Epistatic selection on a selfish *Segregation Distorter* supergene: drive, recombination, and genetic load

Beatriz Navarro-Domínguez, Ching-Ho Chang, Cara L. Brand, Christina A. Muirhead, Daven C. Presgraves and Amanda M. Larracuente

Department of Biology, University of Rochester, Rochester, New York, 14610, USA

Correspondence to: daven.presgraves@rochester.edu, alarracu@bio.rochester.edu

Abstract

Meiotic drive supergenes are complexes of alleles at linked loci that together subvert Mendelian segregation to gain preferential transmission. In males, the most common mechanism of drive involves the disruption of sperm bearing alternative alleles. While at least two loci are important for male drive— the driver and the target— linked modifiers can enhance drive, creating selection pressure to suppress recombination. In this work, we investigate the evolution and genomic consequences of an autosomal multilocus, male meiotic drive system, *Segregation Distorter* (*SD*) in the fruit fly, *Drosophila melanogaster*. In African populations, the predominant *SD* chromosome variant, *SD-Mal*, is characterized by two overlapping, paracentric inversion on chromosome arm *2R* and nearly perfect (~100%) transmission. We study the *SD-Mal* system in detail, exploring its components, chromosomal structure, and evolutionary history. Our findings reveal a recent chromosome-scale selective sweep mediated by strong epistatic selection for haplotypes carrying *Sd*, the main driving allele, and one or more factors within the double inversion. While most *SD-Mal* chromosomes are homozygous lethal, *SD-Mal* haplotypes can recombine with other, complementing haplotypes via crossing over and with wildtype chromosomes only via gene conversion. *SD-Mal* chromosomes have nevertheless accumulated lethal mutations, excess non-synonymous mutations, and excess transposable element insertions. Therefore, *SD-Mal* haplotypes evolve as a small, semi-isolated subpopulation with a history of strong selection. These results may explain the evolutionary turnover of *SD* haplotypes in different populations around the world and have implications for supergene evolution broadly.

Introduction

Supergenes are clusters of linked loci that control complex phenotypes. Some supergenes mediate adaptive polymorphisms that are generally maintained by some form of frequency- or density-dependent natural selection, as in, *e.g.*, mimicry in butterflies, self-incompatibility in plants, plumage polymorphisms in birds, and heteromorphic sex chromosomes (see SCHWANDER *et al.* 2014; THOMPSON AND JIGGINS 2014 for review). Other supergenes are maintained by selfish social behaviors that enhance the fitness of carriers at the expense of non-carriers, as in some ant species (KELLER AND ROSS 1998; WANG *et al.* 2013). Still other supergenes are maintained by their ability to achieve selfish, better-than-Mendelian transmission during gametogenesis, as in so-called meiotic drive complexes found in fungi, insects, and mammals (LYON 2003; LARRACUENTE AND PRESGRAVES 2012; LINDHOLM *et al.* 2016; SVEDBERG *et al.* 2018).

Meiotic drive complexes gain transmission advantages at the expense of other loci and their hosts. In heterozygous carriers of male drive complexes, the driver disables spermatids that bear drive-sensitive target alleles (LARRACUENTE AND PRESGRAVES 2012; LINDHOLM *et al.* 2016). To spread in the population, the driver must be linked in a *cis*-arrangement to a drive-resistant (insensitive) target allele (CHARLESWORTH AND HARTL 1978). Recombination between the driver and target results in a "suicide" haplotype that distorts against itself (SANDLER AND CARPENTER 1972; HARTL 1974). These epistatic interactions between driver and target lead to selection for modifiers of recombination that tighten linkage, such as chromosomal inversions (SCHWANDER *et al.* 2014; THOMPSON AND JIGGINS 2014; CHARLESWORTH 2016). Like most supergenes (TURNER 1967; CHARLESWORTH AND CHARLESWORTH 1975), meiotic drive complexes originate from two or more loci with some degree of initial linkage. Successful drivers thus tend to be located in regions of low recombination, such as non-recombining sex chromosomes (HAMILTON 1967; HURST AND POMIANKOWSKI 1991), centromeric regions, or in chromosomal inversions of autosomes (LYON 2003; LARRACUENTE AND PRESGRAVES 2012; LINDHOLM *et al.* 2016; SVEDBERG *et al.* 2018).

The short-term benefits of reduced recombination can entail long-term costs. Chromosomal inversions that lock supergene loci together also incidentally capture linked loci, which causes large chromosomal regions to segregate as blocks. Due to reduced recombination, the efficacy of natural selection in these regions is compromised: deleterious mutations can accumulate and beneficial ones are more readily lost (MULLER 1964; HILL AND ROBERTSON 1968; FELSENSTEIN 1974). Many meiotic drive complexes are thus homozygous lethal or sterile. The degeneration of drive haplotypes is not however inevitable, as different drive haplotypes that complement one another (DOD *et al.* 2003; PRESGRAVES *et al.* 2009; BRAND *et al.* 2015), they may be able to recombine, if only among themselves. Gene conversion from wildtype chromosomes may also ameliorate the genetic load of supergenes (UYENOYAMA 2005; WANG *et al.* 2013; TUTTLE *et al.* 2016; BRANCO *et al.* 2018; STOLLE *et al.* 2019; BRELSFORD *et al.* 2020). Male meiotic drive complexes thus represent a class of selfish supergenes that evolve and persist via the interaction of drive, recombination, and natural selection.

Here we focus on the evolutionary genetics of *Segregation Distorter* (*SD*), a well-known autosomal meiotic drive complex in *Drosophila melanogaster* (SANDLER *et al.* 1959). In heterozygous males, *SD* disables sperm bearing drive-sensitive wildtype chromosomes via a chromatin condensation defect (HARTL *et al.* 1967 ; TEMIN *et al.* 1991). *SD* has two main components: the driver, *Segregation distorter* (*Sd*), is a truncated duplication of the gene *RanGAP* located in chromosome arm *2L* (POWERS AND GANETZKY 1991; MERRILL *et al.* 1999; KUSANO *et al.* 2001); and the target of drive, *Responder* (*Rsp*), is a block of satellite

DNA in the pericentromeric heterochromatin of *2R*. Previous studies of *SD* chromosomes have detected linked upward modifiers of drive, including *Enhancer of SD* (*E[SD]*) on *2L* and several others on *2R* (SANDLER AND HIRAIZUMI 1960; MIKLOS 1972; GANETZKY 1977; HIRAIZUMI *et al.* 1980; BRITTNACHER AND GANETZKY 1984), but their molecular identities are unknown. *Sd-RanGAP* and *Rsp* straddle the centromere, a region of reduced recombination, and some *SD* chromosomes bear pericentric inversions that presumably further tighten linkage among these loci. Many *SD* chromosomes also bear paracentric inversions on *2R* (reviewed in LYTTLE 1991; LARRACUENTE AND PRESGRAVES 2012). Although recombination between paracentric inversions and the main components of *SD* is possible, their strong association implies a role for epistatic selection in the evolution of these supergenes (LARRACUENTE AND PRESGRAVES 2012).

While *SD* is present at low population frequencies (<5%) around the world (TEMIN *et al.* 1991; LARRACUENTE AND PRESGRAVES 2012), *Sd-RanGAP* appears to have originated in sub-Saharan Africa, the ancestral geographic range of *D. melanogaster* (PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). The predominant *SD* variant in Africa is *SD-Mal*, which recently swept across the entire continent (PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). *SD-Mal* has a pair of rare, African-endemic, overlapping paracentric inversions on *2R*: *In(2R)51B6–11;55E3–12* and *In(2R)44F3–12;54E3–10*, hereafter collectively referred to as *In(2R)Mal* (AULARD *et al.* 2002; PRESGRAVES *et al.* 2009). *SD-Mal* chromosomes are particularly strong drivers, with ~100% transmission. Notably, recombinant chromosomes bearing the *Sd-RanGAP* duplication from this haplotype but lacking the inversions do not drive (PRESGRAVES *et al.* 2009), suggesting that *In(2R)Mal* is essential for *SD-Mal* drive. We therefore expect strong epistatic selection to enforce the association of *Sd-RanGAP* and *In(2R)Mal.* The functional role of *In(2R)Mal* for drive is still unclear: do these inversions function to suppress recombination between *Sd-RanGAP* and a major distal enhancer on *2R*, or do they contain a major enhancer?

Here, we combine genetic and population genomic approaches to study *SD-Mal* haplotypes sampled from a single population in Zambia, the putative ancestral range of *D. melanogaster* (POOL *et al.* 2012). We address four issues. First, we reveal the structural features of the *SD-Mal* haplotype, including the organization of the insensitive *Rsp* allele and the *In(2R)Mal* rearrangements. Second, we characterize the genetic function of *In(2R)Mal* and its role in drive. Third, we infer the population genetic history of the rapid rise in frequency of *SD-Mal* in Zambia. And fourth, we explore the evolutionary consequences of reduced recombination on *SD-Mal* haplotypes. Our results show that *SD-Mal* experienced a recent chromosome-scale selective sweep mediated by epistatic selection and has, as a consequence of its reduced population recombination rate, accumulated excess non-synonymous mutations and transposable element

insertions. The *SD-Mal* haplotype is a supergene that evolves as a small, semi-isolated subpopulation in which complementing *SD-Mal* chromosomes can recombine *inter se* via crossing over and with wildtype chromosomes only via gene conversion. These results have implications for supergene evolution and may explain the enigmatic evolutionary turnover of *SD* haplotypes in different populations around the world.

Results and discussion

To investigate the evolutionary genomics of *SD-Mal*, we sequenced haploid embryos from nine driving *SD-Mal* haplotypes sampled from a single population in Zambia (BRAND *et al.* 2015), the putative ancestral range of *D. melanogaster* (POOL *et al.* 2012). Illumina read depth among samples ranged between ~46-67x, (Supplementary Table S1; BioProject PRJNA649752 in NCBI). Additionally, we obtained ~12x coverage with long-read Nanopore sequencing of one homozygous viable line, *SD-ZI125*, to create a *de novo* assembly of a representative *SD-Mal* haplotype. We use these data to study the evolution of *SD-Mal* structure, diversity, and recombination.

Chromosomal features of the SD-Mal supergene

The *SD-Mal* haplotype has at least three key features: the main drive locus, the *Sd-RanGAP* duplication on *2L*; an insensitive *Responder* (Rsp^{i}) in *2R* heterochromatin; and the paracentric *In(2R)Mal* arrangement on chromosome *2R* (Figure 1). We used our long-read and short-read sequence data for *SD-ZI125* to confirm the structure of the duplication (Fig. 1a) and then validated features in the other *SD-Mal* haplotypes. All *SD-Mal* chromosomes have the *Sd-RanGAP* duplication at the same location as the parent gene on chromosome *2L* (see also BRAND *et al.* 2015). The *Rsp* locus, the target of *SD,* corresponds to a block of ~120-bp satellite repeats in *2R* heterochromatin (Fig. 1b; WU *et al.* 1988). The reference genome, *Iso-1,* has a Rsp^{s} allele corresponding to a primary *Rsp* locus containing two blocks of tandem *Rsp* repeats— *Rsp-proximal* and *Rsp-major*— with ~1000 copies of the *Rsp* satellite repeat interrupted by transposable elements (KHOST *et al.* 2017). A small number of *Rsp* repeats exist outside of the primary *Rsp* locus, although they are not known to be targeted by *SD*. There are three of these additional *Rsp* loci in *Iso-1*: ~10 copies in *2R*, distal to the major *Rsp* locus (*Rsp-minor*); a single copy at the distal end of *2R* (60A); and ~12 copies in *3L* (HOUTCHENS AND LYTTLE ; LARRACUENTE 2014; KHOST *et al.* 2017). The genomes of *SD* flies carry ~20 copies of *Rsp* (WU *et al.* 1988; PIMPINELLI AND DIMITRI 1989), but the organization of the primary *Rsp* locus on *SD* chromosomes is unknown. To characterize the *Rsp* locus of the *SD-Mal* haplotype, we mapped *SD-Mal* reads to an *Iso-1* reference genome (see KHOST *et al.* 2017). As expected, reads from *Iso-1* reference are distributed across the whole *Rsp-major* region. For *SD-Mal* chromosomes, however,

very few reads map to the *Rsp* repeats at the *Rsp-major* (Fig. 1b). This suggests that all *SD-Mal* have a complete deletion of the primary *Rsp* locus containing *Rsp-proximal* and *Rsp-major* and that the only *Rsp* copies in the *SD-Mal* genomes are the minor *Rsp* loci in chromosomes *2R* and *3L* (Suppl. Fig. S1).

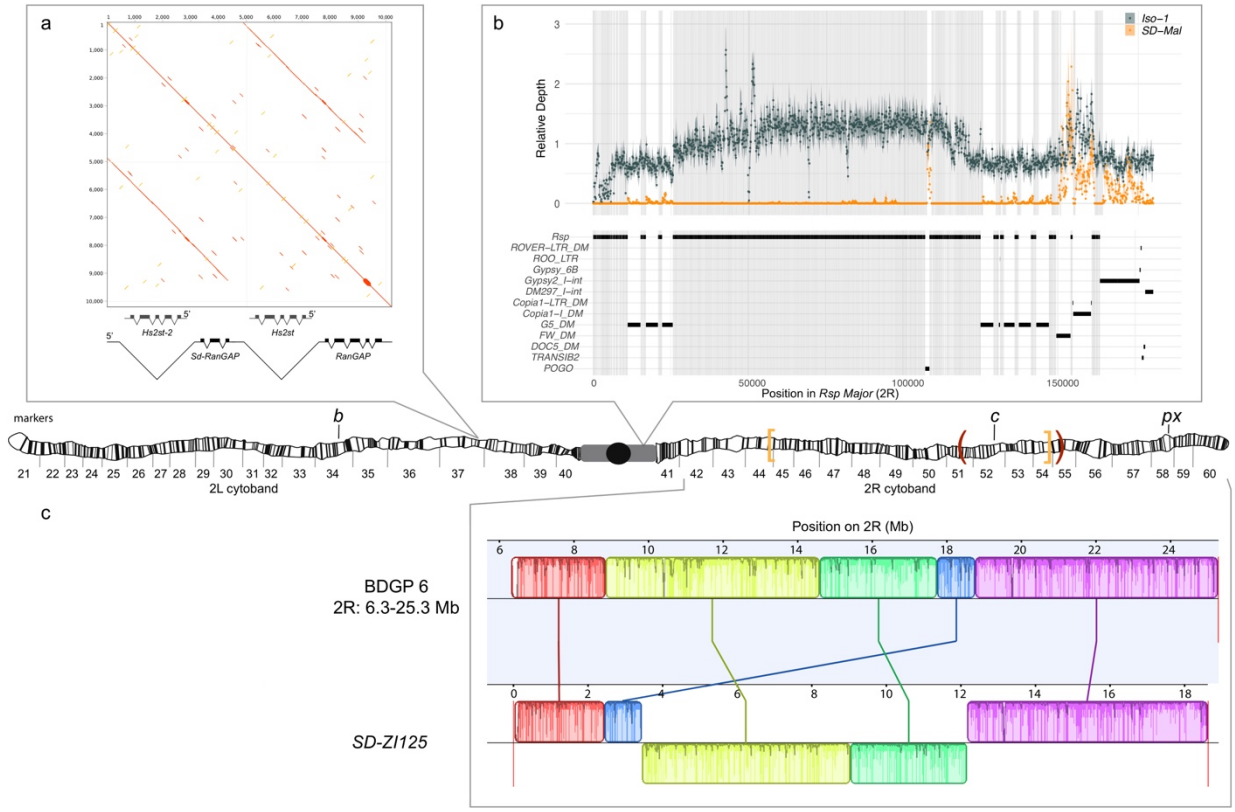

Figure 1. Map depicting the chromosomal features of the *SD-Mal* chromosome. The schematic shows the cytogenetic map of chromosomes *2L* and *2R* (redrawn based on images in (LEFEVRE 1976)) and the major features of the chromosome. (a) Dotplot showing that the *Sd* locus is a partial duplication of the gene *RanGAP* (in black), located at band *37D2-6*. (b) The *Rsp-major* locus is an array of tandem repeats located in the pericentric heterochromatin (band *h39*). Read mapping shows that *SD-Mal* chromosomes do not have any *Rsp* repeats in the *Rsp-major* locus, consistent with being insensitive to distortion by *Sd* (*Rsp*^{i}) (orange, high relative coverage regions correspond to transposable element interspersed), in contrast with *Iso-1*, which is sensitive (*Rsp*^{s}). (c) Two paracentric, overlapping inversions constitute the *In(2R)Mal* arrangement: *In(2R)51BC;55E* (*In(2R)Mal-p*) in orange brackets and *In(2R)44F;54E* (*In(2R)Mal-d*) in red parentheses). Pericentromeric heterochromatin and the centromere are represented by a grey rectangle and black circle, respectively. Our assembly based on long-read sequencing data provide the exact breakpoints of *In(2R)Mal* and confirms that the distal inversion (*Dmel.r6*, *2R*:14,591,034-18,774,475) occurred first, and the proximal inversion (*Dmel.r6*, *2R*:8,855,601-15,616,195) followed, overlapping ~1Mb with the distal inversion. The colored rectangles correspond to locally collinear blocks of sequence. Blocks below the center black line indicate regions that align in the reverse complement orientation. Vertical red lines indicate the end of the assembled chromosomes. Visible marker locations used for generating recombinants (*b* (34D1), *c* (52D1), and *px* (58E4-58E8)) are indicated on the cytogenetic map.

The complex *In(2R)Mal* inversion is distal to the *Rsp* locus on chromosome *2R* (Fig 1c). We used our *SD-ZI125* assembly to determine the precise breakpoints of these inversions. Relative to the standard *D. melanogaster 2R* scaffold (BDGP6), *SD-ZI125* has three large, rearranged blocks of sequence corresponding to *In(2R)Mal* (Fig. 1c): a 1.03-Mb block collinear with the reference but shifted proximally; a second inverted 5.74-Mb block; and a third inverted 3.16-Mb block. From this organization, we infer that the distal inversion, which we refer to as *In(2R)Mal-d,* occurred first and spanned 4.18 Mb (approx. *2R*:14,591,003-18,774,475). The proximal inversion, which we refer to as *In(2R)Mal-p,* occurred second and spanned 6.76 Mb, with 1.02 Mb overlapping with the proximal region of *In(2R)Mal-d* (approx. *2R*:8,855,602- 17,749,310; see Suppl. Fig. S2). All four breakpoints of the *In(2R)Mal* rearrangement involve simple joins of unique sequence. Three of these four breakpoints span genes: *sns* (*2R*:8,798,489 - 8,856,091), *CG10931* (*2R*:17,748,935 -17,750,136), and *Mctp* (*2R*:18,761,758 - 18,774,824). The CDSs of both *sns* and *Mctp* remain intact in the *In(2R)Mal* arrangement, with the inversion disrupting their 3'UTRs. Neither of these two genes is expressed in testes (https://flybase.org/reports/FBgn0024189; https://flybase.org/reports/FBgn0034389; CHINTAPALLI *et al.* 2007; FB2021_06; LARKIN *et al.* 2021), so it is unlikely that they are related to drive. *In(2R)Mal-p* disrupted the CDS of *CG10931*, which is a histone methyltransferase with high expression levels in testis (https://flybase.org/reports/FBgn0034274; CHINTAPALLI *et al.* 2007; FB2021_06; LARKIN *et al.* 2021). Future work is required to determine if this gene has a role in the *SD-Mal* drive phenotype.

In African populations, *In(2R)Mal* appears essential for *SD* drive: recombinant chromosomes bearing *Sd* but lacking *In(2R)Mal* do not drive (PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). The functional role of *In(2R)Mal* in drive is however unclear. As expected, *In(2R)Mal* suppresses recombination: in crosses between a multiply marked chromosome *2*, *b c px*, and *SD-Mal*, we find that *In(2R)Mal* reduces the *b – c* genetic distance by 54.6% (from 26.6 to 11.8 cM) and the *c - px* genetic distance by 92.4% (from 23.2 to 1.8 cM), compared with control crosses between *b c px* and Oregon-R. Our crosses also confirm that *In(2R)Mal* is indeed required for drive: all *Sd, In(2R)Mal* haplotypes show strong drive (Table 1, rows 1 and 2), whereas none of the recombinants that separate *Sd* and *In(2R)Mal* drive (Table 1, rows 3 and 4). We conclude that *SD-Mal* drive requires both *Sd* and *In(2R)Mal*, which implies that one or more essential enhancers, or co-drivers, is located within or distal to *In(2R)Mal*.

Table 1. Strength of segregation distortion in recombinants of *SD-ZI125*.

	Genotype	Markers	*N*	*n*	*+/- SE*	*k*	*+/- SE*	*k**	*+/- SE*	*p-value* (*k** = 0.5)
1	*Sd,In(2R)Mal*	*+++; b++*	112	90.3	6.03	0.99	0.00	0.98	0.00	< 0.0001
2	*Sd,In(2R)Mal*	*++px*	71	118.8	9.48	0.97	0.01	0.96	0.01	< 0.0001
3	*Sd,In(2R)Mal*$^{+}$	*+c px*	19	147.6	14.39	0.54	0.01	0.51	0.01	0.3082
4	*Sd*$^{+}$*,In(2R)Mal*	*b++*	24	124.8	10.31	0.68	0.02	0.55	0.03	0.0572
5	*Sd*$^{+}$*,In(2R)Mal*$^{+}$	*+c px*	65	120.4	8.32	0.53	0.01	0.51	0.01	0.3586

Chromosome *2* markers are *black* (*b*)*, curved* (*c*)*,* and *plexus* (*px*). *N*, number of crosses; *n*, average number of progeny from the crosses; *SE*, standard error; *k*, average proportion of progeny inheriting the partial SD_r chromosome from SD_r*/bcpx* males; *k**, average proportion of progeny inheriting the partial SD_r chromosome from SD_r*/bcpx* males, corrected for viability. *P-values* reported by a single sample t-test with a null hypothesis of *k**=0.5, as expected for Mendelian segregation.

The temporal order of inversions (first *In(2R)Mal-d*, then *In(2R)Mal-p*) suggests two possible scenarios. *In(2R)Mal-d,* occurring first, may have captured the essential enhancer, with the subsequent *In(2R)Mal-p* serving to further reduce recombination between *Sd* and the enhancer. Alternatively, an essential enhancer is located distal to *In(2R)Mal-d*, and the role of both *In(2R)Mal* inversions is to reduce recombination with *Sd*. To distinguish these possibilities, we measured drive in b^{+} *Sd* c^{+} *In(2R)Mal px* recombinants, which bear *Sd* and *In(2R)Mal* but recombine between the distal breakpoint of *In(2R)Mal* (*2R*: 18,774,475) and *px (2R: 22,494,297)*. All of these recombinants show strong drive (n=71; Table 1, row 2). Assuming that recombination is uniformly distributed throughout the 3.72-Mb interval between the *In(2R)Mal-d* distal breakpoint and *px*, the probability of failing to separate an essential codriver or distal enhancer among any of our 71 recombinants is <0.014. Furthermore, using molecular markers, we detected two recombinants within 100 kb of the distal breakpoint of *In(2R)Mal*, both with strong drive (*k*>0.99). We therefore infer that the co-driver likely resides within the *In(2R)Mal* arrangement. More specifically, we speculate that the *In(2R)Mal-d* inversion both captured the co-driver and reduced recombination with *Sd*, whereas *In(2R)Mal-p* tightened linkage between centromere-proximal components of *SD-Mal* and *In(2R)Mal-d.*

While recombination occurs readily between *Sd* and the proximal break of *In(2R)Mal* (Table 1; PRESGRAVES *et al.* 2009; BRAND *et al.* 2015), long-range linkage disequilibrium nevertheless exists between *Sd* and *In(2R)Mal*. Using 204 haploid genomes from Zambia (LACK *et al.* 2016; see Methods), we identified 198 wildtype haplotypes (Sd^{+} $In(2R)Mal^{+}$), 3 *SD-Mal* haplotypes (*Sd In(2R)Mal*), 3 recombinant haplotypes (all *Sd* $In(2R)Mal^{+}$, none Sd^{+} *In(2R)Mal*). Despite the individually low frequencies of *Sd* (frequency = 0.0294) and *In(2R)Mal* (frequency = 0.0147), they tend to co-occur on the same haplotypes (Fisher's exact P=1.4 x 10^{-5}). Given a conservative recombination frequency between *Sd* and *In(2R)Mal* of ~5 cM (FlyBase; FB2021_06; LARKIN *et al.* 2021), the observed estimated coefficient of linkage disequilibrium, D = 0.0143, has a half-life of just ~14 generations (1.4 years, assuming 10 generations per year) and should decay to negligible within 100 generations (10 years). We therefore conclude that strong epistatic selection maintains the *SD-Mal* supergene haplotype.

Rapid increase in frequency of the SD-Mal supergene

We used population genomics to infer the evolutionary history and dynamics of *SD-Mal* chromosomes. We called SNPs in our Illumina reads from nine complete *SD-Mal* haplotypes from Zambia (see Methods). For comparison, we also analyzed wildtype (SD^{+}) chromosomes from the same population in Zambia (LACK *et al.* 2016), including those with chromosome *2* inversions: 10 with the *In(2L)t* inversion and 10 with the *In(2R)NS* inversion (see Methods). Table 2 shows that nucleotide diversity (π) is significantly lower on *SD-Mal* haplotypes compared to uninverted SD^{+} chromosome arms (Table 2, rows 1 and 4; Figure 2a). The relative reduction in diversity on *SD-Mal* haplotypes is distributed heterogeneously: π is sharply reduced for a large region that spans ~25.8 Mb, representing 53% of chromosome 2 and extending from *Sd-RanGAP* on *2L* (*2L*: 19,441,959; Suppl. Fig. S3), across the centromere, and to ~2.9 Mb beyond the distal breakpoint of *In(2R)Mal* (*2R*: 18,774,475; Table 2, rows 3, 5 and 6; Fig. 2a). Thus, the region of reduced nucleotide diversity on *SD-Mal* chromosomes covers all of the known essential loci for the drive phenotype: *Sd-RanGAP*, Rsp^{i} and *In(2R)Mal*.

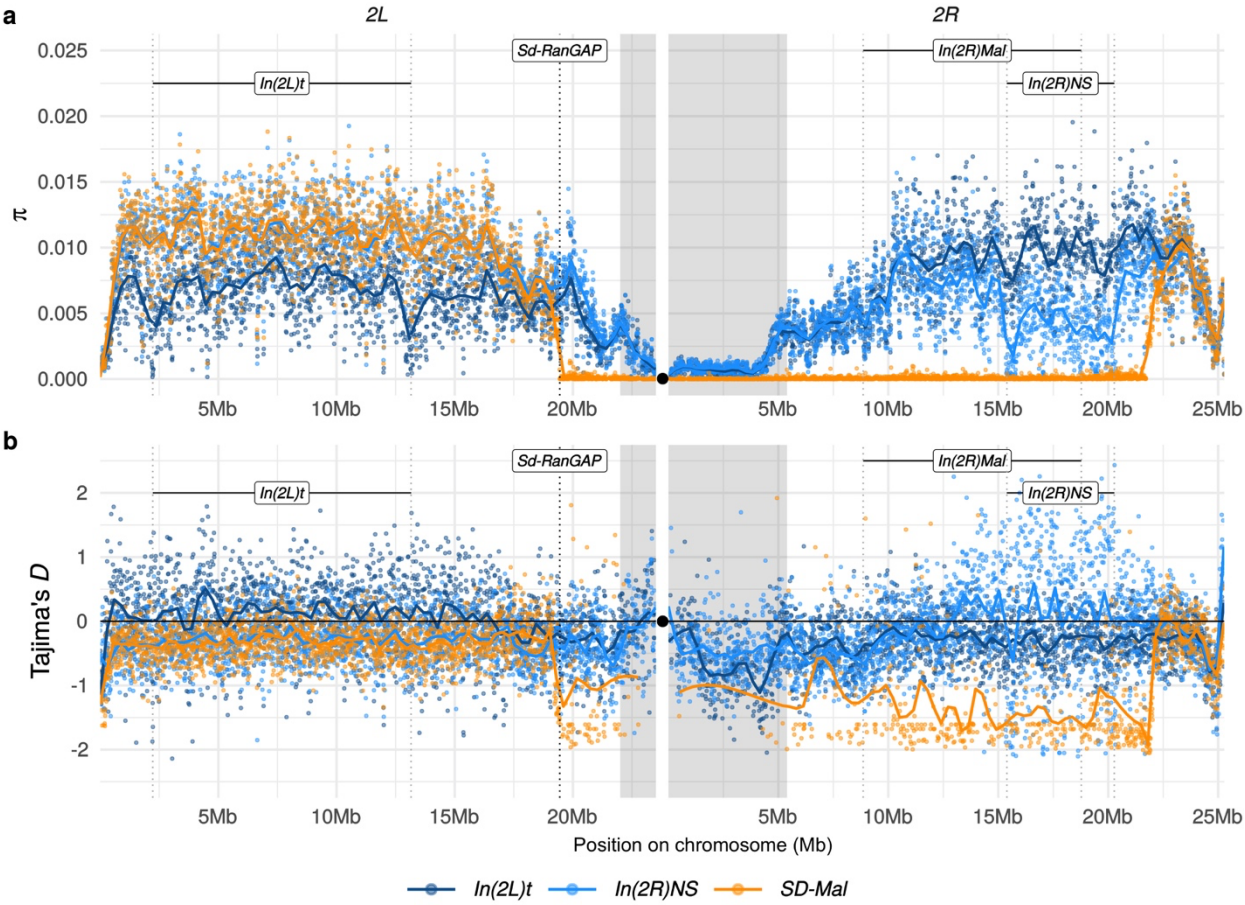

Figure 2. Diversity on *SD-Mal* chromosomes. (a) Average pairwise nucleotide diversity per site (π) and (b) Tajima's *D* estimates in non-overlapping 10-kb windows along chromosome *2*, in Zambian *SD-Mal* chromosomes (n=9, orange) and SD^{+} chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10, dark blue) and *In(2R)NS* (n=10, light blue). Regions corresponding to pericentric heterochromatin are shaded in grey and the centromere location is marked with a black circle. *SD-Mal* chromosomes show a sharp decrease in nucleotide diversity and skewed frequency spectrum from the *Sd* locus (*Sd-RanGAP*, *2L*:19.4Mb) to ~2.9 Mb beyond the distal breakpoint of *In(2R)Mal*.

247

Table 2. Nucleotide diversity (π) on *SD-Mal* an SD^+ chromosomes.

	Chr.	Region	π (+/- st.dev)			p-value	
			SD^+	*SD-Mal*	$SD^+ \times f$	*SD-Mal vs.* SD^+	*SD-Mal vs.* $SD^+ \times f$
1	*2L*	Whole arm	9.41E-03 (+/- 3.65E-03)	8.69E-03 (+/- 4.63E-03)	1.38E-04 (+/- 5.36E-05)	2.68E-08	0.00E+00
2	*2L*	Distal to *Sd-RanGAP*	1.03E-02 (+/- 3.01E-03)	1.03E-02 (+/- 3.09E-03)	1.52E-04 (+/- 4.43E-05)	0.5727	0.00E+00
3	*2L*	Proximal to *Sd-RanGAP*	4.44E-03 (+/- 2.75E-03)	9.39E-05 (+/- 1.66E-04)	6.52E-05 (+/- 4.04E-05)	5.84E-90	0.0027
4	*2R*	Whole arm	6.96E-03 (+/- 4.03E-03)	1.11E-03 (+/- 2.73E-03)	1.02E-04 (+/- 5.93E-05)	0.00E+00	2.82E-63
5	*2R*	*In(2R)Mal*	8.94E-03 (+/- 2.95E-03)	7.97E-05 (+/- 1.18E-04)	1.31E-04 (+/- 4.33E-05)	0.00E+00	1.42E-33
6	*2L-2R*	*SD-Mal* supergene	6,42E-03 (+/- 4.03E-03)	7.98E-05 (+/- 1.32E-04)	9.43E-05 (+/- 5.92E-05)	0.00E+00	2.60E-06

Average nucleotide diversity (π) per nucleotide and standard deviation estimated in 10-kb windows along chromosome 2, for SD^+, *SD-Mal* and SD^+ scaled by the estimated frequency of *SD-Mal* chromosomes ($SD^+ \times f$; f=1.47%). *P-values* reported by paired t-test between 10 kb windows.

248

249

250 The reduced nucleotide diversity among *SD-Mal* may not be surprising, given its low frequency in natural
251 populations (see below; PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). *SD* persists at low frequencies in
252 populations worldwide, presumably reflecting the balance between drive, negative selection, and genetic
253 suppression and/or resistance (HARTL 1975; CHARLESWORTH AND HARTL 1978; LARRACUENTE AND
254 PRESGRAVES 2012). If the *SD-Mal* supergene has been maintained at stable drive-selection-suppression
255 equilibrium frequency for a long period of time, then its neutral nucleotide diversity may reflect a mutation-
256 drift equilibrium appropriate for its effective population size. Under this scenario, we expect diversity at
257 the supergene to be similar to wild type (SD^+) diversity scaled by the long-term equilibrium frequency of
258 *SD*. We estimated *SD-Mal* frequency to be 1.47% by identifying the *Sd* duplication and *In(2R)Mal*
259 breakpoints in 204 haploid genomes from Zambia (3/204, comparable to PRESGRAVES *et al.* 2009; BRAND
260 *et al.* 2015; data from LACK *et al.* 2016; see Methods). To approximate our expectation under mutation-
261 drift equilibrium, we scaled average π from the SD^+ sample by 1.47% in 10-kb windows across the region

corresponding to the *SD-Mal* supergene, defined as the region from *Sd-RanGAP* to the distal breakpoint of *In(2R)Mal*. Table 2 (row 6) shows that diversity in the *SD-Mal* supergene region is still significantly lower than expected, suggesting that the low frequency of *SD-Mal* cannot fully explain its reduced diversity. This observation suggests two possibilities: the *SD-Mal* supergene historically had an equilibrium frequency less than 1.47% in Zambia; or the *SD-Mal* supergene, having reduced recombination, has experienced hitchhiking effects due to background selection and/or a recent selective sweep.

To distinguish between these possibilities, we analyzed summaries of the site frequency spectrum. We find strongly negative Tajima's *D* mirroring the distribution of reduced diversity, indicating an excess of rare alleles (Figure 2b). Such a skew in the site frequency spectrum suggests a recent increase in frequency of the *SD-Mal* supergene in Zambia. The high differentiation of *SD-Mal* from SD^+ chromosomes from the same population similarly suggests a large shift in allele frequencies. Wright's fixation index, F_{ST}, in the *SD-Mal* supergene region is unusually high for chromosomes from the same population (Figure 3a). Neither of the SD^+ chromosomes with cosmopolitan inversions show such high differentiation within their inversions, and mean nucleotide differences (d_{XY}) between *SD-Mal* and SD^+ is comparable to the other inversions, implying that the differentiation of the *SD-Mal* supergene is recent. Our results are thus consistent with a rapid increase in frequency of the *SD-Mal* haplotype that reduced nucleotide diversity within *SD-Mal* and generated large differences in allele frequencies with SD^+ chromosomes.

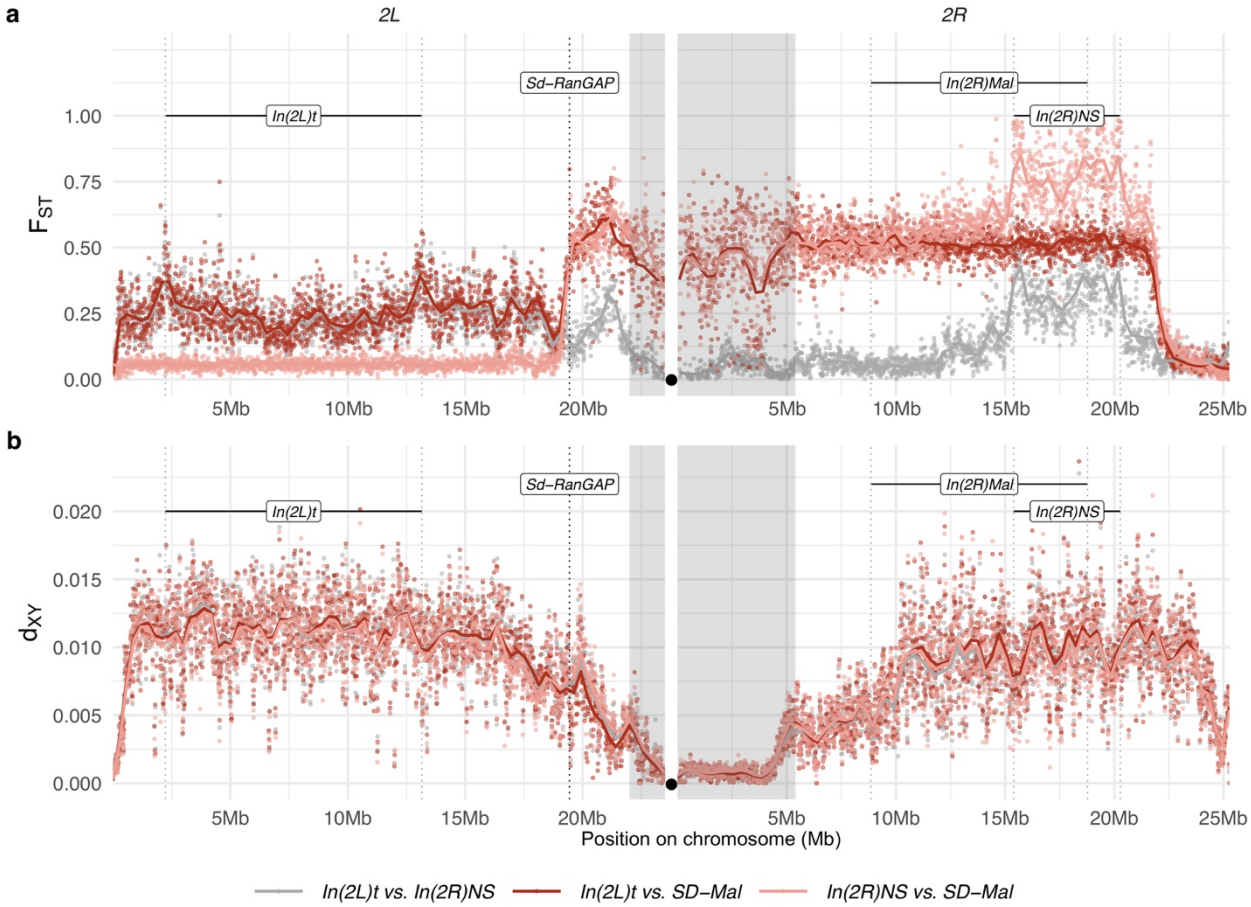

Figure 3. Differentiation between *SD-Mal* and wildtype chromosomes. (a) Pairwise F_{ST} and (b) d_{XY} per base pair in non-overlapping 10-kb windows along chromosome *2*, between Zambian *SD-Mal* haplotypes (n=9) and wildtype chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10) and *In(2R)NS* (n=10). Regions corresponding to pericentric heterochromatin are shaded in grey and the centromere location is marked with a black circle.

To estimate the timing of the recent expansion of the *SD-Mal* supergene, we used an Approximate Bayesian Computation (ABC) method with rejection sampling in neutral coalescent simulations. We do not know if *SD* chromosomes acquired *In(2R)Mal* in Zambia or if the inversions occurred *de novo* on an *SD* background. For our simulations, we assume that the acquisition of these inversion(s) was a unique event that enhanced drive strength and/or efficiency. Under this scenario, extant *SD-Mal* chromosomes have a single origin. We therefore simulated this history in a coalescent framework as an absolute bottleneck to a single chromosome. We performed simulations considering a sample size of $n = 9$ and assumed no recombination in the ~9.92-Mb region of *In(2R)Mal*. We simulated with values of S drawn from a uniform distribution ±5% of the observed number of segregating sites in *In(2R)Mal*. We considered a prior uniform

297 distribution of the time of the expansion (t) ranging from 0 to $4N_e$ generations (0 – 185,836 years ago),
298 assuming that *D. melanogaster* N_e in Zambia 3,160,475 (KAPOPOULOU *et al.* 2018), a *In(2R)Mal* frequency
299 of 1.47%, and 10 generations per year (LI AND STEPHAN 2006; THORNTON AND ANDOLFATTO 2006;
300 LAURENT *et al.* 2011; KAPOPOULOU *et al.* 2018). Using the ABC with rejection sampling conditional on
301 our observed estimates of π and Tajima's D for *In(2R)Mal* ($\pi_{In(2R)Mal}$ = 760.49, D = -1.27; note that $\pi_{In(2R)Mal}$
302 is an overall, unscaled estimate of nucleotide diversity for the whole *In(2R)Mal* region), we infer that the
303 *SD-Mal* expansion began ~0.096 (95% credibility intervals 0.092 - 0.115) x $4N_e$ generations ago or,
304 equivalently, ~1792 years ago (0.88% rejection sampling acceptance rate; Figure 4). To account for possible
305 gene conversion (see below), we discarded SNPs shared with SD^+ chromosomes (see below), and
306 recalculated π and Tajima's D using only private SNPs ($\pi_{In(2R)Mal}$ = 563.34, D = -1.41). Based on these
307 parameters, the estimated *SD-Mal* expansion occurred ~0.0737 (95% credibility intervals 0.070 - 0.092)
308 $4N_e$ generations ago, ~1368 years (0.86% rejection sampling acceptance rate Figure 4). To calculate the
309 posterior probability of the model, we performed 100,000 simulations under a model assuming a stable
310 frequency of *SD-Mal* and under sweep models (assuming t_{all} = 0.096 and t_{shared_excl}= 0.0737) (Suppl. Fig.
311 S4). The simulated data are inconsistent with a long-term stable frequency of *SD-Mal* (All SNPs, p_π =
312 0.0526, p_D = 0.1084; Private, p_π = 0.0285, p_D = 0.0755). Instead, our simulations suggest that a recent
313 selective sweep is more consistent with the data (All SNPs, p_π = 0.3106, p_D = 0.5929; Private, p_π = 0.3091,
314 p_D = 0.6092). Taken together, evidence from nucleotide diversity, the site frequency spectrum, population
315 differentiation, and coalescent simulations suggest a rapid non-neutral increase in frequency of the *SD-Mal*
316 supergene that began < 2,000 years ago.
317

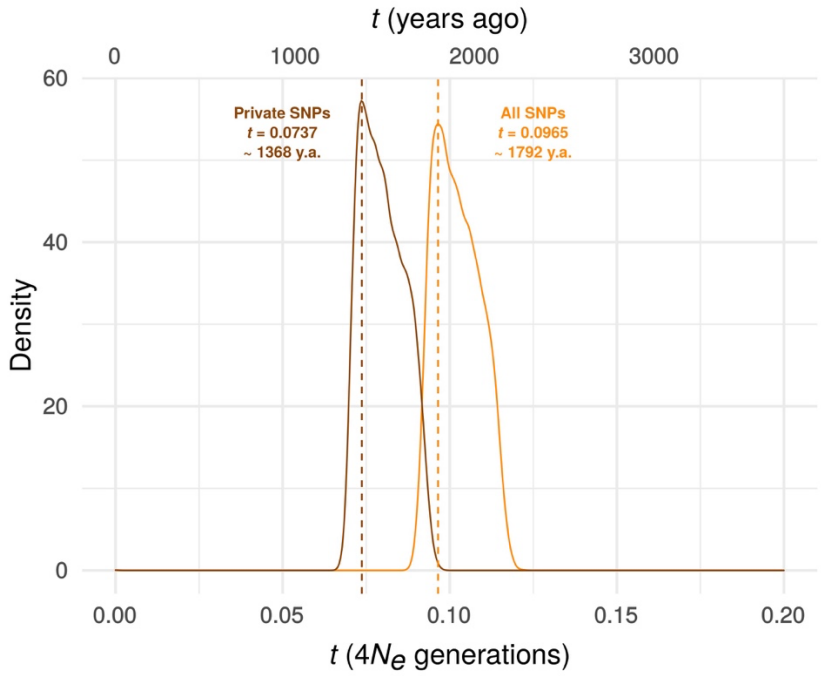

Figure 4. Estimating the time since the *SD-Mal* selective sweep. ABC estimates based on 10,000 posterior samples place the onset of the selective sweep between 0.096 (95% C.I. 0.092 - 0.115) and 0.0737 (0.070 - 0.092) x 4 N_e generations, i.e. ~1,368-1,792 years ago, considering recent estimates of N_e in Zambia from (KAPOPOULOU *et al.* 2018), frequency of *SD-Mal* in Zambia 1.47% and 10 generations per year). Estimates were done considering only *In(2R)Mal*, where crossing over is rare and only occurs between *SD-Mal* chromosomes using all SNPs and excluding shared SNPs in order to account for gene conversion from SD^+ chromosomes.

The sweep signal on the *SD-Mal* haplotypes begins immediately distal to *Sd-RanGAP* on *2L* and extends ~3 Mb beyond the distal boundary of *In(2R)Mal* on *2R*. To understand why the sweep extends so far beyond the *In(2R)Mal-d* distal breakpoint, we consider three, not mutually exclusive, possibilities. First, chromosomal inversions can suppress recombination ~1-3 Mb beyond their breakpoints (STEVISON *et al.* 2011; MILLER *et al.* 2016; CROWN *et al.* 2018; MILLER *et al.* 2018), extending the size of the sweep signal. To determine the extent of recombination suppression caused by *In(2R)Mal*, we estimated recombination rates in the region distal to the inversion. The expected genetic distance between the distal breakpoint of *In(2R)Mal* (*2R*: 18.77 Mb) and *px* (*2R*: 22.49 Mb) is ~13.87 cM (FISTON-LAVIER *et al.* 2010). Measuring recombination between *SD-Mal* and standard arrangement chromosomes for the same (collinear) interval, we estimate a genetic distance of ~1.76 (see above), an 87.3% reduction. *In(2R)Mal* strongly reduces recombination beyond its bounds. Second, although we have inferred that the essential enhancer resides within the *In(2R)Mal* inversion (see above), we have not excluded the possibility of weak enhancers distal to the inversion which might contribute to the sweep signal. We find that *SD-Mal* chromosomes with *In(2R)Mal*-distal material recombined away (b^+ *Sd* c^+ *In(2R)Mal px*) have modestly but significantly lower

drive strength (k = 0.96 *versus* 0.98; Table 1, lines 1-2), suggestive of one or more weak distal enhancers. Third, there may be mutations distal to *In(2R)Mal* that contribute to the fitness of *SD-Mal* haplotypes but without increasing the strength of drive, *e.g.*, mutations that compensate for the effects of *SD-Mal*-linked deleterious mutations.

Recombination on SD supergenes

While nearly all *SD-Mal* haplotypes are individually homozygous lethal and do not recombine with wild type chromosomes in and around *In(2R)Mal*, ~90% of pairwise combinations of different *SD-Mal* chromosomes (SD_i/SD_j) are viable and fertile in complementation tests (PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). Therefore, recombination may occur between *SD-Mal* chromosomes in SD_i/SD_j heterozygous females. To determine if *SD-Mal* chromosomes recombine, we estimated mean pairwise linkage disequilibrium (r^2) between SNPs located within the *In(2R)Mal* arrangement. We found that mean r^2 between pairs of SNPs declines as a function of the physical distance separating them (Figure 5a), a hallmark of recombination via crossing over (HILL AND ROBERTSON 1968; MIYASHITA AND LANGLEY 1988; SCHAEFFER AND MILLER 1993; AWADALLA *et al.* 1999; CONWAY *et al.* 1999). Pairwise LD is higher and extends further in *In(2R)Mal* than in the equivalent region of SD^+ chromosomes or in any of the other two cosmopolitan inversions, *In(2L)t* and *In(2R)NS* (Figure 5a). This pattern is not surprising: the low frequency of *SD-Mal* makes SD_i/SD_j genotypes, and hence the opportunity for recombination, rare. To further characterize the history of recombination between *SD-Mal* haplotypes, we used 338 non-singleton, biallelic SNPs in *In(2R)Mal* to trace historical crossover events. From these SNPs, we estimate that Rm (HUDSON AND KAPLAN 1985), the minimum number of recombination events, in this sample of *SD-Mal* haplotypes is 15 (Figure 5c). Thus, assuming that these *SD-Mal* haplotypes are ~17,929 generations old (Figure 4), we estimate that recombination events between *SD-Mal* haplotypes occur a minimum of once every ~1,195 generations. We can thus confirm that crossover events are relatively rare, likely due to the low population frequency of *SD-Mal* and the possibly reduced fitness of SD_i/SD_j genotypes.

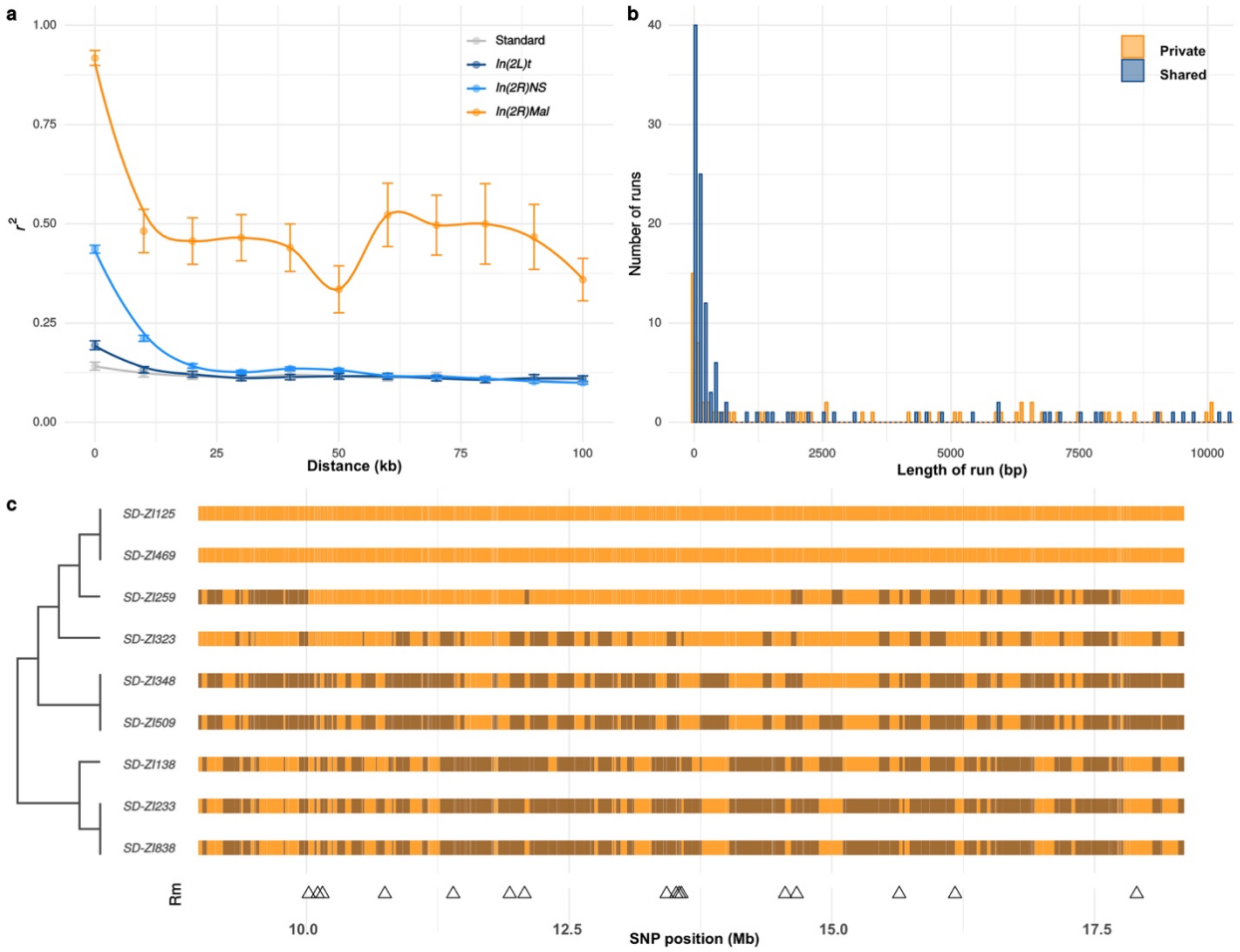

Figure 5. Recombination on *SD-Mal* haplotypes. (a) Linkage disequilibrium (r^2) as a function of distance in 10-kb windows, measured in *In(2R)Mal*, *In(2L)t*, *In(2R)NS*, and the corresponding region of *In(2R)Mal* in a standard, uninverted *2R* chromosome. (b) Histogram of length of runs of SNPs in *In(2R)Mal* shows that a high proportion of shared SNPs concentrate in runs shorter than 1 kb. (c) Chromosomal configuration of the 338 non-singleton SNPs in nine different *SD-Mal* lines. Color coded for two states (same in light orange or different in dark orange) using *SD-ZI125* as reference. Locations of minimal number of recombination events are labeled as triangles at the bottom. Maximum likelihood tree is displayed on the left.

While crossing over is suppressed in *SD-Mal/SD*$^+$ heterozygotes, gene conversion or double crossover events may still occur, accounting for the shared SNPs between *SD-Mal* and *SD*$^+$ chromosomes within *In(2R)Mal*. As both events exchange tracts of sequence, we expect shared SNPs to occur in runs of sites at higher densities than private SNPs, which should be distributed randomly. Accordingly, in *In(2R)Mal*, SNP density is five times higher for runs of shared SNPs (0.63 SNPs/kb) than for runs of *SD*-private SNPs (0.12 SNPs/kb), as expected if *SD*$^+$ chromosomes, which have higher SNP densities, were donors of conversion tract sequences. Additionally, 62.2% (89 out of 143) of the shared SNP runs are <1 kb (Figure 5b), consistent with current estimates of gene conversion tract lengths in *D. melanogaster* (COMERON *et al.*

2012). Surprisingly, these inferred gene conversion events are unevenly distributed across *In(2R)Mal*, being more frequent in the *In(2R)Mal-p* than in *In(2R)Mal-d* (Suppl. Table S2). Our discovery that *SD-Mal* haplotypes can recombine with each other distinguishes the *SD-Mal* supergene from other completely genetically isolated supergenes (WANG *et al.* 2013; CHARLESWORTH 2016; TUTTLE *et al.* 2016). The lack of crossing over with SD^+ chromosomes, however, means that *SD-Mal* haplotypes evolve as a semi-isolated subpopulation, with a nearly 100-fold smaller N_e and limited gene flow from SD^+ via gene conversion events. The reduced recombination, low N_e, and history of epistatic selection may nevertheless lead to a higher genetic load on *SD-Mal* than SD^+ chromosomes. We therefore examined the accumulation of deleterious mutations, including non-synonymous mutations and transposable elements, on the *SD-Mal* supergene.

Consequences of reduced recombination, small effective size, and epistatic selection

We first studied the effects of a reduced efficacy of selection on SNPs in *In(2R)Mal.* As many or most non-synonymous polymorphisms are slightly deleterious (OHTA 1976; FAY *et al.* 2001; EYRE-WALKER *et al.* 2002), relatively elevated ratios of non-synonymous to synonymous polymorphisms (*N/S* ratio) can indicate a reduced efficacy of negative selection. For the SNPs in *In(2R)Mal*, the overall *N/S* ratio is 2.3-fold higher than that for the same region of SD^+ chromosomes (Table 3). Notably, the *N/S* ratio for private SNPs is 3.1-fold higher (Table 3), whereas the N/S ratios for shared SNPs do not significantly differ from SD^+ chromosomes (Table 3, Suppl. Fig. S5). These findings suggest that gene conversion from SD^+ ameliorates the accumulation of potentially deleterious non-synonymous mutations on *SD-Mal* chromosomes.

Table 3. Synonymous and non-synonymous SNPs

	Genotype	***N***	***S***	***N/S***	**Fold change**	***p-value***
All SNPs	*SD-Mal*	79	114	0.69	2.27	<0.0001
	SD^+	10,470	34,301	0.31		
Private SNPs	*SD-Mal*	61	55	1.11	3.10	<0.0001
	SD^+	6,782	18,938	0.36		
Shared SNPs	*SD-Mal*	18	59	0.31	1.27	0.3722
	SD^+	3,688	15,363	0.24		

Counts of non-synonymous (*N*) and synonymous (*S*) SNPs in the *In(2R)Mal* region of *SD-Mal* chromosomes, and the equivalent region of uninverted, SD^+ chromosomes. *N/S* ratio per genotype, fold-change of *N/S* ratios between *SD-Mal* and SD^+. *P-values* reported by Pearson's χ^2 test of independence.

Gene conversion may not, however, rescue *SD-Mal* from deleterious transposable elements (TE) insertions, as average TE length exceeds the average gene conversion tract length (KAMINKER *et al.* 2002). TEs accumulate in regions of reduced recombination, such as centromeres (CHARLESWORTH *et al.* 1994) and inversions, especially those at low frequency (EANES *et al.* 1992; SNIEGOWSKI AND CHARLESWORTH 1994). Indeed, TE densities for the whole euchromatic region of chromosome *2R* are significantly higher for *SD-Mal* compared to SD^+ chromosomes (Figure 6a). This increased TE density on *SD-Mal* is driven by the non-recombining regions of the haplotype: *In(2R)Mal* has significantly higher TE density than SD^+ whereas the distal region of *2R* outside of the sweep region, does not (Figure 6b). The most overrepresented families in *In(2R)Mal* relative to standard chromosomes are *M4DM*, *MDG*1, *ROO_I,* and *LINE* elements (Suppl. Fig.S6)— TEs that are currently or recently active (KAMINKER *et al.* 2002; KOFLER *et al.* 2015; DÍAZ-GONZÁLEZ AND DOMÍNGUEZ 2020)— consistent with the recent origin of the *SD-Mal* haplotype. Thus, the differences in shared *versus* private SNPs suggests that gene conversion from SD^+ chromosomes may provide a mechanism to purge deleterious point mutations but not TEs. Despite occasional recombination, the small N_e of *SD-Mal* haplotypes has incurred a higher genetic load.

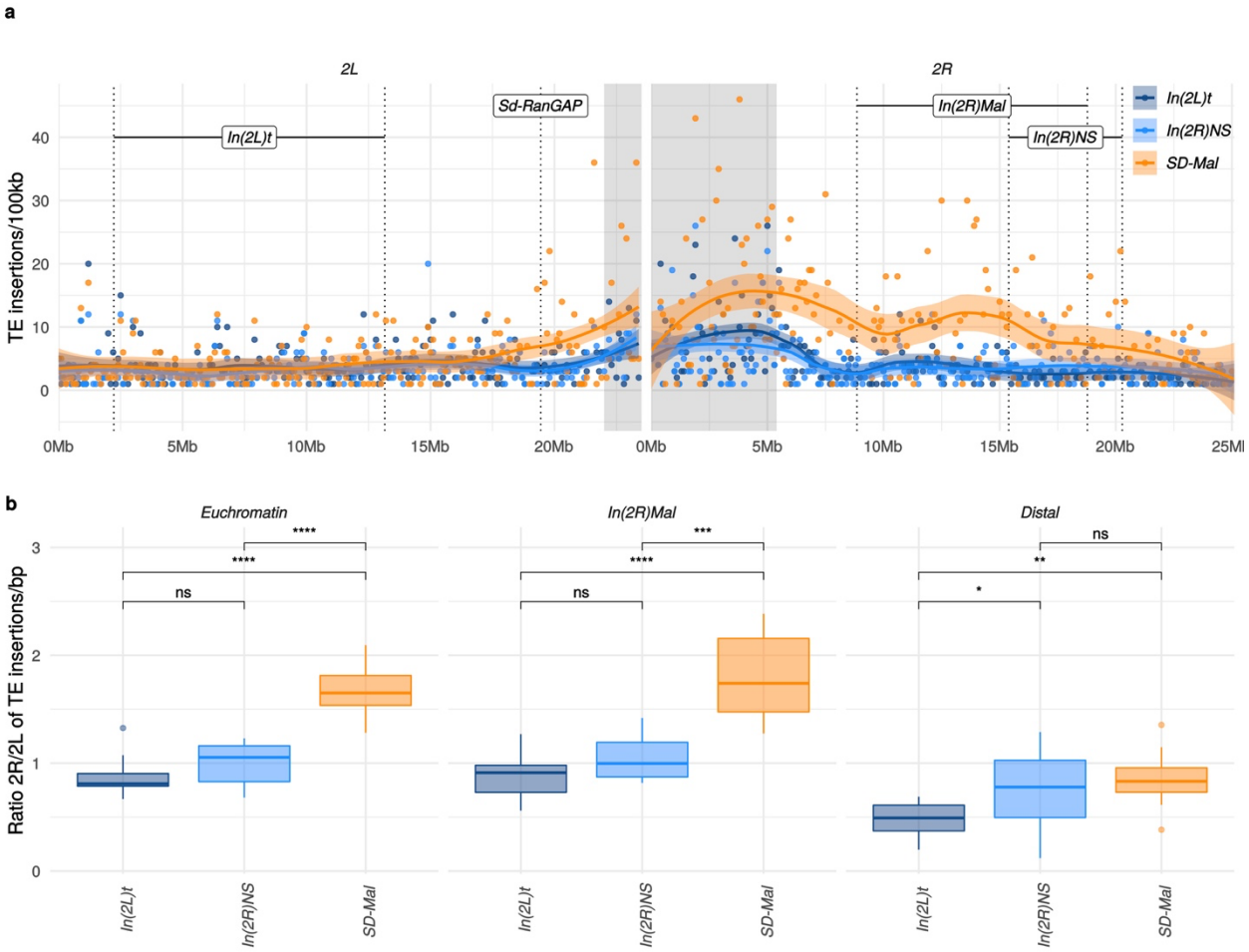

Figure 6. Transposable elements on *SD-Mal* haplotypes. (a) Number of *TE* insertions per 100-kb windows along chromosome 2, in Zambian *SD* chromosomes (n=9, orange) and wildtype chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10, dark blue) and *In(2R)NS* (n=10, light blue). (b) Ratio of the number of insertions in the euchromatin of *2R* to *2L* per library. The relative enrichment in *TE*s in *2R* of *SD-Mal* haplotypes is mostly due to an increase of TE insertions in non-recombining regions of the chromosome.

Conclusions

Supergenes are balanced, multigenic polymorphisms in which epistatic selection among component loci favors the recruitment of recombination modifiers that reinforce the linkage of beneficial allelic combinations. The advantages of reduced recombination among strongly selected loci can however compromise the efficacy of selection at linked sites. Supergenes thus provide opportunities to study the interaction of recombination and natural selection. We have studied a population of *selfish* supergenes, the *SD-Mal* haplotypes of Zambia, to investigate the interplay of recombination, selection, and meiotic drive. Our findings demonstrate, first, that the *SD-Mal* supergene extends across ~25.8 Mb of *D. melanogaster*

chromosome *2*, a region that comprises the driving *Sd-RanGAP*, a drive-insensitive deletion at the major *Rsp* locus, and the *In(2R)Mal* double inversion. Second, using genetic manipulation, we show that *SD-Mal* requires *Sd-RanGAP* and an essential co-driver that localizes almost certainly within the *In(2R)Mal* rearrangement, and probably within the distal inversion. These data provide experimental evidence for epistasis between *Sd-RanGAP* and *In(2R)Mal*: neither allele can drive without the other. Third, we provide population genomics evidence that epistatic selection on loci spanning the *SD-Mal* supergene region drove a very recent, chromosome-scale selective sweep. These patterns are consistent with recurrent episodes of replacement of one *SD* haplotype by others (PRESGRAVES *et al.* 2009; BRAND *et al.* 2015). Fourth, despite rare crossovers among complementing *SD-Mal* haplotypes and gene conversion from wildtype chromosomes, the relative genetic isolation and low frequency of *SD-Mal* results in the accumulation of deleterious mutations including, especially, TE insertions. From these findings, we conclude that the *SD-Mal* supergene population is of small effective size, semi-isolated by from the greater population of wildtype chromosomes, and subject to bouts of very strong selection.

Non-recombining supergenes that exist exclusively in heterozygous state tend to degenerate, as in the case of Y chromosomes (reviewed in CHARLESWORTH AND CHARLESWORTH 2000) and some autosomal supergenes which, for different reasons, lack any opportunity for recombination (UYENOYAMA 2005; WANG *et al.* 2013; TUTTLE *et al.* 2016; BRANCO *et al.* 2018; STOLLE *et al.* 2019 ; BRELSFORD *et al.* 2020). But not all supergenes are necessarily expected to degenerate. In *SD-Mal*, for instance, complementing *SD-Mal* haplotypes can recombine via crossing over, if rarely, and gene flow from wildtype SD^{+} to *SD-Mal* chromosomes can occur via gene conversion. In the mouse *t*-haplotype, there is similar evidence for occasional recombination between complementing *t*-haplotypes (DOD *et al.* 2003) and with standard chromosomes, probably via gene conversion (HERRMANN *et al.* 1987; ERHART *et al.* 2002; WALLACE AND ERHART 2008; KELEMEN AND VICOSO 2018). Despite the many parallels characterizing supergenes, their ultimate evolutionary fates depend on the particulars of the system.

Material and methods

Fly lines, library construction and sequencing

We sequenced haploid embryos using the scheme in Langely et al (LANGLEY *et al.* 2011), which takes advantage of a mutation, *ms(3)K81* (FUYAMA 1984), which causes the loss of the paternal genome early in

472 embryonic development. We crossed *SD-Mal/CyO* stocks generated in (BRAND *et al.* 2015) to homozygous
473 *ms(3)K81* males and allowed them to lay eggs overnight. We inspected individual embryos under a
474 dissecting scope for evidence of development and then isolated them for whole genome amplification using
475 the REPLI-g Midi kit from Qiagen (catalog number 150043). For each WGA DNA sample, we tested for
476 the presence of *Sd-RanGAP* using PCR (primers from (PRESGRAVES *et al.* 2009). We prepared sequencing
477 libraries for Illumina sequencing with TruSeq PCR free 350bp. We assessed library quality using a
478 BioAnalyzer and sequenced with HiSeq2500 2x150bp reads (TruSeq) or 2x125bp reads (Nextera). To trim
479 reads, we used Trimgalore v0.3.7 and the parameters: *q 28 --length 20 --paired -a*
480 *GATCGGAAGAGCACACGTCTGAACTCCAGTCAC -a2*
481 *GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT --phred33 --fastqc --retain_unpaired -r1 21 -r2 21 --*
482 *dont_gzip --length 20.* Trimmed reads are available in SRA (Bioproject PRJNA649752, SRA accession
483 numbers in Table S1)
484
485 We sequenced a total of 10 *SD-Mal* genomes. One of these genomes (*SD-ZI157*) was *Sd-In(2R)Mal*$^{+}$, non-
486 driving, and therefore excluded from further analysis. Out of the remaining 9 driving *SD-Mal* chromosomes,
487 one of them (*SD-ZI138*) had lower depth than the other 8 (Table S1, sheet 2) in the main chromosome arms
488 but unusually high depth in the mitochondrial genome. We ran additional analyses dropping *SD-ZI138* and
489 show that including this sample does not affect our main conclusions (Supp. Table S5; sheet 2).
490
491 For the Nanopore library, we extracted High-Molecular-Weight DNA from ~200 frozen female *SD-*
492 *ZI125/SD-ZI125* virgins. We extracted DNA using a standard phenol-chloroform method and spooled DNA
493 using capillary tubes. We constructed a library with ~1 ug DNA using RAD004 kit and the ultra-long read
494 sequencing protocol (QUICK). We sequenced the library using R9.4 flow cells and called bases with the
495 ONT Albacore Sequencing Pipeline Software version v2.2.10.

496

497 *Estimating Rsp copy number*

498 We mapped Zambian *SD* reads to an assembly containing *2R* pericentric heterochromatin (CHANG AND
499 LARRACUENTE 2019), including the *Rsp* locus detailed in Khost et al. (KHOST *et al.* 2017) with bowtie2
500 v2.3.5 (LANGMEAD AND SALZBERG 2012). We estimated mean per-window and per-*Rsp* repeat depth using
501 mosdepth v0.2.9 (PEDERSEN AND QUINLAN 2018). Coordinates for *Rsp* repeats were based on annotations
502 in Khost et al. (2017).

503

In(2R)Mal breakpoints

To assemble *SD-ZI125*, we filtered Nanopore reads using Porechop (v0.2.3) and Filtlong (--min_length 500) to remove adapters and short reads (https://github.com/rrwick/Porechop; (WICK *et al.* 2017) and https://github.com/rrwick/Filtlong). We were left with a total of 1,766,164,534 bases in 327,248 filtered reads. We performed *de novo* assemblies with the Nanopore reads using Flye v2.3.7 (KOLMOGOROV *et al.* 2019) with parameters "-t 24 -g 160m --nano-raw" and wtdbg v2.2 (RUAN AND LI 2020) with parameters -p 19 -AS 1 -s 0.05 -L 0 -e 1". We independently polished these two assemblies 10 times with Pilon v1.22 (WALKER *et al.* 2014) using paired-end Illumina reads. Lastly, we reconciled these two polished assemblies using quickmerge v0.3 (CHAKRABORTY *et al.* 2016) using the flye assembly as the reference with the command "python merge_wrapper.py wtdbg assembly flye assembly". We aligned the contig containing the euchromatin on *SD-ZI125* to chromosome *2R* of the *D. melanogaster* (BDGP6) genome using Mauve (DARLING *et al.* 2010). We defined the breakpoints according to the block rearrangement shown in Figure 1. To validate these breakpoints, we designed primers anchored at both sides of the most external breakpoints of *In(2R)Mal* (Suppl.Table S3) for PCR.

Measuring genetic distances along SD-Mal and strength of distortion in the recombinants

To estimate recombination frequencies and obtain *SD* recombinant genotypes, we used a stock *(al[1] dpy[ov1] b[1] pr[1] c[1] px[1] speck[1]*, BDSC156, Bloomington Drosophila Stock Center), which has three visible, recessive markers on chromosome 2: *black* (*b*, 2L: 13.82), *curved* (*c*, 2R:15.9) and *plexus* (*px*, 2R:22.49). All crosses were transferred to a fresh vial after 5 days, and then adults were removed from the second vial after 5 days. Progeny emerging from the crosses were scored for up to 20 days following the cross.

To generate *SD-Mal* recombinant chromosomes, we crossed 8-10 *b,c,px*/*b,c,px* virgin females to 3-5 *SD-ZI125* males, recovered *SD-ZI125*/*b,c,px* virgins, then backcrossed 8-10 of them to 3-5 *b,c,px* homozygous males. To estimate genetic distance between the visible markers, we scored the number of recombinants in 11 crosses (n=1820). To compare genetic distance in *SD-Mal* to wild-type chromosomes, we estimated the number of recombinants from 15 crosses between *OregonR*/*b,c,px* females to *b,c,px*/*b,c,px* males (n=1716).

We recovered three types of recombinant chromosomes from *SD-ZI125*/*b,c,px* x *b,c,px*/*b,c,px* crosses: *b,Sd*$^{+}$*,c*$^{+}$*,In(2R)Mal,px*$^{+}$; *b*$^{+}$*,Sd,c,In(2R)Mal*$^{+}$*,px* and *b*$^{+}$*,Sd,c*$^{+}$*,In(2R)Mal,px*. We crossed 3-5 virgin *b,c,px/b,c,px* females to individual recombinant males of each type, and scored the proportion of progeny carrying the recombinant chromosome (k=$n_{recombinant}/n_{total}$). To distinguish distortion from viability effects, we also measured transmission of recombinant chromosomes through females, as drive is male-specific.

We used these crosses to measure relative viability (w=$n_{recombinant}$/n_{bcpx}). We then used w to calculate a viability-corrected strength of distortion in males (k*=$n_{recombinant}$/($w n_{bcpx}$+$n_{recombinant}$) (POWERS AND GANETZKY 1991).

Estimate of the frequency of SD-Mal in the DPGP3 dataset

To estimate the frequency of *In(2R)Mal* in a random sample of Zambian chromosomes, we mapped the 204 Illumina paired-end libraries from the DPGP3 dataset (LACK *et al.* 2016) to the *D. melanogaster* (BDGP6) genome, using bwa-mem (v0.7.9a), and we visually looked for an accumulation of discordant read pairs surrounding the estimated breakpoints of *In(2R)Mal*. To test the reliability of this method, we also applied it to detect cosmopolitan inversions *In(2L)t* and *In(2R)NS* and compared our inversion calls with the most recent inversion calls for the DPGP3 dataset (http://johnpool.net/Updated_Inversions.xls, last accessed 07/13/2020), getting a 98% and 99% of concordance for *In(2L)t* and *In(2R)NS,* respectively. To determine the frequency of the *Sd-RanGAP* duplication in the DPGP3 dataset we applied a similar method around the breakpoints of the *Sd-RanGAP* duplication (see Suppl.Table S4).

SNP calling and annotation

For SNP calling, we mapped the Illumina reads from our *SD-Mal* libraries and the 20 SD^+ libraries from the *DPGP3* dataset to *D. melanogaster* (BDGP6) genome (ftp://ftp.ensembl.org/pub/release-88/fasta/drosophila_melanogaster/dna/; last accessed 6/25/20) using BWA mem (v0.7.9a). We removed duplicated reads with Picard (2.0.1) and applied the GATK (3.5) "best practices" pipeline for SNP calling. We did local realignment and base score recalibration using SNPs data from DPGP1 ensembl release 88 (ftp://ftp.ensembl.org/pub/release-88/variation/vcf/drosophila_melanogaster/; last accessed 6/25/20). To call SNPs and indels, we used HaplotypeCaller and performed joint genotyping for each of the five genotypes using GenotypeGVCFs. SNPs filtered with following parameters: 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0'. We annotated SNPs as synonymous or nonsynonymous using SNPeff (4.3, CINGOLANI *et al.* 2012) with the integrated *D. melanogaster* database (dmel_r6.12) database and parsed these annotations with SNPsift (CINGOLANI *et al.* 2012). To classify the SNPs as 'shared' between *SD-Mal*, SD^+*In(2L)t* and SD^+*In(2R)NS*, or 'private' to each one of them, we used BCFtools intersect (1.6; DANECEK *et al.* 2021).

Population genomics analysis

We wrote a Perl script to estimate S, π, Tajima's D, F_{ST} and d_{XY} in windows across the genome (available here: https://github.com/bnavarrodominguez/sd_popgen). To calculate F_{ST} values, we used the Weir-Cockerham estimator (WEIR AND COCKERHAM 1984). Only those sites with a minimum sample depth of 8 were included in the F_{ST} and Tajima's D calculations. We determined window size by the number of 'acceptable sample depth' sites (and not, for example, a particular range of chromosome coordinates). To confirm that repeats were not interfering with our results, we ran our population genomics pipeline masking SNPs in repetitive elements identified by RepeatMasker (SMIT *et al.* 2013), which yielded equivalent results (Suppl. Table S5, sheet 1).

Age of the sweep

We calculated overall $S_{In(2R)Mal}$, $\pi_{In(2R)Mal}$ and Tajima's $D_{In(2R)Mal}$ from the *SD-Mal* SNP set using our same Perl script (available here: https://github.com/bnavarrodominguez/sd_popgen), using a single window of 9.5Mb within the boundaries of *In(2R)Mal*. To account for gene conversion, we calculated an additional set of summary statistics masking the SNPs annotated as shared by at least one of the SD^+ libraries. We estimated the time since the most recent selective sweep using an ABC method with rejection sampling. We modeled the selective sweep as an absolute bottleneck (N_t=1) at some time (t, $4N_e$ generations) in the past. We performed simulations in *ms* (HUDSON 2002), considering a sample size of 9 and assuming no recombination in the ~9.92 Mb of *In(2R)Mal*. We simulated with values of S_{Sim} drawn from a uniform distribution ±5% of $S_{In(2R)Mal}$. We considered a prior uniform distribution of time of the sweep (t) ranging from 0 to $4N_e$ generations, *i.e.*, 0 – 185,836 years ago, considering *D. melanogaster* N_e in Zambia 3,160,475 (KAPOPOULOU *et al.* 2018), frequency of *In(2R)Mal* 1.47% and 10 generations per year. The rejection sampling algorithm is as follows: (1) draw S_{Sim} and t from prior distributions; (2) simulate 1000 samples using the coalescent under a selective sweep model; (3) calculate average summary statistics for drawn S_{Sim} and t; (4) accept or reject chosen parameter values conditional on $|\pi_{In(2R)Mal} - \pi_{Sim}| \leq \varepsilon$, $|D_{In(2R)Mal} - D_{Sim}| \leq \varepsilon$; (5) return to step 1 and continue simulations until m desired samples from the joint posterior probability distribution are collected. For estimates of t, ε was set to 5% of the observed values of the summary statistics (in step 4) and m was set to 10,000. These simulations were performed with parameters calculated using all the SNPs in *In(2R)Mal* and excluding SNPs shared with SD^+ chromosomes to account for gene conversion. We simulated 100,000 samples with the resulting estimated t and $S_{In(2R)Mal}$, under our sweep model and under a constant size population model, and calculated two-sided p values for π and Tajima's D using an empirical cumulative probability function (ecdf) in R (TEAM 2019). We estimated the maximum *a posteriori* estimate (MAP) as the posterior mode and 95% credibility intervals (CI) in R (TEAM 2019).

599

600 *Recombination*

601 For estimates of recombination, we filtered the SNPs in *In(2R)Mal* to variable positions genotyped in all of
602 the 9 *ZI-SD* samples and excluded singletons, resulting in a total of 338 SNPs. We estimated pairwise
603 linkage disequilibrium (r^2) using PLINK v1.9 (PURCELL *et al.* 2007). We discarded r^2 data calculated for
604 pairs of SNPs flanking the internal *In(2R)Mal* breakpoints. For comparison, we estimated pairwise linkage
605 disequilibrium in the same region of *In(2R)Mal* for SD^+ uninverted *2R* chromosome arms and, for
606 comparison, in SD^+ *In(2R)NS* inversion and in SD^+ *In(2L)t* inversions. For SD^+ chromosomes we applied
607 the same filters (variable, non-singleton SNPs), plus a SNP 'thinning' to 1 SNPs/kb to get a manageable
608 set of results. To investigate the possibility of crossing over between *SD-Mal* chromosomes, we used
609 RecMin (MYERS AND GRIFFITHS 2003) to estimate the minimum number of crossovers between the 338
610 bi-allelic, non-singleton SNPs in *In(2R)Mal*. RecMin input is a binary file, which we generated using *SD-*
611 *ZI125* as an arbitrary reference for *SD*, assigning 0 or 1 on each position depending on if it was the same
612 base or different. Maximum likelihood trees to establish relationships between *SD-Mal* haplotypes based
613 on these 338 SNPs were estimated using RAxML-NG (KOZLOV *et al.* 2019).

614 Runs of shared and private SNPs were identified in R, using all SNPs (including singletons). A run of SNPs
615 is defined as a region from 5' to 3' where all the SNPs are in the same category (shared or private). Distance
616 between the first and the last SNP of a category is considered length of the run. The region between the last
617 SNP of a category and the first SNP of the alternative category is considered distance between runs. Because
618 our sample size is small, we may underestimate the number of shared SNPs, as some private SNPs may be
619 shared with some SD^+ chromosomes that we have not sampled.

620

621 *Transposable element calling*

622 We used a transposable element (TE) library containing consensus sequences of *Drosophila* TE families
623 (CHANG AND LARRACUENTE 2019). With this library, we ran RepeatMasker (SMIT *et al.* 2013) to annotate
624 reference TEs in the *D. melanogaster* (BDGP6) genome. To detect genotype specific TE insertions in our
625 Illumina libraries, we used the McClintock pipeline (NELSON *et al.* 2017), which runs six different programs
626 with different strategies for TE calling. We collected the redundant outputs from RetroSeq (KEANE *et al.*
627 2013), PoPoolationTE (KOFLER *et al.* 2012), ngs_te_mapper (LINHEIRO AND BERGMAN 2012), TE-Locate
628 (PLATZER *et al.* 2012) and TEMP (ZHUANG *et al.* 2014), discarded the calls produced by TEMP based on
629 non-evidence of absence, and then merged the insertions detected by all different programs, considering
630 the same insertion those of the same TE closer than a distance of +/- 600 bp, as described in (BAST *et al.*

2019). To reduce false positives, we only considered TE insertion calls that were predicted by more than one of the methods. To account for differences in library read number and/or length between datasets, we report the TE counts for *2R* normalized by the TE count for chromosome *2L* for the same library (Figure 6b). To assess whether library differences qualitatively affect our results, we repeated the above TE analysis on a set of 3 million randomly selected paired-end reads, trimmed to a fixed length of 75 bp, from each library and report TE count for chromosomes *2R* and *2L* separately (Suppl. Fig. S7).

Acknowledgements

This work was funded by the National Institutes of Health (NIH) National Institute of General Medical Sciences (R35GM119515 and NIH-NRSA F32GM105317to A.M.L.), Stephen Biggar and Elisabeth Asaro fellowship in Data Science to A.M.L, a David and Lucile Packard Foundation grant and University of Rochester funds to D.C.P. We thank Dr. Danna Eickbush for assistance with genomic DNA preparation for nanopore sequencing. We also thank the University of Rochester CIRC for access to computing cluster resources and UR Genomics Research Center for the library construction and sequencing.

Data availability

Raw sequence data are deposited in NCBI's short read archive under project accession PRJNA649752. All code for data analysis and figure generation is available in Github (https://github.com/bnavarrodominguez/sd_popgen).

Figure Legends

Figure 1. Map depicting the chromosomal features of the *SD-Mal* chromosome. The schematic shows the cytogenetic map of chromosomes *2L* and *2R* (redrawn based on images in LEFEVRE 1976) and the major features of the chromosome. (a) Dotplot showing that the *Sd* locus is a partial duplication of the gene *RanGAP* (in black), located at band *37D2-6*. (b) The *Rsp-major* locus is an array of tandem repeats located in the pericentric heterochromatin (band *h39*). Read mapping shows that *SD-Mal* chromosomes do not have any *Rsp* repeats in the *Rsp-major* locus, consistent with being insensitive to distortion by *Sd* (Rsp^{i}) (orange, high relative coverage regions correspond to transposable element interspersed), in contrast with *Iso-1*, which is sensitive (Rsp^{s}). (c) Two paracentric, overlapping inversions constitute the *In(2R)Mal*

arrangement: *In(2R)51BC;55E* (*In(2R)Mal-p*) in orange brackets and *In(2R)44F;54E* (*In(2R)Mal-d*) in red parentheses). Pericentromeric heterochromatin and the centromere are represented by a grey rectangle and black circle, respectively. Our assembly based on long-read sequencing data provide the exact breakpoints of *In(2R)Mal* and confirms that the distal inversion (*Dmel.r6*, *2R*:14,591,034-18,774,475) occurred first, and the proximal inversion (*Dmel.r6*, *2R*:8,855,601-15,616,195) followed, overlapping ~1Mb with the distal inversion. The colored rectangles correspond to locally collinear blocks of sequence. Blocks below the center black line indicate regions that align in the reverse complement orientation. Vertical red lines indicate the end of the assembled chromosomes. Visible marker locations used for generating recombinants (*b* (34D1), *c* (52D1), and *px* (58E4-58E8)) are indicated on the cytogenetic map.

Figure 2. Diversity on *SD-Mal* chromosomes. (a) Average pairwise nucleotide diversity per site (π) and (b) Tajima's *D* estimates in non-overlapping 10-kb windows along chromosome *2*, in Zambian *SD-Mal* chromosomes (n=9, orange) and SD^+ chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10, dark blue) and *In(2R)NS* (n=10, light blue). Regions corresponding to pericentric heterochromatin are shaded in grey and the centromere location is marked with a black circle. *SD-Mal* chromosomes show a sharp decrease in nucleotide diversity and skewed frequency spectrum from the *Sd* locus (*Sd-RanGAP*, *2L*:19.4Mb) to ~2.9 Mb beyond the distal breakpoint of *In(2R)Mal*.

Figure 3. Differentiation between *SD-Mal* and wildtype chromosomes. (a) Pairwise F_{ST} and (b) d_{XY} per base pair in non-overlapping 10-kb windows along chromosome *2*, between Zambian *SD-Mal* haplotypes (n=9) and wildtype chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10) and *In(2R)NS* (n=10). Regions corresponding to pericentric heterochromatin are shaded in grey and the centromere location is marked with a black circle.

Figure 4. Estimating the time since the *SD-Mal* selective sweep. ABC estimates based on 10,000 posterior samples place the onset of the selective sweep between 0.096 (95% C.I. 0.092 - 0.115) and 0.0737 (0.070 - 0.092) x 4 N_e generations, i.e. ~1,368-1,792 years ago, considering recent estimates of N_e in Zambia from (KAPOPOULOU *et al.* 2018), frequency of *SD-Mal* in Zambia 1.47% and 10 generations per year). Estimates were done considering only *In(2R)Mal*, where crossing over is rare and only occurs between *SD-Mal* chromosomes using all SNPs and excluding shared SNPs in order to account for gene conversion from SD^+ chromosomes.

Figure 5. Recombination on *SD-Mal* haplotypes. (a) Linkage disequilibrium (r^2) as a function of distance in 10-kb windows, measured in *In(2R)Mal*, *In(2L)t*, *In(2R)NS*, and the corresponding region of *In(2R)Mal* in a standard, uninverted *2R* chromosome. (b) Histogram of length of runs of SNPs in *In(2R)Mal* shows that a high proportion of shared SNPs concentrate in runs shorter than 1 kb. (c) Chromosomal configuration of the 338 non-singleton SNPs in nine different *SD-Mal* lines. Color coded for two states (same in light orange or different in dark orange) using *SD-ZI125* as reference. Locations of minimal number of recombination events are labeled as triangles at the bottom. Maximum likelihood tree is displayed on the left.

Figure 6. Transposable elements on *SD-Mal* haplotypes. (a) Number of *TE* insertions per 100-kb windows along chromosome 2, in Zambian *SD* chromosomes (n=9, orange) and wildtype chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10, dark blue) and *In(2R)NS* (n=10, light blue). (b) Ratio of the number of insertions in the euchromatin of *2R* to *2L* per library. The relative enrichment in *TE*s in *2R* of *SD-Mal* haplotypes is mostly due to an increase of TE insertions in non-recombining regions of the chromosome.

Supplemental figure legends

Suppl. Fig. S1. Estimated abundance of *Rsp* repeats at each *Rsp* locus in the reference *Iso-1* genome and *SD-Mal.* For each locus annotated in the reference *D. melanogaster* genome (KHOST *et al.* 2017), we plot estimated *Rsp* abundance as the sum of average depth of repeats at each locus normalized by average depth of chromosome 2 on the y-axis. *SD-Mal* has very few reads mapping to the primary *Rsp* locus (*Rsp-proximal* and *Rsp-major*), suggesting a complete deletion of the target of drive.

Suppl. Fig. S2. Model of the *In(2R)Mal* rearrangement. (a) Wild-type arrangement of chromosome *2R*. Pericentromeric heterochromatin and the centromere are represented by a grey rectangle and black circle, respectively. (b) *In(2R)Mal-d*: inversion of 4.18 Mb of 2R (*2R*:14,591,003-18,774,475), which disrupted the 3' UTR of the *Mctp* gene (*2R*:18,761,758 - 18,774,824). (c) *In(2R)Mal-p*: Inversion of 6.76 Mb of 2R (*2R*:8,855,602- 17,749,310), with 1.02 Mb overlapping with the now proximal segment of *In(2R)Mal-d*. This inversion disrupted the 3' UTR of the *sns* gene (*2R*:8,798,489 - 8,856,091) and the CDS of the *CG10931* gene (*2R*:17,748,935 -17,750,136).

Suppl. Fig. S3. Average pairwise nucleotide diversity per site (π) in non-overlapping, 1-kb windows, for SD^+ and *SD-Mal* chromosomes, around the *Sd-RanGAP* locus.

Suppl. Fig. S4. Neutral coalescent simulations under a constant size population model and a sweep and expansion model (absolute bottleneck at a time *t*, between 0 and $4N_e$ generations). Simulations were done with *S* estimated using all SNPs in *In(2R)Mal* (a) and excluding SNPs shared with SD^+ chromosomes (b) to account for gene conversion. Blue horizontal line marks observed $\pi_{In2RMal}$ (estimated for the entire region) and Tajima's *D* in *In(2R)Mal* in Zambian *SD* chromosomes.

Suppl. Fig. S5. Frequency spectra of synonymous and non-synonymous SNPs in the *In(2R)Mal* chromosome region, in Zambian *SD* chromosomes (n=9, orange) and wildtype chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (n=10, dark blue) and *In(2R)NS* (n=10, light blue). *N/S* ratio for each of the frequency categories.

Suppl. Fig. S6. Number of insertions per TE family in *SD-Mal* compared to uninverted SD^+ chromosome *2R*, both in *In(2R)Mal (2R:8.85-18.77,* top panel) and the region distal to it *(2R:18.77-25.29,* bottom panel). The families *DNA/M4DM*, *LTR/MDG1, LTR/ROO_I* and *Non-LTR/LINE* are highly overrepresented in *In(2R)Mal*.

Suppl. Fig. S7 Abundance of TEs in down sampled (3M reads, 75-bp) libraries for Zambian *SD* chromosomes (n=9, orange) and SD^+ chromosomes from the same population, bearing the cosmopolitan inversions *In(2L)t* (N=10, dark blue) and *In(2R)NS* (N=10, light blue).

References

Aulard, S., J. R. David and F. Lemeunier, 2002 Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. Genetics Research 79**:** 49-63.

Awadalla, P., A. Eyre-Walker and J. M. Smith, 1999 Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286**:** 2524-2525.

753 Bast, J., K. S. Jaron, D. Schuseil, D. Roze and T. Schwander, 2019 Asexual reproduction reduces
754 transposable element load in experimental yeast populations. Elife 8.
755 Branco, S., F. Carpentier, R. C. R. de la Vega, H. Badouin, A. Snirc *et al.*, 2018 Multiple convergent
756 supergene evolution events in mating-type chromosomes. Nature communications 9**:** 1-13.
757 Brand, C. L., A. M. Larracuente and D. C. Presgraves, 2015 Origin, evolution, and population genetics of
758 the selfish *Segregation Distorter* gene duplication in European and African populations of
759 *Drosophila melanogaster*. Evolution 69**:** 1271-1283.
760 Brelsford, A., J. Purcell, A. Avril, P. T. Van, J. Zhang *et al.*, 2020 An ancient and eroded social supergene
761 is widespread across *Formica* ants. Current Biology.
762 Brittnacher, J. G., and B. Ganetzky, 1984 On the components of segregation distortion in *Drosophila*
763 *melanogaster*. III. Nature of enhancer of *SD*. Genetics 107**:** 423-434.
764 Chakraborty, M., J. G. Baldwin-Brown, A. D. Long and J. J. Emerson, 2016 Contiguous and accurate de
765 novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 44**:** e147.
766 Chang, C. H., and A. M. Larracuente, 2019 Heterochromatin-Enriched Assemblies Reveal the Sequence
767 and Organization of the *Drosophila melanogaster* Y Chromosome. Genetics 211**:** 333-348.
768 Charlesworth, B., and D. Charlesworth, 2000 The degeneration of Y chromosomes. Philos Trans R Soc
769 Lond B Biol Sci 355**:** 1563-1572.
770 Charlesworth, B., and D. L. Hartl, 1978 Population Dynamics of the *Segregation Distorter* Polymorphism
771 of *Drosophila melanogaster*. Genetics 89**:** 171-192.
772 Charlesworth, B., P. Sniegowski and W. Stephan, 1994 The evolutionary dynamics of repetitive DNA in
773 eukaryotes. Nature 371**:** 215-220.
774 Charlesworth, D., 2016 The status of supergenes in the 21st century: recombination suppression in Batesian
775 mimicry and sex chromosomes and other complex adaptations. Evolutionary Applications 9**:** 74-
776 90.
777 Charlesworth, D., and B. Charlesworth, 1975 Theoretical genetics of Batesian mimicry II. Evolution of
778 supergenes. Journal of Theoretical Biology 55**:** 305-324.
779 Chintapalli, V. R., J. Wang and J. A. Dow, 2007 Using FlyAtlas to identify better *Drosophila melanogaster*
780 models of human disease. Nat Genet 39**:** 715-720.
781 Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and
782 predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
783 *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6**:** 80-92.
784 Comeron, J. M., R. Ratnappan and S. Bailin, 2012 The many landscapes of recombination in *Drosophila*
785 *melanogaster*. PLoS genetics 8.
786 Conway, D. J., C. Roper, A. M. Oduola, D. E. Arnot, P. G. Kremsner *et al.*, 1999 High recombination rate
787 in natural populations of *Plasmodium falciparum*. Proceedings of the National Academy of
788 Sciences 96**:** 4506-4511.
789 Crown, K. N., D. E. Miller, J. Sekelsky and R. S. Hawley, 2018 Local inversion heterozygosity alters
790 recombination throughout the genome. Curr Biol 28**:** 2984-2990 e2983.
791 Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and
792 BCFtools. Gigascience 10.
793 Darling, A. E., B. Mau and N. T. Perna, 2010 progressiveMauve: Multiple genome alignment with gene
794 gain, loss and rearrangement. PLoS One 5**:** e11147.
795 Díaz-González, J., and A. Domínguez, 2020 Different structural variants of roo retrotransposon are active
796 in *Drosophila melanogaster*. Gene 741**:** 144546.
797 Dod, B., C. Litel, P. Makoundou, A. Orth and P. Boursot, 2003 Identification and characterization of t
798 haplotypes in wild mice populations using molecular markers. Genetics Research 81**:** 103-114.
799 Eanes, W. F., C. Wesley and B. Charlesworth, 1992 Accumulation of P elements in minority inversions in
800 natural populations of *Drosophila melanogaster*. Genetics Research 59**:** 1-9.
801 Erhart, M. A., S. Lekgothoane, J. Grenier and J. H. Nadeau, 2002 Pattern of segmental recombination in
802 the distal inversion of mouse t haplotypes. Mammalian genome 13**:** 438.

Eyre-Walker, A., P. D. Keightley, N. G. Smith and D. Gaffney, 2002 Quantifying the slightly deleterious mutation model of molecular evolution. Molecular Biology and Evolution 19**:** 2142-2149.

Fay, J. C., G. J. Wyckoff and C.-I. Wu, 2001 Positive and negative selection on the human genome. Genetics 158**:** 1227-1234.

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics 78**:** 737-756.

Fiston-Lavier, A. S., N. D. Singh, M. Lipatov and D. A. Petrov, 2010 Drosophila melanogaster recombination rate calculator. Gene 463**:** 18-20.

Fuyama, Y., 1984 Gynogenesis in *Drosophila-Melanogaster*. Japanese Journal of Genetics 59**:** 91-96.

Ganetzky, B., 1977 On the components of segregation distortion in *Drosophila melanogaster*. Genetics 86**:** 321-355.

Hamilton, W. D., 1967 Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. Science 156**:** 477-488.

Hartl, D. L., 1974 Genetic dissection of segregation distortion. I. Suicide combinations of *SD* genes. Genetics 76**:** 477-486.

Hartl, D. L., 1975 Genetic dissection of segregation distortion II. Mechanism of suppression of distortion by certain inversions. Genetics 80**:** 539-547.

Hartl, D. L., Y. Hiraizumi and J. F. Crow, 1967 Evidence for sperm dysfunction as mechanism of segregation distortion in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 58**:** 2240-+.

Herrmann, B. G., D. P. Barlow and H. Lehrach, 1987 A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal t complex inversion. Cell 48**:** 813-825.

Hill, W., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theoretical and applied genetics 38**:** 226-231.

Hiraizumi, Y., D. W. Martin and I. A. Eckstrand, 1980 A modified model of segregation distortion in *Drosophila melanogaster*. Genetics 95**:** 693-706.

Houtchens, K., and T. W. Lyttle, 2003 *Responder* (*Rsp*) alleles in the *Segregation Distorter* (*SD*) system of meiotic drive in Drosophila may represent a complex family of satellite repeat sequences. Genetica 117**:** 291-302.

Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics 18**:** 337-338.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111**:** 147-164.

Hurst, L. D., and A. Pomiankowski, 1991 Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics 128**:** 841-858.

Kaminker, J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas *et al.*, 2002 The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome biology 3**:** research0084. 0081.

Kapopoulou, A., S. P. Pfeifer, J. D. Jensen and S. Laurent, 2018 The demographic history of African *Drosophila melanogaster*. Genome biology and evolution 10**:** 2338-2342.

Keane, T. M., K. Wong and D. J. Adams, 2013 RetroSeq: transposable element discovery from next-generation sequencing data. Bioinformatics 29**:** 389-390.

Kelemen, R. K., and B. Vicoso, 2018 Complex history and differentiation patterns of the t-haplotype, a mouse meiotic driver. Genetics 208**:** 365-375.

Keller, L., and K. G. Ross, 1998 Selfish genes: a green beard in the red fire ant. Nature 394**:** 573-575.

Khost, D. E., D. G. Eickbush and A. M. Larracuente, 2017 Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. Genome research 27**:** 709-721.

Kofler, R., A. J. Betancourt and C. Schlotterer, 2012 Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genet 8**:** e1002487.

Kofler, R., V. Nolte and C. Schlotterer, 2015 Tempo and Mode of Transposable Element Activity in Drosophila. PLoS Genet 11: e1005406.

Kolmogorov, M., J. Yuan, Y. Lin and P. A. Pevzner, 2019 Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37: 540-546.

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel and A. Stamatakis, 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35: 4453-4455.

Kusano, A., C. Staber and B. Ganetzky, 2001 Nuclear mislocalization of enzymatically active *RanGAP* causes segregation distortion in *Drosophila*. Dev Cell 1: 351-361.

Lack, J. B., J. D. Lange, A. D. Tang, R. B. Corbett-Detig and J. E. Pool, 2016 A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. Mol Biol Evol 33: 3308-3313.

Langley, C. H., M. Crepeau, C. Cardeno, R. Corbett-Detig and K. Stevens, 2011 Circumventing heterozygosity: sequencing the amplified genome of a single haploid *Drosophila melanogaster* embryo. Genetics 188: 239-246.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357-359.

Larkin, A., S. J. Marygold, G. Antonazzo, H. Attrill, G. Dos Santos *et al.*, 2021 FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res 49: D899-D907.

Larracuente, A. M., 2014 The organization and evolution of the *Responder* satellite in species of the *Drosophila melanogaster* group: dynamic evolution of a target of meiotic drive. BMC evolutionary biology 14: 233-233.

Larracuente, A. M., and D. C. Presgraves, 2012 The selfish *Segregation Distorter* gene complex of *Drosophila melanogaster*. Genetics 192: 33-53.

Laurent, S. J., A. Werzner, L. Excoffier and W. Stephan, 2011 Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. Molecular biology and evolution 28: 2041-2051.

Lefevre, G., 1976 A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands. , pp. 31-66 in *The Genetics and Biology of Drosophila*, edited by Ashburner and Novitski.

Li, H. P., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive substitution in *Drosophila*. Plos Genetics 2: 1580-1589.

Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier *et al.*, 2016 The ecology and evolutionary dynamics of meiotic drive. Trends in Ecology & Evolution 31: 315-326.

Linheiro, R. S., and C. M. Bergman, 2012 Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. PloS one 7: e30008.

Lyon, M. F., 2003 Transmission ratio distortion in mice. Annual review of genetics 37: 393-408.

Lyttle, T. W., 1991 Segregation distorters. Annual review of genetics 25: 511-581.

Merrill, C., L. Bayraktaroglu and A. Kusano, *et al.* 1999 Truncated RanGAP encoded by the *Segregation Distorter* locus of Drosophila.

Miklos, G. L. G., 1972 The genetic structure of chromosomes carrying segregation-distorter. Canadian Journal of Genetics and Cytology 14: 235-243.

Miller, D. E., K. R. Cook, A. V. Arvanitakis and R. S. Hawley, 2016 Third Chromosome Balancer Inversions Disrupt Protein-Coding Genes and Influence Distal Recombination Events in Drosophila melanogaster. G3 (Bethesda) 6: 1959-1967.

Miller, D. E., K. R. Cook, E. A. Hemenway, V. Fang, A. L. Miller *et al.*, 2018 The Molecular and Genetic Characterization of Second Chromosome Balancers in *Drosophila melanogaster*. G3 (Bethesda) 8: 1161-1171.

Miyashita, N., and C. H. Langley, 1988 Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics 120: 199-212.

Muller, H. J., 1964 The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1: 2-9.

Myers, S. R., and R. C. Griffiths, 2003 Bounds on the minimum number of recombination events in a sample history. Genetics 163**:** 375-394.

Nelson, M. G., R. S. Linheiro and C. M. Bergman, 2017 McClintock: an integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. G3: Genes, Genomes, Genetics 7**:** 2763-2778.

Ohta, T., 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theoretical population biology 10**:** 254-275.

Pedersen, B. S., and A. R. Quinlan, 2018 Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34**:** 867-868.

Pimpinelli, S., and P. Dimitri, 1989 Cytogenetic analysis of segregation distortion in Drosophila melanogaster: the cytological organization of the *Responder* (*Rsp*) locus. Genetics 121**:** 765-772.

Platzer, A., V. Nizhynska and Q. Long, 2012 TE-Locate: a tool to locate and group transposable element occurrences using paired-end next-generation sequencing data. Biology 1**:** 395-410.

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet 8**:** e1003080.

Powers, P. A., and B. Ganetzky, 1991 On the components of segregation distortion in *Drosophila melanogaster*. V. Molecular analysis of the Sd locus. Genetics 129**:** 133-144.

Presgraves, D. C., P. R. Gerard, A. Cherukuri and T. W. Lyttle, 2009 Large-scale selective sweep among *Segregation Distorter* chromosomes in African populations of *Drosophila melanogaster*. PLoS Genet 5**:** e1000463.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81**:** 559-575.

Quick, J.

Ruan, J., and H. Li, 2020 Fast and accurate long-read assembly with wtdbg2. Nat Methods 17**:** 155-158.

Sandler, L., and A. Carpenter, 1972 A note on the chromosomal site of action of SD in Drosophila melanogaster, pp. 247-268 in *Ediburgh Symposium on the Genetics of the Spermatozoon*.

Sandler, L., and Y. Hiraizumi, 1960 Meiotic drive in natural populations of *Drosophila melanogaster.* IV. Instability at the *Segregation-Distorter* locus. Genetics 45**:** 1269.

Sandler, L., Y. Hiraizumi and I. Sandler, 1959 Meiotic Drive in Natural Populations of *Drosophila Melanogaster.* I. the Cytogenetic Basis of Segregation-Distortion. Genetics 44**:** 233-250.

Schaeffer, S. W., and E. L. Miller, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics 135**:** 541-552.

Schwander, T., R. Libbrecht and L. Keller, 2014 Supergenes and complex phenotypes. Current Biology 24**:** R288-R294.

Smit, A., R. Hubley and P. Green, 2013 RepeatMasker, pp. in *Open-4.0*.

Sniegowski, P. D., and B. Charlesworth, 1994 Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. Genetics 137**:** 815-827.

Stevison, L. S., K. B. Hoehn and M. A. F. Noor, 2011 Effects of Inversions on Within- and Between-Species Recombination and Divergence. Genome Biol Evol 3**:** 830-841.

Stolle, E., R. Pracana, P. Howard, C. I. Paris, S. J. Brown *et al.*, 2019 Degenerative expansion of a young supergene. Molecular biology and evolution 36**:** 553-561.

Svedberg, J., S. Hosseini, J. Chen, A. A. Vogan, I. Mozgova *et al.*, 2018 Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. Nature communications 9**:** 1-13.

team, R. c., 2019 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, Vienna, Austria, R Foundation for Statistical Computing, Vienna, Austria.

Temin, R. G., B. Ganetzky, P. A. Powers, T. W. Lyttle, S. Pimpinelli *et al.*, 1991 Segregation distortion in *Drosophila melanogaster*: genetic and molecular analyses. The American Naturalist 137**:** 287-331.

Thompson, M. J., and C. D. Jiggins, 2014 Supergenes and their role in evolution. Heredity 113**:** 1-8.

954 Thornton, K., and P. Andolfatto, 2006 Approximate Bayesian inference reveals evidence for a recent,
955 severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics 172**:** 1607-
956 1619.
957 Turner, J. R., 1967 On supergenes. I. The evolution of supergenes. The American Naturalist 101**:** 195-221.
958 Tuttle, E. M., A. O. Bergland, M. L. Korody, M. S. Brewer, D. J. Newhouse *et al.*, 2016 Divergence and
959 Functional Degradation of a Sex Chromosome-like Supergene. Curr Biol 26**:** 344-350.
960 Uyenoyama, M. K., 2005 Evolution under tight linkage to mating type. New phytologist 165**:** 63-70.
961 Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: an integrated tool for
962 comprehensive microbial variant detection and genome assembly improvement. PLoS One 9**:**
963 e112963.
964 Wallace, L. T., and M. A. Erhart, 2008 Recombination within mouse t haplotypes has replaced significant
965 segments of t-specific DNA. Mammalian Genome 19**:** 263.
966 Wang, J., Y. Wurm, M. Nipitwattanaphon, O. Riba-Grognuz, Y.-C. Huang *et al.*, 2013 A Y-like social
967 chromosome causes alternative colony organization in fire ants. Nature 493**:** 664-668.
968 Weir, B. S., and C. C. Cockerham, 1984 Estimating F-Statistics for the Analysis of Population Structure.
969 Evolution 38**:** 1358-1370.
970 Wick, R. R., L. M. Judd, C. L. Gorrie and K. E. Holt, 2017 Completing bacterial genome assemblies with
971 multiplex MinION sequencing. Microb Genom 3**:** e000132.
972 Wu, C.-I., T. W. Lyttle, M.-L. Wu and G.-F. Lin, 1988 Association between a satellite DNA sequence and
973 the responder of segregation distorter in *D. melanogaster*. Cell 54**:** 179-189.
974 Zhuang, J., J. Wang, W. Theurkauf and Z. Weng, 2014 TEMP: a computational method for analyzing
975 transposable element polymorphism in populations. Nucleic Acids Res 42**:** 6826-6838.
976