficient of approximately 0.50, or about 80% of recombination occurring in 50% of the sequence (**Fig. 3**). This is similar to the Gini coefficient estimated for *D. melanogaster* (0.47), falling between *C. elegans* (0.28) and *S. cerevisiae* (0.64) (59). This supports the hypothesis that recombination in *Drosophila* is more evenly distributed across the genome, with no region experiencing a near-complete absence of recombination, as compared to humans which exhibit large stretches of essentially no recombination events (77).

Figure 3: The concentration of fine-scale recombination rate in *D. pseudoobscura*.

Another way to examine the distribution of recombination is to examine recombination rate near genomic features, such as transcription start sites (TSS). As previously discussed, organisms without *Prdm9* exhibit an increase in recombination around TSS, most likely due to reduced nucleosome occupancy in these regions. However, in *D. pseudoobscura*, recombination is reduced around TSS and adjacent regions up to 35 kb away (**Fig. 4**). This is supported by findings in *D. melanogaster* as well (64), although the reduction in rates appears to persist for greater distances in *D. pseudoobscura*. Similarly, we see a slight reduction in recombination rate at the 5' end of genes, which marginally increases with

distance from the start of the coding region (**S11 Fig.**). Recombination rate within a gene (0.060 ρ /bp) is notably decreased from the average recombination rate of 0.098 ρ /bp.

Figure 4: Recombination near transcription start sites in *D. pseudoobscura*. Recombination rate was averaged for 5kb windows for 200kb flanking transcription start sites (TSS) across the genome.

Patterns of linked selection

One of the most ubiquitous patterns in molecular evolution is the association between recombination rate and nucleotide diversity, first observed in *Drosophila melanogaster* and since reproduced in many species, including several additional species of *Drosophila* (6, 79). The association with diversity, but not nucleotide divergence, has been interpreted as evidence that selective sweeps and/or background selection are eliminating diversity in regions of low recombination (80-85). Previous research has observed these associations at scales ranging from 50kb to 1.5 Mb, but failed to detect an association below 50kb (18, 58, 79, 86-88). This could potentially be explained by the size of the footprint of a selective event, such as a selective sweep, which may be undetectable at fine scales. To characterize the correlation between recombination and diversity at varying scales, we assessed diversity at four-fold degenerate positions binned into windows ranging from 5kb to 1Mb (**S12 Fig.**). Similar to Singh *et al.* (58), we detect a weak correlation between recombination and diversity at very fine scales, but this correlation increases in strength with window size so that at 1 Mb, recombination explains 36% of the variance in diversity. This pattern can be explained in part by the smoothing of stochasticity in diversity and recombination rate measures in large windows, and is potentially confounded by the method of recombination estimation depending on SNP density, although this method has been used for this purpose many times (64, 83, 89). We were able to re-test the correlation for a sample of windows using thinned SNP data (excluding 50% of SNPs), and did not find it to alter our

results. We therefore conclude that patterns of linked selection are indeed apparent in *D. pseudoobscura*, and the power of this selection is strongest at large scales.

Conclusions

With our genome wide fine scale recombination maps for two closely related species, we were able to identify both shared and exclusive patterns of recombination across different taxa. For example, without hotspots, it was unclear whether *Drosophila* would recapitulate the pattern observed in several other species, in which broad scale recombination rates are more conserved than fine scale recombination rates. While we could not compare rates at the 1-2kb level, by examining correlations at levels from 500kb to 50kb, we can conclude that recombination rates are less conserved at fine scales across the genome, one of the only studies to show this phenomenon exists outside of mouse and human. This raises interesting implications about selection pressures on recombination that persist across distantly related taxa with widely different recombination landscapes, supporting a relaxed constraint on rates at finer scales.

To illuminate potential strain variation or local changes in recombination rate over time, we can compare historical and present day recombination rates between species (**Tables 1-3**). Present recombination rates between species are moderately well correlated (**Table 3**), suggesting that recombination rate has not experienced drastic changes since these species diverged, although this analysis is limited to chromosome 2 and parts of XRgroup6 and XRgroup8. Comparing historical rates (LD-based) between species at a similar scale (**Table 2**), we see a decrease in correlation strength compared to present day, some regions quite dramatically. Generally, this likely represents variability across strains, as the LD-based recombination rates portray a population average. In specific regions like 4group1 and XLgroup1e, which seem to exhibit very different historical recombination patterns between species, this may indicate species specific local changes to recombination rate. As discussed above, one

potential driver of this change is chromosomal rearrangements, such as inversions. Fittingly, an inversion between *D. miranda* and *D. pseudoobscura* overlaps almost exactly with the region of greatest divergence on chromosome arm XLgroup1e. This phenomenon has been previously documented in human-chimpanzee broad scale correlations, where inverted regions showed a lower correlation in recombination rate than non-inverted regions (21), supporting the conclusion that inversions influence broad scale patterns of recombination across multiple species.

Another potential modifier of recombination, originally implicated in empirical work (63), was confirmed here by observing genome wide elevated recombination in historical recombination rates of *D. miranda* despite a much lower effective population size. This new data provides compelling evidence that the predicted recombination modifier is a species wide trait. Of course, modifiers of recombination are well known to *Drosophila* (90, 91), and the theory of conditions that favor increased (or decreased) recombination is rich (92-99). The selection pressures that *D. miranda* has faced are unknown, and unfortunately, determining the locus or loci responsible for an increase in recombination in this species is not possible through traditional mapping methods, as *D. pseudoobscura* and *D. miranda* do not interbreed.

Of course one of the most recent, well-studied determinants of recombination, *Prdm9*, is absent in *Drosophila*, just one of many other differences that set *Drosophila* recombination apart from other organisms. As has already been discussed, *Drosophila* lack hotspots of recombination, and exhibit a more consistent distribution of recombination than yeast, mouse, and human. Additionally, recombination is decreased around the start of genes, in contrast to other organisms without *Prdm9*. These results add further evidence to long known differences in *Drosophila* meiosis, which can largely be attributed to the loss of recombination in males. Whether there's a link between the evolution of sex-specific recombination and the patterns observed more recently is unclear, but the decreased cost of

sequencing and computational approaches like LDhelmet may help elucidate this question by characterizing the recombination landscape of other sex-specific recombination systems, among other questions such as why sex-specific recombination has evolved so many times, and why recombination hotspots were lost in the lineage leading to *Drosophila* when they persist across fungi, plants, and metazoans.

In conclusion, the creation of recombination maps will continue to illuminate new perspectives on the evolution of recombination, shedding light on novel determinants of recombination, patterns of selection, and genome evolution.

Materials and Methods

Population sequencing and variant calling

We used 11 whole genome sequences of *D. pseudoobscura* (MV2-25 (reference), Mather32, MSH24, MSH9, TL, PP1134, PP1137, AFC12, FLG14, FLG16, FLG18) previously sequenced and analyzed by McGaugh et al. (63) and which are available at <http://pseudobase.biology.duke.edu> and GenBank (accession numbers: AFC12 SRX091462, Flagstaff14 SRX091308, Flagstaff16 SRX091303, Flagstaff18 SRX091310, Mather32 SRX091461, MatherTL SRX091324, MSH9 SRX091465, MSH24 SRX091463, PP1134 SRX091323, PP1137 SRX091311). Additionally, we used 11 whole genome sequences of *D. miranda* (Lines: MA28, MAO101.4, MAO3.3, MAO3.4, MAO3.5, MAO3.6, ML14, ML16, ML6f, SP138, and SP235), previously sequenced and provided by Doris Bachtrog's lab. There is little to no population structure present in *D. pseudoobscura*, allowing us to treat flies collected from various regions across the North American west as panmictic (100-102). All sequences of *D. pseudoobscura* and *D. miranda* were aligned to the *D. pseudoobscura* reference sequence v2.9 using bwa (103). Variants were called for

the *D. miranda* genome sequences using the Genome Analysis Tool Kit (GATK) v2.7-2 to remove duplicates and locally realign reads (104, 105), and samtools v0.1.18 to call single nucleotide variants (106). We used custom Perl scripts to convert the variant output files into aligned species-specific individual chromosome FASTA files for input to LDhelmet, see below (2, 4group1, 4group2, 4group3, 4group4, XLgroup1a, XLgroup1e, XLgroup3a, XLgroup3b, XRgroup3a, XRgroup5, XRgroup6, XRgroup8). Note that chromosome 3 was not included in the analysis due to known segregating inversions in *D. pseudoobscura* (107), and chromosome 4group5 was excluded due to poor sequence coverage.

Recombination estimates

Empirical estimates of recombination for *D. pseudoobscura* and *D. miranda* were obtained from McGaugh et al. (63)(**Table 2**). Population sequencing based estimates of recombination were determined using the LDhelmet program (64), a statistical approach designed for *Drosophila* which estimates the population recombination parameter ρ using a reversible jump Markov Chain Monte Carlo mechanism (rjMCMC), as in LDhat (3, 4). The program calculates $\rho=4N_e r$ (where N_e is the effective population size and r is the recombination rate per generation) to estimate the amount of recombination needed in the population to produce the observed levels of linkage disequilibrium under a given model. LDhelmet is specifically tailored to address issues in *Drosophila* that may be problematic for programs such as LDhat, like a magnitude fold higher background recombination rate, higher SNP density, and a large portion of the genome influenced by positive selection (82, 108). The program was run individually for each chromosome of each species using default parameters, with the exception that we estimated theta from our dataset ($\theta_{pse} = 0.01$, $\theta_{mir} = 0.0058$) and created our own mutation transition matrices. Following the procedure of Chan et al., we ran the program for 1,000,000 iterations after a burn-in of 100,000 iterations, and used a conservative block penalty of 50.

To compare empirical recombination estimates to LD-based recombination estimates, the recombination estimate from LDhelmet was corrected for distance and averaged over a given interval. The empirical chromosomal recombination average (cM/Mb) was divided by the total average LD-based recombination rate for a chromosome to get a conversion factor. Each interval's average LD-based estimate was multiplied by the conversion factor to get an approximation of recombination rates in the units centiMorgan per Megabase (cM/Mb) per Chan *et. al* (64).

Analyses of genomic correlates

Transcription start sites (TSS): Locations of *D. pseudoobscura* transcription start sites (TSS) were obtained from Main *et al.* (109). The recombination estimates from LDhelmet were averaged over 5000bp intervals for the 200kb regions flanking either side of a TSS. Genome wide data was aggregated to create **Figure 3**. We confirmed that the effect seen around TSS is not due to a lower SNP density caused by conservation of promoter regions, by comparing SNP density in 20 kb regions flanking TSS to 20 kb regions where no genes are present.

Proportion recombination: A Perl script was used to sort recombination rates and distances to build a list of the proportion of recombination occurring in the proportion of sequence (Laurie Steverson, unpublished). This data was plotted using a Lorenz curve from the R package "ineq" (Achim Zeileis, Christian Kleiber). The Gini coefficient was also calculated using this package as in Kaur *et al.* (59).

Introns and Exons: Locations of exons and introns and relative positions in a gene were extracted from *D. pseudoobscura* v2.9 annotations from FlyBase (110). The recombination estimate from LDhelmet was corrected for distance and averaged over the given interval, then aggregated to give genome wide totals for each exon and intron position within a gene.

Nucleotide diversity: We calculated pairwise nucleotide diversity (π) at four-fold degenerate sites using custom Perl scripts, excluding sites where an insertion or deletion was found in any line.

Data archiving: All datasets used to create the figures or tables in this study are archived in Dryad, doi:10.5061/dryad.sh5c3.

Acknowledgements

Thanks To L. Stevison for helpful comments on the preparation of this manuscript. We are grateful to D. Bachtrog for sharing the *D. miranda* whole genome sequences with us. Our sincere thanks to Andrew H. Chan for his generous guidance and assistance with LDhelmet throughout the project.

References

1. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. *Nature reviews Genetics*. 2013;14(11):794-806.
2. Coop G, Przeworski M. An evolutionary view of human recombination. *Nature reviews Genetics*. 2007;8(1):23-34.
3. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*. 2002;160(3):1231-41.
4. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science*. 2004;304(5670):581-4.
5. Slatkin M. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*. 2008;9(6):477-85.
6. Smukowski CS, Noor MAF. Recombination rate variation in closely related species. *Heredity*. 2011;107(6):496-508.
7. Nachman MW. Variation in recombination rate across the genome: evidence and implications. *Current opinion in genetics & development*. 2002;12(6):657-63.
8. Dumont BL, Payseur BA. Evolution of the genomic rate of recombination in mammals. *Evolution*. 2008;62(2):276-94.
9. Paigen K, Petkov P. Mammalian recombination hot spots: properties, control and evolution. *Nature reviews Genetics*. 2010;11(3):221-33.
10. Lichten M, Goldman ASH. Meiotic recombination hotspots. *Annual Review of Genetics*. 1995;29:423-44.
11. Petes TD. Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*. 2001;2(5):360-9.
12. Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: recombination distributions in mammals. *Nature reviews Genetics*. 2004;5(6):413-24.
13. Backstrom N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, et al. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*. 2010;20(4):485-95.
14. Meznar ER, Gadau J, Koeniger N, Rueppell O. Comparative Linkage Mapping Suggests a High Recombination Rate in All Honeybees. *Journal of Heredity*. 2010;101:S118-S26.
15. Paigen K, Szatkiewicz JP, Sawyer K, Leahy N, Parvanov ED, Ng SHS, et al. The Recombinational Anatomy of a Mouse Chromosome. *PLoS genetics*. 2008;4(7):-.
16. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, et al. Fine-scale recombination patterns differ between chimpanzees and humans. *Nature genetics*. 2005;37(4):429-34.
17. McGaugh SE, Noor MA. Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philos Trans R Soc Lond B Biol Sci*. 2012;367(1587):422-9.
18. Stevison LS, Noor MAF. Genetic and Evolutionary Correlates of Fine-Scale Recombination Rate Variation in *Drosophila persimilis*. *Journal of Molecular Evolution*. 2010;71(5-6):332-45.
19. Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M. Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS Biol*. 2004;2(6):e155.
20. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, et al. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*. 2005;308(5718):107-11.
21. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, Street T, et al. A fine-scale chimpanzee genetic map from population sequencing. *Science*. 2012;336(6078):193-8.

22. Pardo-Manuel de Villena F, Sapienza C. Recombination is proportional to the number of chromosome arms in mammals. *Mamm Genome*. 2001;12(4):318-22.
23. Roeder GS. Meiotic chromosomes: it takes two to tango. *Genes Dev*. 1997;11(20):2600-21.
24. Fledel-Alon A, Wilson DJ, Broman K, Wen XQ, Ober C, Coop G, et al. Broad-Scale Recombination Patterns Underlying Proper Disjunction in Humans. *PLoS genetics*. 2009;5(9):-.
25. Rattray A, Santoyo G, Shafer B, Strathern JN. Elevated Mutation Rate during Meiosis in *Saccharomyces cerevisiae*. *PLoS genetics*. 2015;11(1):e1004910.
26. Boulton A, Myers RS, Redfield RJ. The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(15):8058-63.
27. Jeffreys AJ, Neumann R. Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature genetics*. 2002;31(3):267-71.
28. Jeffreys AJ, Neumann R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Human molecular genetics*. 2005;14(15):2277-87.
29. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*. 2010;327(5967):836-40.
30. Parvanov ED, Petkov PM, Paigen K. Prdm9 controls activation of mammalian recombination hotspots. *Science*. 2010;327(5967):835.
31. Berg IL, Neumann R, Lam KWG, Sarbajna S, Odenthal-Hesse L, May CA, et al. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature genetics*. 2010;42(10):859-+.
32. Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(30):12378-83.
33. Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. Genetic recombination is directed away from functional genomic elements in mice. *Nature*. 2012;485(7400):642-5.
34. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, et al. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*. 2010;327(5967):876-9.
35. Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, et al. Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. *PLoS genetics*. 2009;5(12):-.
36. Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, et al. PRDM9 Drives Evolutionary Erosion of Hotspots in *Mus musculus* through Haplotype-Specific Initiation of Meiotic Recombination. *PLoS genetics*. 2015;11(1):e1004916.
37. Hinch AG, Tandon A, Patterson N, Song YL, Rohland N, Palmer CD, et al. The landscape of recombination in African Americans. *Nature*. 2011;476(7359):170-U67.
38. Jeffreys AJ, Neumann R. The rise and fall of a human recombination hot spot. *Nature genetics*. 2009;41(5):625-9.
39. Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M. Absence of the TAP2 human recombination hotspot in chimpanzees. *Plos Biology*. 2004;2(6):849-55.
40. Smagulova F, Gregoret IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*. 2011;472(7343):375-8.
41. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, et al. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*. 2005;308(5718):107-11.
42. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. *Science*. 2005;310(5746):321-4.

43. Ponting CP. What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*. 2011;27(5):165-71.
44. Axelsson E, Webster MT, Ratnakumar A, Consortium L, Ponting CP, Lindblad-Toh K. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Res*. 2012;22(1):51-63.
45. Lichten M. Meiotic chromatin: the substrate for recombination initiation. In: R. Egel DHL, editor. *Recombination and Meiosis: Models, Means, and Evolution*. Berlin: Springer-Verlag; 2008. p. 165-93.
46. Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, et al. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nature genetics*. 2013;45(11):1327-36.
47. Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, et al. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(48):19478-82.
48. Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*. 2008;454(7203):479-U1.
49. Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, et al. A Hierarchical Combination of Factors Shapes the Genome-wide Topography of Yeast Meiotic Recombination Initiation. *Cell*. 2011;144(5):719-31.
50. Auton A, Rui Li Y, Kidd J, Oliveira K, Nadel J, Holloway JK, et al. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS genetics*. 2013;9(12):e1003984.
51. Grey C, Barthes P, Chauveau-Le Fric G, Langa F, Baudat F, de Massy B. Mouse PRDM9 DNA-Binding Specificity Determines Sites of Histone H3 Lysine 4 Trimethylation for Initiation of Meiotic Recombination. *Plos Biology*. 2011;9(10).
52. Hayashi K, Yoshida K, Matsui Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*. 2005;438(7066):374-8.
53. Heil CS, Noor MA. Zinc finger binding motifs do not explain recombination rate variation within or between species of *Drosophila*. *PLoS One*. 2012;7(9):e45055.
54. Hey J. What's so hot about recombination hotspots? *Plos Biology*. 2004;2(6):730-3.
55. Manzano-Winkler B, McGaugh SE, Noor MA. How hot are drosophila hotspots? examining recombination rate variation and associations with nucleotide diversity, divergence, and maternal age in *Drosophila pseudoobscura*. *Plos One*. 2013;8(8):e71582.
56. Miller DE, Takeo S, Nandan K, Paulson A, Gogol MM, Noll AC, et al. A whole-chromosome analysis of meiotic recombination in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics*. 2012;2(2):249-60.
57. Singh ND, Aquadro CF, Clark AG. Estimation of Fine-Scale Recombination Intensity Variation in the white-echinus Interval of *D-melanogaster*. *Journal of Molecular Evolution*. 2009;69(1):42-53.
58. Singh ND, Stone EA, Aquadro CF, Clark AG. Fine-scale heterogeneity in crossover rate in the garnet-scalloped region of the *Drosophila melanogaster* X chromosome. *Genetics*. 2013;194(2):375-87.
59. Kaur T, Rockman MV. Crossover heterogeneity in the absence of hotspots in *Caenorhabditis elegans*. *Genetics*. 2014;196(1):137-48.
60. Hawley R, Harris DT, Cui W, Kramer JJ, Page SL. Meiotic chromosome segregation in *Drosophila*. *Molecular Biology of the Cell*. 2002;13:150a-a.
61. Hawley RS, Irick H, Zitron AE, Haddox DA, Lohe A, New C, et al. There Are 2 Mechanisms of Achiasmate Segregation in *Drosophila* Females, One of Which Requires Heterochromatic Homology. *Developmental Genetics*. 1992;13(6):440-67.
62. McKim KS, Jang JK, Manheim EA. Meiotic recombination and chromosome segregation in *Drosophila* females. *Annu Rev Genet*. 2002;36:205-32.

63. McGaugh SE, Heil CS, Manzano-Winkler B, Loewe L, Goldstein S, Himmel TL, et al. Recombination modulates how selection affects linked sites in *Drosophila*. *PLoS Biol.* 2012;10(11):e1001422.
64. Chan AH, Jenkins PA, Song YS. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS genetics.* 2012;8(12):e1003090.
65. Coop G, Wen XQ, Ober C, Pritchard JK, Przeworski M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science.* 2008;319(5868):1395-8.
66. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P. Human recombination hot spots hidden in regions of strong marker association. *Nature genetics.* 2005;37(6):601-6.
67. Khil PP, Camerini-Otero RD. Genetic crossovers are predicted accurately by the computed human recombination map. *PLoS genetics.* 2010;6(1):e1000831.
68. Reed FA, Tishkoff SA. Positive selection can create false hotspots of recombination. *Genetics.* 2006;172(3):2011-4.
69. McQuilton P, St Pierre SE, Thurmond J, Consortium F. FlyBase 101-the basics of navigating FlyBase. *Nucleic Acids Research.* 2012;40(D1):D706-D14.
70. Loewe L, Charlesworth B, Bartolome C, Noel V. Estimating selection on nonsynonymous mutations. *Genetics.* 2006;172(2):1079-92.
71. Jensen JD, Bachtrog D. Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. *Genome Biology and Evolution.* 2011;3:687-701.
72. Bachtrog D, Andolfatto P. Selection, recombination and demographic history in *Drosophila miranda*. *Genetics.* 2006;174(4):2045-59.
73. Anderson WW, Ayala FJ, Michod RE. Chromosomal and allozymic diagnosis of three species of *Drosophila*. *Drosophila pseudoobscura*, *D. persimilis*, and *D. miranda*. *J Hered.* 1977;68(2):71-4.
74. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, et al. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics.* 2008;179(3):1601-55.
75. Schweitzer MD. An Analytical Study of Crossing over in *Drosophila Melanogaster*. *Genetics.* 1935;20(5):497-527.
76. Cirulli ET, Kliman RM, Noor MAF. Fine-scale crossover rate heterogeneity in *Drosophila pseudoobscura*. *Journal of Molecular Evolution.* 2007;64(1):129-35.
77. Myers S, Spencer CCA, Auton A, Bottolo L, Freeman C, Donnelly P, et al. The distribution and causes of meiotic recombination in the human genome. *Biochemical Society Transactions.* 2006;34:526-30.
78. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature.* 2010;467(7319):1099-103.
79. Begun DJ, Aquadro CF. Levels of Naturally-Occurring DNA Polymorphism Correlate with Recombination Rates in *Drosophila-Melanogaster*. *Nature.* 1992;356(6369):519-20.
80. Charlesworth B, Morgan MT, Charlesworth D. The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics.* 1993;134(4):1289-303.
81. Smith JM, Haigh J. Hitch-Hiking Effect of a Favorable Gene. *Genetical research.* 1974;23(1):23-35.
82. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. *PLoS genetics.* 2011;7(2):e1001302.
83. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS genetics.* 2009;5(5):e1000471.

84. Birky CW, Jr., Walsh JB. Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 1988;85(17):6414-8.
85. Elyashiv E, Sattath S, Hu TT, Strustovsky A, McVicker G, Andolfatto P, et al. A genomic map of the effects of linked selection in *Drosophila*. *arXiv:14085461*. 2014.
86. Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(29):10051-6.
87. Sackton TB, Kulathinal RJ, Bergman CM, Quinlan AR, Dopman EB, Carneiro M, et al. Population genomic inferences from sparse high-throughput sequencing of two populations of *Drosophila melanogaster*. *Genome Biol Evol*. 2009;1:449-65.
88. Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW, et al. Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *Plos Biology*. 2007;5(11):2534-59.
89. Spencer CCA, Deloukas P, Hunt S, Mullikin J, Myers S, Silverman BW, et al. The influence of recombination on human genetic diversity. *PLoS genetics*. 2006;2(9):1375-85.
90. Cattani MV, Kingan SB, Presgraves DC. Cis- and trans-acting genetic factors contribute to heterogeneity in the rate of crossing over between the *Drosophila simulans* clade species. *J Evol Biol*. 2012;25(10):2014-22.
91. Baker BS, Carpenter AT, Esposito MS, Esposito RE, Sandler L. The genetic control of meiosis. *Annu Rev Genet*. 1976;10:53-134.
92. Feldman MW, Otto SP, Christiansen FB. Population genetic perspectives on the evolution of recombination. *Annual Review of Genetics*. 1996;30:261-95.
93. Lenormand T, Otto SP. The evolution of recombination in a heterogeneous environment. *Genetics*. 2000;156(1):423-38.
94. Martin G, Otto SP, Lenormand T. Selection for recombination in structured populations. *Genetics*. 2006;172(1):593-609.
95. Otto SP, Barton NH. Selection for recombination in small populations. *Evolution*. 2001;55(10):1921-31.
96. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. *Nature reviews Genetics*. 2002;3(4):252-61.
97. Otto SP, Michalakis Y. The evolution of recombination in changing environments. *Trends in Ecology & Evolution*. 1998;13(4):145-51.
98. Barton NH. A General-Model for the Evolution of Recombination. *Genetical research*. 1995;65(2):123-44.
99. Barton NH. Mutation and the evolution of recombination. *Philosophical Transactions of the Royal Society B-Biological Sciences*. 2010;365(1544):1281-94.
100. Noor MA, Schug MD, Aquadro CF. Microsatellite variation in populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetical research*. 2000;75(1):25-35.
101. Schaeffer SW, Miller EL. Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics*. 1992;132(2):471-80.
102. Prakash S, Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. *Genetics*. 1969;61(4):841-58.
103. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
104. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011;43(5):491-8.

105. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303.
106. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
107. Fuller ZL, Haynes GD, Zhu D, Batterton M, Chao H, Dugan S, et al. Evidence for stabilizing selection on codon usage in chromosomal rearrangements of *Drosophila pseudoobscura*. *G3.* 2014;4(12):2433-49.
108. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive Natural Selection in the *Drosophila* Genome? *PLoS genetics.* 2009;5(6):-.
109. Main BJ, Smith AD, Jang H, Nuzhdin SV. Transcription start site evolution in *Drosophila*. *Molecular Biology and Evolution.* 2013;30(8):1966-74.
110. St Pierre SE, Ponting L, Stefancsik R, McQuilton P. FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Research.* 2014;42(1):D780-8.

Supporting Information

S1 Fig. A comparison of the LDhelmet estimate of recombination rate with an alternative simple direct

measure of linkage disequilibrium decay. a) Average r^2 as a function of distance between SNPs was generated for all polymorphic base pairs on chromosome 2. The best-fit line (red) was generated using a five-parameter log-logistic function. LD was inferred to have decayed when the best-fit line came to a value 10% higher than background r^2 (blue line, calculated as the average r^2 of SNPs at distance 1500 – 2000bp from each other), b) Correlation between linkage disequilibrium calculated through LDhelmet and the square root of the nucleotide distance at which LD decayed as estimated through the method described in a).

S2 Fig. A comparison of empirical and LD-based estimations of recombination rate for *D.*

***pseudoobscura* Pikes Peak, Chr2.** Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) across chromosome 2 for *D. pseudoobscura* Pikes Peak. All recombination rates are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (black, solid).

S3 Fig. A comparison of empirical and LD-based estimations of recombination rate for ChrXL.

Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) across portions of chromosome XL group1a for *D. pseudoobscura* Flagstaff (top), *D. pseudoobscura* Pikes Peak (middle), and *D. miranda* (bottom). All recombination rates are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (blue, solid).

S4 Fig. A comparison of empirical and LD-based estimations of recombination rate for ChrXR.

Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) across portions of chromosome XR group 6 and 8 for *D. pseudoobscura* Flagstaff (top), *D. pseudoobscura* Pikes Peak

(middle), and *D. miranda* (bottom). All recombination rates are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (blue, solid).

S5 Fig. A comparison of empirical and LD-based estimations of fine scale recombination rate. Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) for three regions of 20kb intervals on chromosome 2 of *D. pseudoobscura* Flagstaff. All recombination rates are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (blue, solid).

S6 Fig. A comparison of empirical and LD-based estimations of super-fine scale recombination rate. Empirical recombination rates (red) are plotted with LDhelmet derived recombination rates (blue) for one region of 5kb intervals on chromosome 2 of *D. pseudoobscura* Flagstaff. All recombination rates are reported in cM/Mb. Error bars depict 95% Confidence Intervals for the empirical (red, dashed), and LD-based (blue, solid).

S7 Fig. A comparison of LD-based estimations of recombination rate between two species at 50kb. LD-based *D. pseudoobscura* recombination rates (red) are plotted with LD-based *D. miranda* recombination rates (blue) across the genome. Average rates were calculated for 50kb window sizes. Chromosomes 4, XL, and XR are split in to multiple groups, each labeled on the x-axis, according to the reference assembly for *D. pseudoobscura* (denoted with alternating background color). All recombination rates are reported in μ /bp. Error bars are not shown for clarity.

S8 Fig. A comparison of LD-based estimations of recombination rate between two species at 250kb. LD-based *D. pseudoobscura* recombination rates (red) are plotted with LD-based *D. miranda* recombination rates (blue) across the genome. Average rates were calculated for 250kb window sizes. Chromosomes 4, XL, and XR are split in to multiple groups, each labeled on the x-axis, according to the reference assembly for *D. pseudoobscura* (denoted with alternating background color). Smaller

chromosome groups are excluded due to lack of data. All recombination rates are reported in ρ /bp.

Error bars are not shown for clarity.

S9 Fig. A comparison of LD-based estimations of recombination rate between two species at 500kb.

LD-based *D. pseudoobscura* recombination rates (red) are plotted with LD-based *D. miranda* recombination rates (blue) across the genome. Average rates were calculated for 500kb window sizes.

Chromosomes 4, XL, and XR are split in to multiple groups, each labeled on the x-axis, according to the reference assembly for *D. pseudoobscura* (denoted with alternating background color). Smaller

chromosome groups are excluded due to lack of data. All recombination rates are reported in ρ /bp.

Error bars are not shown for clarity.

S10 Fig. Recombination hotspots in *D. pseudoobscura*. Two recombination “hotspots” from

chromosome 2 of *D. pseudoobscura* plotted with the background recombination rate (red, dashed). The region of increased recombination is shown with 100kb flanking sequence. Recombination rates are reported in cM/Mb.

S11 Fig. Recombination rate within genes of *D. pseudoobscura*. Recombination rates are slightly

decreased at the start of genes. Genes were characterized by exon and intron number, then recombination rate was averaged over the genomic position of the exon or intron. Recombination rates were aggregated to give genome wide totals for each exon (orange) and intron (green) position within a gene. Recombination rates are reported in ρ /bp.

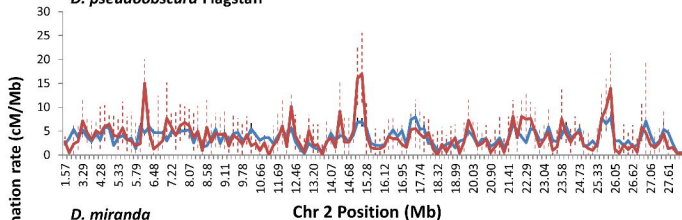
S12 Fig. Linked selection over varying physical scales in *D. pseudoobscura*. Recombination rate was

plotted with pairwise nucleotide diversity at four-fold degenerate positions across the genome of *D. pseudoobscura* with window sizes of 5kb, 10kb, 50kb, 100kb, 250kb, 500kb, and 1Mb. The coefficient of determination (r^2) for each of these associations is plotted (5kb=0.0262, 10kb=0.0326, 50kb=0.0664,

100kb=0.1019, 250kb=0.2721, 500kb=0.3192, and 1Mb=0.3628), showing recombination increasingly explains the variance in nucleotide diversity at large physical scales.

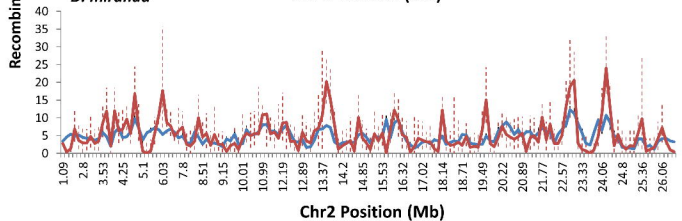
— LD-based — Empirical

D. pseudoobscura Flagstaff

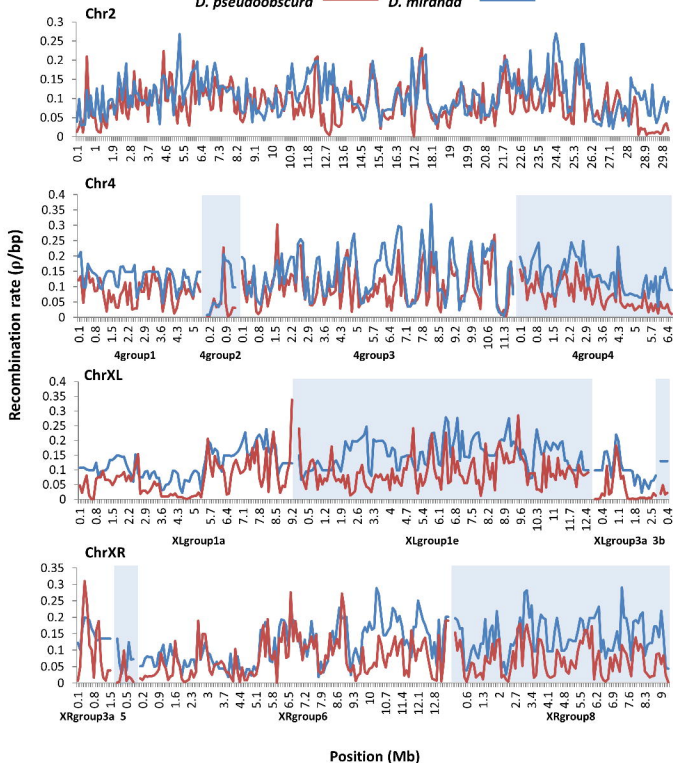


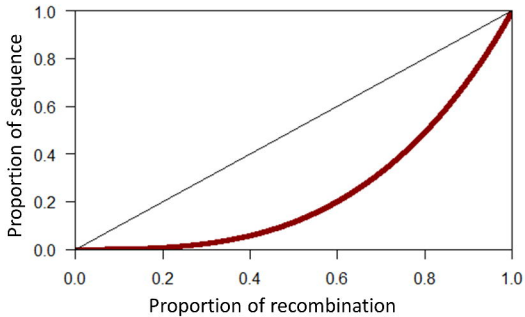
D. miranda

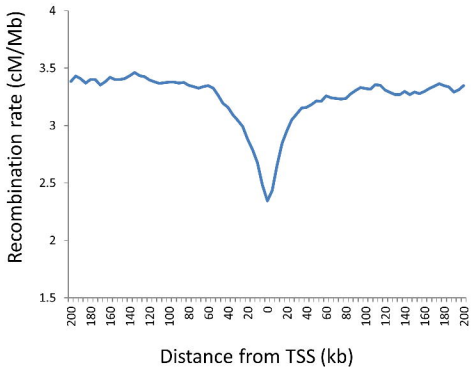
Chr 2 Position (Mb)



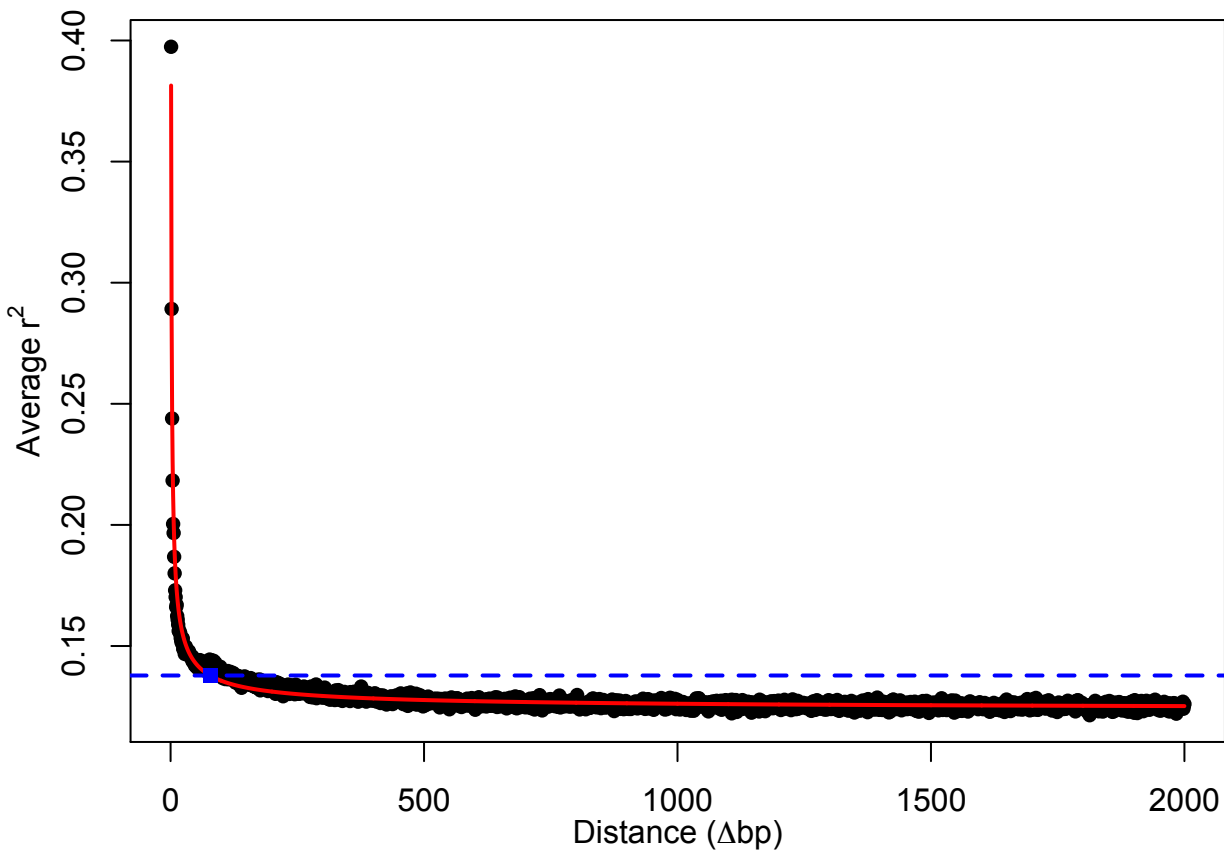
D. pseudoobscura — *D. miranda*



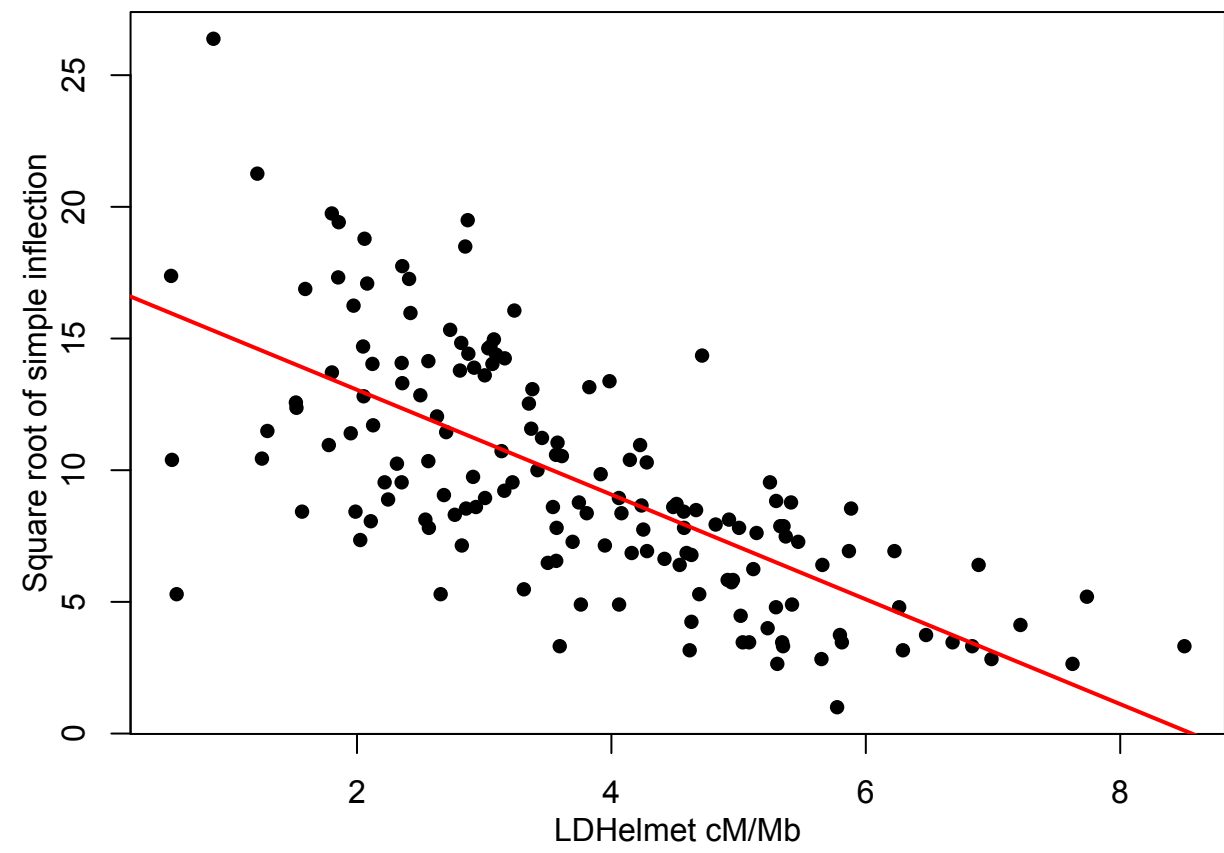




a)

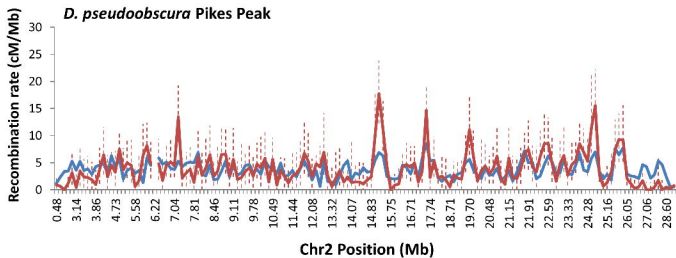


b)



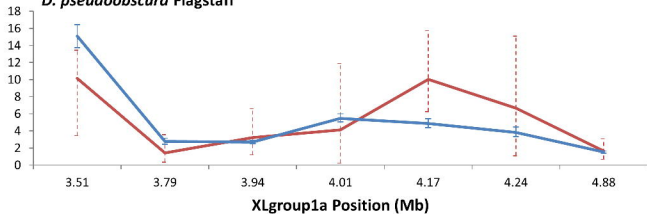
— LD-based — Empirical

D. pseudoobscura Pikes Peak

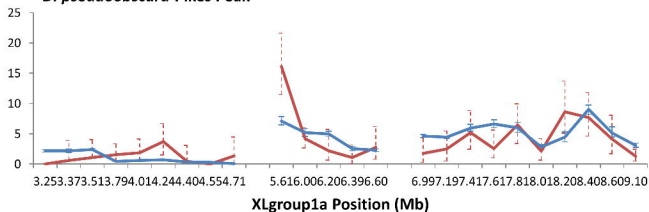


— Empirical — LD-based

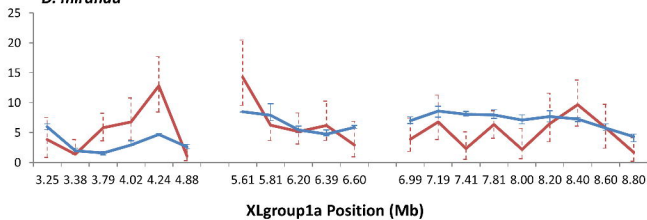
D. pseudoobscura Flagstaff



D. pseudoobscura Pikes Peak

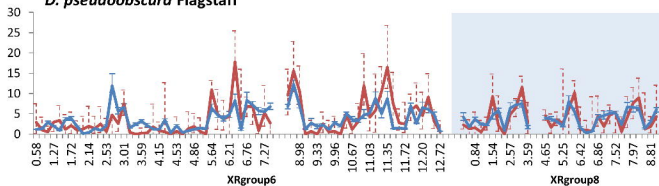


D. miranda



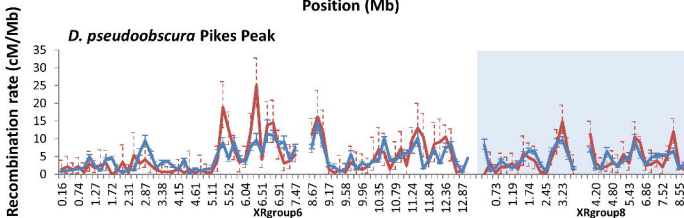
— Empirical — LD-based

D. pseudoobscura Flagstaff



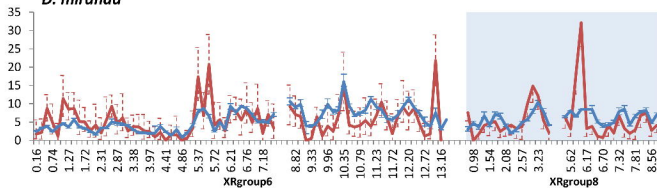
Position (Mb)

D. pseudoobscura Pikes Peak

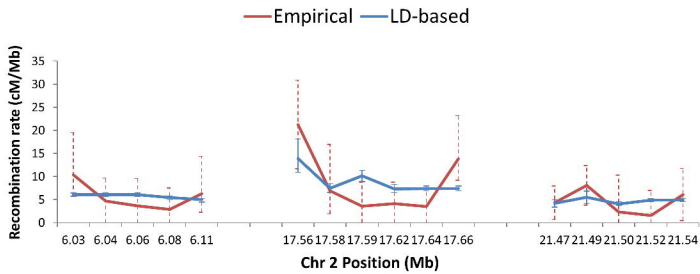


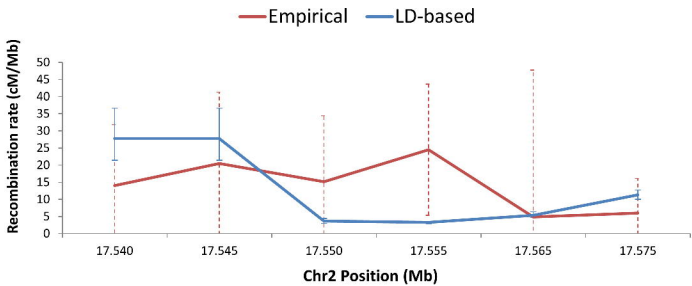
Position (Mb)

D. miranda

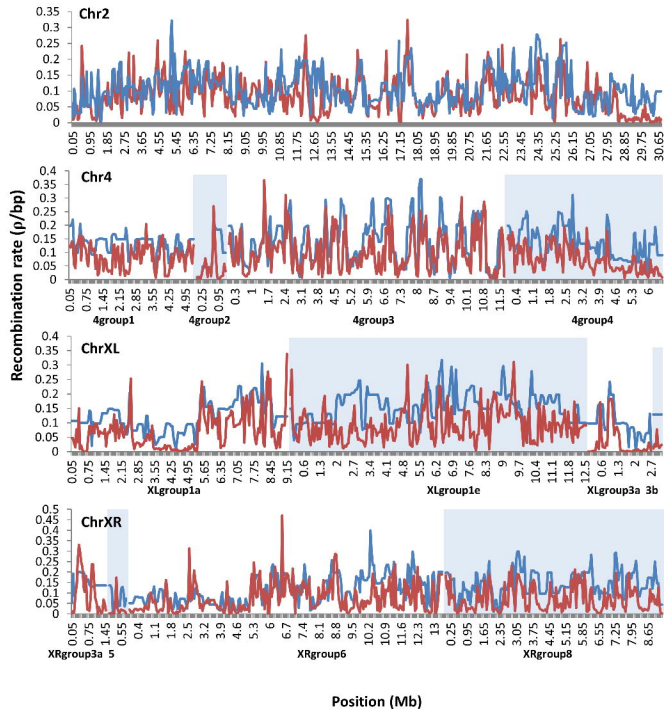


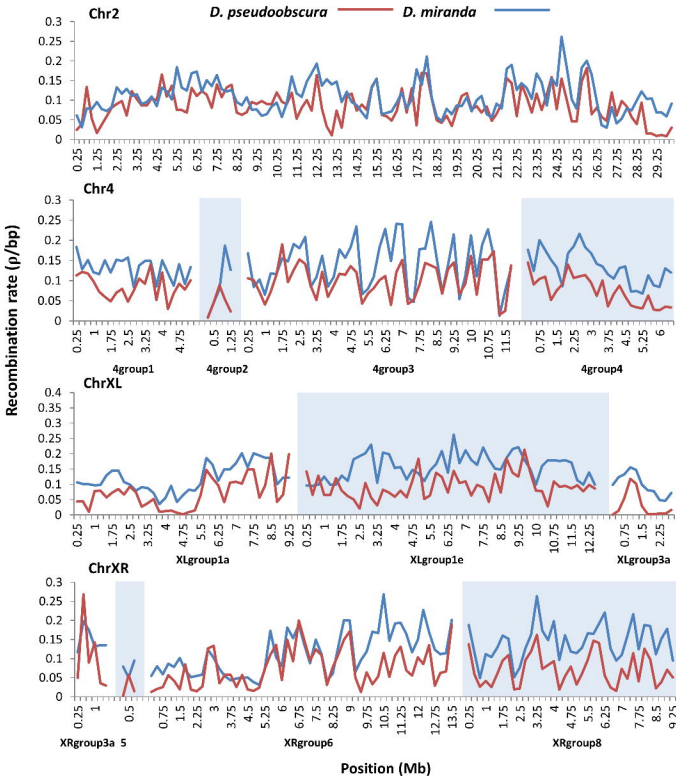
Position (Mb)





D. pseudoobscura — *D. miranda* —





D. pseudoobscura — *D. miranda* —

