DNA motifs are not general predictors of recombination in two *Drosophila* sister species.

James M. Howie [a+],

Rupert Mazzucco [a+],

Thomas Taus [a,b+],

Viola Nolte [a],

Christian Schlötterer* [a].

[a] *Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria*

[b] *Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria*

[+] Equal first author.

*Author for Correspondence: Christian Schlötterer, Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria; Fax: +43 1 25077-4390, Phone: +43 1 25077-4300, Email: christian.schloetterer@vetmeduni.ac.at

Article type: Research Article

Data Deposition: Accession numbers to European Nucleotide Association and Dryad provided upon acceptance of the article.

ABSTRACT

Meiotic recombination is crucial for chromosomal segregation, and facilitates the spread of beneficial and removal of deleterious mutations. Recombination rates frequently vary along chromosomes and *Drosophila melanogaster* exhibits a remarkable pattern. Recombination rates gradually decrease towards centromeres and telomeres, with dramatic impact on levels of variation in natural populations. Two close sister species, *D. simulans* and *D. mauritiana* do not only have higher recombination rates, but also exhibit a much more homogeneous recombination rate that only drops sharply close to centromeres and telomeres. Because certain sequence motifs are associated with recombination rate variation in *D. melanogaster*, we tested whether the difference in recombination landscape between *D. melanogaster* and *D. simulans* can be explained by the genomic distribution of recombination-rate associated sequence motifs. We constructed the first high resolution recombination map for *D. simulans*, and searched for motifs linked with high recombination in both sister species. We identified five consensus motifs, present in either species. While the association between motif density and recombination is strong and positive in *D. melanogaster*, the results are equivocal in *D. simulans*. Despite the strong association in *D. melanogaster*, we do not find a decreasing density of these repeat motifs towards centromeres and telomeres. We conclude that the density of recombination-associated repeat motifs cannot explain the large-scale recombination landscape in *D. melanogaster*, nor the differences to *D. simulans*. The strong association seen for the sequence motifs in *D. melanogaster* likely reflects their impact influencing local differences in recombination rates along the genome.

Keywords: *D. simulans*, Genomic Correlation, Linkage Disequilibrium, Motif Density, Motif Model, Recombination Map

INTRODUCTION

Meiotic recombination rate variation impacts on multiple important biological processes in sexual eukaryotes. It is crucial for chromosomal segregation (John 2005; Roeder 1997), but is also itself a powerful factor influencing genome organisation and sequence variability (Aquadro, et al. 1994; True, et al. 1996). Meiotic recombination arises when a double-stranded break leads to crossing over between homologous chromatids (Bergerat, et al. 1997; Hughes, et al. 2018; Keeney, et al. 1997; Schwacha and Kleckner 1995; Szostak, et al. 1983). Higher rates of recombination break up genetic linkage and can increase the efficacy of natural selection (Charlesworth and Charlesworth 2010; Haddrill, et al. 2007) and so affect the evolution of numerous genomic features. The reduction of transposable element density (Charlesworth and Lapid 1992; Charlesworth, et al. 1994; Kofler, et al. 2012; Petrov, et al. 2011; Rizzon, et al. 2002) and the increased levels of DNA polymorphism (Aquadro, et al. 1994; Begun and Aquadro 1992; Begun, et al. 2007; Kulathinal, et al. 2008) in regions of high recombination are probably the clearest examples.

Yet while the eukaryotic meiotic machinery is generally highly conserved (Keeney 2001), rates of recombination have been observed to vary dramatically across species and populations, between individuals, and across sexes (Stapley, et al. 2017), apparently due to a combination of interacting environmental, epigenetic, and genetic factors (Detlefsen and Roberts 1921; Neel 1941; Parsons 1958; Stapley, et al. 2017; Stern 1926). Moreover, the distribution of meiotic recombination rates among and along chromosomes varies markedly across taxa (Choi and Henderson 2015; Hey 2004; Hunter, et al. 2016a; Lichten and Goldman 1995; Petes 2001; Stapley, et al. 2017). Large-scale recombination suppression is often observed towards centromeres, the so called “centromere effect” (Beadle 1932; Choulet, et al.

2014; Hughes, et al. 2018; Szauter 1984). Depending on the species, either suppression or enhancement of recombination has been observed towards the telomeres (Broman, et al. 1998; Chan, et al. 2012; Comeron, et al. 2012; Myers, et al. 2005). Heterochromatin, which is often associated with these regions, tends also to exhibit lower recombination rates than euchromatin (Baker 1958; Roberts 1965; Sturtevant and Beadle 1936; Szauter 1984; Termolino, et al. 2016). Yet, in addition to these large-scale features of recombination landscapes, fast-evolving (Jeffreys, et al. 2001) finer-scale variation can also be observed (Comeron, et al. 2012; Myers, et al. 2005).

It has been proposed that short sequence motifs are a key factor shaping the recombination landscape. For example, in humans a 13-mer, CCNCCNTNNCCNC motif is targeted by the PRDM9 protein (Billings, et al. 2013; Grey, et al. 2011; Myers, et al. 2010), via its zinc-finger array (Baudat, et al. 2010; Parvanov, et al. 2010), where it promotes histone methylation and meiotic crossover, reorganising the nucleosome around it and driving double stranded break formation (Baker, et al. 2014; Brick, et al. 2012; Mihola, et al. 2009; Pratto, et al. 2014). These highly localized recombination events in 500–2000bp sections of chromosome have been called recombination "hotspots" (Lam and Keeney 2014). They are observed in a multitude of species including yeast, mice, humans among many others (Lam and Keeney 2014).

Hotspots are, however, no universal feature of recombination landscapes, and are not observed in a range of species groups including *Caenorhabditis elegans* and *Drosophila* (Aquadro, et al. 2001; Chan, et al. 2012; Hey 2004; Manzano-Winkler, et al. 2013; Miller, et al. 2016; Nachman 2002; Smukowski Heil, et al. 2015). *Drosophila spp*., exhibit a large heterogeneity in recombination across their chromosomes, as demonstrated in *D. persimilis*

(Stevison and Noor 2010), *D. pseudoobscura* (Cirulli, et al. 2007; Kulathinal, et al. 2008), and *D. melanogaster* (Adrian, et al. 2016; Comeron, et al. 2012; Singh, et al. 2009). Still, *D. melanogaster* exhibits only a handful of mild "hotspots" relative to the ~30,000, often very strong hotspots observed in humans (International HapMap Consortium 2007). Instead the *D. melanogaster* recombination landscape is characterised by recombination "peaks" and "valleys" on a 5kb – 500kb scale (Adrian, et al. 2016; Chan, et al. 2012; Comeron, et al. 2012; Singh, et al. 2009) with which short "recombination motifs" are associated; as is also seen in *D. pseudoobscura, D. persimilis*, and other species (Adrian, et al. 2016; Chan, et al. 2012; Cirulli, et al. 2007; Comeron, et al. 2012; Heil and Noor 2012; Kulathinal, et al. 2008; Miller, et al. 2012; Singh, et al. 2009; Singh, et al. 2013; Stevison and Noor 2010). These motifs, which often reside in transcription-associated euchromatic regions (Comeron, et al. 2012; Petes 2001), are thought to increase the accessibility of DNA chromatin to double-stranded cleavage (Comeron, et al. 2012) and de-stabilize DNA sequences, potentially in a stress, environmental or epigenetically dependent manner (Hunter, et al. 2016b; Kohl and Singh 2018; Neel 1941; Petes 2001; Redfield 1966; Stern 1926).

D. melanogaster, D. simulans, and *D. mauritiana* are sister species which are ecologically and karyotypically similar (LEMEUNIER AND ASHBURNER 1976; TRUE *et al.* 1996), but differ dramatically in their recombination landscapes. While *D. melanogaster* exhibits a characteristic gradual decrease in recombination rate towards centromeres and to a lesser extent also telomeres, the recombination landscape in *D. simulans* and *D. mauritiana* is much flatter with a rather constant recombination rate almost to the end of the chromosome arm, where it drops very quickly (True, et al. 1996). Furthermore, these two species also have a higher recombination rate than *D. melanogaster* (True, et al. 1996), which has been attributed, in *D. mauritiana*, to the MEI-218 protein which has highly diverged between

126 *D. melanogaster* and *D. mauritiana*, promoting recombination to a lesser extent in the former
127 (Brand, et al. 2018).
128
129 Here, to test the hypothesis that differences in genome-wide motif distributions can
130 explain the observed differences in recombination (Adrian, et al. 2016), we take a multi-step
131 approach. First, we produce a high-resolution recombination map for *D. simulans*. Next, we
132 run a motif discovery in each species and construct a consensus motif set. We confirm the
133 clear differences in recombination landscapes between the two species, but find a similar set
134 and distribution of recombination associated motifs in each. Our results suggest that
135 recombination associated motifs cannot explain the large-scale differences in recombination
136 landscapes between the two species but may have a significant impact on recombination on a
137 local scale, in particular in *D. melanogaster*.
138
139 **MATERIAL AND METHODS**
140
141 **Recombination Map**
142
143 *Recombination Map Production*
144
145 A total of 202 isofemale lines were established from a natural *D. simulans* population in
146 Tallahassee, Florida, USA in 2010 (Barghi, et al. 2017). From each of the 189 lines that were
147 still alive in 2016, an individual male was selected and crossed with a virgin "reference"
148 female from the M252 strain that was used to produce the *D. simulans* reference genome
149 (Palmieri, et al. 2015). Paired-end libraries were generated for a single F1 female as described
150 in Barghi, et al. (2017) and sequenced on an Illumina HiSeq XTEN to obtain an average

151 sequence coverage of 30x. Single-nucleotide polymorphisms (SNPs) were called with
152 FreeBayes (v1.1.0-46-g8d2b3a0, Garrison and Marth 2012), requiring a minimum sequencing
153 coverage of 10x and a variant quality of at least 50. All SNPs that were polymorphic in the
154 M252 reference strain were masked. Based on line-specific haplotype information, the
155 genome-wide recombination map was estimated with LDJump (v0.1.4, Hermann, et al. 2018),
156 specifying a segment size of 1kb, with an $\alpha = 0.05$ and an $\Theta = 0.04$. We disabled LDJump's
157 segmentation analysis and worked with raw recombination rate estimates. Recombination
158 rates were converted from ρ to units of cM/Mb by normalising them so as to have a genetic
159 map length between a set of marker genes equivalent to that which has been previously
160 reported (True, et al. 1996).
161
162 The resultant *D. simulans* recombination map was used in parallel with the
163 *D. melanogaster* recombination map produced by Comeron, et al. (2012), downloaded from
164 the *Drosophila melanogaster* Recombination Rate Calculator (Fiston-Lavier, et al. 2010).
165
166 *Recombination Map Scaling*
167
168 As the raw recombination map output by LDJump is noisy, we smoothed each recombination
169 map at several scales. In *D. melanogaster*, the raw map (Comeron, et al. 2012) contained
170 information on recombination rate at a 100kb resolution, in *D. simulans* raw information was
171 generated at a 1kb scale. For smoothing, we used a moving median approach, using window
172 sizes of 5, 25, 101, 501 and 2501 kb for *D. simulans*, and a 101, 501, 2501 kb for
173 *D. melanogaster*, respectively. Advantages of the moving median as a smoothing method
174 include low sensitivity to outliers, and a direct relationship to underling data, in the sense that
175 only values present in the raw data set can be present in the smoothed set if the median is

taken based on an odd number of input values, which in our case it always was. Because this approach is also computationally expensive, and prone to deleting map features when there are long runs of identical values, we investigated as an alternative approach, smoothing via LOESS local regression (Cleveland, et al. 1992), which produces qualitatively equivalent results (Figure S2). The smoothing scales chosen reflect those in Adrian, et al. (2016), relevant to potential motif explanatory power. The "correct" scale on which motifs may function is *a priori* unclear.

DNA Motif Identification

Motif Discovery

For each species, we ran a genome-wide motif discovery using MEME (Bailey and Elkan 1994), from the MEME suite of motif-based sequence analysis tools (Bailey, et al. 2009, version 5.0.1pl, accessible at http://meme-suite.org; Bailey, et al. 2015), a software designed to detect DNA sequence motifs in genetic data. After dividing each of the five large chromosomes (X, 2L, 2R, 3L, 3R) into high- and low-recombining regions based on the chromosome median recombination rate, we used this software in the "differential enrichment" mode to detect motifs enriched in high-recombining areas of the genome. For *D. melanogaster*, we ran MEME on the release 5 reference genome (v. 5.36), for concordance with our recombination information from Comeron et al. (2012). For *D. simulans*, we used the M252 Madagascar reference genome (Palmieri, et al. 2015), to align with our recombination map. Motif discovery searches were run with species specific Markov Background Models, simple matrices of background base frequencies obtained using the MEME *fasta-get-model* command, for each reference genome in turn. The full procedure was

repeated with all smoothed maps (Methods: *Recombination Map Production*). For completeness, a raw 1 kb window motif discovery run was also conducted for *D. simulans*. A similar search for motifs associated with lower recombination areas returned no results.

Motif Consensus Set

MEME motif discovery runs returned a set of 5, 4 and 3 motifs in *D. melanogaster* and 1, 2, 4, 1, 1 and 1 significant motifs in *D. simulans*, at the 101, 501, and 2501, and 1, 5, 25, 101, 501, and 2501 kb scales, respectively (SI.3, $E \leq 0.01$). It was noticed that, while individually distinct, numerous motifs contained similar core patterns whilst varying, for example, only in repeat number. As such, we constructed a set of 5 consensus motifs that captured the core variation in all motifs significantly associated with increased recombination, across both species, and over all scales. This core set of motifs C1–5, was determined via a two-step method. First, we contrasted the motifs across each of our recombination map smoothing scales in both species, retaining only motifs that occurred in at least one scale with a minimum significance of $E \leq 0.01$ in at least one species. Motifs were then simplified by allowing only the most likely base at each position, and motif lengths were fixed as the longest sequence length that could be represented in both species (as lengths were by tendency longer in *D. melanogaster*). This resulted in the following set of consensus motifs: C1=$[A]_{11}$; C2=$[GCA]_4$; C3=$[CA]_6$; C4=$[TA]_5$; C5=$[G]_8$. We note that *D. melanogaster* made the dominant contribution to the consensus motifs, as the motifs in *D. simulans* were less significant than those observed in *D. melanogaster* (SI.3), and that the number of consensus motifs was informed by the data, and not decided *a priori*. As our consensus motifs turned out to be simplified versions of the most predictive motifs that were identified by Adrian et al.

(2016), we quantitatively confirmed this similarity using the MEME Suite tool TomTom (Gupta, et al. 2007), under default parameters (see SI.4).

Genome-Wide Motif Densities

Motif Locations

We converted the 5 consensus motifs into letter-probability matrices, to be used as input to FIMO, a MEME Suite tool designed to find genome-wide motif occurrences (Grant, et al. 2011). Matrices were compiled in a hard, and a softer, version; with the expected base given a probability of 1 and unexpected bases probabilities of 0, or the expected base a probability of 0.97, and unexpected bases a probability of 0.01. FIMO was then run for each species, taking the reference sequences and Markov Background Models as noted in Methods*: Motif Discovery*, and using parameter *max-stored-scores* = 50000000, and all others at default. Results of the hard and soft motif probability runs were qualitatively identical, so hard coded motif probabilities were used for follow-up analysis (soft runs not reported).

Motif Densities

FIMO output provides, per motif, the genomic locations (chromosome, start and stop position) at which a motif was found, as well as a *p*-value and a *q*-score (Benjamini and Hochberg 1995) per record, which show how well the motif was matched to the underlying reference sequence, both before and after correction for multiple testing (Benjamini and Hochberg 1995). To obtain genome-wide motif densities in each species, we calculated for each motif the sum of $1 - q$, across a sliding window of 1 kb, where q refers to the per record

250 *q*-score, such that per window motif densities are discounted in relation to the quality of the
251 motif match, with higher quality matches counting more. A total, genome-wide count (of 1 –
252 *q*) of each motif was also obtained from the raw FIMO output.
253
254 **Motif - Recombination Correlations and Models**
255
256 *Motif Density – Recombination Rate Correlations*
257
258 To investigate the relationship between recombination rates and genome-wide abundances of
259 individual motifs, we calculated the correlations between motif densities, binned at 1 kb, and
260 corresponding recombination rates (cM/Mb), per motif, for *D. simulans* and *D. melanogaster*,
261 respectively. As there was no clear *a priori* expectation for the genomic scale at which motifs
262 would have most impact on recombination, the analysis was repeated for all smoothing scales
263 noted in Methods: *Recombination Map Scaling* for *D. melanogaster* and *D. simulans* (and
264 was repeated on the raw 1 kb scale in for *D. simulans*, not shown). Spearman's roe, ρ, was
265 used as a non-parametric estimator of the correlation between the test variables, and both the
266 direction and significance of all correlations were extracted. To investigate the overall
267 predictive power of motif densities, irrespective of chromosomal background, the analysis
268 was repeated on the total genomic data, pooling across all of the 5 major chromosomes, with
269 the analysis repeated per motif and species.
270
271 Finally, to test for explicit directional effects of each consensus motif on
272 recombination, a linear regression model was fitted, per motif, species, scale, and
273 chromosome, for the effect of motif density on local recombination rate, and repeated for the
274 genome average.

275

276 A schematic representation of this analytic pipeline is presented in Figure 1. All statistical

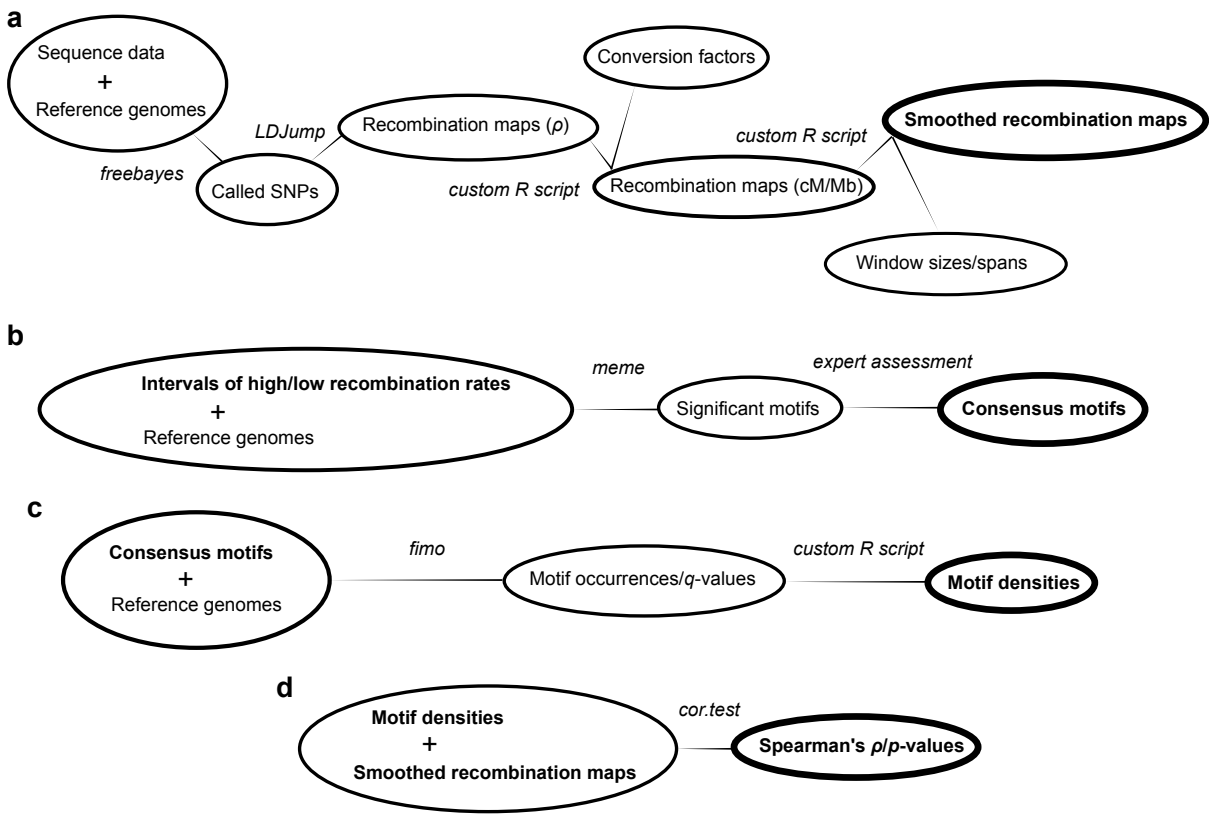
277 analyses were run in R, version 1.1383, using in house scripts (see SI.5).

278

279

280 **Figure 1.**

281



282

283

284

285

286 **Fig. 1.** A schematic representation of the bioinformatic pipeline used. Ovals represent

287 physical data sets, lines represent tools used to derive them; see the Methods for details.

288

289

RESULTS

Recombination rates in *D. simulans* are more uniform across chromosomes, than in *D. melanogaster*

We present the first high-resolution recombination map for *Drosophila simulans*, and contrast it to that of *D. melanogaster* (Comeron, et al. 2012). Across a range of smoothing parameters, the *D. simulans* recombination map is more uniform than that of *D. melanogaster* (Figures 2, 3). The level of recombination suppression is lower towards the centromere in *D. simulans*. As in *D. melanogaster*, the main broad-scale features of the *D. simulans* map hold across the full range of genomic scales, while finer resolution peaks and troughs become visible only at higher resolutions, at the 5 – 501 kb scale (Figure 3). The finer scale peaks (on a kb scale), as with the broader features (on a Mb scale), differ between these two sister *s*pecies, and persist across smoothing scales (Figures 2, 3).

Motif density landscapes are similar in *D. simulans* and *D. melanogaster*

We identify 5 consensus motifs based on motifs recovered in each of the two species (Methods: *Motif Consensus Set*) and obtain their genome-wide densities. The consensus motifs were: C1=$[A]_{11}$; C2=$[GCA]_4$; C3=$[CA]_6$; C4=$[TA]_5$; C5=$[G]_8$. Across all chromosomes and consensus motifs, motif density landscapes were similar in *D. melanogaster* and *D. simulans* (Figure 4). This was especially true for intermediate size landscape features, such as humps and wider valleys (e.g. motif C2 on X, 7.5 Mb position, or 2L at the 8 and 12 Mb positions, Figure 4). Therefore, motif density cannot explain the differences in the broad recombination landscape between both species.

Figure 2.

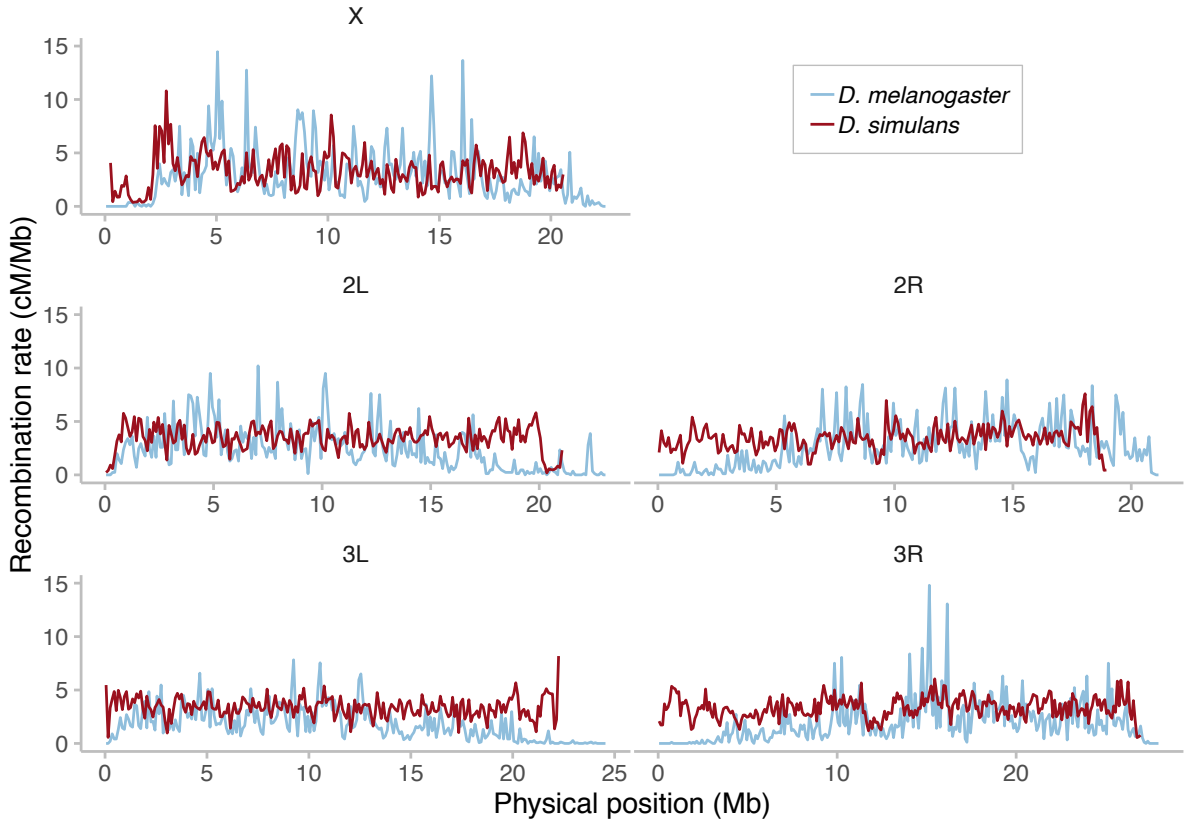


Fig. 2. Recombination rates in *D. simulans* are more uniform across chromosomes than in *D. melanogaster*. Red lines show the recombination rate in *D. simulans* for each of the major chromosomes (name labels in top margin), smoothed at a 101 kb window size with a moving median. For comparison, blue lines show the recombination rate in *D. melanogaster* (with data taken from Comeron *et al*. 2012); Figure 3 for other resolutions.

Figure. 3

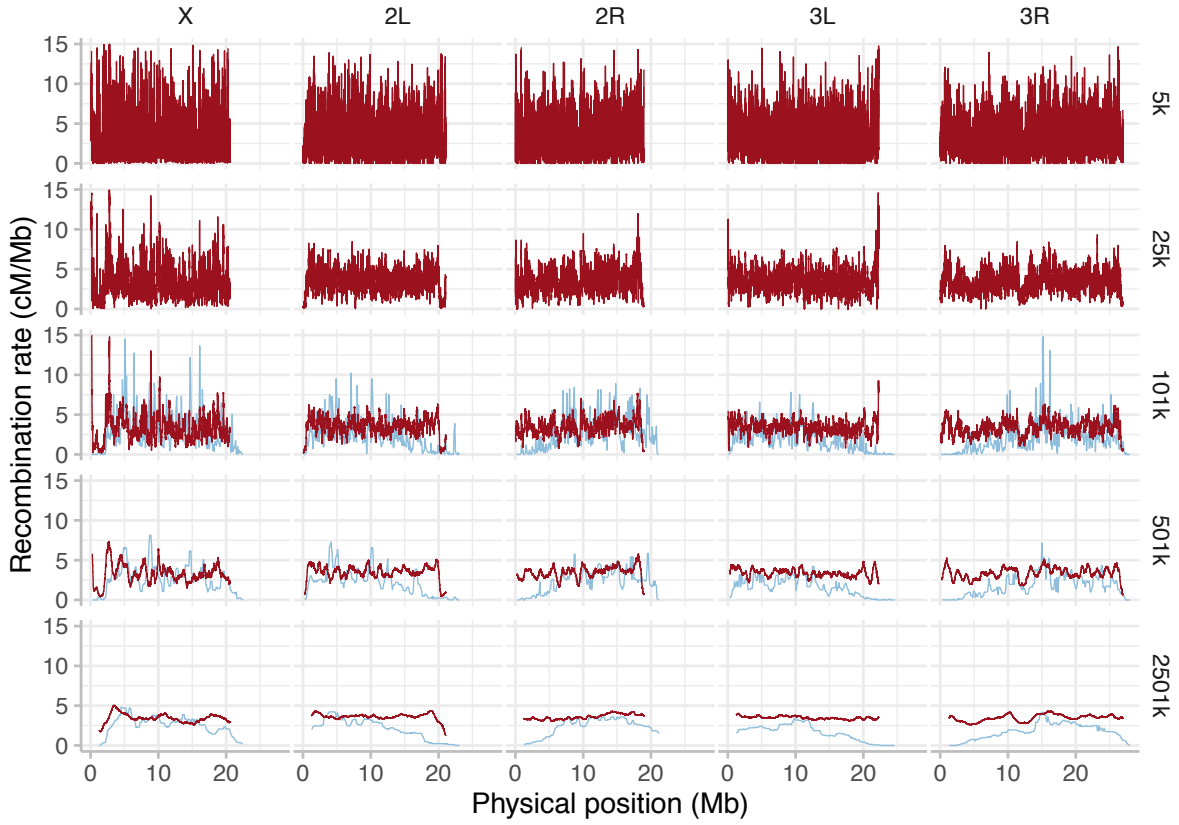


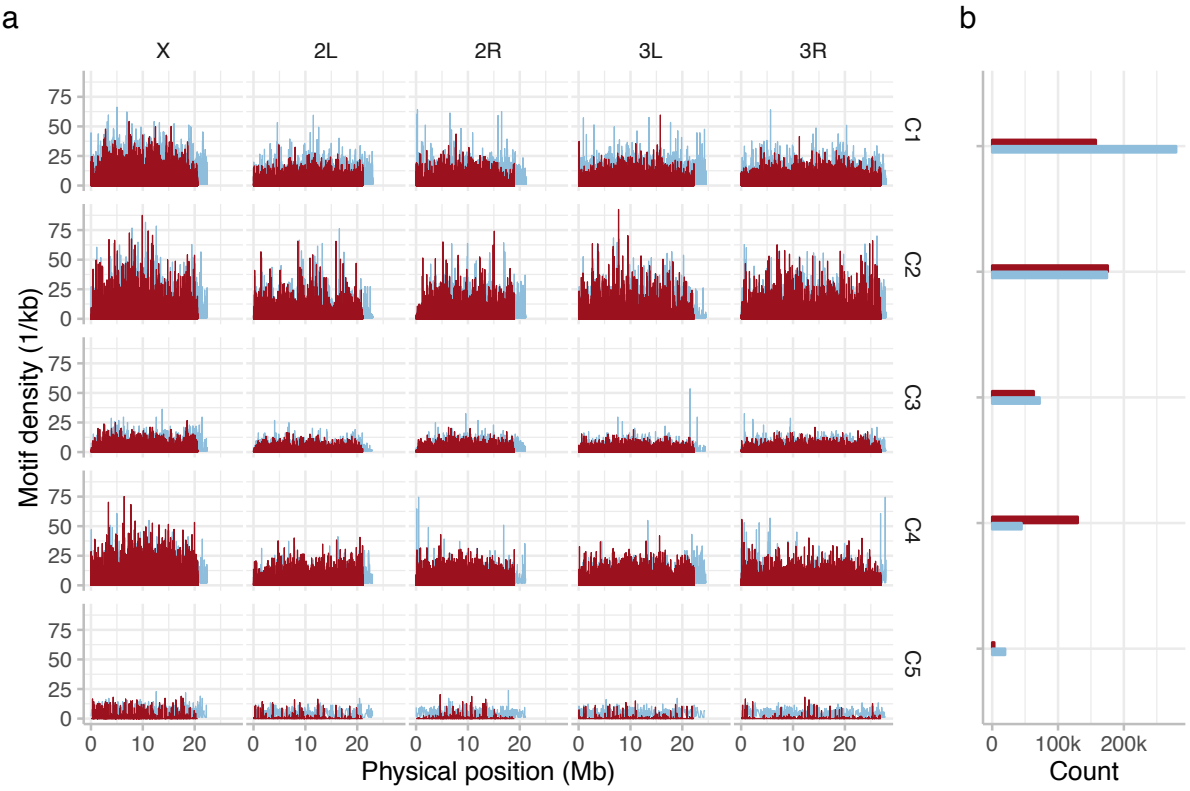
Fig. 3. Recombination rates in *D. simulans* are more uniform across chromosomes than in *D. melanogaster*, at all smoothing scales. Red lines show the recombination rate in *D. simulans* for each of the major chromosomes (names in top margin), smoothed at 5 window sizes (right margin, in bp) with a moving median. For comparison, blue lines show the recombination rate in *D. melanogaster* (data taken from Comeron *et al*. 2012 at 101k; and smoothed at 501k and 2501k; with data not available at smaller resolutions).

343

344

345 **Figure 4.**

346



347

348

349

350 **Fig. 4.** Motif densities are similar in *D. simulans* and *D. melanogaster*. (a) Red (*D. simulans*)

351 and blue (*D. melanogaster*) lines show motif densities across major chromosomes (top

352 margin) as reported by FIMO, with motif occurrences discounted by 1 – *q* (see main text)

353 and binned into 1kb windows for each consensus motif (C1-5). (b) Total motif counts across

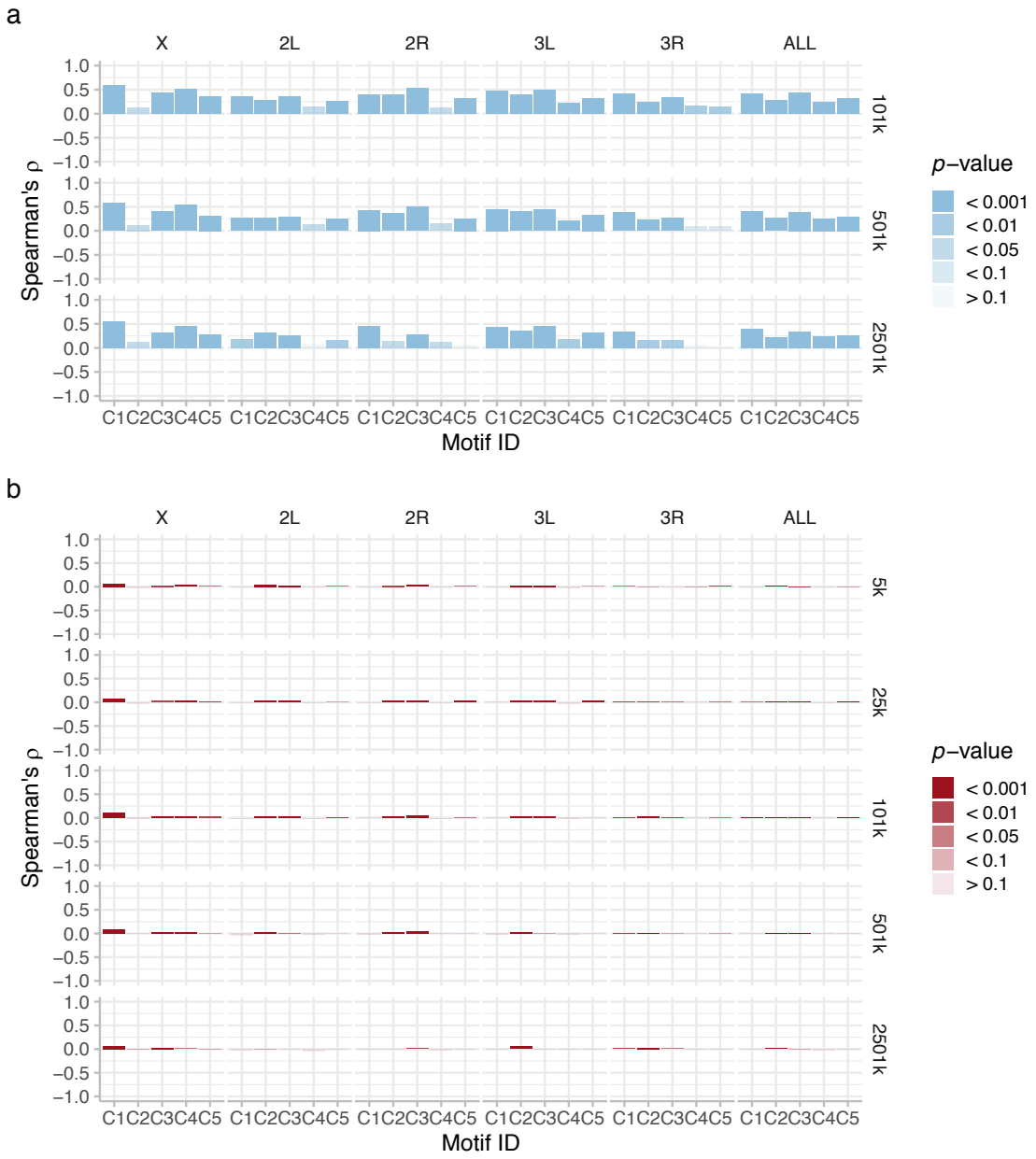
354 all five large chromosomes. FIMO threshold: *p*-value of 1e-4 (default setting).

355

356

357

358 **Figure 5**.



359

360 **Fig. 5.** Associations between motif densities and recombination rates are generally weaker
361 and less significant in *D. simulans* than in *D. melanogaster*. For (a) *D. melanogaster*, and (b)
362 *D. simulans*, bars indicate Spearman's rho, ρ, (height) and the corresponding *p*-value
363 (transparency), from tests of the correlation between motif densities (as shown in Figure. 4 for
364 *D. simulans*, but re-binned for *D. melanogaster* to account for the resolution of the available
365 data) and recombination rates, across individual chromosomes and for the five large
366 chromosomes together (top margin), at all smoothing levels (see right margin).

Finer resolution peaks and troughs varied more between species (e.g. motif C4 on X, 5–15 Mb position, Figure 4). Further, although the different motifs, C1–5, displayed similar broad patterns in each species – per chromosome and genome-wide – some species-specific patterns were seen. Motifs C1, $[A]_{11}$ and C5, $[G]_{8}$ were far less common in *D. simulans*, which had a lower total motif count, while the opposite was true for motif C4, $[TA]_{5}$. Nonetheless, genome-wide motif distributions were similar in each species.

Associations between motif densities and recombination rates are generally weaker and less significant in *D. simulans* than in *D. melanogaster*

We examined correlations between motif densities and recombination rates in each species, both per chromosome, and genome-wide, and at a range of genomic scales. A clear difference was observed between the species. In *D. melanogaster*, all but one correlation was positive, most were highly significant both genome-wide and per chromosome, and the correlation coefficients (Spearman's ρ) were generally large; with a range of ~ 0.4 – 0.6 for the most associated motifs per chromosome (and genome-wide, Figure 5a). In contrast, the associations observed in *D. simulans* were heterogeneously positive or negative, had lower significances than those observed in *D. melanogaster*, and were in all cases weak; with a range of ~ 0.01 – 0.04 for the most associated motifs per chromosome (and genome-wide, Figure 5b). In both species, there was also variation in the importance of different motifs on different chromosomes (below). However, while in *D. melanogaster* the patterns of motif association held across all scales for each chromosome and genome-wide, in *D. simulans* there were occasional exceptions to this rule. For instance, on 2L, 2R, 3L, and genome-wide, the positive correlations for C1 and C4 switched direction at scales larger than 25 – 101 kb. Given that these correlations were very weak with low significance, we attribute these discrepancies stochastic noise, rather than biological signals. We finally note that motifs C1,

C2, and C3 were the most associated with recombination across most major chromosomes in both species (though to a far lesser extent in *D. simulans*), but that an exception is observed for the X chromosome. Here, motif C2 had a very weak association with recombination rate in both species, and motif C4 instead had a high association, relative to its weak association on most autosomes in both species. Very similar observations were seen for the linear regressions (Figure S1), with more models being significant and positive for *D. melanogaster*.

DISCUSSION

We present the first high resolution recombination map for *Drosophila simulans*, and a comparative analysis of recombination motifs and their association with recombination in two sister species, *D. melanogaster* and *D. simulans*. We tested the hypothesis that such motifs predict recombination rates within the *D. melanogaster* species subgroup.

Our *D. simulans* recombination map confirms the results of previous, lower resolution work in this species (Ohnishi and Voelker 1981, 1979; Stuktevanat 1929; True, et al. 1996). We find that the *D. simulans* recombination landscape is far flatter than in *D. melanogaster* (Figures 2, 3). While centromeric recombination suppression on the X, and to some extent on 2L and 3R, is observed in *D. simulans*, it is restricted to a small genomic region, whereas in *D. melanogaster* the recombination rate decreases only gradually over a much larger region in proximity to the centromeres (Comeron, et al. 2012). In *D. simulans*, a similarly sharp teleomeric suppression is also observed on 2L (and to some extent on X, 2R, 3L and 3R) at most smoothing scales (this pattern is less clear at 2501 kb). Unlike in *D. melanogaster*, overall recombination rates in *D. simulans* appear similar between X and the autosomes (Figures 2, 3) (Comeron, et al. 2012). We caution however, that recombination rate estimates

417 from population polymorphism data are sensitive to demographic events and particular the
418 ratio of X-chromosomal and autosomal variation differs widely between populations (Kauer,
419 et al. 2002; Schöfl and Schlötterer 2004). In *D. simulans* and *D. melanogaster*, mid-to-large
420 scale recombination features clearly persist over the 101, 501 and 2501 kb smoothed maps. In
421 *D. simulans*, our high-resolution map shows that such features also persist down to the 25 and
422 5 kb scale (e.g. the dip on 3R at 12.5 Mb position, Figures 3). As with the centromeric
423 differences however, mid-scale and narrower landscape features differ between the species,
424 especially at the 101 and 501 kb resolutions. In short, at all genomic scales tested the two
425 species differ dramatically in recombination rates, over broad- and finer-scale recombination
426 features.
427
428 Direct implications from these differences in genetic maps are that linkage-
429 disequilibrium should be both lower and less variable across the *D. simulans* chromosomes
430 relative to those of *D. melanogaster*. It is important to keep in mind that the *D. simulans*
431 reference genome includes less repetitive DNA at the centromeric and telomeric ends of the
432 chromosomes, so a comparison of recombination rates in not possible at the extremes of these
433 regions. Nonetheless, our results bolster the current understanding of *D. simulans*
434 recombination as less heterogeneous than that of *D. melanogaster* (Comeron, et al. 2012;
435 True, et al. 1996), and indicate that selection will be generally more efficient in *D. simulans*,
436 as genes that are uncoupled by recombination selection may result in more distinct signals, in
437 particular in Evolve and Resequence experiments (Barghi, et al. 2017; Kofler and Schlötterer
438 2014; Tobler, et al. 2014). Hence, adaptive evolutionary changes may occur more rapidly in
439 *D. simulans*, all else being equal, because Hill-Robertson effects are reduced by the higher
440 recombination (Hill and Robertson 1966).
441

442 Turning to the causes of this recombination variation, we ran a MEME motif search to
443 identify short DNA sequence motifs associated with regions of higher than average
444 recombination, repeating this search in both *D. melanogaster*, and *D. simulans*. The first point
445 of note was that a larger number of motifs were returned in *D. melanogaster*, and that those in
446 *D. simulans* were by tendency both shorter and showed a less significant association with
447 recombination rate, with lower quality matches. Nonetheless, a generally similar set of motifs
448 was recovered in each species, and across each recombination map smoothing scale
449 investigated. In short, we obtained a subset of the *D. melanogaster* motifs in *D. simulans*;
450 motifs C1, C5 and by trend, motifs C2 and C4, providing some confidence in the impact of
451 these motifs on the recombination rate. The motif sharing between the two *Drosophila*
452 species provides some evidence that recombination motifs may to some degree be universal
453 across *Drosophila* species. This idea builds upon prior work, which has shown that there is
454 some overlap in motifs between more distant *Drosophila* species, such as *D. pseudoobscura*,
455 which exhibits CACAC (Cirulli, et al. 2007), CCCCACCCC and CCTCCCT motifs
456 (Kulathinal, et al. 2008), and *D. persimilis*, which exhibits a CCNCCNTNNCCNC motif
457 (Stevison and Noor 2010). This led Comeron, et al. (2012) to speculate that *Drosophila* has a
458 stable set of recombination motifs of universal function, which they confirmed in part by
459 showing that *D. melanogaster* also exhibit the CACAC and CCTCCCT motifs, though not the
460 CCCCACCCC motif. Our study builds on this result, showing that a larger degree of motif
461 overlap can be seen both when contrasting consensus motifs and when comparing between
462 more closely related species, and that the $[CA]_n$ motif is universal to all *Drosophila* species
463 studied. However, it is immediately notable that no complex, multi-part motifs were
464 recovered in our study.
465

466 The genome-wide distribution of motifs (Figure 4) revealed, somewhat surprisingly,
467 that there are also clear parallels between the two species motif landscapes. Not only do
468 motifs with higher density in *D. melanogaster* generally have a higher density in *D. simulans*,
469 but the patterns of motif distribution genome-wide are also remarkably similar. For instance, a
470 similar "hump" and "peak" can be observed at the 8 and 9 Mb positions of chromosomes X
471 and 2L respectively, for motif C2, in both species, while a density "trough" can be seen at 15
472 Mb on chromosome 2L for this motif (Figure 4). Motifs C1, C3 and C4 likewise exhibit very
473 limited differences between species, on all chromosomes (Figure 4), despite clear differences
474 in recombination rates (Figure 3). A few differences do exist. Motif C1 is more common in
475 *D. melanogaster*, even if the "landscape" is similar to *D. simulans*; Motif C5 is less common
476 in *D. simulans*, and exhibits a distinct landscape on all autosomes; and, any narrow-scale
477 features rarely overlap between species, mirroring patterns of distinct recombination peaks
478 and similar landscapes seen in *D. melanogaster* populations (Chan, et al. 2012; Smukowski
479 Heil, et al. 2015). Consequently, while it might be tempting to speculate that subtle
480 differences in motif densities can explain the flatter recombination landscape of *D. simulans*
481 and its unique recombination peak set, it is difficult to reconcile the distinctive patterns of
482 recombination rate variation in the two species with their exceptionally similar motif density
483 landscapes, that are almost identical between species, especially when focusing on the large-
484 scale differences in centromeric and telomeric regions.
485
486 The similar motif density patterns between the two species cast doubt on the
487 hypothesis that differences in motif distribution can account for differences in recombination
488 variation in these species. If divergent motif densities really account for the species
489 differences in recombination rates, how can we explain the lack of concordance between
490 reduced recombination towards the centromeres in *D. melanogaster*, the lack of this reduction

491 in *D. simulans*, and the similar motif distributions over these regions in both species? To
492 investigate this observation quantitatively, we calculated Spearman's ρ as an estimator of the
493 correlation between genome-wide motif density and recombination rate (cM/Mb), for each
494 motif, in each species, across a range of smoothing scales. This revealed a striking difference
495 between the two species. In *D. melanogaster*, all associations (aside one) were positive, for all
496 motifs at all scales tested, with low *P*-values observed in most cases (Figure 5). These results
497 accord well with those of Adrian, et al. (2016), who found positive associations between
498 motif densities and recombination rate in *D. melanogaster*, using a similar set of motifs. In
499 contrast, the associations observed in *D. simulans* were far smaller, and far more
500 heterogeneous across chromosomes and motifs (Figure 5). This observation was confirmed by
501 our linear regression models, fitted to explicitly test the predictive power of each motif to
502 explain recombination rate variation, which showed an almost identical pattern (Figure S1).
503 The correlations and model fits were similar within each species across all smoothing scales,
504 in that the level of correlation did not increase with higher or lower resolution recombination
505 maps. The clear implication is that motif densities do not universally predict recombination
506 rates across the *Drosophila* clade, and are in particular not responsible for the large-scale
507 differences observed between our two species. It is therefore pertinent to ask what alternative
508 mechanisms could explain such differences.
509
510 A strong candidate is the dicistronic meiosis gene *mei-217*/*mei-218* and its protein
511 product, MEI-218 (Brand, et al. 2018), which is involved in the resolution of crossing over
512 into double stranded breaks and recombination (Brand, et al. 2018). Divergent forms have
513 recently been identified in *D. mauritiana* and *D. melanogaster*, species that diverged 0.6 – 0.9
514 Ma. Like *D. simulans*, *D. mauritiana* exhibits a higher and flatter recombination rate
515 landscape than *D. melanogaster* (True, et al. 1996), with the difference especially pronounced

516 in the centromeric and telomeric regions (True, et al. 1996), and with this pattern expressed to
517 an even larger extent than is seen in *D. simulans* (True, et al. 1996). Intriguingly then, Brand,
518 et al. (2018) also found a high divergence in DNA and protein structure in the *mei-217/mei-*
519 *218* gene and MEI-218 protein between *D. mauritiana* and *D. melanogaster*. The
520 *D. mauritiana* form was far more effective in promoting recombination, increasing
521 recombination assurance and reducing crossover interference (Brand, et al. 2018). It explained
522 a large portion of the variance in crossover rates between *D. mauritiana* and *D. melanogaster*,
523 especially that in the centromeric and telomeric regions (Brand, et al. 2018), and so could be a
524 primary mechanistic variant explaining the differences in recombination between *D. simulans*
525 and *D. melanogaster*. The clear parallel differences between the recombination maps of
526 *D. melanogaster* versus *D. mauritiana* and *D. melanogaster* versus *D. simulans* imply that
527 *mei-217/mei-218* may be responsible for the heterogeneity in recombination landscape that
528 we have observed.
529
530 What then might explain the clear correlations between motif density and
531 recombination seen in *D. melanogaster*, but not *D. simulans*? A simple explanation is that
532 motifs are responsible for variation in recombination rate on a local scale. Hence, the lower
533 density in *D. simulans*, results also in less micro-scale variation in recombination rate.
534 Alternatively, this pattern could be explained if the recombination motifs are recognised
535 directly by cleavage proteins, similar to PRDM9, that differ in function or effectiveness
536 between *D. simulans* and *D. melanogaster*. Recent evidence shows that a zinc-finger gene and
537 protein of this type exists in *D. melanogaster* (Hunter, et al. 2016a). Yet, such proteins tend to
538 bind to complex, rather than short-repeat motifs, making this explanation unlikely. Another
539 possibility relates to chromatin structure, because short-repeat DNA recombination motifs are
540 thought to play roles in loosening chromatin structure, increasing access for double-stranded

541 break (DSB) inducing proteins (Adrian, et al. 2016, and references therein; Comeron, et al.
542 2012). This could account for micro-variation in recombination rates genome-wide between
543 species, for instance because the motifs were generally shorter and so presumably less
544 effective at chromatin loosening in *D. simulans*, genome-wide. Circumstantial evidence in
545 favour of this hypothesis includes that in both species motif correlation patterns varied cross
546 chromosomes – for instance, C4 was a good predictor only on X – suggesting that motifs can
547 operate in a context dependent manner. Likewise, the removal of subcentromeric and
548 subtelomeric region recombination data has been found not to alter correlational patterns in
549 *D. melanogaster* (Adrian, et al. 2016), suggesting that if motifs densities explain some
550 recombination rate genome wide, they cannot explain centromeric and telomeric differences.
551
552 In short, we present the hypothesis that while short-repeat DNA motifs may affect
553 recombination at a micro-scale, genome-wide, for instance in relation to euchromatic
554 structure context, they cannot explain the large differences in recombination landscape
555 differences between species, especially in the centromeric and telomeric regions. This
556 variation seems far more likely to be explained by a mechanism such as *mei-217*/*mei-218*.
557
558
559 **SUPPLEMENTARY MATERIAL**
560 Supplementary Figures 1-2, and supplements 3-5 are available at [insert location here]. All
561 data generated and used in this study are available (see Data Accessibility below for details).
562
563 **DATA ACCESSIBILITY**
564 Raw sequence reads for the 189 isofemale line haplotypes are available to download at the
565 European Nucleotide Archive (ENA) under the accession numbers: [[added on acceptance]].

566 Phased haplotypes are available from Dryad via accession numbers: [[added on acceptance]].
567 Finally, CSV files for the *D. simulans* recombination map, at each resolution, are available for
568 download from Dryad via accession numbers [[added on acceptance]].
569

570 **ACKNOWLEDGMENTS**

571 The authors thank all members of the Institute of Population Genetics for discussion and
572 support on this project. This work was supported by the European Research Council (ERC)
573 grant "ArchAdapt" to CS, an Austrian Academy of Sciences DOC fellowship to TT, and the
574 Austrian Science Fund (FWF, W1225).
575

576 **AUTHOR CONTRIBUTIONS**

577 RM, JMH and CS conceived the study and interpreted the results. TT produced the
578 recombination map. VN generated the NGS libraries. RM performed the analysis, with input
579 from JMH. JMH and C.S. wrote the manuscript with input from all authors.
580

581 **REFERENCES**

582 Adrian AB, Corchado JC, Comeron JM 2016. Predictive models of recombination rate
583 Variation across the *Drosophila melanogaster* genome. Genome Biol Evol 8: 2597-2612. doi:
584 10.1093/gbe/evw181
585 Aquadro CF, Bauer DuMont V, Reed FA 2001. Genome-wide variation in the human and
586 fruitfly: a comparison. Curr Opin Genet Dev 11: 627-634.
587 Aquadro CF, Begun DJ, Kindahl EC. 1994. Selection, Recombination, and DNA
588 Polymorphism in *Drosophila*. In. Non-neutral evolution: Springer. p. 46-56.
589 Bailey TL, et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids
590 Res 37: W202-208. doi: 10.1093/nar/gkp335

591 Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover
592 motifs in biopolymers. Proceedings of the Second International Conference on Intelligent
593 Systems for Molecular Biolog. Menlo Park, California.
594 Bailey TL, Johnson J, Grant CE, Noble WS 2015. The MEME Suite. Nucleic Acids Res 43:
595 W39-49. doi: 10.1093/nar/gkv416
596 Baker CL, Walker M, Kajita S, Petkov PM, Paigen K 2014. PRDM9 binding organizes hotspot
597 nucleosomes and limits Holliday junction migration. Genome Res 24: 724-732. doi:
598 10.1101/gr.170167.113
599 Baker WK 1958. Crossing over in heterochromatin. Am. Nat. 92: 59-60. doi: 10.1086/282010
600 Barghi N, Tobler R, Nolte V, Schlötterer C 2017. *Drosophila simulans*: a species with improved
601 resolution in evolve and resequence studies. G3 (Bethesda) 7: 2337-2343. doi:
602 10.1534/g3.117.043349
603 Baudat F, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in
604 humans and mice. Science 327: 836-840.
605 Beadle GW 1932. A possible influence of the spindle fibre on crossing-over in *Drosophila*.
606 Proc Natl Acad Sci U S A 18: 160-165. doi: 10.1073/pnas.18.2.160
607 Begun DJ, Aquadro CF 1992. Levels of naturally occurring DNA polymorphism correlate with
608 recombination rates in *D. melanogaster*. Nature 356: 519-520. doi: 10.1038/356519a0
609 Begun DJ, et al. 2007. Population genomics: whole-genome analysis of polymorphism and
610 divergence in *Drosophila simulans*. PLoS Biol 5: e310. doi: 10.1371/journal.pbio.0050310
611 Benjamini Y, Hochberg Y 1995. Controlling the false discovery rate - a practical and powerful
612 approach to multiple testing. J R Stat Soc Series B Stat Methodol 57: 289-300.
613 Bergerat A, et al. 1997. An atypical topoisomerase II from Archaea with implications for
614 meiotic recombination. Nature 386: 414-417. doi: 10.1038/386414a0

615 Billings T, et al. 2013. DNA binding specificities of the long zinc-finger recombination protein
616 PRDM9. Genome Biol. 14: R35. doi: 10.1186/gb-2013-14-4-r35
617 Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC 2018. Molecular evolution at
618 a meiosis gene mediates species differences in the rate and patterning of recombination. Curr
619 Biol 28: 1289-1295 e1284. doi: 10.1016/j.cub.2018.02.056
620 Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV 2012. Genetic
621 recombination is directed away from functional genomic elements in mice. Nature 485: 642-
622 645. doi: 10.1038/nature11089
623 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL 1998. Comprehensive human
624 genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:
625 861-869. doi: 10.1086/302011
626 Chan AH, Jenkins PA, Song YS 2012. Genome-wide fine-scale recombination rate variation in
627 *Drosophila melanogaster*. PLoS Genet 8: e1003090. doi: 10.1371/journal.pgen.1003090
628 Charlesworth B, Charlesworth D. 2010. Elements of Evolutionary Genetics. Colorado: Roberts
629 and Company Publishers.
630 Charlesworth B, Lapid A 1992. The distribution of transposable elements within and between
631 chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and
632 distribution. Genet Res 60: 103-114.
633 Charlesworth B, Sniegowski P, Stephan W 1994. The evolutionary dynamics of repetitive DNA
634 in eukaryotes. Nature 371: 215-220. doi: 10.1038/371215a0
635 Choi K, Henderson IR 2015. Meiotic recombination hotspots - a comparative view. Plant J 83:
636 52-61. doi: 10.1111/tpj.12870
637 Choulet F, et al. 2014. Structural and functional partitioning of bread wheat chromosome 3B.
638 Science 345: 1249721. doi: 10.1126/science.1249721

639 Cirulli ET, Kliman RM, Noor MA 2007. Fine-scale crossover rate heterogeneity in *Drosophila*
640 *pseudoobscura*. J Mol Evol 64: 129-135. doi: 10.1007/s00239-006-0142-7
641 Cleveland WS, Grosse E, Shyu WM. 1992. Local regression models. In: Chambers JM, Hastie
642 TJ, editors. Statistical Models Wadsworth & Brooks/Cole.
643 Comeron JM, Ratnappan R, Bailin S 2012. The many landscapes of recombination in
644 *Drosophila melanogaster*. PLoS Genet 8: e1002905. doi: 10.1371/journal.pgen.1002905
645 Detlefsen JA, Roberts E 1921. Studies on crossing over—I. The effect of selection on crossover
646 values. J Exp Zool A Ecol Integr Physiol 32: 333–354.
647 Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA 2010. *Drosophila melanogaster*
648 recombination rate calculator. Gene 463: 18-20. doi: 10.1016/j.gene.2010.04.015
649 Garrison E, Marth G 2012. Haplotype-based variant detection from short-read sequencing.
650 arXiv: 1207.3907.
651 Grant CE, Bailey TL, Noble WS 2011. FIMO: scanning for occurrences of a given motif.
652 Bioinformatics 27: 1017-1018. doi: 10.1093/bioinformatics/btr064
653 Grey C, et al. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3
654 lysine 4 trimethylation for initiation of meiotic recombination. PLoS Biol 9: e1001176. doi:
655 10.1371/journal.pbio.1001176
656 Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS 2007. Quantifying similarity between
657 motifs. Genome Biol. 8: R24. doi: ARTN R24
658 10.1186/gb-2007-8-2-r24
659 Haddrill PR, Halligan DL, Tomaras D, Charlesworth B 2007. Reduced efficacy of selection in
660 regions of the *Drosophila* genome that lack crossing over. Genome Biol. 8: R18. doi:
661 10.1186/gb-2007-8-2-r18

662 Heil CS, Noor MA 2012. Zinc finger binding motifs do not explain recombination rate variation
663 within or between species of *Drosophila*. PLoS One 7: e45055. doi:
664 10.1371/journal.pone.0045055
665 Hermann P, Heissl A, Tiemann-Boege I, Futschik A 2018. LDJump: Estimating variable
666 recombination rates from population genetic data. bioRxiv. doi: 10.1101/190876
667 Hey J 2004. What's so hot about recombination hotspots? PLoS Biology 2: e190.
668 Heil CS, Ellison C, Dubin M, Noor MA 2015. Recombining without hotspots: a comprehensive
669 evolutionary portrait of recombination in two closely related species of *Drosophila*. Genome
670 Biol Evol 7: 2829-2842.
671 Hill WG, Robertson A 1966. The effect of linkage on limits to artificial selection. Genet Res 8:
672 269-294.
673 Hughes SE, Miller DE, Miller AL, Hawley RS 2018. Female meiosis: snapsis, recombination,
674 and segregation in *Drosophila melanogaster*. Genetics 208: 875-908.
675 Hunter CM, Huang W, Mackay TF, Singh ND 2016a. The genetic architecture of natural
676 variation in recombination rate in *Drosophila melanogaster*. PLoS Genet 12: e1005951. doi:
677 10.1371/journal.pgen.1005951
678 Hunter CM, Robinson MC, Aylor DL, Singh ND 2016b. Genetic background, maternal age,
679 and interaction effects mediate rates of crossing over in *Drosophila melanogaster* females. G3
680 (Bethesda) 6: 1409-1416. doi: 10.1534/g3.116.027631
681 International HapMap Consortium 2007. A second generation human haplotype map of over
682 3.1 million SNPs. Nature 449: 851.
683 Jeffreys AJ, Kauppi L, Neumann R 2001. Intensely punctate meiotic recombination in the class
684 II region of the major histocompatibility complex. Nat Genet 29: 217-222. doi:
685 10.1038/ng1001-217
686 John B. 2005. Meiosis. New York: Cambridge University Press.

687 Kauer M, Zangerl B, Dieringer D, Schlötterer C 2002. Chromosomal patterns of microsatellite
688 variability contrast sharply in African and non-African populations of *Drosophila*
689 *melanogaster*. Genetics 160: 247-256.
690 Keeney S 2001. Mechanism and control of meiotic recombination initiation. Curr Top Dev Biol
691 52: 1-53.
692 Keeney S, Giroux CN, Kleckner N 1997. Meiosis-specific DNA double-strand breaks are
693 catalyzed by Spo11, a member of a widely conserved protein family. Cell 88: 375-384.
694 Kofler R, Betancourt AJ, Schlötterer C 2012. Sequencing of pooled DNA samples (Pool-Seq)
695 uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*.
696 PLoS Genet 8: e1002487. doi: 10.1371/journal.pgen.1002487
697 Kofler R, Schlötterer C 2014. A guide for the design of evolve and resequencing studies. Mol
698 Biol Evol 31: 474-483. doi: 10.1093/molbev/mst221
699 Kohl KP, Singh ND 2018. Experimental evolution across different thermal regimes yields
700 genetic divergence in recombination fraction but no divergence in temperature associated
701 plastic recombination. Evolution 72: 989-999. doi: 10.1111/evo.13454
702 Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA 2008. Fine-scale mapping of
703 recombination rate in *Drosophila* refines its correlation to diversity and divergence. Proc Natl
704 Acad Sci U S A 105: 10051-10056. doi: 10.1073/pnas.0801848105
705 Lam I, Keeney S 2014. Mechanism and regulation of meiotic recombination initiation. Cold
706 Spring Harb Perspect Biol 7: a016634. doi: 10.1101/cshperspect.a016634
707 Lichten M, Goldman AS 1995. Meiotic recombination hotspots. Annu Rev Genet 29: 423-444.
708 doi: 10.1146/annurev.ge.29.120195.002231
709 Manzano-Winkler B, McGaugh SE, Noor MA 2013. How hot are *Drosophila* hotspots?
710 Examining recombination rate variation and associations with nucleotide diversity, divergence,
711 and maternal age in *Drosophila pseudoobscura*. PLoS One 8: e71582.

712 Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J 2009. A mouse speciation gene
713 encodes a meiotic histone H3 methyltransferase. Science 323: 373-375. doi:
714 10.1126/science.1163601
715 Miller DE, et al. 2016. Whole-genome analysis of individual meiotic events in *Drosophila*
716 *melanogaster* reveals that noncrossover gene conversions are insensitive to interference and the
717 centromere effect. Genetics 203: 159-171. doi: 10.1534/genetics.115.186486
718 Miller DE, et al. 2012. A whole-chromosome analysis of meiotic recombination in *Drosophila*
719 *melanogaster*. G3 (Bethesda) 2: 249-260. doi: 10.1534/g3.111.001396
720 Myers S, Bottolo L, Freeman C, McVean G, Donnelly P 2005. A fine-scale map of
721 recombination rates and hotspots across the human genome. Science 310: 321-324. doi:
722 10.1126/science.1117196
723 Myers S, et al. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in
724 meiotic recombination. Science 327: 876-879. doi: 10.1126/science.1182363
725 Nachman MW 2002. Variation in recombination rate across the genome: evidence and
726 implications. Curr Opin Genet Dev 12: 657-663.
727 Neel JV 1941. A relation between larval nutrition and the frequency of crossing over in the
728 third chromosome of *Drosophila melanogaster*. Genetics 26: 506-516.
729 Ohnishi S, Voelker RA 1981. Comparative studies of allozyme loci in *Drosophila simulans* and
730 *Drosophila melanogaster*. I. Three dipeptidase loci. Biochem Genet 19: 75-85.
731 Ohnishi S, Voelker RA 1979. Comparative studies of allozyme loci in *Drosophila-simulans*
732 and *D-melanogaster* .2. Gene arrangement on the 3rd chromosome. Jpn J Genet 54: 203-209.
733 doi: DOI 10.1266/jjg.54.203
734 Palmieri N, Nolte V, Chen J, Schlötterer C 2015. Genome assembly and annotation of a
735 *Drosophila simulans* strain from Madagascar. Mol Ecol Resour 15: 372-381. doi:
736 10.1111/1755-0998.12297

737 Parsons PA 1958. Selection for increased recombination in *Drosophila melanogaster*. Am Nat
738 92: 255-256. doi: Doi 10.1086/282033
739 Parvanov ED, Petkov PM, Paigen K 2010. *Prdm9* controls activation of mammalian
740 recombination hotspots. Science 327: 835. doi: 10.1126/science.1181495
741 Petes TD 2001. Meiotic recombination hot spots and cold spots. Nat Rev Genet 2: 360-369.
742 doi: 10.1038/35072078
743 Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J 2011. Population genomics
744 of transposable elements in *Drosophila melanogaster*. Mol Biol Evol 28: 1633-1644. doi:
745 10.1093/molbev/msq337
746 Pratto F, et al. 2014. DNA recombination. Recombination initiation maps of individual human
747 genomes. Science 346: 1256442. doi: 10.1126/science.1256442
748 R Core Development Team. 2018. R: A language and environment for statistical computing: R
749 Foundation for Statistical Computing.
750 Redfield H 1966. Delayed mating and the relationship of recombination to maternal age in
751 *Drosophila melanogaster*. Genetics 53: 593-607.
752 Rizzon C, Marais G, Gouy M, Biemont C 2002. Recombination rate and the distribution of
753 transposable elements in the *Drosophila melanogaster* genome. Genome Res 12: 400-407. doi:
754 10.1101/gr.210802
755 Roberts PA 1965. Difference in the behaviour of eu- and hetero-chromatin: crossing-over.
756 Nature 205: 725-726.
757 Roeder GS 1997. Meiotic chromosomes: it takes two to tango. Genes Dev 11: 2600-2621.
758 Schöfl G, Schlötterer C 2004. Patterns of microsatellite variability among X chromosomes and
759 autosomes indicate a high frequency of beneficial mutations in non-African *D. simulans*. J Mol
760 Biol Evol 21: 1384-1390.

761 Schwacha A, Kleckner N 1995. Identification of double Holliday junctions as intermediates in
762 meiotic recombination. Cell 83: 783-791. doi: 10.1016/0092-8674(95)90191-4
763 Singh ND, Aquadro CF, Clark AG 2009. Estimation of fine-scale recombination intensity
764 variation in the white-echinus interval of *D. melanogaster*. J Mol Evol 69: 42-53. doi:
765 10.1007/s00239-009-9250-5
766 Singh ND, Stone EA, Aquadro CF, Clark AG 2013. Fine-scale heterogeneity in crossover rate
767 in the garnet-scalloped region of the *Drosophila melanogaster* X chromosome. Genetics 194:
768 375-387. doi: 10.1534/genetics.112.146746
769 Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM 2017. Variation in
770 recombination frequency and distribution across eukaryotes: patterns and processes. Philos
771 Trans R Soc Lond B Biol Sci 372: 20160455. doi: 10.1098/rstb.2016.0455
772 Stern C 1926. An effect of temperature and age on crossing-over in the first chromosome of
773 *Drosophila melanogaster*. Proc Natl Acad Sci U S A 12: 530-532.
774 Stevison LS, Noor MA 2010. Genetic and evolutionary correlates of fine-scale recombination
775 rate variation in *Drosophila persimilis*. J Mol Evol 71: 332-345. doi: 10.1007/s00239-010-
776 9388-1
777 Stuktevanat H. 1929. The Genetics of *Drosophila simulans*. Carnegie Institute: Washington
778 Publishing Company.
779 Sturtevant AH, Beadle GW 1936. The relations of inversions in the X chromosome of
780 *Drosophila melanogaster* to crossing over and cisjunction. Genetics 21: 554-604.
781 Szauter P 1984. An analysis of regional constraints on exchange in *Drosophila melanogaster*
782 using recombination-defective meiotic mutants. Genetics 106: 45-71.
783 Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW 1983. The double-strand-break repair
784 model for recombination. Cell 33: 25-35.

785 Termolino P, Cremona G, Consiglio MF, Conicella C 2016. Insights into epigenetic landscape
786 of recombination-free regions. Chromosoma 125: 301-308. doi: 10.1007/s00412-016-0574-9
787 Tobler R, et al. 2014. Massive habitat-specific genomic response in *D. melanogaster*
788 populations during experimental evolution in hot and cold environments. Mol Biol Evol 31:
789 364-375. doi: 10.1093/molbev/mst205
790 True JR, Mercer JM, Laurie CC 1996. Differences in crossover frequency and distribution
791 among three sibling species of *Drosophila*. Genetics 142: 507-523.
792
793 **SI. Legends**
794
795 **Fig. S1.** Linear models predict some of the variance in recombination rate in
796 *D. melanogaster*, but not in *D. simulans*. Scatterplot of recombination rate vs. motif density
797 for (a) *D. melanogaster* and (b) *D. simulans* (species also indicated by blue and red colour,
798 respectively). Gray lines represent single-motif linear model fits, inset numbers the
799 corresponding r^2 values, and appended asterisks indicate the *p*-values of the model fits at * <
800 0.05, ** < 0.01, *** < 0.001. For purposes of this comparison only, smoothed *D. simulans*
801 data at 101k is shown here, with the same resolution of the *D. melanogaster* data. The
802 recognisable correlation features in (b) are unaffected by this downsampling step (not shown).
803
804 **Fig. S2.** Loess-smoothed recombination maps. Red lines show the recombination rate in
805 *D. simulans* for each of the major chromosomes (name labels in top margin), smoothed at 4
806 window sizes (see right margin, in bp) with the LOESS span parameter. LOESS span
807 parameters correspond to 25, 101, 501, and 2501 kb, as span parameters equivalent to 5 kb
808 can't be implemented. For comparison, blue lines show the recombination rate in

809 *D. melanogaster* (with data taken from Comeron *et al*. 2012 at 101k; and then smoothed at

810 501k and 2501k; with data not available at smaller resolutions).

811

812 **S3.** MEME motif discovery output for each *Drosophila* species at each genomic resolution.

813

814 **S4.** TomTom contrast of motifs from Adrian et al. (2016) to our set of 5 consensus motifs.

815

816 **S5.** R-Markdown document with script to reproduce our results [on acceptance of the article].