The Recombination Landscape of *Drosophila virilis* is Robust to Transposon Activation in Hybrid Dysgenesis

Lucas W. Hemmer

Justin P. Blumenstiel

Department of Ecology and Evolutionary Biology

University of Kansas

Lawrence, KS 66045 USA

Correspondence: lhemmer@ku.edu, jblumens@ku.edu

Abstract

DNA damage in the germline is a double-edged sword. Induced double-strand breaks establish the foundation for meiotic recombination and proper chromosome segregation but can also pose a significant challenge for genome stability. Within the germline, transposable elements are powerful agents of double-strand break formation. How different types of DNA damage are resolved within the germline is poorly understood. For example, little is known about the relationship between the frequency of double-stranded breaks, both endogenous and exogenous, and the decision to repair DNA through one of the many pathways, including crossing over and gene conversion. Here we use the *Drosophila virilis* hybrid dysgenesis model to determine how recombination landscapes change under transposable element activation. In this system, a cross between two strains of *D. virilis* with divergent transposable element profiles results in the hybrid dysgenesis phenotype, which includes the germline activation of diverse transposable elements, reduced fertility, and male recombination. However, only one direction of the cross results in hybrid dysgenesis. This allows the study of recombination in genetically identical F1 females; those with baseline levels of programmed DNA damage and those with an increased level of DNA damage resulting from transposable element proliferation. Using multiplexed shotgun genotyping to map crossover events, we compared the recombination landscapes of hybrid dysgenic and non-hybrid dysgenic individuals. The frequency and distribution of meiotic recombination appears to be robust during hybrid dysgenesis. However, hybrid dysgenesis is also associated with occasional clusters of recombination derived from single dysgenic F1 mothers. The clusters of recombination are hypothesized to be the result of mitotic crossovers during early germline development. Overall, these results show that meiotic recombination in *D. virilis* is robust to the damage caused by transposable elements during early development.

Introduction

Meiosis produces haploid gametes and meiotic recombination plays a critical role in ensuring proper chromosome segregation. However, recombination also serves a benefit to populations since it creates new genotypic combinations that can facilitate adaptation to changing environments (Burt 2000). While meiosis is a conserved process important for fitness, it can be exploited by selfish genetic elements. Transposable elements (TEs) are one selfish genetic element that can increase in abundance within the genome, but at the detriment of the host. TEs can exploit sexual reproduction because fertilization provides an opportunity for TEs to invade new genomes (Hickey 1982). Upon genome invasion, TE proliferation can result in mutation, misregulation of gene expression, chromosomal rearrangement, and improper recombination (Kidwell et al. 1977; Kazazian et al. 1988; Bennetzen 2000; Slotkin et al. 2007; Zhang et al. 2011). Moreover, TEs can activate the DNA damage response within developing germline stem cells resulting in apoptosis and sterility (Levine et al. 2016).

The harmful effects of TEs are especially evident in syndromes of hybrid dysgenesis, when sterility can arise through intraspecific crosses of males carrying TEs absent in females (Bingham et al. 1982; Bucheton et al. 1984; Yannopoulos et al. 1987; Lozovskaya et al. 1990). Hybrid dysgenesis is the result of TE activation in the absence of maternal repression by PIWI-interacting RNAs or piRNAs (Aravin et al. 2007; Brennecke et al. 2008). The piRNA system of genome defense requires recognition of native TEs and maternal deposition of piRNA to successfully silence TEs within the progeny germline. The combination of unrecognized TEs introduced to a naive genome without the accompanying piRNA deposition result in TE activation and hybrid dysgenesis (Brennecke et al. 2007; Brennecke et al. 2008). A hybrid dysgenesis system is known within *D. virilis* in intraspecific crosses between two strains, the inducing strain 160 and the reactive strain 9 (Lozovskaya et al. 1990). The primary TE family responsible for inducing dysgenesis remains unknown but sterility appears to be due to the mass activation of several

TE families abundant in strain 160 but not strain 9. Up to four elements are proposed to contribute significantly to dysgenesis; *Penelope*, *Helena*, *Paris*, and *Polyphemus* (Petrov et al. 1995; Evgen'ev et al. 1997; Vieira et al. 1998; Blumenstiel 2014).

The relationship between TEs and recombination rates is complicated and varies between organisms and TE families. Typically, TEs accumulate in low-recombining regions of the genome, including heterochromatic regions and centromeres (Bartolomé et al. 2002; Rizzon et al. 2002; Jensen-Seaman et al. 2004; Kent et al. 2017). However, some TEs, such as *Alu* elements in humans and DNA transposons in wheat and potato, accumulate in gene rich areas and their density can be positively correlated with recombination rates (Witherspoon et al. 2009; Daron et al. 2014; Marand et al. 2017). The differences might lie in the coevolution between TEs and their host genomes over evolutionary time to drive TEs into low-recombining regions in contrast to other forces such as insertion bias associated with certain TE families (Kent et al. 2017).

TEs may also modulate recombination rates directly through transposition. Transposition itself can induce illegitimate recombination. For example, in the *P-M* system, some males experiencing hybrid dysgenesis undergo recombination (Hiraizumi 1971; Kidwell and Kidwell 1975; Kidwell et al. 1977). This is abnormal since meiotic recombination is absent in *D. melanogaster* males. However, those rates of recombination are typically low, at approximately 2-3% of meiotic events (Preston and Engels 1996). These have been usually attributed to an increased rate of mitotic exchange induced by DNA damage (Isackson et al. 1981). Likewise, many of these recombination events occur closely near the locations of *P* elements, require transposase, and are likely the product of transposition (McCarron et al. 1994; Sved et al. 1995; Gray et al. 1996; Preston et al. 1996; Preston and Engels 1996).

The effects of hybrid dysgenesis on female recombination rates in *D. melanogaster* are less clear. Changes in recombination rates were not initially observed in the *P* element system (Hiraizumi 1971;

Chaboissier et al. 1995) but later studies found a slight increase in female recombination rates during hybrid dysgenesis (Broadhead et al. 1977; Kidwell 1977; Sved et al. 1991). These increases in recombination are often within pericentric heterochromatin. This could indicate loss of control in recombination mechanisms since meiotic recombination within the centromere may impair the centromere functions role in segregation during meiosis (Koehler et al. 1996; Hassold and Hunt 2001; Rockmill et al. 2006). Slightly increased rates were also observed within the *I-R* element system but these were not localized specifically to the centromere (Chaboissier et al. 1995). However, others have identified no increase in recombination rates but rather a redistribution of towards lower recombining and heterochromatic regions near the centromere (Slatko 1978; Hiraizumi 1981). Since studies in male *Drosophila* indicate that aberrant recombination occurs at low frequencies, differences in female recombination levels are often difficult to detect. Genotyping high numbers of progeny or constructing fine-scale genetic maps are needed to detect changes in the recombination landscape in dysgenic females.

How hybrid dysgenesis impacts the recombination landscape during meiosis remains unclear. Several non-mutually exclusive mechanisms could impact meiotic recombination upon transposable element activation and transposition. These include the following: 1) Increased rates of double-stranded break (DSBs) formation arising transposition, 2) feedback between TE activation of the DNA damage response and meiotic recombination, and 3) modulation of epigenetic marks leading to heterochromatin formation.

The transposition of mobile elements frequently produces DSBs at insertion and excision sites. Coincidentally, DSBs are also the necessary substrate for the initiation of meiotic recombination. In meiosis, this is a controlled, programmed process. Spo11 protein initiations DSBs that become the substrate for repair through homologous recombination. Additional DSBs produced by TEs during meiosis may similarly be repaired by same homologous recombination machinery resulting in crossovers at the sites of TE insertion and excision. However, TE-induced DSBs that occur in meiosis may also be repaired by homologous recombination without subsequent crossing. This may occur during synthesis-

dependent strand annealing or single-strand annealing. Alternately, non-homologous end joining would also preclude crossover generation. In the *P* element system, crossing over is rare (< 1%), occurring at rates indistinguishable from non-dysgenic crosses (Preston et al. 2006; Johnson-Schlitz et al. 2007).

Transposon activity and DSBs often activate a suite of genes regulating the DNA damage response pathway (Joyce et al. 2011; Shim et al. 2014; Wylie et al. 2016; Ma et al. 2017; Tasnim and Kelleher 2018). Key regulators of the DNA damage response pathway that become activated include *p53*, *chk2*, and ATM (*tefu* in *D. melanogaster*). Additionally, *p53* and *chk2* are activated during hybrid dysgenesis to determine the fate of germline stem cells and the severity of the dysgenic phenotype (Ma et al. 2017; Tasnim and Kelleher 2018). Many of the DNA damage-response genes, including *brca2* and *p53*, are also necessary for regulation of meiotic recombination (Klovstad et al. 2008; Lu et al. 2010; Joyce et al. 2011). Meiotic recombination begins with programmed DSBs created by Spo11 to activate the DNA damage response pathway in similar manner to non-programmed DSBs or TE activity. Induction of the DNA damage response pathway via transposon activation and hybrid dysgenesis could create feedback between the pathway and typical meiotic recombination. This could lead to global effects on meiotic recombination frequency and distribution in dysgenic progeny.

Transposon activity is often suppressed through heterochromatin formation (Josse et al. 2007; Phalke et al. 2009) and many piRNA genes responsible for directly suppressing TE activity modify heterochromatin formation in progeny by maternal transmission of piRNA (Brower-Toland et al. 2007; Sienski et al. 2012; Le Thomas et al. 2014). Likewise, loss of epigenetic silencing marks upon disruption of piRNA genes leads to de-repression of TEs via reduced heterochromatin formation (Klenov et al. 2007; Sienski et al. 2012; Le Thomas et al. 2013). Differences in maternally-provisioned piRNA profiles from *Drosophila* strains in hybrid dysgenesis crosses could lead to differences in the establishment of heterochromatin at TE sites. Likewise, chromatin accessibility is a predictor for crossover locations as recombination hotspots are associated with open heterochromatin marks in different species (Berchowitz et al. 2009;

Getun et al. 2010; Wang and Elgin 2011; Choi et al. 2013; Shilo et al. 2015; Marand et al. 2017). Even though *Drosophila* do not have recombination hotspots, motifs associated with open chromatin are still reliable predictors for recombination distribution in *D. melanogaster* (Adrian et al. 2016). Thus, differences in the establishment of heterochromatin associated with dysgenesis or TE profiles between strains could lead to the changes in the frequency and distribution of recombination along the length of chromosomes. In *Drosophila*, TEs suppressed by heterochromatin-associated epigenetic marks exhibit epigenetic suppression of genes up to 20 kb from the site of the TE (Lee and Karpen 2017). Meanwhile, loss of epigenetic silencing pathways in *Arabidopsis* increases recombination rates in pericentromeric regions resembling the changes in recombination associated with the loss of position-effect variegation genes in *D. melanogaster* (Westphal and Reuter 2002; Underwood et al. 2017).

Previous studies on recombination during hybrid dysgenesis were all in either the *P-M* or *I-R* in *D. melanogaster* using phenotypic markers to detect crossovers (Broadhead et al. 1977; Kidwell 1977; Chaboissier et al. 1995). The influence of hybrid dysgenesis on recombination within *D. virilis* is still unknown. Likewise, little is known about recombination in *D. virilis*. Previous studies constructed genetic maps with a limited number of phenotypic or genotypic markers and the estimated recombination rates are highly varied between studies (Weinstein 1920; Gubenko and Evgen'ev 1984; Huttunen et al. 2004) (Table 3.1). This study produces the first fine-scaled genetic map for *D. virilis* using thousands of genotypic markers for further recombination studies. It is also the first to investigate differences in crossover frequency and distribution in the hybrid dysgenesis syndrome of *D. virilis*. I find no detectable differences in recombination between dysgenic and non-dysgenic progeny except in two cases of mitotic recombination produced during dysgenesis to conclude there is no effect of TE activity on meiotic recombination landscapes.

Materials and Methods

Fly Stocks and Crosses

Each strain and all subsequent crosses were maintained on standard media at 25°C. Strain 9 and strain 160 were previously inbred for 10 generations from sibling crosses to form two highly inbred lines for accurate genotyping. Approximately 20 virgin females of one strain and 20 younger males 2-10 days old of the other strain were crossed for six days. Strain 9 females crossed to strain 160 males induced dysgenesis in the F1 generation while the cross in the other direction produced non-dysgenic F1 flies. Individual F1 females four days post-emergence were backcrossed to two or three 2-10 day old strain 9 males in vials for six days. Non-dysgenic females were often allowed to lay embryos for 4-5 days because their high fertility. Some dysgenic F1 females were transferred to another vial after ten days and allowed to mate for an additional four days to obtain greater numbers of progeny and test whether fertility may be restored with age in hybrid dysgenesis. F2 females were collected once per day and immediately frozen at -20°C. Only 12-20 of the early emerging flies from non-dysgenic F1 backcrosses were collected while all progeny of the dysgenic F1 backcrosses were collected to keep record of the dysgenic F1 parents' fertility.

DNA Extraction and Library Preparation

I extracted DNA with the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) following the Insect Tissue Protocol and stored at -20°C. Prior to DNA extraction, flies were homogenized by 3.5 mm glass grinding balls (BioSpec) placed into a U-bottom polypropylene 96-well plate with lysis buffer from the kit placed into a MiniBeadBeater-96 at 2,100 rpm for 45 seconds. DNA extraction yields varied from <0.1 ng/µl to 5 ng/µl. DNA was quantified using a Qubit fluorometer (Invitrogen). DNA quantification was not performed on the majority of samples but was assumed to average 1-2 ng/µl necessary for library preparation based on the measured samples.

Library preparations for 192 samples were performed following the protocol outlined in Andolfatto et al. (2011) with some minor changes. Assuming a DNA concentration around 1 ng/µl, 10 µl of genomic DNA were added to a clean 96-well PCR plate and digested with 3.3 U of MseI in 20 µl of reaction volume for 4 h at 37°C with heat inactivation at 65°C for 20 min. Bar-coded adapters (5 µM) were attached to the digested DNA with 1 U of T4 DNA ligase (New England Biolabs) in 50 µl of reaction volume at 16°C for 5 h and inactivated at 65°C for 10 minutes. The samples were pooled and concentrated using isopropanol precipitation (1/10 vol NaOAc at pH 5.2, 1 vol of 100% isopropanol, and 1 µl glycogen). The library was resuspended in 125 µl of 1X Tris-EDTA (pH 8). Adapter dimers were removed 1.5X vol AMPure XP Beads (Agencourt) and resuspended in 32 µl of 1X Tris-EDTA (pH 8). I selected for 200-400 bp DNA fragments using a BluePippin (Sage Science). Size-selected fragments were cleaned using 2X vol of AMPure XP beads and resuspended 20 µl of 1X elution buffer (10 µM Tris, pH 8). The libraries were quantified using a Qubit fluorometer before an 18-cycle PCR amplification on bar-coded fragments with Phusion high-fidelity PCR Kit (New England Biolabs). The adaptors used were the FC1 and FC2 adaptors specifically noted in Andolfatto et al. (2011). The PCR products were cleaned using 1X vol of AMPure XP Beads.

Library preparations for 768 samples were performed using in-house produced Tn5 transposase produced following the procedure outlined in Picelli et al. (2014) following a similar tagmentation protocol. I extracted DNA using the same Agencourt DNAdvance Genomic DNA isolation kit and protocols. Assuming an average of 1-2 ng/µl DNA concentration per sample in a 96-well plate, 1 µl of DNA was tagmented with the in-house Tn5 transposase at a concentration of 1.6 mg/ml with pre-annealed oligonucleotides in a 20 µl reaction volume for 55°C for 7 min and stopped by holding at 10°C. The reaction volume also contained 2 µl of 5X TAPS-DMF buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5), 50% v/v DMF) and 2 µl of 5x TAPS-PEG buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5), 60% v/v PEG 8000) for the desired DNA fragment lengths. The in-house Tn5 transposase was inactivated

with an addition of 5 µl of 0.2% SDS and heating the total reaction to 55°C for 7 min. Only 2.5 µl of tagmentation reaction is needed for the PCR amplification with KAPA HiFi HotStart ReadyMix PCR Kit (Thermo Fisher Scientific), 1 µl of 4 µM Index 1 (i7) primers (Appendix 16), and 1 µl of 4 µM Index 2 (i5) primers (Appendix 17) in 9 µl of reaction volume. The PCR occurred as follows: 3 min at 72°C, 2 min 45 sec at 98°C, and then 14 cycles of 98°C for 15 sec, 62°C for 30 sec, 72°C for 1 min 30 sec. The PCR-amplified samples were pooled and cleaned using 0.8 X vol AMPure XP Beads. I size-selected DNA fragments 250-400 bp on a BluePippin and cleaned using 1X vol of AMPure XP Beads.

Sequencing and Crossover Quantification

All libraries were sequenced at the University of Kansas Genomics Core on an Illumina HiSeq 2500 Sequencer with 100 bp single-end sequencing. The first 192 samples were sequenced on two lanes using the Rapid-Run Mode resulting in 120 million reads per lane while the Tn5-produced libraries were sequenced on two lanes using the High-Output Mode producing 180 million reads per lane. FASTQ files were parsed before using the multiplex shotgun genotyping (MSG) bioinformatic pipeline for identifying reliable genotype markers and determining ancestry at those markers using a Hidden Markov Model. Samples with a minimum of 10,000 reads provided reliable genotype calls with high confidence; all samples with less were discarded. Crossover events identified by this pipeline were manually curated for errors. Double crossovers less than 750 kb apart were discarded as these events are unlikely to occur within 1 Mb and most were due to mapping errors. Due to low quality sequences leading to erroneous mapping calls on the ends of chromosomes, crossovers located within 500 kb of the telomere end for the X and 4th chromosome and crossovers within 700 kb on the 2nd, 3rd, and 5th chromosomes were removed. Crossovers near the centromere ends of each chromosome were removed as follows: within 3.5 Mb on the X chromosome, within 1.1 Mb on the 2nd chromosome, within 1.5 Mb on the 3rd chromosome, within 2.4 Mb on the 4th chromosome, and 2.3 Mb on the 5th chromosome.

Data Analysis

Crossover data was parsed and analyzed within the R Version 3.4.2 programming environment (R Core Team 2017). The following packages were also used in genetic map construction, model testing, and general data analysis: R/qtl (Broman et al. 2003), lme4 (Bates et al. 2015), lsmeans (Lenth 2016), and Biostrings (Pagès et al. 2017). Figures were produced using ggplot2 (Wickham 2016). Maker (Cantarel et al. 2008) was used to annotate the genome assemblies of strains 9 and 160 with the most up-to-date gff file for D. virilis (1.6) downloaded from Flybase (Gramates et al. 2017). The genome was annotated for TEs annotations using Repeatmasker (Tarailo-Graovac and Chen 2009) with the current catalog of TE sequences in *D. virilis* from Repbase (Bao et al. 2015). Repeatmasker was also used to identify intact *Polyphemus*, *Penelope*, and *Helena* sites that were less than 5% diverged from the annotated sequence and within 100 bp of the annotated length in strain 160 sequences. Strain 160 sequences were recently sequenced using PacBio although not completely assembled in contig-level scaffolds.

Results

Crosses and Genotyping

The hybrid dysgenesis syndrome in *D. virilis* is induced in crosses between strain 9 females and strain 160 males but the severity of dysgenesis varies in the resulting progeny (Lozovskaya et al. 1990; Erwin et al. 2015). The F1 females from the dysgenic and reciprocal crosses were backcrossed to strain 9 males. The F2 progeny were sequenced to quantify crossover events produced within F1 female germline. F1 females collected in the non-dysgenic reciprocal crosses were highly fertile and capable of producing large number of progeny. By nature of dysgenesis, most dysgenic F1 females have reduced fertility with many producing few or no offspring. Therefore, all the progeny produced by the termed "low-fecund" dysgenic flies were sequenced and the first 12-20 F2 flies produced from non-dysgenic females were

sequenced to balance the number of progeny from a single mother. As previously mentioned, up to 30-50% of the F1 females produced in dysgenic crosses show no outward signs of dysgenesis. The termed "high fecund" F1 females are capable of producing as many progeny as the non-dysgenic flies. Approximately 40 F2 progeny from several highly fecund dysgenic females were sequenced to obtain greater resolution on the effects of transposition on recombination within a single maternal germline. The variation in dysgenesis provides an additional comparison in the analysis of recombination landscapes between two outcomes of TE activation: TE activation with deleterious effects to fertility and TE activation with no observable negative effects.

F2 female progeny were sequenced on an Illumina HiSeq 2500 and genotypes were called following multiplexed shotgun genotyping (MSG) protocol for indexing (Andolfatto et al. 2011). In total, 828 F2 female flies were sequenced to map recombination breakpoints. Out of the total, 275 F2 flies were sequenced from 20 F1 non-dysgenic females, 311 F2 flies collected from 66 lowly fecund F1 dysgenic females, and 242 F2 flies were collected from seven highly fecund F1 dysgenic females. The MSG pipeline identified a total of 1,150,592 quality-filtered SNPs between the two parental genomes with an average of distance of 136 bp between SNPs. The median crossover localization interval for identified crossovers was approximately 18 kb with 84% of all crossovers localized within 50 kb or less. There were 17 crossovers with interval range at ~ 1 Mb but those were in samples with low read counts near the cutoff for reliable CO detection (10,000-20,000 reads).

A High-Resolution Genetic Map of D. virilis

The few previous studies on recombination within *D. virilis* indicate much higher rates of recombination than *D. melanogaster*. This has been shown in a genetic map approximately three times the size of the genetic map of *D. melanogaster* (Gubenko and Evgen'ev 1984; Huttunen et al. 2004). Critically, the genetic map lengths estimated between these two studies is highly variable due to the limited number of

physical and genetic markers available (Table 3.1). In grouping my samples, I provide the first high-resolution recombination map for *D. virilis*. Of particular interest, the dot chromosome of *D. virilis* occasionally undergoes recombination unlike the dot chromosome in *D. melanogaster* (Chino and Kikkawa 1933; Fujii 1940; Gubenko and Evgen'ev 1984). The rate of recombination on the *D. virilis* dot chromosome is low but detectable occurring at a frequency of 1%. However, I was unable to accurately detect crossovers on the dot chromosome, likely due to its repetitive nature. There were a limited number of markers for the MSG pipeline on the dot chromosome with 4,013 quality SNPs within the 2 Mb chromosome identified and only 21 genotypic makers shared between all samples. COs identified on the dot chromosome in the MSG pipeline were primarily on the distal ends of the chromosome. CO events on the dot chromosome failed to pass inspection because few high quality genotypic markers were located on the chromosome ends resulting in false recombination events inferred by the Hidden Markov Model.

In support of previous studies, the total genetic map length of *D. virilis* in my reciprocal crosses between strain 9 and 160 is 732 cM (centiMorgans) or 2.5 times longer than the genetic map length of *D. melanogaster* (Comeron et al. 2012) (Table 3.1). The total genetic map length in this study is over 100 cM shorter than the first detailed genetic map of *D. virilis* (Table 3.1). This may be due to my stringent criteria of excluding crossovers less than 700 kb in effort to reduce genotyping errors from the MSG pipeline. The differences in recombination rates between *D. virilis* and *D. melanogaster* may appear to be comparable in accordance with genome size. The estimated genome size of *D. virilis* is roughly twice the size of the *D. melanogaster* genome, 404 Mb to 201 Mb respectively (Bosco et al. 2007). Thus, across the entire genome, the average rate of recombination in *D. virilis* is 1.8 cM/Mb and similar to the average recombination rate of 1.4 cM/Mb in *D. melanogaster*. However, close to half of the *D. virilis* genome is composed of satellite DNA where little or no recombination takes place (Bosco et al. 2007). Thus, the *D. virilis* euchromatic assembly, where most COs take place, is 206 Mb in length and is only half of the total genome size. Accounting for satellite DNA in both species, the average rate of euchromatic recombination in *D. virilis* is twice as high as *D. melanogaster* based on euchromatic assembly genome

size (3.5 cM/Mb to 1.8 cM/Mb respectively). The genetic map length of each chromosome correlates with physical length of the chromosome in *D. virilis* (R^2 = 0.851, p = 0.025).

Crossover interference reduces the probability of an additional CO in proximity to other COs. I calculated interference in *D. virilis* using the Housworth-Stahl model to calculate *nu*, a unitless measure of interference, with a maximum likelihood function based on intercrossover distances (Housworth and Stahl 2003). If COs are not subject to interference, intercrossover distances are Poisson distributed and the collection of distances from many crosses resembles a Poisson distribution with a *nu* = 1 (Broman and Weber 2000). Each chromosome in *D. virilis* has detectable interference between crossovers with an average *nu* of ~ 3 (Table 3.2). The Houseworth-Stahl model also calculates the percentage of COs produced through an alternative pathway not subject to interference as the escape parameter *P* (de los Santos et al. 2003). Less than 1% of the COs in my study are estimated to be produced through the alternative CO pathway (Table 3.2). In contrast to interference, crossover assurance is the recombination control mechanism to ensure the minimal number of COs on each chromosome for proper chromosome segregation during meiosis. In the absence of CO assurance and interference, the distribution of COs within a collection of progeny should resemble a Poisson distribution in that the variance in CO number is equal to the mean. I find the CO mean and variance are not equal; the mean CO number amongst all F2 progeny is 7.3 and variance is 4.9. The distribution of COs is significantly different from a Poisson distribution (χ^2 = 53.6, p = 5.74E-08). Moreover, *D. virilis* has a higher than expected frequency of individuals with CO numbers close to the mean (5-8 COs) with a lower than expected frequency of extreme CO numbers (2-3 COs on the low end, 12-15 COs on the high end) (Appendix 18). This indicates that the collective action of CO assurance and interference ensures an appropriate number of COs on average produced during meiosis.

In many organisms, the total number of crossovers created in a single tetrad is unavailable because crossovers are detected only single chromatid transmitted to the progeny. This is known as random spore

analysis. Tetrad analysis uses the number of crossovers for each chromosome to calculate the frequency of non-exchange tetrads (E_0), single-exchange tetrads (E_1), or multiple-exchange tetrads (E_n) (Weinstein 1918). I used the classic Weinstein method to determine the frequency of E_0 tetrads for each chromosome in *D. virilis*. The X chromosome and third chromosome E_0 tetrad frequencies were 1.2% and 2.1% respectively (Appendix 19). The calculated E_0 tetrad frequencies in *D. virilis* are lower in comparison to *D. melanogaster* previously estimated to be 5-10% (Zwick et al. 1999a; Hughes et al. 2018). However, the second, fourth, and fifth chromosomes each had biologically meaningless negative E_0 tetrad frequencies ranging from -2.6% to -3.9%. There were also several negative tetrad frequencies estimated for single-exchange and multiple-exchange tetrads as well (Appendix 19). Negative tetrad frequencies are a drawback to using the classic Weinstein method (Zwick et al. 1999a). Nonetheless, these results indicate there are fewer non-exchange tetrads in *D. virilis* compared to *D. melanogaster*.

Recombination rates are often correlated with certain sequence features, such as GC content, simple motifs, and nucleotide polymorphism (Begun and Aquadro 1992; Kong et al. 2002; Comeron et al. 2012). In *D. virilis*, recombination rates are appear to be weakly correlated with GC content and gene density as not all chromosomes show significant correlations to either sequence parameter (Table 3.3). This may be due to unusual patterns of recombination along the length of the chromosome (discussed later). Simple repeats and SNP density show strong positive correlations amongst all chromosomes even after removal of non-recombining regions. Nucleotide diversity is frequently correlated with recombination rates (Begun and Aquadro 1992; Kong et al. 2002) and the strong correlation between SNP density and recombination in my data is consistent with this pattern (Appendix 20). TE density shows a strong negative correlation until non-recombining regions are removed (p = 0.037 after Bonferroni correction). The similar pattern of weak or no correlation between TE density and recombination is also seen in *D. melanogaster* because the majority of TEs are found in the non-recombining heterochromatin regions near centromeres (Kofler et al. 2012; Adrion et al. 2017).

No Differences in Recombination Rates nor Frequency in Hybrid Dysgenesis

To compare and contrast the sum of crossovers in the F2 progeny sired by dysgenic and non-dysgenic females, I constructed a full mixed-effects likelihood model using the lme4 R package for the data (Bates et al. 2015). The state of dysgenesis and brood were treated as fixed effects in a Poisson link model. The mother of origin and the fecundity of the mother were treated as random effects. In the full model, I find no difference in the total number of COs between the two broods (Type III Wald χ^2, $p = 0.171$). The effect of dysgenesis and the interaction between dysgenesis and brood were nearly significant (Type III Wald χ^2, $p = 0.060$, 0.075 respectively) despite nearly identical mean crossover number between the dysgenic and non-dysgenic flies (7.3 and 7.2 mean COs respectively, Kruskal-Wallis χ^2, $p = 0.622$, Figure 3.1A); random effects of F1 mother and fecundity of the F1 did not have a significant contribution to variance within groups. The progeny of a single highly fecund dysgenic mother termed 701 had a significantly larger mean CO number in the than the F2 progeny of the other high fecund dysgenic mothers (8.3 COs, least squares mean contrast, $p = 0.021$, Figure 3.1B). Without the F2 progeny sired by the 701 mother, the difference in recombination between dysgenic and non-dysgenic, broods, and their interaction is negligible (Type III Wald χ^2, p= 0.874, 0.515, and 0.803 respectively). Likewise, there were no major differences in E_0 tetrad frequency and interference between dysgenic and non-dysgenic mothers (data not shown).

The higher recombination rates in *D. virilis* in comparison to *D. melanogaster* are due to a higher number of COs on any given chromosome. In *D. melanogaster*, chromosome arms typically have zero, one, or two COs with three COs on a single chromosome arm being incredibly rare (Miller et al. 2016). A chromosome with three or more COs is a common observation in *D. virilis* in both dysgenic and non-dysgenic directions of the cross with as many as five COs on a single chromosome (Figure 3.2). The proportion of chromosomes with zero, one, two, three, or more COs is not different between the progeny of dysgenic and non-dysgenic mothers ($\chi^2 = 0.529$, $p = 0.97$). There was also no difference between the

progeny non-dysgenic mothers and dysgenic mothers if they were highly fecund ($\chi^2 = 3.70$, $p = 0.45$) nor if the mothers were low fecund ($\chi^2 = 3.45$, $p = 0.49$). Additionally, the contribution of any one chromosome, X or autosome, to the proportion of total chromosomes based on CO number is not different between the progeny of non-dysgenic mothers, low fecund dysgenic mothers, and high fecund dysgenic mothers (Figure 3.2).

I also examined the distribution of recombination along the length of each chromosome between non-dysgenic flies, high fecund dysgenic flies, and low fecund dysgenic flies. There were no major changes in the distribution of recombination along the length of the chromosomes among any of the chromosomes (Figure 3.3). The recombination rates between all three groups are strongly correlated amongst all chromosomes (Appendix 21). Overall, I find no differences in the recombination landscape between dysgenic and non-dysgenic F1 mothers in *D. virilis* at the global level. It appears there is no feedback between activation of the DNA damage response by transposition and the modulation of meiotic recombination. I also examined differences in CO number at the fine scale near intact transposons. CO counts examined are near TEs hypothesized to be the main inducers of dysgenesis including the retrotransposons *Helena* and *Penelope* as well as *Polyphemus*, a DNA transposon. There was a significant increase in the CO number in close proximity to *Polyphemus* sites in the progeny of dysgenic F1 mothers but not at the sites of the two retrotransposons (Table 3.4). This suggests that there is no influence of heterochromatin on recombination near the sites of retrotransposons. However, it is unclear if the increase in recombination at *Polyphemus* sites is due to differences in heterochromatin state or the method of transposition; DNA transposons produce DSBs upon their excision while retrotransposons do not excise themselves.

Evidence for Mitotic Crossing Over in Dysgenic Progeny

One *Polyphemus* site exhibited a much higher number of COs among the dysgenic progeny, specifically from progeny of the highly fecund F1 mothers. This cluster of recombination is located near a *Polyphemus* site 9.7 MB away from the telomere on the third chromosome. The 500 kb interval containing the cluster has an increased recombination rate of 26 cM/Mb, nearly twice as high as any other interval within the genome (Figure 3.3C). The region contained 32 COs in the progeny of the F1 dysgenic mothers compared to a single crossover in the progeny of non-dysgenic mothers within a 250 kb interval. Of those dysgenic mothers, 21 COs were derived from a single F1 mother labeled 5011. Within a 500 kb window, there are 28 COs in proximity to this *Polyphemus* site attributed to the progeny of F1 mother 5011. Reciprocal products of the crossover were observed with equal frequency with no transmission distortion (Binomial test, all markers $p > 0.05$, Figure 3.4B). In addition to the CO near the *Polyphemus* site, there were additional COs detected along the entire length of the third chromosome in the progeny (Figure 3.4B). Only four progeny out of 32 did not have a CO within the interval on the third chromosome.

An additional cluster of recombination was identified on the X chromosome approximately 21.7 Mb from the telomere with an effective recombination rate of 15.7 cM/Mb (Figure 3.2A). While the rate of recombination within this interval was not as extreme as the interval on the third chromosome, the vast majority of COs at this site are attributed to progeny from a single highly fecund dysgenic F1 female labeled 4029. Half of the progeny of the 4029 female had the crossover and no crossovers were detected on the X chromosome distal to the cluster of recombination in all progeny (Figure 3.4C). All progeny obtained were heterozygous for all markers distal to the cluster of recombination. Since F1 females were heterozygous for all SNPs, only half of the progeny are expected to be heterozygous for a given SNP. The extreme excess of heterozygosity in the F2s indicates an extreme transmission distortion of the strain 160 genome from the 4029 mother for this chromosomal region (227 markers between 0.5 - 21.4 MB, Binomial test, $p < 1.6E-08$). However, the proximal region of the chromosome from the cluster of recombination shows no transmission distortion (86 markers between 21.5 - 29.0 Mb Binomial test, $p >$

0.5). Any additional COs on the X chromosome in the 4029 progeny were within this proximal region regardless of the presence or absence of a CO within the cluster of recombination.

I propose both clusters of recombination originated from a mitotic CO event during the early development of germline stem cells (Figure 3.5). Transposons can become mobile in the germline during early development (Engels and Preston 1979; Sokolova et al. 2010). DSBs produced as an outcome of transposition are repaired by one of several mechanisms including homologous recombination via mitotic crossing over. The mass action of TEs from dysgenesis, specifically a *Polyphemus* DNA transposon on the third chromosome in F1 mother 5011, likely produced a DSB to be repaired through homologous recombination in the mitotic germline. One possibility is that mitotic crossover may have occurred prior to DNA duplication, within an early developing germline stem cell in the 5011 mother (Fig 3.5A). The CO would appear in any daughter cells derived from this germline stem cell and reciprocal products would be observed in equal frequency on average. Alternately, a mitotic crossing over may have occurred after DNA replication prior to mitosis in the 5011 mother (Figure 3.5B). During mitosis, the chromatids segregated resulting in both crossover products in one daughter cell while the other daughter cell receives the non-crossover chromatids. Other germline stem cells must have been present within the 5011 mother because there are several progeny without the common crossover product. Nonetheless, this indicates a severe depletion of intact germline stem cells in the F1 dysgenic mother with a recovery of germline stem cells that experienced a mitotic recombination event.

Another mitotic crossover event may have occurred in an early developing germline stem cell on the X chromosome of the 4029 mother (Figure 3.5C). A mitotic CO event occured after DNA replication and the chromatids segregated during metaphase resulting in each daughter cell receiving one chromatid with the CO and one without. One average, the mitotic CO would be transmitted to half of the progeny as seen in the data. The pattern of segregation of crossover and noncrossover chromatids results in a loss-of-heterozygosity in the region of the chromosome distal to the CO. The loss of heterozygosity is responsible

for failure to detect additional meiotic COs derived from the homozygous distal region as seen the data. The complete transmission distortion of one genome in the distal region is the result of the depletion of germline stem cells containing the reciprocal mitotic CO products attributed to hybrid dysgenesis. However, the cluster of recombination in the 4029 mother is not in proximity to the three transposons initially investigated nor any other intact TEs. The mitotic CO may have been the product of a TE insertion but it is unclear from the present data.

Discussion

A High-Resolution Genetic Map of D. virilis

There were no major differences in global recombination between non-dysgenic and dysgenic mothers with the exception of mothers 4029 and 5011, as outlined. The recombination data was combined together to produce the first high-resolution genetic map for *D. virilis*. The genus *Drosophila* contributes significantly to our understanding of meiotic recombination. Recombination was first discovered in *D. melanogaster* over 100 years ago and continues to serve as a model for understanding the mechanisms and consequences of recombination. There is also significant work on recombination in *D. pseudoobscura* and genetic maps or recombination studies in *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. persimilis*, *D. miranda*, *D. serrata*, *D. mojavensis*, and others (True et al. 1996; Takano-Shimizu 2001; Staten et al. 2004; Kulathinal et al. 2008; Stevison and Noor 2010; Stocker et al. 2012; Smukowski Heil et al. 2015). A high-resolution genetic map of *D. virilis* will continue to add to the growing number of genetic maps of species of *Drosophila* for future studies of recombination. Of note is the high rate of recombination in *D. virilis* in comparison to other species, especially *D. melanogaster*. Typically in *Drosophila*, recombination rates frequently peak in the middle of the chromosome arm and decrease towards the centromere and telomere to resemble a bell curve (True et al. 1996). However, the distribution of

recombination on the second, third, and fourth chromosomes in *D. virilis* resembles a bimodal distribution (Figure 3.3). The high rate of recombination in *D. virilis* may explain why some chromosomes have higher recombination rates near the chromosome ends. Even though a single chromosome may have two or more COs, interference prevents COs forming too closely. Interference would favor CO formation on opposite ends of the chromosome when recombination rates are high. Additionally, *D. virilis* has the highest amount of satellite DNA amongst *Drosophila* species (Bosco et al. 2007) and most of the satellite sequences are not in the current genome assembly of *D. virilis*. The distribution of satellite DNA, associated with low rates of recombination, may be shaping patterns of recombination along the lengths of the chromosomes of *D. virilis*.

Meiotic Recombination in Light of Hybrid Dysgenesis in D. virilis

The vast majority of studies on the association between TEs and meiotic recombination focus on sites where TEs accumulate on an evolutionary time frame. There are also a number of studies in the field of molecular biology elucidating the mechanisms of TE activity and its association with inducing aberrant recombination through a variety of mechanisms including non-homologous end joining and single strand annealing. This study is one of few to examine differences in the meiotic recombination landscape upon TE activation by using the hybrid dysgenesis syndrome in *D. virilis*. Previous studies of hybrid dysgenesis in *D. melanogaster* either are conflicting as some found no effect (Hiraizumi 1971; Chaboissier et al. 1995), increases in recombination rates (Broadhead et al. 1977; Kidwell 1977; Sved et al. 1991), or changes in the distribution of recombination (Slatko 1978; Hiraizumi 1981). While my findings are in the syndrome of a different species, it is the first study to investigate recombination differences using high-throughput genotyping rather than phenotypic markers. This allows a greater insight in the fine-scale changes in recombination rates and distribution that may have escaped unnoticed before.

I found no major differences in the distribution and frequency of recombination in *D. virilis* overall under hybrid dysgenesis. There is no evidence for feedback between the activation of the DNA damage response to TE mobilization and the response to meiotic recombination. DNA damage response regulators such as p53 may determine the fates of germline stem cells to either undergo atrophy or tolerate TE mobilization (Tasnim and Kelleher 2018). The incomplete penetrance of hybrid dysgenesis in *D. virilis* may be due to undiscovered differences in the modulation of the DNA damage response and germline fate. Differences in the DNA damage response may lead to differences in meiotic recombination between high and low fecund dysgenic mothers. However, the null effect of fecundity on recombination is further evidence the DNA damage pathway activated by dysgenesis does not feedback into meiotic recombination. This is presumably due to the early effects of hybrid dysgenesis in *D. virilis*.

DNA Transposon Inducing Mitotic Rather than Meiotic Recombination

Increases in CO number in proximity to the DNA transposon *Polyphemus* could either be attributed to modulation of heterochromatin at *Polyphemus* sites or the mechanism of transposition. No difference in CO number at retrotransposons suggests the latter is more likely. Why would heterochromatin formation at *Polyphemus* be any different? *Polyphemus*, *Helena*, and *Penelope* are all more highly abundant in strain 160, are provisioned more piRNA in the strain 160 germline, and mobilize during dysgenesis (Petrov et al. 1995; Blumenstiel 2014; Erwin et al. 2015). *Helena* is interesting in that it continues to be highly expressed in the germlines of the dysgenic progeny in comparison to non-dysgenic progeny (Erwin et al. 2015) but that does not result in higher CO numbers near it. Additionally, there is evidence that while maternally-provisioned piRNA profiles responsible for heterochromatin formation at TE sites differ between strains 9 and 160 this does not translate to major differences in the heterochromatin modulation between dysgenic and non-dysgenic progeny (Evgen'ev, personal communication). It is likely *Polyphemus* increases CO number by producing a DSB upon activation, and thusly, excision from the site as an active DNA transposon. However, a DSB produced by a *Polyphemus* excision leads to a mitotic CO

rather than a meiotic CO in the case of progeny of mother 5011. Mother 5011 was capable of producing high numbers of progeny to provide better power to detect a mitotic CO. Other mitotic COs, especially in low fecund dysgenic flies, can occur without detection in my data analysis because of the low power to detect those events in a limited sample of progeny. Previous studies indicate TEs mobilize during hybrid dysgenesis in the early developing germline within embryos (Engels and Preston 1979; Sokolova et al. 2010). In *D. virilis*, TE suppression resumes by adulthood in dysgenic progeny via production of piRNAs and the negative impacts of dysgenesis disappear in the following generations (Erwin et al. 2015). This indicates TEs rarely produce DSBs when germline cells are undergoing meiosis and homologous recombination to repair transposon-induced DNA damage occurs prior to meiosis. Meiotic recombination appears robust to TE activity and recombination is observed only in rare cases when homologous recombination produces a cluster of recombination or loss of heterozygosity as a result of a CO in the early developing germline.

Tables

Table 3.1: Genetic map lengths of *D. virilis* chromosomes reported in previous studies and this study.

Source	**Chromosome**					
	X	2	3	4	5	6
Gubenko & Evgen'ev 1984	170	257	145	108	203	1
Huttunen et al 2004	-	118	125	147	60	-
This study	143.5	160.9	139.6	148.3	140.0	-

Table 3.2: Interference values (*nu*) and frequency of crossovers created in the non-interference pathway (*P*) for all chromosomes. Both values were estimated with the Housworth-Stahl model for the entire dataset.

Variable	**Chromosome**				
	X	2	3	4	5
nu	3.22	3.17	2.68	3.09	3.37
P	4.81E-02	1.05E-02	2.73E-08	6.91E-03	9.75E-03

Table 3.3: Correlations between rates of recombination and sequence parameters in 250 kb intervals along each chromosome in *D. virilis*. Pearson's correlation coefficients (R) and significance (*p*-value) are listed. Significant values are bolded.

Sequence Parameter		Chromosome					Total	Total minus zero recomb
		X	2	3	4	5		
GC Content	R	0.08	**0.35**	0.11	0.18	**0.33**	**0.23**	**0.15**
	p	0.372	**1.29E-05**	0.263	0.079	**5.03E-04**	**1.08E-08**	**03.68E-04**
Gene density	R	**0.31**	**0.21**	0.12	**0.32**	**0.33**	**0.19**	0.03
	p	**2.89E-04**	**1.16E-02**	0.222	**3.27E-04**	**5.443E-04**	**1.88E-06**	0.506
Simple repeats	R	**0.44**	**0.43**	**0.31**	**0.32**	**0.54**	**0.39**	**0.177**
	p	**2.04E-07**	**3.11E-08**	**1.31E-03**	**03.57E-04**	**1.14E-09**	**3.03E-24**	**2.91E-05**
SNP Density	R	**0.64**	**0.553**	**0.60**	**0.67**	**0.65**	**0.62**	**0.49**
	p	**4.73E-16**	**2.49E-13**	**7.31E-12**	**4.70E-17**	**1.69E-14**	**5.34E-66**	**6.42E-34**
TE Density	R	**-0.47**	**-0.47**	**-0.33**	**-0.44**	**-0.49**	**-0.44**	**-0.14**
	p	**1.58E-08**	**1.27E-09**	**6.11E-04**	**4.08E-07**	**5.18E-08**	**4.05E-30**	**1.07E-03**

Table 3.4: Counts of F2 progeny with a crossover present or absent near intact transposons in dysgenic and non-dysgenic crosses. *Helena*, *Penelope*, and *Polyphemus* are TE families hypothesized to induce hybrid dysgenesis in *D. virilis*. COs within 250 kb of an intact transposon in the strain 160 genome were counted as present. Only full length TEs with less than 95% divergence from the TE consensus sequence were included. Significant *p*-values are bolded.

TE	Helena		Penelope		Polyphemus	
	Dys	Non-Dys	Dys	Non-Dys	Dys	Non-Dys
CO Absent	432	217	393	179	296	170
CO Present	121	58	160	96	257	105
	$p = 0.859$		$p = 0.094$		$\mathbf{p = 0.026}$	

Figures

Figure 3.1: The distribution of the total CO count per F2 progeny with the mean and standard deviation. A) The distribution of the total CO count per F2 progeny of low fecund dysgenic, high fecund and non-dysgenic F1 mothers. B) The distribution of CO count per F2 progeny of each high fecund dysgenic mother with mean and standard deviation. Asterisks denotes statistical significance by least square means ($p < 0.05$). Progeny from mother 701 had a higher average CO count than progeny from other mothers while progeny from mother 4029 exhibited a lower average CO count.

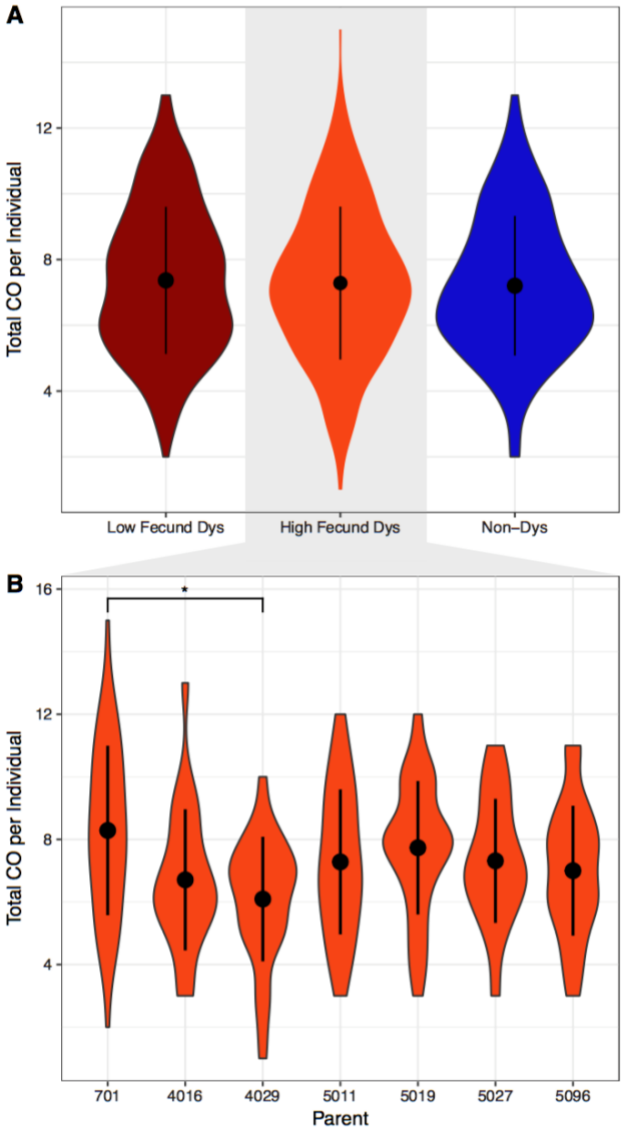

Figure 3.2: The proportion of chromosomes grouped by CO count in F2 progeny of high fecund dysgenic, low fecund dysgenic, and non-dysgenic F1 mothers. 95% confidence intervals were calculated by sampling F2 progeny by bootstrapping 1000 times.

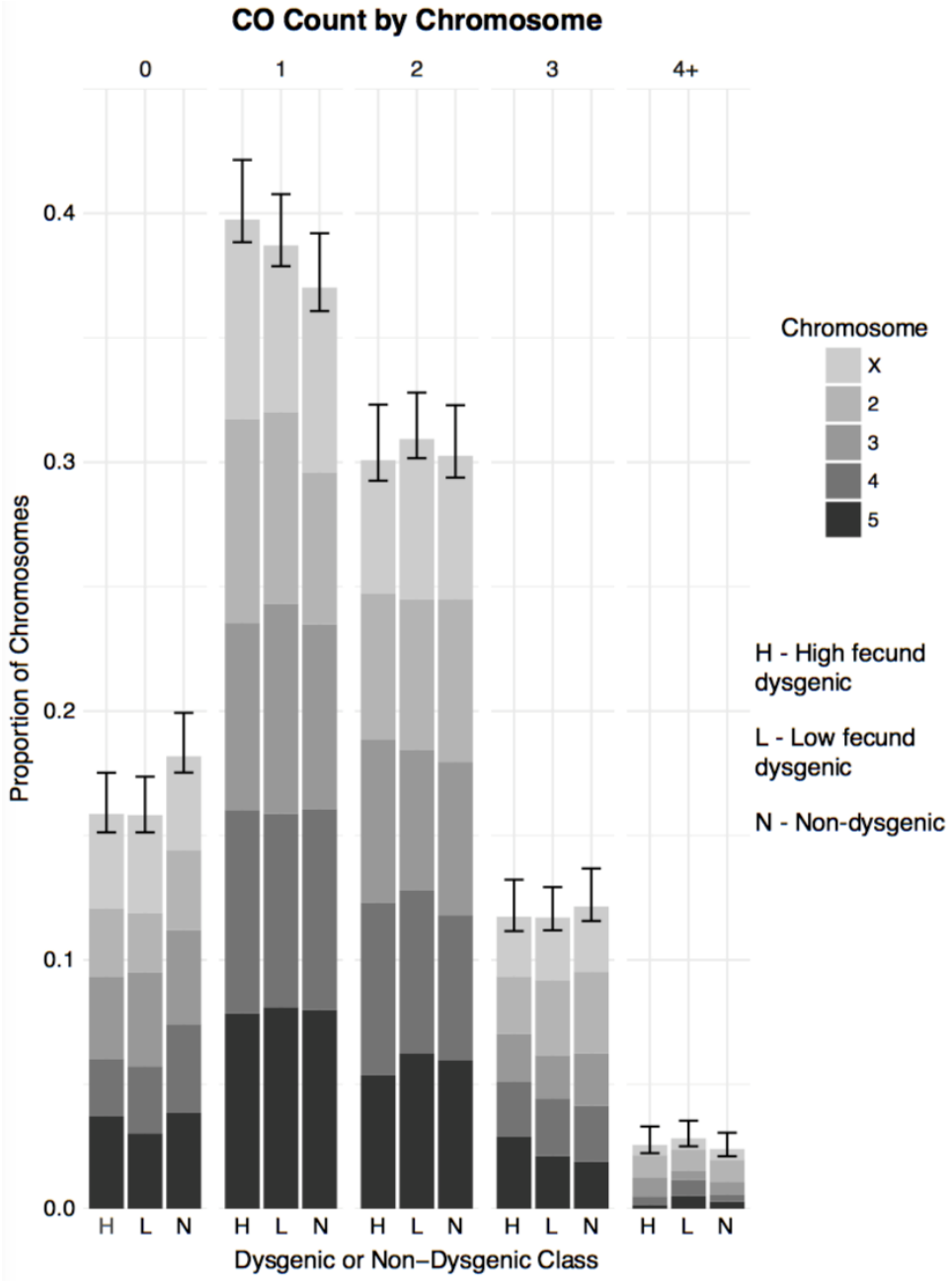

Figure 3.3: Loess smoothed splines of the recombination rate along the length of each chromosome in D. virilis from the telomere (left) to the centromere (right) with standard error. The dotted line represents the centromere effect of recombination suppression as recombination = 0 from the line to the end of the sequence. The rate of recombination was calculated in 500 kb intervals in F2 progeny of low fecund dysgenic, high fecund and non-dysgenic F1 mothers for the A) X chromosome, B) 2nd chromosome, C) 3rd chromosome, D) 4th chromosome, and E) 5th chromosome.

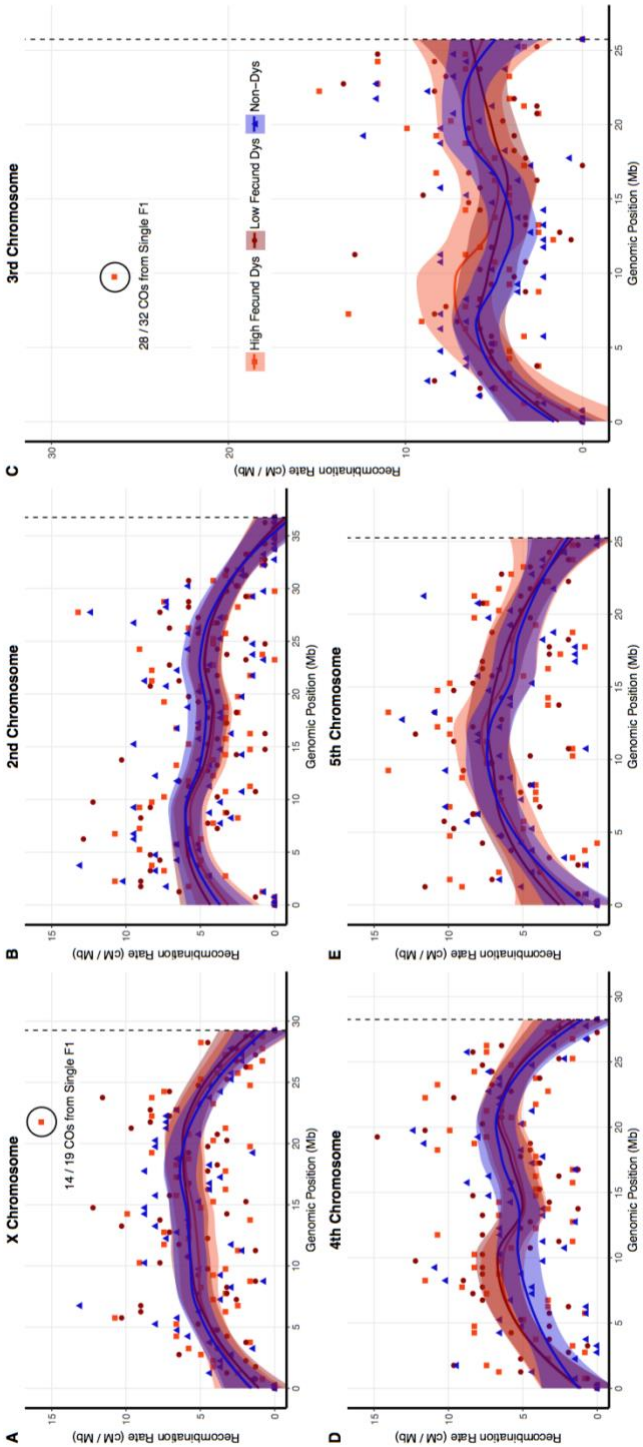

Figure 3.4: Haplotypes of F2 progeny from a single high-fecund dysgenic mother. A) Haplotypes of the third chromosome in progeny of the 4029 F1 mother is typical of most chromosomes with no cluster of recombination. B) Haplotypes of the third chromosome in progeny of the 5011 F1 mother identify a cluster of recombination in most of the progeny and reciprocal products of recombination in equal frequency (Binomial test, $p > 0.05$). C) Haplotypes of the X chromosome in progeny of the 4029 F1 mother indicate a cluster of recombination in half of the progeny and extreme segregation distortion of the distal portion of the chromosome (227 markers 0.5 - 21.4 MB, Binomial test, $p < 1.6E-08$). The proximal region of the chromosome shows no segregation distortion (86 markers 21.5 - 29.0 Mb Binomial test, $p > 0.5$).

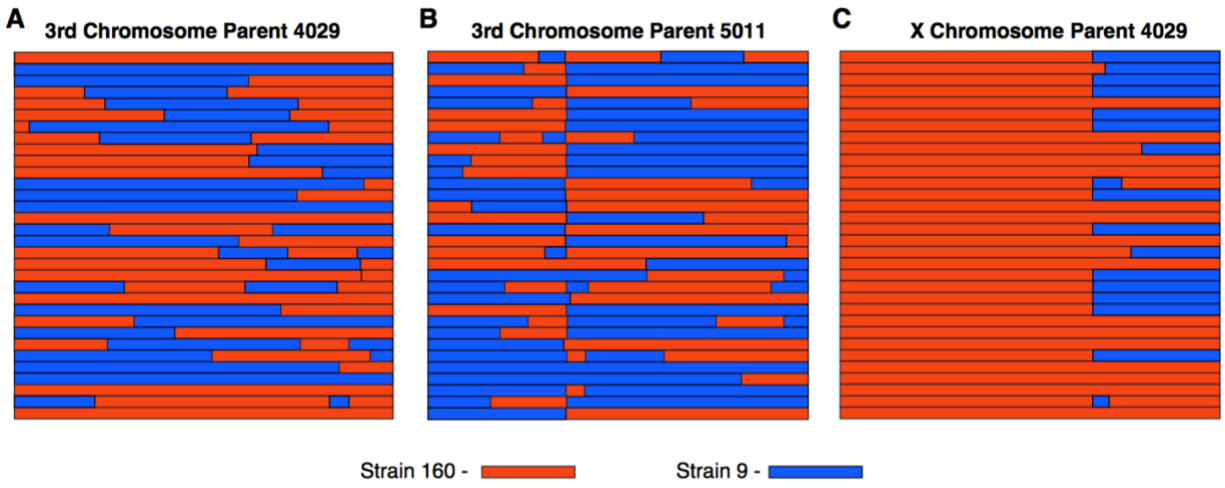

Figure 3.5: A model to explain the clusters of recombination on the third and X chromosomes in the progeny of two highly fecund dysgenic mothers. In the 5011 F1 female, a mitotic CO either occured A) prior to DNA replication in the early developing germline resulting in two daughter cells with the CO or B) after DNA replication and segregating so that one daughter cell has both CO chromatids. Oocytes produced by these germline stem cells will transmit the CO and the reciprocal products. C) A mitotic CO in the 4029 F1 female occurred after DNA replication in the developing germline and each daughter cell received one CO chromatid and one non-crossover. This results in a loss of heterozygosity in the distal portion of the chromosome to resemble segregation distortion and recombination events are not detectable.