Transposable element control disrupted by meiotic drive in a stalk-eyed fly genome

Josephine A Reinhardt[1*], Richard H. Baker[2], Aleksey V. Zimin[3], Chloe Ladias[1], Kimberly A Paczolt[4], John Werren[5], Cheryl Hayashi[2], Gerald S. Wilkinson[4]

1. Biology Department, State University of New York at Geneseo, Geneseo, New York, USA.
2. Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA
3. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA
4. Department of Biology, University of Maryland, College Park, Maryland, USA
5. University of Rochester, Rochester, New York, USA

*corresponding author
Direct correspondence to:
Josie Reinhardt
reinhardt@geneseo.edu
1 College Circle
Department of Biology
SUNY Geneseo
Geneseo, NY 14454
585-245-5413

Running title: TE control disrupted by meiotic drive

Keywords: meiotic drive, selfish genetic elements, transposable elements, stalk-eyed fly, diptera

Abstract

Some stalk-eyed flies in the genus *Teleopsis* carry selfish genetic elements that induce sex ratio (SR) meiotic drive and impact the fitness of male and female carriers. Here, we produce a chromosome-level genome assembly of the stalk-eyed fly, *T. dalmanni,* to elucidate the pattern of genomic divergence associated with the presence of drive elements. We find evidence for multiple nested inversions along the sex ratio haplotype and widespread differentiation and divergence between the inversion types along the entire X chromosome. In addition, the genome contains tens of thousands of transposable element (TE) insertions and hundreds of transcriptionally active TE families that have produced new insertions. Moreover, we find that many TE families are expressed at a significantly higher level in SR male testis, suggesting a molecular connection between these two types of selfish genetic elements in this species. We identify *T. dalmanni* orthologs of genes involved in genome defense via the piRNA pathway, including core members *maelstrom, piwi* and *Argonaute3,* that are diverging in sequence, expression or copy number between the SR and standard (ST) chromosomes, and likely influence TE regulation in flies carrying a sex ratio X chromosome.

Introduction

The genome was once thought to be little more than a blueprint needed to accomplish the biological functions and, ultimately, reproduction of an organism. Research in past decades has demonstrated that genomes of most organisms are colonized by Selfish Genetic Elements (SGEs) with their own evolutionary interests (Werren et al. 1988; Burt and Trivers 2006; Werren 2011; McLaughlin and Malik 2017). The most well-studied category of SGE - and arguably the most impactful - are transposable elements (TEs). TE's comprise at least 50% of the human genome (International Human Genome Sequencing Consortium 2001), but most are inactive. In the genome of the model dipteran *Drosophila melanogaster*, TEs make up about 20% of the genome (Hubley et al. 2016), but up to one-third actively produce new insertions (McCullers and Steiniger 2017). Transposable elements negatively impact hosts in a variety of ways, such as disrupting functional elements in the genome, causing (by their repetitive nature) ectopic recombination (Langley et al. 1988), repressing nearby gene expression (Sienski et al. 2012; Lee 2015; Lee and Karpen 2017), and increasing costs of DNA replication and storage (Badge and Brookfield 1997). Of course, like any type of mutation, TE insertions can occasionally be beneficial (Miller et al. 1997; Sinzelle et al. 2009; Jangam et al. 2017), but the overall negative impact of TEs on organismal fitness is so important that hosts have repeatedly evolved diverse genomic defenses against them (Selker 1990; Aravin et al. 2001; Muckenfuss et al. 2006; Vagin et al. 2006; Wolf and Goff 2009). Genes in the piwi-associated RNA (piRNA) pathway are a primary line of defense against TEs in the germline of Metazoa, and adapt to target active elements by incorporating transcriptionally active repetitive sequences into piRNA loci, which are then expressed and processed into piRNAs that silence their parent elements (reviewed in (Aravin et al. 2007)). Transgenic constructs carrying novel TE's can quickly become incorporated into piRNA loci (Le Thomas et al. 2013), and during the recent spread of P-elements in *Drosophila*, new insertions into piRNA loci have occurred many times in response to the new genomic threat (Zhang et al. 2020).

Another well-studied type of SGE are meiotic drivers, variously known as segregation distorters or transmission-ratio distorters. These elements spread by manipulating gametogenesis in their favor –at a cost to the homologous locus – leading to greater than 50% representation of the driver in mature gametes (Lyttle 1991). If drivers are sex-linked, a skew in the sex ratio of offspring will result. Such sex ratio distortion (SRD) may cause population collapse or even

extinction (Hamilton 1967), but can be maintained stably (reviewed in (Jaenike 2001; Lindholm et al. 2016)) if impacts on carrier fitness counterbalance the advantage of drive to the element (Curtsinger and Feldman 1980). Drive elements can also persist if their action is suppressed (eliminating SRD); however, enhancers may also evolve, potentially leading to a cyclical arms-race between enhancers and suppressors of drive (Hall 2004).

If segregation distortion occurs due to the combined effects of multiple loci, inversions will be favored to maintain linkage among them. Such drive-associated inversions are common, particularly for autosomal drivers where linkage disequilibrium between the driver and its target must be maintained. Some X-linked distorters also involve large regions of recombination suppression and differentiation relative to the standard (ST) chromosome arrangement (Wu and Beckenbach 1983; Dyer et al. 2007; Paczolt et al. 2017; Fuller et al. 2020) while others are freely recombining single loci (Montchamp-Moreau et al. 2006). Like any inversion, drive-associated inversions may recombine when homozygous, but depending on their frequency or effects on fitness, recombination may occur rarely or not at all. A reduction in the rate of recombination slows or prevents purging of new deleterious variation (Muller 1964) and may also reduce overall levels of nucleotide diversity, as the hitchhiking effects of positive or background selection are stronger when recombination rates are low (Smith and Haigh 1974; Charlesworth et al. 1993). Ultimately, like TE's, meiotic drivers persist despite typically reducing fitness of individuals carrying them, although theoretical work suggests that populations with X-linked drive could prevail during interspecific competition due to a higher reproductive rate associated with the production of excess females (Unckless and Clark 2014).

Multiple lines of evidence raise the possibility that the spread of these two categories of selfish genetic elements may not always occur independently. Most well-characterized meiotic drive systems involve regions of repetitive or satellite DNA either in the target of drive (Wu et al. 1988; Tao et al. 2007b,a; Helleu et al. 2016) or in both the driver and its target (Hurst 1996; Cocquet et al. 2012). For example, the element targeted during distortion by Segregation Distorter (SD) in *Drosophila melanogaster* (*responder*) is composed of complex arrays of 240 bp repeats, with the presence of more copies leading to stronger segregation distortion (Wu et al. 1988; Houtchens and Lyttle 2003; Larracuente 2014). In addition to TE's, piRNA may also target such satellite DNA for silencing: small RNA matching the *responder* repeats has been shown to be associated with piRNA pathway genes, *Aubergine* and *Argonaute3* (Nagao et al.

2010), and when *Aubergine* was knocked out in an SD background, the strength of distortion by the driver was enhanced (Gell and Reenan 2013), suggesting that proper silencing of the *responder* satellite DNA may act to suppress drive. The first piRNA locus to be characterized, *Suppressor of Stellate* (Aravin et al. 2001) likely first originated as a suppressor of meiotic drive (Hurst 1996), but now acts to silence TEs. Drivers and TEs, then, may share a common enemy in genome defense, and could facilitate one another's propagation at the expense of the host, similar to processes that may occur among co-infecting organismal parasites (Krasnov et al. 2005; Karvonen et al. 2009).

Here, we analyze the impacts of selfish genetic elements within the genome of a stalk-eyed fly (*Teleopsis dalmanni*). In this species 10-30% of X chromosomes actively drive against the Y chromosome and result in carrier males producing 90% or more daughters (Presgraves et al. 1997). This *sex ratio* (SR) X chromosome causes an array of positive and negative impacts on fitness (Wilkinson et al. 2006; Finnegan et al. 2019; Meade et al. 2019) including reduced sexual ornament (eyespan) size in SR males (Wilkinson et al. 1998; Johns et al. 2005; Cotton et al. 2013). Based on a three locus comparison to an outgroup species, the SR X chromosome (X^{SR}) is old, having originated approximately 500 kya (Paczolt et al. 2017), and hundreds of mostly X-linked genes are differentially expressed in SR males (Reinhardt et al. 2014). X^{SR} contains at least one large chromosomal inversion compared to the standard arrangement (X^{ST}) and likely more, as recombination has not been detected in X^{SR} / X^{ST} females (Johns et al. 2005; Paczolt et al. 2017). Although recombination occurs between X^{SR} haplotypes in homozygous females, the rate of recombination is about half that in females homozygous for X^{ST} (Paczolt et al. 2017), which has led to drastically reduced polymorphism on X^{SR} (Christianson et al. 2011).

To provide a detailed view of how the presence of drive elements have influenced patterns of divergence and patterns of genome evolution, such as TE distribution, we combined a comprehensive genomic assembly with resequencing of sex ratio and standard individuals. Using a chromosome-level assembly of *T. dalmanni* carrying X^{ST} and both short and long read resequencing, we find inversions span the X chromosome, and that *expression* of multiple TE families is higher in testes from sex ratio males, indicating less TE control in SR males. In addition, multiple genes associated with piRNA control and heterochromatin regulation are also differentially expressed, acquiring novel duplicates, or accumulating sequence differences,

including core factors involved in TE defense - *Argonaute3* (Brennecke et al. 2007) and *maelstrom* (Findley et al. 2003).

Results

The *Teleopsis dalmanni* genome contains 10x more transposable element insertions than *Drosophila*

The genome of *Teleopsis dalmanni,* a stalk-eyed fly from southeast Asia, was assembled with MaSuRCA (Zimin et al. 2013) from hybrid sequencing data containing Pacific Biosciences RSII long reads sequenced with P6C4 chemistry, and Illumina PCR-free short-read sequences (**Table S1**). The combined MaSuRCA assembly was then scaffolded using chromatin conformation information (Hi-C technology, Lieberman-Aiden et al. 2009). This assembly produced three chromosome length scaffolds with a total size of 583.1MB, comprising 93.9% of the MaSuRCA assembly. We validated this assembly by comparison to an independently generated Multiplexed Shotgun Genotyping (Andolfatto et al. 2011) linkage map produced using a backcross family from a QTL study crossing two diverged populations (Wilkinson et al. 2014). The order of X-linked scaffolds that were mapped using both methods was strongly correlated ($R^2 = 0.94$ $p < 0.0001$, **Figure S1**), providing high confidence in the integrity of the Hi-C scaffolding. We identified a 12MB region that appears to be inverted between the assembly and the MSG map, which could be due to assembly error or an inversion difference between the two populations used for the backcross. BUSCO (Seppey et al. 2019) analysis confirmed the presence of 96% of 2,799 conserved dipteran genes in the entire assembly (**Table S2**), and 95.8% within the three chromosomal scaffolds, with 9.3% of BUSCO genes duplicated. Consistent with previous analysis (Baker and Wilkinson 2010), we find extensive synteny between chromosome arms (**Figure 1A**) and the Muller elements when comparing the positions of 1-to-1 *Drosophila melanogaster* orthologs. Overall, 88.2% of genes are located on the same Muller element in the two species (**Figure 1B**). As previously reported (Baker and Wilkinson 2010; Vicoso and Bachtrog 2015), the *Teleopsis* X chromosome is orthologous to chromosome 2L in *D. melanogaster* (Muller element B). The smaller autosome, previously referred to as "C1", consists of Muller D and A (chromosomes 3L and X in *D. melanogaster*), while the larger autosome ("C2") contains Muller C, F, and E in that order (chromosomes 2R, 3R and 4 in *D. melanogaster*). We also produced a draft assembly of a closely-related undescribed cryptic

Figure 1

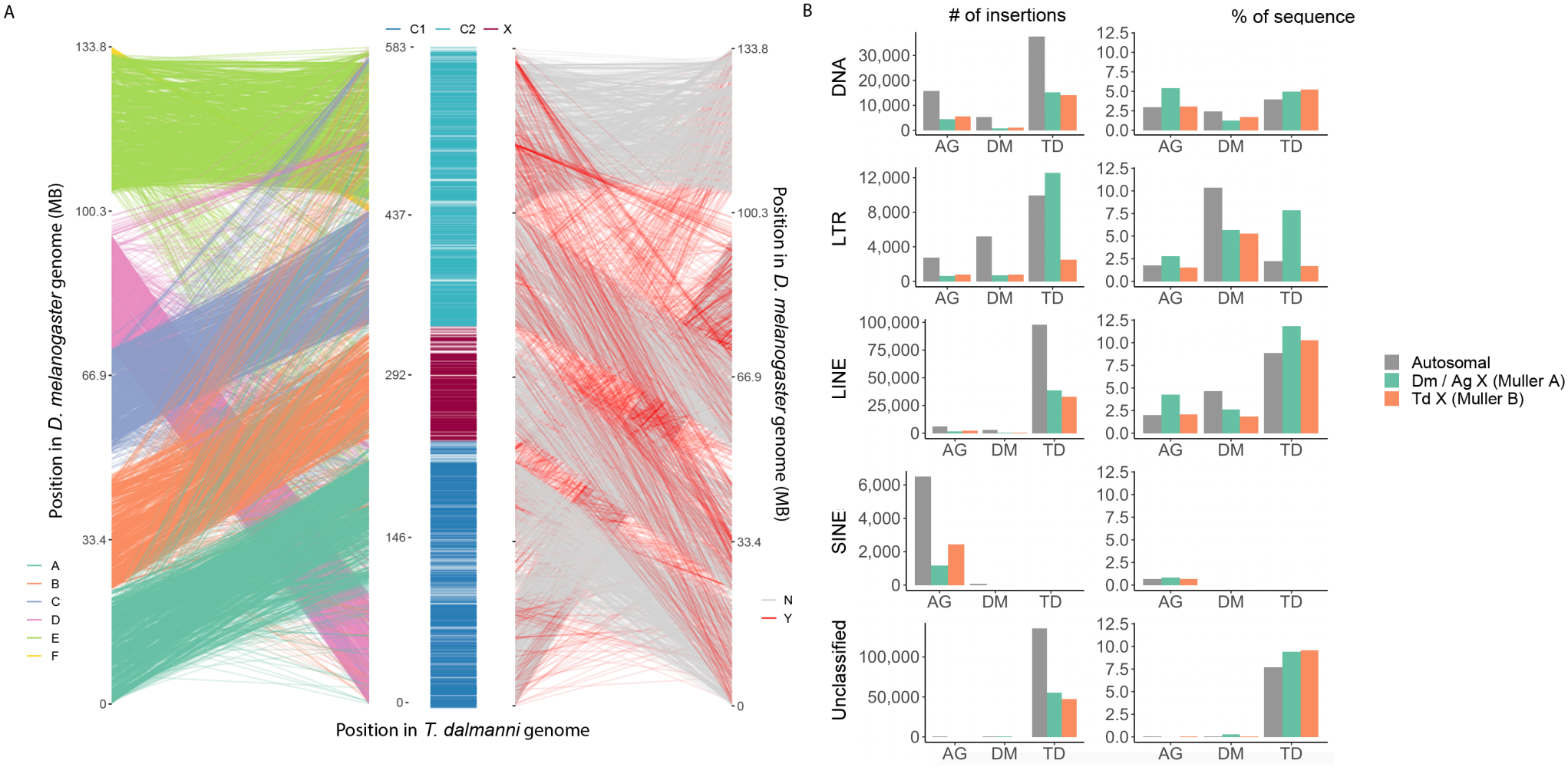
A
C1
C2
X
133.8
100.3
66.9
33.4
0
583
437
292
146
0
Position in D. melanogaster genome (MB)
Position in T. dalmanni genome
A
B
C
D
E
F
N
Y
B
# of insertions
% of sequence
DNA
LTR
LINE
SINE
Unclassified
AG
DM
TD
Autosomal
Dm / Ag X (Muller A)
Td X (Muller B)

Figure 1. Synteny and gene movement between Muller elements and transposable element diversity. (A) 1-to-1 orthologs from *D. melanogaster* were mapped to the *T. dalmanni* genome and assessed for their relative position compared to the *D. melanogaster* genome (ordered by Muller element – X, 2L, 2R, 3L, 3R). As expected, the X chromosome in *T. dalmanni* carries mostly genes from *D. melanogaster* 2L (Muller B). Orthologs on chromosome C1 correspond to genes from *D. melanogaster* chromosomes 3L and X, Chromosome C2 orthologs correspond to genes from *D. melanogaster* 2R and 3R, and chromosome 4. Overall ~88% of genes were on the same chromosome arm in *T. dalmanni* and *D. melanogaster* (grey) while about 12% were on different chromosome arms (red). (B) *T. dalmani* has thousands of insertions of LTR and nonLTR retrotransposons and Class I (DNA) elements, but no SINE elements. Transposable elements cover a slightly higher percentage of the *T. dalmanni* genome than either fruit fly (DM) or mosquito (AG); however, given the size of the *T. dalmanni* genome this represents vastly more total insertions. Muller Element A (part of chromosome C1 in *T. dalmanni*) appears to be enriched for TE's, especially LTRs

species (Paczolt et al. 2017) we refer to as *Teleopsis dalmanni sp 2* (*Td2*). This assembly contains 50,545 scaffolds and is less complete (BUSCO = 88.5%, N50 = 35,545) than the *T. dalmanni s.s.* assembly.

We identified TEs in the *T. dalmanni s.s.* genome using Repeatmodeler (Smit and Hubley 2008) followed by annotation by Repeatmasker (Smit et al. 2013) (**Figure 1B**). Compared to *Drosophila melanogaster* putative TE's cover more of the stalk-eyed fly genome (~26.1% versus ~13.5%). About 8.4% of the genome comprises unclassified interspersed elements, while the rest includes 351 families from 27 superfamilies of Class I (DNA) elements (4.4% of genome) and Class II elements including LINE (9.8%) and LTR (3.4%) elements, but no SINE elements. Given the size of the *T. dalmanni* genome, this amounts to *27-fold* as many TE insertions as in *D. melanogaster*, and 9.5-fold as many as the malaria mosquito *Anopheles gambiae*. The most abundant TE superfamilies in *T. dalmanni* include LOA, Jockey, and RTE-BovB non-LTR (LINE) elements (86552, 39370, and 26607 copies, respectively), TcMar-Mariner Class I (DNA) elements (33668 copies), and Gypsy LTR Class-II elements (19292 copies). We aligned male and female genomic resequencing data (SRS2309195-SRS2309198) to our TE families to identify any transposable elements that show a pattern of male-biased abundance and thus putative Y-linkage. A specific Penelope family element had *77-fold* more normalized sequence coverage in the male library. This LINE element is also found on the X and autosomes, so this pattern represents an expansion of male-specific (Y-linked) copies. We found an unclassified interspersed element that was similarly amplified on the Y (**Table S5**).

The sex ratio X is differentiating due to multiple overlapping inversions

Using PacBio long reads collected from *sex ratio* male siblings, we identified six large chromosomal inversions specific to the SR haplotype (**Figure 2**). These inversions (which we validated by comparison with short-read poolseq data, **Table S3**) spanned the entire 100MB chromosome except for a small region at the proximal end (0-1.76MB) and the region between 61MB and 64MB. Many of the inversions overlap, particularly near the proximal end, likely suppressing crossover between X^{SR} and X^{ST}. Alignment of the *Td2* draft genome across the inversion breakpoints revealed that the *Td2* scaffolds were consistent with the ST karyotype in five out of six cases, indicating the derived arrangement appeared and spread in the SR lineage.

Figure 2

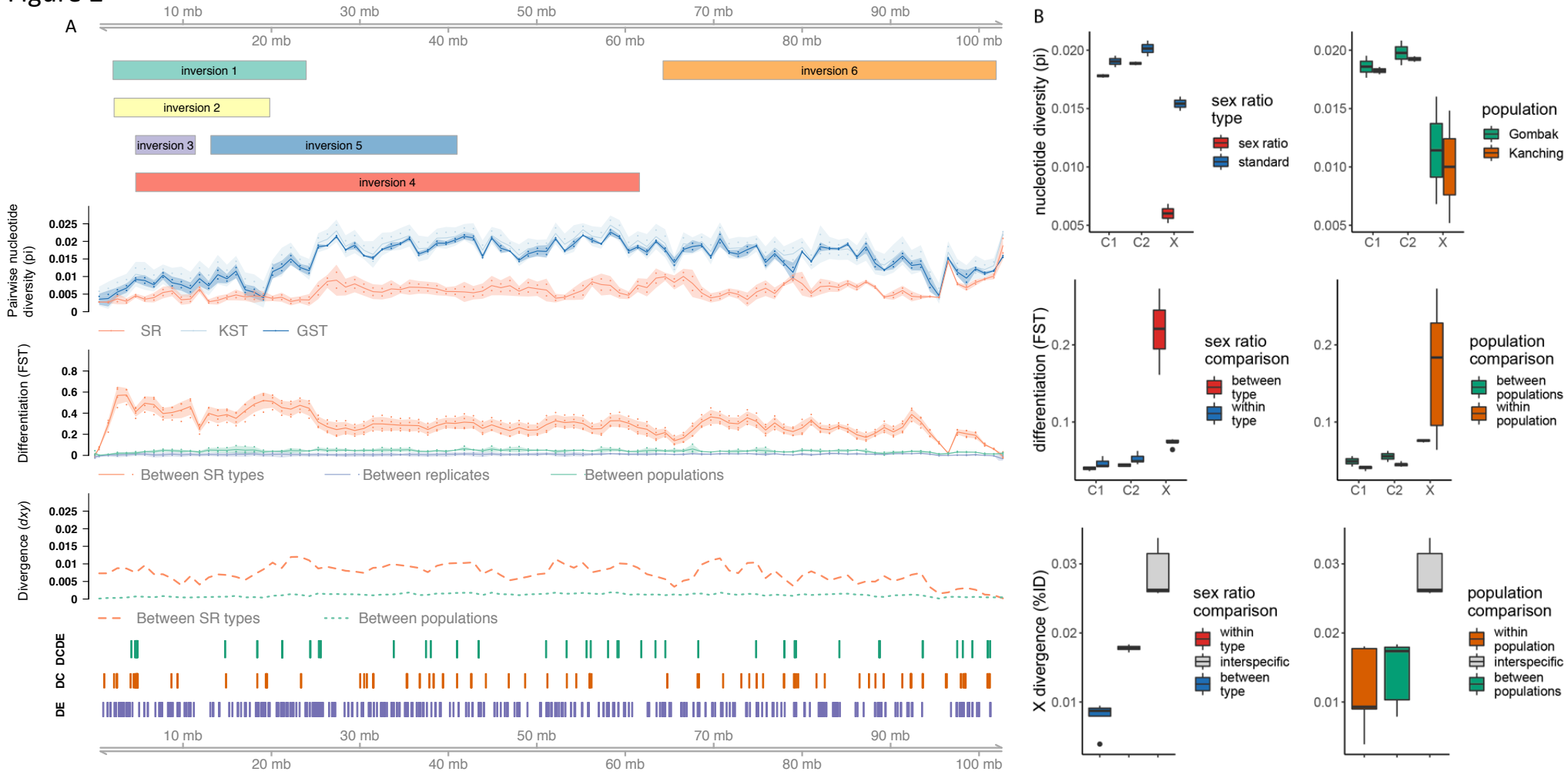
A
10 mb
20 mb
30 mb
40 mb
50 mb
60 mb
70 mb
80 mb
90 mb
100 mb
inversion 1
inversion 2
inversion 3
inversion 4
inversion 5
inversion 6
Pairwise nucleotide diversity (pi)
SR
KST
GST
Differentiation (FST)
Between SR types
Between replicates
Between populations
Divergence (dxy)
Between SR types
Between populations
DCDE
DC
DE
B
nucleotide diversity (pi)
C1
C2
X
sex ratio type
sex ratio
standard
population
Gombak
Kanching
differentiation (FST)
sex ratio comparison
between type
within type
population comparison
between populations
within population
X divergence (%ID)
sex ratio comparison
within type
interspecific
between type
population comparison
within population
interspecific
between populations

Figure 2. Six inversions (A) span the *T. dalmanni* X chromosome, leading to extensive differentiation between SR and ST chromosome haplotypes. Overall nucleotide diversity (B) is lower on the X chromosome than on autosomes, but is drastically lower in sex-ratio samples, while F_{ST} and divergence is higher in comparisons between X^{SR} and X^{ST} than in comparisons between populations but within sex ratio type. F_{ST} and divergence across the X chromosome (mean values for each comparison with 95% CI) was elevated between *sex-ratio* types (e.g. SR vs ST), but not between populations, and is lower on the proximal and distal ends of the chromosomes. Raw sequence divergence between X^{SR} and X^{ST} is elevated compared to X chromosomes of the same sex ratio type, though not as high as divergence relative to sister species *T. dalmanni sp 2*. Differentiation (FST) but not divergence (K2P) is elevated in the area of multiple overlapping inversions. Locations of genes that were found to be differentially expressed (DE) in testes between SR and ST RNAseq samples, have differential Coverage (DC) between SR and ST resequencing pools, or both (DCDE) are also shown.

Only the SR arrangement for inversion 2 (2.2 MB to 19.8MB) appears to be present in *Td2*, implying that the ST arrangement is the derived state (**Table S3**).

Using resequencing data from pools of field-derived and sex-ratio screened sex ratio and standard males (Paczolt et al. 2017) we analyzed patterns of genetic variation on the autosomes and X chromosomes. As expected, nucleotide diversity and differentiation **(Figure 2, Figure S2)** did not differ on the autosomes between pools of SR and ST males from the same collection sites, though there was minor genome-wide differentiation between the two collection sites. Genomic coverage in females was approximately twice that in males across the X, but not the autosomes, also as expected (**Figure S2**). All three chromosomes contain regions with reduced pairwise nucleotide diversity in all pools. On the two autosomes, these are near the center of the chromosome, where the Muller elements transition and so presumably are centromeres of these metacentric chromosomes. Given a similar pattern at the proximal end of the X chromosome **(Figure 2, Figure S2)** we infer this region also represents the centromere of the telocentric X chromosome. Reduced polymorphism in other fly genomes is associated with reduced rates of crossover near the centromeres (Begun and Aquadro 1992; Begun et al. 2007). X^{ST} has less nucleotide diversity (0.01547+/- 0.000053 CI) than the autosomes (0.01960+/-0.000028). However, consistent with previous findings (Christianson et al. 2011), X^{SR} has drastically lower diversity (0.00606 +/- 0.000039) **(Figure 2).**

Genetic differentiation (F_{ST}) is elevated across the X chromosome *between sex ratio* types (X^{SR} vs X^{ST}) relative to differentiation between collection sites or between replicate pools from the same collection site (**Figure 2**). Differentiation was elevated near the centromere where there is a higher density of overlapping inversions. Estimates of F_{ST} should be elevated where diversity is low and rare variants are common, as is the case near the centromere. Across the X, overall sequence similarity (%ID) of X-chromosome consensus sequences from each pool is higher between *sex ratio* types (X^{SR} *vs.* X^{ST} comparisons, 1.78%) than it is within *sex ratio* types (X^{SR} *vs.* X^{SR} and X^{ST} *vs.* X^{ST} comparisons, 0.80%), though both are lower than comparisons to the sister species *T. dalmanni sp2* (*Td2*) (2.76%)). Divergence (d_{xy} with K2P correction) in the X^{SR} lineage relative to ST (Paczolt et al. 2017) is elevated across the X chromosome, but unlike with F_{ST}, no substantial elevation in divergence occurs near the centromere (**Figure 2**). We did not detect significantly elevated divergence near inversion breakpoints (X^{ST} *vs.* X^{SR} d_{xy} = 0.0119 within 1MB of inversions, d_{xy} = 0.0118 in other regions),

as has been reported in other systems (Machado et al. 2007; Korunes and Noor 2019). Reduced divergence nearer the center of inversions is thought to occur due to gene flux between heterokaryotypes, facilitated by double crossover. However, for overlapping and nested inversions, double crossover without loss of genetic material should be extremely rare. This may explain why X^{SR} *vs* X^{ST} divergence remains elevated across the *T. dalmanni* X.

Hundreds of genes have sequence, copy number and expression differences between X^{SR} and X^{ST}

Using the poolseq data from X^{SR} and X^{ST} males described above in combination with published X^{SR} and X^{ST} RNAseq data (Reinhardt et al. 2014), we jointly analyzed patterns of genetic divergence (copy number variation and *dN/dS* between X^{SR} and X^{ST} pools) and differential expression (DE) due to *sex ratio*. We identified 652 DE genes (43.7% were more highly expressed in X^{SR}) and 185 genes with significant differential genomic coverage (DC genes, 63.2% with higher coverage on X^{SR}). Among 58 genes that were both DE and DC, there was a highly significant association in the direction of DE and differential genomic coverage (**Table S4**, FET $p < 0.0001$) with only 6 showing the opposing pattern (e.g., higher genomic coverage in SR but higher expression in ST). This result indicates differences in copy number are likely impacting expression levels – even though most DE genes (91.2%) show no difference in genomic coverage. We attempted to calculate *dN/dS* between X^{SR} and X^{ST} for 3,525 predicted X-linked protein-coding genes. Among these, 766 were identical while another 1210 had only synonymous (873) or only nonsynonymous (337) differences so *dN/dS* could not be calculated. We were able to calculate *dN/dS* between X^{SR} and X^{ST} for 1,549 genes and found 78 genes were putatively evolving under positive selection (*dN/dS* > 1) between X^{SR} and X^{ST}. *dN/dS* between X^{SR} and X^{ST} was significantly correlated (Pearson correlation = 0.466, $p < 2.2e-16$) with *dN/dS* between *Td2* and X^{ST} (**Figure S3**). Across genes for which we had all three measures (*dN/dS*, DC, and DE), we found that *dN/dS* was best predicted by a model including both DE and DC (nested within DE) (**Table 1**), with both predicting higher *dN/dS* (**Figure S3**). Genes that are DE have higher *dN/dS* than non-DE genes, and there was an additional impact (higher *dN/dS*) among genes that were also DC. *dN/dS* was not > 1 for most of these genes, so this effect could be caused by reduced selective constraint, adaptive evolution, or a combination. Among the genes that were both DE and DC (**Table S4**) were two X-linked orthologs of a gene called *Jasper* that

Table 1. Results of generalized linear model predicting *dN/dS* as a function of DE and DC

			Log-Ratio	
Term	**Estimate**	**SE**	**X^2**	**P**
Intercept	2.7429687	0.0964218	809.26659	<0.0001*
DE[0]	0.8104615	0.0964218	70.650405	4.27E-12
DE[1]:DC[0]	0.8460263	0.1622121	27.202023	7.50E-6

regulates chromatin (Albig et al. 2019), as well as two paralogous copies of *versager* and *Mcm10*, both of which are involved in chromosome condensation phenotypes. Jasper additionally appears to have undergone extensive tandem duplication on X^{SR} (**Figure S4**), and the two tandem duplicated copies of Jasper are truncated at the 5' end relative to their *Drosophila* orthologs and a full-length, but weakly expressed copy of Jasper found between them (**Figure S4**). Another DE and DC gene, *Tetraspanin29efb* a plasma membrane scaffolding protein, also has $dN/dS > 1$.

Transposable element activity is influenced by the *sex ratio* X

We found TEs have inserted more frequently on the *T. dalmanni* X than on the two autosomes (483.1 TEs/MB *vs.* 439.8 TEs/MB). However, X enrichment varies by element type (**Figure 1B**), with more DNA elements (20.6%) and LINE elements (18.3%) but fewer LTR elements (8.4%) than expected given the X comprises 17.5% of the genome (102MB / 586MB). The paucity of LTR elements on the *T. dalmanni* X is driven by an enrichment of LTR elements on *T. dalmanni* Muller element A, which is the larger arm of C1 in *T. dalmanni* and is the X chromosome in *D. melanogaster* and *A. gambiae* (**Figure 1**). We found that fifteen of 75 Class I (DNA) families have more copies (**Figure 3A**) than expected on the X chromosome ($X^2 = 11.59$, $P < 0.0001$) including 11 different TcMariner elements, while six of 180 LINE element families have fewer copies than expected on the X ($X^2 = 11.59$, $P < 0.0001$). Out of 80 LTR element families - including 32 different Gypsy elements - 40 have fewer insertions on the X than expected ($X^2 = 65.10$, $P < 0.0001$) none are more abundant, and 40 occur in expected proportions.

To investigate potential causes of TE accumulation on the X chromosome, we assessed the expression of 350 classified TE families (Lerat et al. 2016) using testis RNAseq data from SR and ST testis (Reinhardt et al. 2014). Of these 350 TE families, 48 are differentially expressed (DE) between SR and ST testis, 85% with higher expression in testes from sex ratio males compared to testes from standard males (**Figure 3).** These families include a variety of TE superfamilies from both Class I and II elements (**Table S5**) as well as unclassified families.

Using the pool-seq data we identified transposable element insertions (Kofler et al. 2016) that were not present in the genome assembly, likely representing polymorphisms enriched for relatively recent insertions. The pools contained 986 total insertions, of which 320 were from

Figure 3

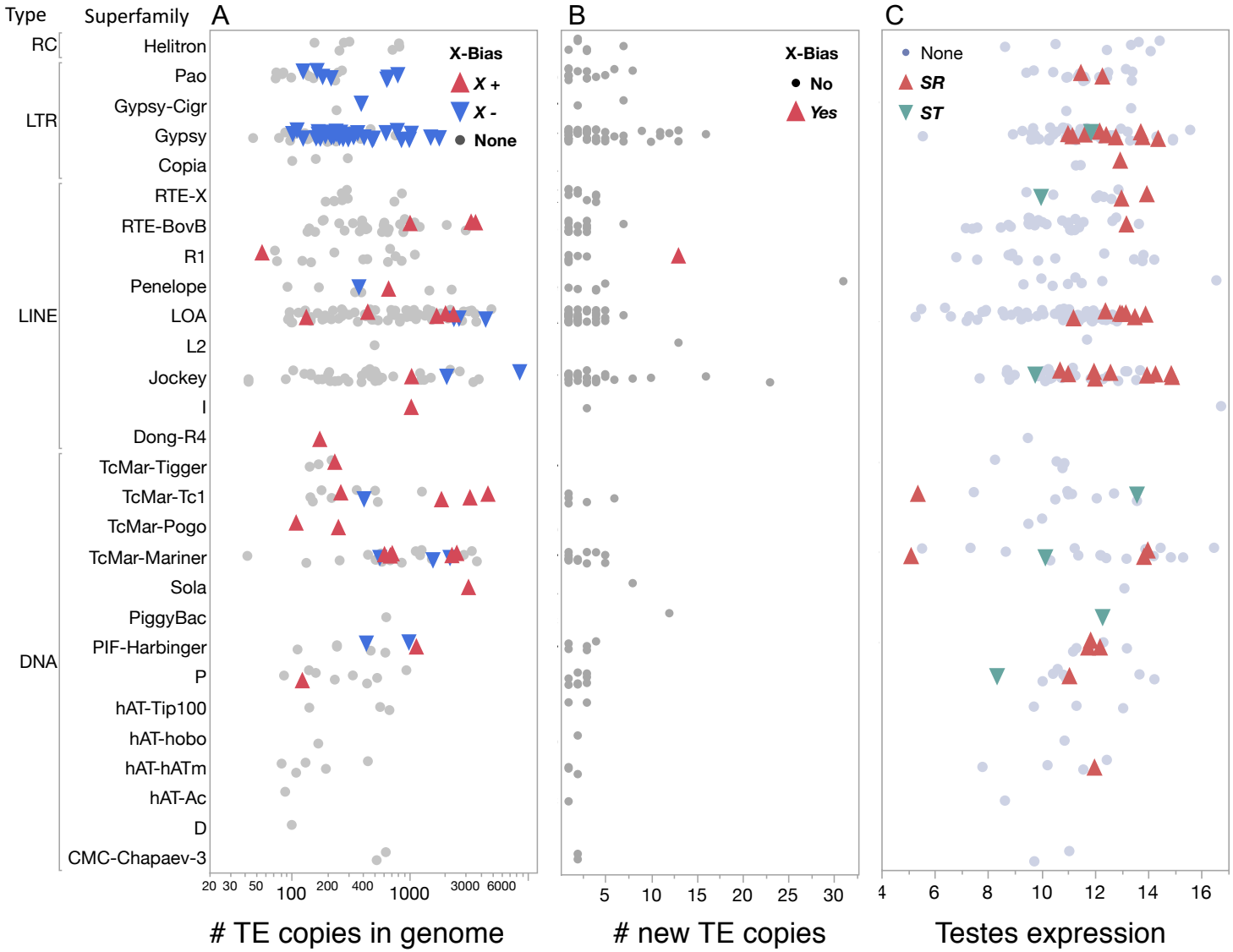
Type
Superfamily
A
B
C
RC
LTR
LINE
DNA
Helitron
Pao
Gypsy-Cigr
Gypsy
Copia
RTE-X
RTE-BovB
R1
Penelope
LOA
L2
Jockey
I
Dong-R4
TcMar-Tigger
TcMar-Tc1
TcMar-Pogo
TcMar-Mariner
Sola
PiggyBac
PIF-Harbinger
P
hAT-Tip100
hAT-hobo
hAT-hATm
hAT-Ac
D
CMC-Chapaev-3
X-Bias
X +
X -
None
X-Bias
No
Yes
None
SR
ST
20 30 50 100 200 400 1000 3000 6000
5 10 15 20 25 30
4 6 8 10 12 14 16
TE copies in genome
new TE copies
Testes expression

Figure 3. TE control is disrupted in the X^{SR} background. Transposable elements are distributed in tens of thousands of copies across the genome (A), with some families (Red triangles) over-represented on the X and others (blue down triangles) underrepresented. (B) Recent insertions are generally found at expected proportions on the X by genome size and are generally from highly expressed (C) families. Among 349 classified families, 27 were upregulated in expression in the X^{SR} background (up red triangles) and 7 were downregulated (green down triangles). The expression level of an element, genomic abundance, and the extent of differential expression between SR and ST testes all predicted the rate of insertions.

unclassified families, leaving 666 insertions that could be classified. These 666 were from 349 different transposable element families. Recent DNA element insertions were more likely to be on the X (20.2%), recent LTR insertions were less likely to be on the X (8.0%), and nonLTR retroelement insertions occurred at expected proportions (17.9%). 47.8% of all insertions on the X and autosomes are present in all six sample pools, but the number of insertions that are restricted to either X^{SR} or X^{ST} differs depending on drive status. The SR pools contain three X-linked insertions not found in ST pools – a P DNA element and an unclassified element each found in one pool and an R1 LINE element found in both SR pools. In contrast, the ST pools contain 38 X-linked insertions, including nineteen unknown, eight DNA, one Helitron, one LTR and ten non-LTR elements with just one element found in all four ST pools. Thus, X^{SR} has fewer recently inserted TEs than might be expected, even after considering the ST sample pools contained twice as many individuals in total. This implies that TE insertions are either less common or TEs are more easily lost from X^{SR} than from X^{ST}. It is worth noting that the effective population size of X^{SR} is much smaller than X^{ST} and, therefore, represents a much smaller target for mobile elements.

New insertions of TE's could be due to TE expression or to other mechanisms, such as unequal crossover or errors during replication. To determine if the number of TE insertions are influenced by expression in the germline, genomic abundance, and/or the influence of X^{SR} on expression, we used the log number of elements annotated in the genome, the $\log_2$ normalized (Love et al. 2014) testis expression of each element, and a categorical variable (SR, ST, none) to indicate presence or absence of differential expression between SR and ST testis to predict the number of novel insertions in each family, after separating by element Type (Class I – nonLTR retroelements, Class I – LTR retroelements, and Class II – DNA elements). We fit generalized regression models by maximum likelihood assuming a negative binomial distribution to predict the number of copies for each class of transposable element in all six poolseq samples (**Table 2**). We find that genome copy number predicts the number of new Class I LTR insertions and new Class II DNA element insertions. In contrast, the number of Class I non-LTR (LINE) insertions is predicted only by transposable element expression.

piRNA related genes show a variety of effects associated with the sex ratio X

Table 2. Results of generalized linear model to predict number of recent transposable element insertions

Class II (DNA) Elements, N = 79, R^2 = 0.28

Effect	**DF**	**Wald X^2**	**P**
Log Genome Copies	**1**	**5.91**	**0.015**
TE expression	1	3.09	0.079
SR/ST diff. expression	2	5.08	0.079

Class I nonLTR (LINE) Elements, N = 180, R^2 = 0.26

Effect	**DF**	**Wald X^2**	**P**
Log Genome Copies	1	0.49	0.4832
TE expression	**1**	**43.68**	**<0.0001**
SR/ST diff. expression	2	4.61	0.0996

Class I LTR Elements, N = 80, R^2 = 0.17

Effect	**DF**	**Wald X^2**	**P**
Log Genome Copies	**1**	**11.41**	**0.0007**
TE expression	1	0.01	0.9339
SR/ST diff. expression	2	2.21	0.3310

Given the observation that TE control is disrupted on X^{SR}, we were interested to know if genes associated with TE genome defense were likewise impacted. In addition to the canonical piRNA genes, *piwi*, *maelstrom, aubergine,* and *Argonaute3*, we identified an additional 168 piRNA-related genes from three recent studies in *D. melanogaster* (Handler et al. 2013; Palmer et al. 2018; Ozata et al. 2019) for further analysis. In addition, the *Heterochromatin protein 1* family is expanding within diptera in association with meiotic drive (Helleu and Levine 2018) and is associated with *de novo* heterochromatic silencing near novel TE insertions (Lee and Karpen 2017). In *Drosophila* the HP1E family includes *rhino* (HP1E) which is a key piRNA gene in *Drosophila* (Klattenhoff et al. 2009) but is a young duplicate (Helleu and Levine 2018) not found *T. dalmanni.*

We identified *T. dalmanni* ortholog(s) and new paralogs of each piRNA gene, and assessed differences in gene expression, rates of nonsynonymous change and number of X-linked duplication events between X^{SR} and X^{ST} (**Table 3**). *Maelstrom* (*mael*) is present in six copies within the stalk-eyed fly genome, four of which arose within the genus (**Figure S5**). One paralog is X-linked and testis expressed (*mael* is ovary-expressed in *D. melanogaster*). This X-linked paralog is evolving under positive selection between X^{SR} and X^{ST} (dN/dS = 1.58) and has six nonsynonymous X^{SR}- X^{ST} differences, including several within its self-named functional domain (**Figure S6**). *Argonaute3* similarly has accumulated five nonsynonymous changes between X^{SR} and X^{ST} (**Figure S7)**, and *piwi* has two nonsynonymous changes between X^{SR} and X^{ST} **(Figure S8**). Six different genes (**Table 3**) identified in screens for piRNA function were significantly differentially expressed (DE) between X^{SR} and X^{ST} – all but one were overexpressed in SR testis. Eight of these piRNA associated genes (including two that are also DE) are found in multiple copies (2-8) in *T. dalmanni* (**Table 3**). Overall, about 10% of piRNA associated genes we analyzed – including core genes – are in some way impacted by the presence of X^{SR} in this species. We also assessed positive selection for a group of core piRNA associated genes previously found to be under positive selection in other species using the McDonald-Kreitman (MK) test and found that *Hen1*, three of the five *maelstrom* paralogs, and *zucchini* showed signs of adaptive protein evolution within *T. dalmanni* (**Table S6**).

Discussion

Table 3. Impacts of X^{SR} and X^{ST} divergence on piRNA associated proteins

Gene	**Reference**	**Copies by Chromosome**	**Expression pattern**	**X^{SR}-X^{ST} Differential Expression**	**X^{SR}-X^{ST} *dN/dS***	**X^{SR}-X^{ST} NS changes**
mael	Core	1 X, 3 C1, 2 C2	2 Testes, and 4 Germline	not DE	1.58	6
Ago3	Core	X	germline	not DE	0.453	5
piwi	Core	X	ovary	not DE	0.213	3
Wdr82	Handler 2013	7 X, 4 C2	7 Testes, 2 Ovary, 2 Germline	2 X copies **up** in SR	NA	NA
papi	Palmer 2018 and Ozata 2019	X	Testes	**up** in SR	0.063	1
Hel25E	Palmer 2018 and Ozata 2019	X	Germline	2 copies **up** in SR	0.132, 0.052	1 in each copy
Lrr47	Handler 2013	X	Germline	2 copies **up** in SR	NA	0
ND-51	Handler 2013	1 X, 3 C1	Germline	1 C1 copy is **down** in SR	NA	0
sbr	Handler 2013	X	Testes	**up** in SR	NA	0

The discovery of genetic elements that persist despite negative impacts on organismal fitness required a reassessment of the traditional idea that selection acts only at the level of the individual. Indeed, such selfish genetic elements overall have divergent evolutionary interests to their hosts, despite an intimate reliance on them, making parasites their best analogs (Burt and Trivers 2006). Here, we analyze the distribution and spread of transposable elements in the context of a long-term co-existence with another selfish genetic element, an X-linked meiotic driver, *sex ratio* (SR). We find that TE's play a large role in genome evolution in *T. dalmanni*, covering ~25% of the genome and totaling a staggering 261,000 copies (plus nearly as many fragments) across the 500MB genome. We found that the presence of the X^{SR} chromosome increased expression of TEs in the genome.

Disruption of TE control in *sex ratio* males occurs in the presence of an old (~500kya) X karyotype which we find carries multiple overlapping inversions relative to the standard arrangement. The consequences for the evolution of the X chromosome are profound - the reduced gene flux between X karyotypes caused by nested inversions (Johns et al. 2005; Paczolt et al. 2017) has led to reduced polymorphism and differences in sequence and gene expression across the X (**Figure 2**), confirming previous findings (Christianson et al. 2011; Reinhardt et al. 2014). It is plausible, therefore, that differences in TE expression could be caused by genetic drift between the two chromosome types – X^{SR} has reduced polymorphism and recombination, making selection less efficient and subjecting the chromosome to effects of Muller's Ratchet (Muller 1964). If higher TE expression generally harms fitness, we might expect this bias as just an example of the accumulation of deleterious variation. Alternatively, a more active relationship between TE's and drive is plausible given repeated connections between meiotic drivers, regulation of chromatin and TE control. So far, all targets of X-linked meiotic drivers involve repetitive Y-linked sequences (Soh et al. 2014; Helleu et al. 2016). Here, we identified Y-ampliconic interspersed elements that were present in high copy number on the Y chromosome, consistent with these examples (**Table S6**). Furthermore, the first piRNA locus identified – *Supressor of Stellate* is likely a relict suppressor of drive (Hurst 1996) and studies of *Drosophila* SD suggest a role of piRNA transport misregulation in disrupting chromatin condensation in *responder*-bearing sperm (Nagao et al. 2010; Gell and Reenan 2013; Larracuente 2014).

Supporting the case for an active connection between *sex ratio* drive and the expansion of TE's, we find that several piRNA genes are accumulating non-synonymous differences between X^{SR} and X^{ST} with some evolving adaptively. We also observe that eight of nine differentially expressed genes associated with piRNA function are upregulated in X^{SR}, implying carriers may be reacting to TE release by upregulating piRNA functions. A missing piece here is the expression of the piRNA themselves, as well as potential divergence between putative piRNA clusters on X^{ST} and X^{SR}. Given that piRNA clusters are formed via new mutations (Zhang et al. 2020), if present on the X, they should accumulate different TE insertions over time due to the lack of recombination between X^{ST} and X^{SR} and so target different TE's.

Prior work has shown that some piRNA pathway proteins, like immune genes, evolve rapidly (Obbard et al. 2009; Kolaczkowski et al. 2011a; Lee and Langley 2012; Simkin et al. 2013; Palmer et al. 2018), though the precise mechanism underlying this rapid evolution is unclear (reviewed in (Blumenstiel et al. 2016), but see (Parhad and Theurkauf 2019)). We, too, find that some of these genes (*zucchini, hen1*, and several *maelstrom* orthologs) show signs of evolving under positive selection within *T. dalmanni* (**Table S6**). Furthermore, we found that X^{SR} *alleles* of key X-linked TE control genes *piwi*, *Argonaute3*, and an X-linked copy of *maelstrom* have accumulated nonsynonymous differences compared to the allele on X^{ST} (**Figure S5-S7**), with *maelstrom* showing evidence of positive selection between X^{SR} and X^{ST} (dN/dS = 1.5). In *Drosophila,* the evidence for ongoing positive selection for *maelstrom, piwi* and *Ago3* is mixed, with signs of positive selection in a minority of comparisons, while other genes like *Aubergine, spn-E,* and *armitage* are more consistently found to be under positive selection (Blumenstiel et al. 2016). *Maelstrom* and *aubergine* have also duplicated within diopsids. We previously (Baker et al. 2016) identified five duplicates of *maelstrom*, and here found an additional near-identical copy. Four of these six copies, including the adaptively evolving X-linked copy appear to be novel within the genus *Teleopsis* (**Figure S8**). During piRNA mediated silencing in *Drosophila*, *piwi* is responsible for establishing a heterochromatic histone marks (Histone 3 Lysine 9 methylation) whereas *maelstrom* promotes the spreading of heterochromatic states to nearby genes (Sienski et al. 2012).

In *Drosophila simulans*, a young duplicate of heterochromatin protein 1 (HP1D) causes *sex ratio* meiotic drive (Montchamp-Moreau et al. 2006; Helleu et al. 2016), and HP1D (also known as *rhino*) is a key player in piRNA genome defense in *D. melanogaster* and presumably

other species (Klattenhoff et al. 2009). Furthermore, recurrent amplification of HP1 copies occurs in lineages where *sex ratio* drive is common (Helleu and Levine 2018). We identified multiple HP1 family members in the *T. dalmanni* genome, and one (the HP1E ortholog) was X-linked, but it did not show signs of SR-associated divergence, differential expression or coverage. We did, however, identify X^{SR}-specific amplification of a gene recently found to be associated with the silencing of heterochromatin. *Jasper* is among the few genes exhibiting strong differential expression and differential copy number between SR and ST males (**Table S4),** and there has been X^{SR}-specific tandem amplification of the PWWP domain of the protein (**Figure S4B**). In *Drosophila* the PWWP domain binds to methylated nucleosomes during heterochromatin silencing by its partner JIL-1 (Albig et al. 2019). Given the tripartite dance between TE control, heterochromatin silencing, and meiotic drive, this will be an intriguing candidate for further analysis.

Finally, meiotic drive has been theorized (Frank 1991; Hurst and Pomiankowski 1991) to be a possible "engine of speciation", by causing rapid evolution of suppressors and enhancers that influence sperm development, though empirical support is mixed (McDermott and Noor 2010; Meiklejohn et al. 2018). Multiple studies have shown that misregulation of heterochromatin is a major underlying cause of hybrid incompatibility in flies (Brideau et al. 2006; Kelleher et al. 2012), and hybrid incompatibility in *D. simulans* and *D. melanogaster* has been linked to a Dobzhansky-Muller incompatibility between interspecific alleles of piRNA pathway gene *rhino* - alleles of which can also cause drive (Helleu et al. 2016) - and another piRNA gene *deadlock* (Parhad et al. 2017). X^{SR} causes many fitness effects, and X-linked loci in *T. dalmanni* have been shown to contribute both to a cryptic meiotic drive and to sperm defects (Wilkinson et al. 2014). Interspecific hybrids often show a series of defects referred to as hybrid dysgenesis, which typically includes misregulation of heterochromatin thought to be caused by a mismatch between repetitive DNA and the machinery regulating it. Given the extensive divergence between X^{SR} and X^{ST} it is possible that the TE misregulation we observe in *sex ratio* males could cause incompatibilities between X^{SR} and autosomal loci. We might also expect X-linked incompatabilities in heterokaryotypic females. Consistent with this idea, X^{SR} causes reduced viability (hatch rate) in embryos of SR males as well as female carriers (Finnegan et al. 2019), although adult X^{SR} / X^{ST} females are slightly more fecund than either homokaryotype (Wilkinson et al. 2006). TE expression in females with different sex ratio karyotypes has not yet

been measured, but could provide valuable insights into evolutionary dynamics of multiple selfish genetic elements.

Material and Methods

Genome assembly of *Teleopsis dalmanni s.s.*

A draft genome assembly for *Teleopsis dalmanni*, NLCU01 was created with a combination of Roche 454, Illumina and PacBio sequence data (**Table S1**) using MaSuRCA (Zimin et al. 2013) and is available on Ensembl genomes (Kersey et al. 2018). This assembly contains ~25k scaffolds with an N50 of ~75k. All DNA sequences were obtained from an inbred population (line "2A") of *Teleopsis dalmanni*. This population is derived from flies that were first collected near the Gombak River in peninsular Malaysia (3 12'N, 101 42'E) in 1989 and then maintained as a control line for an artificial selection study on relative eyespan (Wilkinson 1993; Wolfenbarger, L. L. & Wilkinson 2001). After 50 generations of selection, full-sib mating was conducted for seven generations to establish the line, which has subsequently been maintained without additional inbreeding. This population has been used in several prior studies (Christianson et al. 2005; Wilkinson et al. 2014) and does not carry any drive-associated genetic markers. Contaminating bacterial scaffolds were identified and removed prior to submission using a modification of the Wheeler et al. (2013) DNA-based homology pipeline, as described in Poynton et al. (2018). Male and female genomic short-read resequencing data was aligned to each scaffold using nextgenmap (Sedlazeck et al. 2013) with default parameters, and relative coverage of male and female reads was used to identify X-linked scaffolds (cf. (Vicoso and Bachtrog 2015), with the expectation that the normalized ratio of female to male reads should be approximately 1 to 1 for autosomal and 2 to 1 for X-linked scaffolds.

A chromosome-level assembly was then created by incorporating chromatin conformation information and validated with a high-density linkage map. Chromatin conformation capture data was generated using a Phase Genomics (Seattle, WA) Proximo Hi-C Plant Kit, which is a commercially available version of the Hi-C protocol (Lieberman-Aiden et al. 2009). Following the manufacturer's instructions, intact cells from unsexed pupae from the 2A inbred line were crosslinked using a formaldehyde solution, digested using the Sau3AI restriction enzyme, and proximity-ligated with biotinylated nucleotides to create chimeric molecules composed of fragments from different regions of the genome that were physically

proximal in vivo, but not necessarily genomically proximal. Continuing with the manufacturer's protocol, molecules were pulled down with streptavidin beads and processed into an Illumina-compatible sequencing library. Sequencing was performed on an Illumina HiSeq 4000, generating a total of 202,608,856 100 bp read pairs. Reads were aligned to the draft assembly (NLCU01.30_45_breaks.fasta) following the manufacturer's recommendations. Briefly, reads were aligned using BWA-MEM (Li and Durbin 2009) with the -5SP and -t 8 options specified, and all other options default. SAMBLASTER (Faust and Hall 2014) was used to flag PCR duplicates, which were later excluded from analysis. Alignments were then filtered with samtools (Li et al. 2009) using the -F 2304 filtering flag to remove non-primary and secondary alignments. Putative misjoined contigs were broken using Juicebox (Rao et al. 2014; Durand et al. 2016) based on the Hi-C alignments. A total of 113 breaks in 105 contigs were introduced, and the same alignment procedure was repeated from the beginning on the resulting corrected assembly. The Phase Genomics' Proximo Hi-C genome scaffolding platform was used to create chromosome-scale scaffolds from the corrected assembly as described (Bickhart et al. 2017). As in the LACHESIS method (Burton et al. 2013), this process computes a contact frequency matrix from the aligned Hi-C read pairs, normalized by the number of Sau3AI restriction sites (GATC) on each contig, and constructs scaffolds in such a way as to optimize expected contact frequency and other statistical patterns in Hi-C data. In addition to Hi-C data, chromosomal linkage information (see below) was used as input to the scaffolding process. Linkage groups from a linkage map were used to constrain chromosome assignment during the clustering phase of Proximo by discarding any suggested clustering steps that would incorporate contigs from different linkage groups onto the same chromosome, but linkage map data were not used during subsequent ordering and orientation analyses in Proximo. Approximately 528,000 separate Proximo runs were performed to optimize the number of scaffolds and scaffold construction in order to make the scaffolds as concordant with the observed Hi-C data as possible. This process resulted in a set of three chromosome-scale scaffolds containing a total of 581.7 Mbp of sequence (93.6% of the corrected assembly). Finally, Juicebox was again used to correct remaining scaffolding errors.

Linkage groups used to constrain and validate the Hi-C assembly were created by mating a female hybrid offspring obtained from a cross between a male from the 2A inbred strain and a female from a noninbred population of *T. dalmanni* collected near Bukit Lawang, Sumatra (3

35'N, 98 6'E) to a male from the 2A strain. This backcross produced 249 (131 female and 118 male) individuals which were individually genotyped using multiplex shotgun sequencing (Andolfatto et al. 2011) and multiple STR loci (Wilkinson et al. 2014). Genotypes were determined as either heterozygous or homozygous in each individual for each scaffold by combining all loci present on a scaffold into a single "super locus". Reads were aligned using bwa (Li and Durbin 2009) and genotypes were assessed as either homozygous or heterozygous using samtools v.1.9 (Li et al. 2009) (mpileup -v). Because this was a backcross, for autosomal loci those individuals with the backcross allele (pure "2A") should be homozygous at informative markers whereas individuals with the non-backcross allele should be heterozygous, with an expectation of a 1 to 1 ratio of these genotypes. This results in an overall 3 to 1 ratio of the backcross to the non-backcross allele for autosomal markers and X linked markers in females, and an overall 1 to 1 ratio for X linked markers in hemizygous males. Markers were retained as potentially informative if at least one individual was found to carry the non-backcross Bukit Lawang (Wilkinson et al. 2014) allele, defined as the less common allele across all female individuals. Markers were removed if they violated expected allele ratios using a binomial test against the expectations described above, or if more than 20% of female individuals were found to carry **only** the non-backcross allele. Finally, within each individual all markers from a given scaffold were pooled to give an overall number of reads supporting each genotype, and requiring a minimum coverage of 5 reads per marker. Individuals were assigned in the final matrices as "a" (for 2A/backcross genotypes) or "b" (for Bukit Lawang/foreign genotypes).

Separate genotype matrices were then created for the X-linked scaffolds (as determined by male and female coverage (Vicoso and Bachtrog 2015)) and autosomal scaffolds and rank-ordered by the number of individuals genotyped. We then used JoinMap (Stam 1993) v4.1 to assign the top 1000 autosomal scaffolds into one of two linkage groups (chromosomes). Only scaffolds with a LOD score > 5 were assigned to a chromosome. We used a similar process for the top 250 X-linked scaffolds but used a LOD score > 10 to assign scaffolds to the chromosome. These linkage groups were used to constrain the Hi-C genome assembly as noted above. Then, independent from the Hi-C scaffolding process, we ordered scaffolds within each linkage group by regression mapping using a Haldane mapping function (Rédei 2008). We used regression mapping, rather than maximum likelihood, because it is less sensitive to missing genotype data (Van Ooijen 2006) as is typical for MSG studies. We removed markers from the

final map if there was evidence for significant (after Bonferroni correction) lack of fit to their nearest neighbors. The resulting linkage map included 762 scaffolds spanning 147.1 Mbp. Collinearity between the Hi-C and linkage maps was assessed by comparing the relative location of scaffolds which were found in both maps. Chromosomal synteny of the assembly scaffolds with *Drosophila melanogaster* chromosome arms was assessed by alignment of a set of 8098 previously annotated 1-to-1 *Drosophila* orthologs (Baker et al. 2016) to the genome assembly using GMAP (Wu and Watanabe 2005) (--npaths=1 --format=gff3_gene --min-identity=0.9). Assemblies were assessed for completeness using BUSCO (Seppey et al. 2019).

Draft Genome assembly of *T. dalmanni* species 2

We recently described a cryptic species of stalk-eyed fly (Paczolt et al. 2017), which we refer to as *T. dalmanni sp 2* (Td2) and which corresponds to prior collections of *T. dalmanni* from several sites in peninsular Malaysia, such as Cameron Highlands (Christianson et al. 2005; Swallow et al. 2005). To produce a draft genome of this species, we extracted HMW DNA from a single male from a laboratory population of Td2 using the Gentra Puregene tissue kit (Qiagen cat 158667). 1 ug of DNA was sent to the New York Genome Center (NYGC) where it was prepped with the Chromium Genome linked read kit (10X Genomics) and sequenced on a half lane of an Illumina HiSeqX machine, producing a total of 416 million reads. These reads were assembled at NYGC using Supernova (v2.0.1). The resultant draft genome contained 10,290 scaffolds greater than 10 kb with a N50 of 45.2 kb and total genome size of 355 MB.

Short-read resequencing of sex ratio (SR) and standard (ST) males

To identify sequence and structural variations specific to X^{SR}, we sequenced DNA from replicate pools of SR or ST males either collected in the field from two different sites in peninsular Malaysia or representing the first three generations of sons descended from field-collected females. One SR and two ST sample pools were created from the DNA of males from each of two collection sites (Gombak and Kanching) previously phenotyped for offspring sex ratio (Paczolt et al. 2017). When an excess of individuals was available from a collection and *sex ratio* category, genotype data from 9 X-linked STR loci (Paczolt et al. 2017) was used to avoid over-sampling closely related individuals (e.g. male full-siblings from the same brood). Nevertheless, haplotype diversity within pools was not significantly different to diversity among all candidate

males for that pool (following Christianson et al 2011, Dunnett's t-test, $p>0.05$ for all comparisons, **Table S7**). DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and quantified using PicoGreen Quant-IT dsDNA quantification kit (Thermofisher Q33130). Pools were then assembled using an equimolar amount of total DNA from each sample. Sample size for each pool ranged from 15-18 individuals (**Table S7**). Six barcoded libraries were prepared and multiplexed on two lanes of a HiSeq1500 set to RapidRun mode to generate 150 bp paired-end sequences. Bam-formatted alignments of these libraries to the genome were produced using nextgenmap (Sedlazeck et al. 2013) with default parameters and used in subsequent analyses. Pairwise genetic diversity for each pool was calculated from the pooled resequencing alignments in 5kb non-sliding windows using popoolation (Kofler et al. 2011) and F_{ST} was estimated following (Kolaczkowski et al. 2011b). After poolseq had been completed, it was determined that four pooled ST individuals were actually *Td2* males (genetic markers distinguishing these species had not been identified until after pooling, see **Table S7**, (Paczolt et al. 2017), and previous work had suggested that *Td2* would not be present in the collection sites we visited (Christianson et al. 2005; Swallow et al. 2005)). Both SR pools and one ST pool (Kanching ST2) were entirely composed of *T. dalmanni* s.s. (**Table S7**) so all analyses where excess polymorphism (actually species divergence from the *Td2* individuals) within a pool could impact the results of analysis were repeated using only these three samples, and results of the reduced analysis were found to be qualitatively similar to the full analysis (**Figure S9** and **Table S8**).

Long-read resequencing of a SR haplotype

To identify inversion breakpoints between X^{SR} and X^{ST}, we used long-read sequencing (Pacific Biosciences). A pool of full-sib males bearing a single identical-by-descent (IBD) X^{SR} haplotype was created by mating a SR/SR female to a male from the 2A strain and then backcrossing the female progeny to males from the same strain. We then genotyped 107 sons from this backcross at three X-linked STR loci (ms125, ms395, and CRC) in order to distinguish X^{SR} sons from X^{ST} sons. DNA from 46 X^{SR} sons was then extracted using the Gentra PureGene Tissue Kit (Qiagen) and pooled, followed by a phenol-chloroform extraction and ethanol precipitation. A PacBio long insert (15 Kb) library was then prepared and run on three PacBio Sequel SMRT cells. These runs yielded a total of 15.1 Gb of sequence, with mean read length of 4.9 Kb and maximum read

length of 92.8 Kb. Raw long reads were aligned to the genome using ngmlr with default parameters and structural variants were called using sniffles (Sedlazeck et al. 2018) requiring at least 2 reads to support each variant call. The resulting output was filtered to find inversions that were fixed within the SR PacBio long reads relative to the reference genome. Each putative inversion was then validated as a fixed SR specific inversion by comparison to the read-pair orientation in the SR and ST poolseq data at the breakpoint using IGV (Thorvaldsdottir et al. 2013). An inversion was considered validated if reads from all of the SR poolseq samples but none of the ST poolseq samples agreed with the sniffles call at that position (**Table S3**). Finally, to polarize the direction of the inversion mutation, an alignment of the *T. dalmanni sp 2* genome assembly was performed using blat (Kent 2002) and scaffold alignments near the breakpoints were examined to determine if they 1) support the standard arrangement (span the breakpoint), 2) support the SR arrangement (scaffold breaks and aligns to other end of breakpoint), 3) scaffold present near breakpoint but supports another arrangement, or 4) no scaffolds mapping near the breakpoint (**Table S3**).

Transposable element annotation

Transposable elements were annotated in the *T. dalmanni ss* assembly using RepeatModeler v. 1.0.4 (Smit and Hubley 2008) with default parameters and the NLCU01 assembly as input. Resulting consensus fasta formatted TE sequences were input into RepeatMasker (Smit et al. 2013) with the Hi-C assembly as the reference, producing a repeat-masked reference genome and repeat annotations. The tool "One code to find them all" (Bailly-Bechet et al. 2014) was used with the RepeatMasker output to count the numbers and locations of each type of insertion in the *T. dalmanni* genome. These were compared to the RepeatMasker annotations for two other dipterans (*Drosophila melanogaster* dm6 RepeatMasker open-4.0.6 and *Anopheles gambiae* anoGam1 RepeatMasker open-4.0.5) analyzed using the same procedure.

Novel, polymorphic insertions of TE's were called in each of the six Illumina resequencing pools using PopoolationTE2 (Kofler et al. 2016) running the "separate" analysis mode on the repeat masked assembly and TE sequences. To compare the rate of insertions of TE's between the samples, insertions were inferred to be orthologous if they were an insertion of the same element within 500 bp in multiple pools.

The expression of RepeatModeler TE families in SR and ST males was assessed using TEtools (Lerat et al. 2016), using the default settings and including alignment with bowtie2 (Langmead and Salzberg 2012) using the RepeatModeler TE library and RNAseq reads from two pools of SR male testis and two pools of ST male testis (Reinhardt et al. 2014) (BioProject PRJNA240197). Unannotated repetitive elements ("Unknown" interspersed and simple repeats) were removed after normalization but prior to differential expression analysis with DESeq2 (Love et al. 2014). Differential TE expression between the SR and ST pools was assessed using the negative binomial Wald test on the DESeq-normalized counts for each TE family. TEtools was also used to estimate TE family copy number within the SR and ST genomic resequencing pools and in male and female genomic libraries. DESeq2 was used to normalize read counts prior to comparisons across libraries.

Annotation of gene duplication and expression, and differential coverage

A set of protein-coding genes annotated from a transcriptome assembly (BioProject PRJNA240197) were aligned to the three largest scaffolds using GMAP allowing for up to 10 gene alignments ("paths") per gene (--npaths=10 --format=gff3_gene). Annotations were removed as potential TEs misannotated as genes if they had >50% alignment overlap with any TE annotation from Repmasker. Bedtools (Quinlan and Hall 2010) (intersect -wao) was used to determine the number of bases of overlap for each exon, then the proportion of overlapping bases was calculated across the entire length of each gene alignment ("path" in gmap terminology). RNAseq data from SR and ST male testis (Reinhardt et al. 2014) was aligned to the genome using Hisat2 (Kim et al. 2019) v2.0.1-beta (--dta -X 800) and expression of each gene feature – including newly discovered duplicates found by GMAP – in each library was estimated using featurecounts (Liao et al. 2014). Differential coverage and differential expression were each assessed on a by-feature basis using DESeq2 (Love et al. 2014) with the default Wald test on the negative binomial distribution.

Molecular evolutionary analyses

Divergence was assessed in comparison to the *Teleopsis dalmanni sp 2* (Paczolt et al. 2017) genome described above. The *Td2* draft genome scaffolds were aligned to the three largest (chromosomal) scaffolds in the Hi-C assembly for *T. dalmanni s.s.* using GMAP (Wu and

Watanabe 2005) (--nosplicing --format=samse). A *Td2* consensus was called from the gmap alignment of *Td2* scaffolds to the *Teleopsis dalmanni* genome by sorting and indexing with samtools (Li et al. 2009) v. 1.3.1, then calling the consensus with Bcftools (Li 2011) v 1.9 (bcftools call --ploidy 1 -mA). Regions which did not have an aligned scaffold or align as gaps show up as stretches of "N's" in the consensus when these parameters are used. The bam-formatted pool-seq library alignments were used with bcftools (bcftools call --ploidy 1 -c ; vcfutils.pl vcf2fq), to create a majority-rule X^{SR} consensus sequence for the large X-chromosomal scaffold (PGA_scaffold1) from a bam file combining both X^{SR} pools into a single bam file. In addition, to have a comparable (similar sequencing and allelic coverage) X^{ST} consensus, an X^{ST} alignment (bam) file was produced using 1 replicate from each collection site and consensus called as above. The best alignment of the coding regions of 3,709 X-linked genes previously (Baker et al. 2016) assembled and annotated using data from a multi-tissue RNAseq experiment (BioProject PRJNA240197) were localized to the X-chromosomal scaffold via alignment with GMAP (--npaths=1 --format=gff3_gene --min-identity=0.9). Gene sequences were extracted from the X^{SR} and X^{ST} consensus X chromosomes described above using the gffread utility (Pertea and Pertea 2020) and the mRNA gff annotations from GMAP. A few hundred genes contained in-frame stop codons in one or more lineages - these genes were trimmed to the longest open reading frame present and if what remained was longer than 50 amino acids, were retained. Genes were excluded if they contained only ambiguity sequence ("N") in one or more of the consensus genomes. After exclusions, we counted nonsynonymous divergent sites and calculated pairwise *dN/dS* of X^{SR} vs X^{ST} and for X^{ST} vs Td2 for 3,191 X-linked genes using the SNAP utility (Korber et al. 2000). Pairwise *dN/dS* (X^{ST} vs Td2) was also calculated and the McDonald-Kreitman test (Td1 vs Td2) was performed on a set of 17 core piRNA genes (**Table S6**).

A 13 "taxon" unphased alignment was also produced from the poolseq data. A major and minor allele was called for polymorphic sites from each of the 6 pools and represented as two sequences in the alignment, giving 13 total with the *Td2* outgroup. The 13 taxon alignment was used to identify segregating vs. divergent (relative to *Td2*) sites in each pool across the genome. D_{xy} (K2P corrected) was calculated between either the X^{SR} and X^{ST} chromosome (between SR type), or between the two collection sites (across SR type) in 10kb windows using poly+div_sfs.pl.

Data Access

Raw data and genome assemblies used in this project are available on NCBI BioProjects PRJNA655584, PRJNA391339, PRJNA662429 and PRJNA659474. Transposable element family consensus fasta sequences for *Teleopsis dalmanni*, the msg genotype matrix and gff3 formatted gene annotations are available on the Digital Repository at University of Maryland (DRUM), archive ID's 1903/26380, fgxn-tuaf , and gfqi-iktk

Acknowledgements

The authors thank Melanie Kirk, Nathaniel Lowe, Wyatt Shell, George Ru, and Gabriel Welsh for assistance with analysis, sample preparation, and fly rearing; Philip Johns and Max Brown for fly collections; Shawn Sullivan for HiC analysis; Najib El-Sayed and Suwei Zhao for HiSeq library prep and sequencing assistance; Ellen Martinson for bacterial contamination screening; and Molly Schumer, Peter Andolffato, and Wei Wang for reagents and advice on multiplexed shotgun genotyping (MSG). Funding for this work was provided by National Science Foundation grants DEB-0951816 to R.H.B., DEB-0952260 to G.S.W., by the USDA National Institute of Food and Agriculture grant 2018-67015-28199 to A.V.Z, and by the University of Maryland and the Geneseo Foundation.

Disclosure Declaration

Nothing to disclose

Author Contributions

JR, RB, and GW prepared the manuscript. JR, AZ, KP, CL, JW, GW, and RB analyzed the data. CH, GW provided sequencing data.

References Cited:

Albig, C., C. Wang, G. P. Dann, F. Wojcik, T. Schauer, S. Krause, S. Maenner, W. Cai, Y. Li, J. Girton, T. W. Muir, J. Johansen, K. M. Johansen, P. B. Becker, and C. Regnard. 2019.

JASPer controls interphase histone H3S10 phosphorylation by chromosomal kinase JIL-1 in *Drosophila*. Nat Communications 10:5343.

Andolfatto, P., D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, and D. L. Stern. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research 21:610–617.

Aravin, A. A., G. J. Hannon, and J. Brennecke. 2007. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science 318:761–764.

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, and V. A. Gvozdev. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Current Biology 11:1017–1027.

Badge, R. M., and J. F. Y. Brookfield. 1997. The Role of Host Factors in the Population Dynamics of Selfish Transposable Elements. Journal of Theoretical Biology 187:261–271.

Bailly-Bechet, M., A. Haudry, and E. Lerat. 2014. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. Mobile DNA 5:13.

Baker, R. H., A. Narechania, R. DeSalle, P. M. Johns, J. A. Reinhardt, and G. S. Wilkinson. 2016. Spermatogenesis Drives Rapid Gene Creation and Masculinization of the X Chromosome in Stalk-Eyed Flies (Diopsidae). Genome Biology and Evolution 8:896–914.

Baker, R. H., and G. S. Wilkinson. 2010. Comparative Genomic Hybridization (CGH) Reveals a Neo-X Chromosome and Biased Gene Movement in Stalk-Eyed Flies (Genus Teleopsis). PLoS Genetics 6:e1001121.

Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature 356:519–520.

Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh, M. W. Hahn, P. M. Nista, C. D. Jones, A. D. Kern, C. N. Dewey, L. Pachter, E. Myers, and C. H. Langley. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biology. 5:e310.

Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. A. Ponce de León, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell, and T. P. L. Smith. 2017. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. Nature Genetics 49:643–650.

Blumenstiel, J. P., A. A. Erwin, and L. W. Hemmer. 2016. What Drives Positive Selection in the *Drosophila* piRNA Machinery? The Genomic Autoimmunity Hypothesis. Yale J Biol Med 89:499–512.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. Cell 128:1089–1103.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang, and D. A. Barbash. 2006. Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in *Drosophila*. Science 314:1292–1295.

Burt, A., and R. Trivers. 2006. Genes in conflict: the biology of selfish genetic elements. Belknap Press of Harvard University Press, Cambridge, Mass.

Burton, J. N., A. Adey, R. P. Patwardhan, R. Qiu, J. O. Kitzman, and J. Shendure. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nature Biotechnology 31:1119–1125.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Christianson, S. J., C. L. Brand, and G. S. Wilkinson. 2011. Reduced Polymorphism Associated with X Chromosome Meiotic Drive in the Stalk-Eyed Fly Teleopsis dalmanni. PLoS ONE 6:e27254.

Christianson, S. J., J. G. Swallow, and G. S. Wilkinson. 2005. Rapid evolution of postzygotic reproductive isolation in stalk-eyed flies. Evolution 59:849–857.

Cocquet, J., P. J. I. Ellis, S. K. Mahadevaiah, N. A. Affara, D. Vaiman, and P. S. Burgoyne. 2012. A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. PLoS Genetics 8:e1002900.

Cotton, A., M. Foldvari, S. Cotton, and A. Pomiankowski. 2013. Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (Teleopsis dalmanni). Heredity 112:363-369.

Curtsinger, J. W., and M. W. Feldman. 1980. Experimental and Theoretical Analysis of the "Sex-Ratio" Polymorphism in *Drosophila Pseudoobscura*. Genetics 94:445–466.

Durand, N. C., J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, and E. L. Aiden. 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Systems 3:99–101.

Dyer, K. A., B. Charlesworth, and J. Jaenike. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proceedings of the National Academy of Sciences, U.S.A. 104:1587–1592.

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

Findley, S. D., M. Tamanaha, N. J. Clegg, and H. Ruohola-Baker. 2003. Maelstrom, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development 130:859–871.

Finnegan, S. R., N. J. White, D. Koh, M. F. Camus, K. Fowler, and A. Pomiankowski. 2019. Meiotic drive reduces egg-to-adult viability in stalk-eyed flies. Proc. R. Soc. B 286:20191414.

Frank, S. A. 1991. Divergence of meiotic drive suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution 45:262–267.

Fuller, Z. L., S. A. Koury, C. J. Leonard, R. E. Young, K. Ikegami, J. Westlake, S. Richards, S. W. Schaeffer, and N. Phadnis. 2020. Extensive Recombination Suppression and Epistatic Selection Causes Chromosome-Wide Differentiation of a Selfish Sex Chromosome in *Drosophila pseudoobscura*. Genetics genetics.303460.2020.

Gell, S. L., and R. A. Reenan. 2013. Mutations to the piRNA Pathway Component *Aubergine* Enhance Meiotic Drive of Segregation Distorter in *Drosophila melanogaster*. Genetics 193:771–784.

Hall, D. W. 2004. Meiotic drive and sex chromosome cycling. Evolution 58:925–931.

Hamilton, W. D. 1967. Extraordinary sex ratios. Science 156:477–488.

Handler, D., K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber, and J. Brennecke. 2013. The Genetic Makeup of the *Drosophila* piRNA Pathway. Molecular Cell 50:762–777.

Helleu, Q., P. R. Gérard, R. Dubruille, D. Ogereau, B. Prud'homme, B. Loppin, and C. Montchamp-Moreau. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. Proceedings of the National Academy of Sciences, U.S.A. 113:4110–4115.

Helleu, Q., and M. T. Levine. 2018. Recurrent Amplification of the Heterochromatin Protein 1 (HP1) Gene Family across Diptera. Molecular Biology and Evolution 35:2375–2389.

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.

Hubley, R., R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones, W. Bao, A. F. A. Smit, and T. J. Wheeler. 2016. The Dfam database of repetitive DNA families. Nucleic Acids Research 44:D81–D89.

Hurst, L. D. 1996. Further evidence consistent with Stellate's involvement in meiotic drive. Genetics 142:641–643.

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.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Jaenike, J. 2001. Sex chromosome meiotic drive. Annual Review of Ecology and Systematics 32:25–49.

Jangam, D., C. Feschotte, and E. Betrán. 2017. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends in Genetics 33:817–831.

Johns, P. M., L. L. Wolfenbarger, and G. S. Wilkinson. 2005. Genetic linkage between a sexually selected trait and X chromosome meiotic drive. Proc. R. Soc. B. 272:2097–2103.

Karvonen, A., O. Seppälä, and E. Tellervo Valtonen. 2009. Host immunization shapes interspecific associations in trematode parasites: Host immunization and parasite associations. Journal of Animal Ecology 78:945–952.

Kelleher, E. S., N. B. Edelman, and D. A. Barbash. 2012. *Drosophila* Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. PLoS Biol 10:e1001428.

Kent, W. J. 2002. BLAT--the BLAST-like alignment tool. Genome Res. 12:656–664.

Kersey, P. J., J. E. Allen, A. Allot, M. Barba, S. Boddu, B. J. Bolt, D. Carvalho-Silva, M. Christensen, P. Davis, C. Grabmueller, N. Kumar, Z. Liu, T. Maurel, B. Moore, M. D. McDowall, U. Maheswari, G. Naamati, V. Newman, C. K. Ong, M. Paulini, H. Pedro, E. Perry, M. Russell, H. Sparrow, E. Tapanari, K. Taylor, A. Vullo, G. Williams, A. Zadissia, A. Olson, J. Stein, S. Wei, M. Tello-Ruiz, D. Ware, A. Luciani, S. Potter, R. D. Finn, M. Urban, K. E. Hammond-Kosack, D. M. Bolser, N. De Silva, K. L. Howe, N. Langridge, G. Maslen, D. M. Staines, and A. Yates. 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Research 46:D802–D808.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology 37:907–915.

Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu, J. S. Khurana, F. Zhang, N. Schultz, B. S. Koppetsch, A. Nowosielska, H. Seitz, P. D. Zamore, Z. Weng, and W. E. Theurkauf. 2009. The *Drosophila* HP1 homolog *Rhino* is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138:1137–1149.

Kofler, R., D. Gómez-Sánchez, and C. Schlötterer. 2016. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. Molecular Biology and Evolution 33:2759–2764.

Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol, and C. Schlötterer. 2011. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLoS ONE 6:e15925.

Kolaczkowski, B., D. N. Hupalo, and A. D. Kern. 2011a. Recurrent Adaptation in RNA Interference Genes Across the *Drosophila* Phylogeny. Molecular Biology and Evolution 28:1033–1042.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun. 2011b. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187:245–260.

Korber, B., A. Rodrigo, and G. Learn. 2000. HIV Signature and Sequence Variation Analysis. Pp. 55–72 *in* Computational Analysis of HIV Molecular Sequences. Kluwer Academic Publishers, Dordrecht, Netherlands.

Korunes, K. L., and M. A. F. Noor. 2019. Pervasive gene conversion in chromosomal inversion heterozygotes. Molecular Ecology 28:1302–1315.

Krasnov, B. R., D. Mouillot, I. S. Khokhlova, G. I. Shenbrot, and R. Poulin. 2005. Covariance in species diversity and facilitation among non-interactive parasite taxa: all against the host. Parasitology 131:557.

Langley, C. H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth. 1988. On the role of unequal exchange in the containment of transposable element copy number. Genetics Research. 52:223–235.

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

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.

Le Thomas, A., A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, J. K. Hur, A. A. Aravin, and K. F. Toth. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes & Development 27:390–399.

Lee, Y. C. G. 2015. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in *Drosophila melanogaster*. PLoS Genetics 11:e1005269.

Lee, Y. C. G., and G. H. Karpen. 2017. Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. eLife 6:e25762.

Lee, Y. C. G., and C. H. Langley. 2012. Long-Term and Short-Term Evolutionary Impacts of Transposable Elements on *Drosophila*. Genetics 192:1411–1432.

Lerat, E., M. Fablet, L. Modolo, H. Lopez-Maestre, and C. Vieira. 2016. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Research gkw953.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.

Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, L. Holman, H. Johannesson, U. Knief, H. Kokko, A. M. Larracuente, A. Manser, C. Montchamp-Moreau, V. G. Petrosyan, A. Pomiankowski, D. C. Presgraves, L. D. Safronova, A. Sutter, R. L. Unckless, R. L. Verspoor, N. Wedell, G. S. Wilkinson, and T. A. R. Price.

2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. Trends in Ecology & Evolution 31:315–326.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.

Lyttle, T. W. 1991. Segregation distorters. Ann. Rev. Genetics 25:511–557.

Machado, C. A., T. S. Haselkorn, and M. A. F. Noor. 2007. Evaluation of the Genomic Extent of Effects of Fixed Inversion Differences on Intraspecific Variation and Interspecific Gene Flow in *Drosophila pseudoobscura* and *D. persimilis*. Genetics 175:1289–1306.

McCullers, T. J., and M. Steiniger. 2017. Transposable elements in *Drosophila*. Mobile Genetic Elements 7:1–18.

McDermott, S. R., and M. A. F. Noor. 2010. The role of meiotic drive in hybrid male sterility. Phil. Trans. R. Soc. B 365:1265–1272.

McLaughlin, R. N., and H. S. Malik. 2017. Genetic conflicts: the usual suspects and beyond. J Exp Biol 220:6–17.

Meade, L., S. Finnegan, R. Kad, K. Fowler, and A. Pomiankowski. 2019. Maintenance of fertility in the face of meiotic drive. The American Naturalist 707372.

Meiklejohn, C. D., E. L. Landeen, K. E. Gordon, T. Rzatkiewicz, S. B. Kingan, A. J. Geneva, J. P. Vedanayagam, C. A. Muirhead, D. Garrigan, D. L. Stern, and D. C. Presgraves. 2018. Gene flow mediates the role of sex chromosome meiotic drive during complex speciation. eLife 7:e35468.

Miller, W. J., J. F. McDonald, and W. Pinsker. 1997. Molecular domestication of mobile elements. Genetica 100:261–270.

Montchamp-Moreau, C., D. Ogereau, N. Chaminade, A. Colard, and S. Aulard. 2006. Organization of the sex-ratio Meiotic Drive Region in *Drosophila simulans*. Genetics 174:1365–1371.

Muckenfuss, H., M. Hamdorf, U. Held, M. Perković, J. Löwer, K. Cichutek, E. Flory, G. G. Schumann, and C. Münk. 2006. APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition. Journal of Biological Chemistry 281:22161–22172.

Muller, H. J. 1964. The Relation of Recombination to Mutational Advance. Mutat. Res. 106:2–9.

Nagao, A., T. Mituyama, H. Huang, D. Chen, M. C. Siomi, and H. Siomi. 2010. Biogenesis pathways of piRNAs loaded onto *AGO3* in the *Drosophila* testis. RNA 16:2503–2515.

Obbard, D. J., K. H. J. Gordon, A. H. Buck, and F. M. Jiggins. 2009. The evolution of RNAi as a defence against viruses and transposable elements. Phil. Trans. R. Soc. B 364:99–115.

Ozata, D. M., I. Gainetdinov, A. Zoch, D. O'Carroll, and P. D. Zamore. 2019. PIWI-interacting RNAs: small RNAs with big functions. Nature Reviews Genetics 20:89–108.

Paczolt, K. A., J. A. Reinhardt, and G. S. Wilkinson. 2017. Contrasting patterns of X-chromosome divergence underlie multiple sex-ratio polymorphisms in stalk-eyed flies. Journal of Evolutionary Biology., doi: 10.1111/jeb.13140.

Palmer, W. H., J. D. Hadfield, and D. J. Obbard. 2018. RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics 208:1585–1599.

Parhad, S. S., and W. E. Theurkauf. 2019. Rapid evolution and conserved function of the piRNA pathway. Open Biology. 9:180181.

Parhad, S. S., S. Tu, Z. Weng, and W. E. Theurkauf. 2017. Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev. Cell 43:60-70.e5.

Pertea, G., and M. Pertea. 2020. GFF Utilities: GffRead and GffCompare. F1000Res 9:304.

Presgraves, D., Severance, E. and Wilkinson, G.S. 1997. Sex chromosome drive in stalk-eyed flies. Genetics 147:1169-1180

Poynton, H. C., S. Hasenbein, J. B. Benoit, M. S. Sepulveda, M. F. Poelchau, D. S. T. Hughes, S. C. Murali, S. Chen, K. M. Glastad, M. A. D. Goodisman, J. H. Werren, J. H. Vineis, J. L. Bowen, M. Friedrich, J. Jones, H. M. Robertson, R. Feyereisen, A. Mechler-Hickson, N. Mathers, C. E. Lee, J. K. Colbourne, A. Biales, J. S. Johnston, G. A. Wellborn, A. J. Rosendale, A. G. Cridge, M. C. Munoz-Torres, P. A. Bain, A. R. Manny, K. M. Major, F. N. Lambert, C. D. Vulpe, P. Tuck, B. J. Blalock, Y.-Y. Lin, M. E. Smith, H. Ochoa-Acuña, M.-J. M. Chen, C. P. Childers, J. Qu, S. Dugan, S. L. Lee, H. Chao, H. Dinh, Y. Han, H. Doddapaneni, K. C. Worley, D. M. Muzny, R. A. Gibbs, and S. Richards. 2018. The Toxicogenome of Hyalella azteca: A Model for Sediment Ecotoxicology and Evolutionary Toxicology. Environmental Science and Technology. 52:6009–6022.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rao, S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden. 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159:1665–1680.

Rédei, G. P. 2008. Haldane's Mapping Function. Pp. 836–836 *in* Encyclopedia of Genetics, Genomics, Proteomics and Informatics. Springer Netherlands, Dordrecht.

Reinhardt, J. A., C. L. Brand, K. A. Paczolt, P. M. Johns, R. H. Baker, and G. S. Wilkinson. 2014. Meiotic Drive Impacts Expression and Evolution of X-Linked Genes in Stalk-Eyed Flies. PLoS Genetics 10:e1004362.

Sedlazeck, F. J., P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, and M. C. Schatz. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nature Methods 15:461–468.

Sedlazeck, F. J., P. Rescheneder, and A. von Haeseler. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29:2790–2791.

Selker, E. U. 1990. Premeiotic Instability of Repeated Sequences in *Neurospora crassa*. Annual Review of Genetics 24:579–613.

Seppey, M., M. Manni, and E. M. Zdobnov. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods in Molecular Biology. 1962:227–245.

Sienski, G., D. Dönertas, and J. Brennecke. 2012. Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. Cell 151:964–980.

Simkin, A., A. Wong, Y.-P. Poh, W. E. Theurkauf, and J. D. Jensen. 2013. Recurrent and recent selective sweeps in the piRNA pathway. Evolution 67:1081–1090.

Sinzelle, L., Z. Izsvák, and Z. Ivics. 2009. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cellular and Molecular Life Sciences. 66:1073–1093.

Smit, A., and R. Hubley. 2008. RepeatModeler Open-1.0.

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

Smith, J., and J. Haigh. 1974. The hitchhiking effect of a favourable gene. Genetics Research. 23:23–25.

Soh, Y. Q. S., J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves, P. J. Minx, R. S. Fulton, C. Kremitzki, N. Koutseva, J. L. Mueller, S. Rozen, J. F. Hughes, E. Owens, J. E. Womack, W. J. Murphy, Q. Cao, P. de Jong, W. C. Warren, R. K. Wilson, H. Skaletsky, and D. C. Page. 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. Cell 159:800–813.

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. The Plant Journal 3:739–744.

Swallow, J. G., L. E. Wallace, S. J. Christianson, P. M. Johns, and G. S. Wilkinson. 2005. Genetic divergence does not predict change in ornament expression among populations of stalk-eyed flies. Molecular Ecology 14:3787–3800.

Tao, Y., L. Araripe, S. B. Kingan, Y. Ke, H. Xiao, and D. L. Hartl. 2007a. A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. PLoS Biology. 5:e293.

Tao, Y., J. P. Masly, L. Araripe, Y. Ke, and D. L. Hartl. 2007b. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. PLoS Biology. 5:e292.

Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14:178–192.

Unckless, R. L., and A. G. Clark. 2014. Sex-ratio meiotic drive and interspecific competition. Journal of Evolutionary Biology. 27:1513–1521.

Vagin, V. V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P. D. Zamore. 2006. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science 313:320–324.

Van Ooijen, J. W. 2006. JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen 33:10–1371.

Vicoso, B., and D. Bachtrog. 2015. Numerous Transitions of Sex Chromosomes in *Diptera*. PLoS Biol 13:e1002078.

Werren, J. H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proceedings of the National Academy of Sciences 108:10863–10870.

Werren, J. H., U. Nur, and C. I. Wu. 1988. Selfish genetic elements. Trends in Ecology and Evolution 3:297–302.

Wheeler, D., A. J. Redding, and J. H. Werren. 2013. Characterization of an Ancient Lepidopteran Lateral Gene Transfer. PLoS ONE 8:e59262.

Wilkinson, G. S. 1993. Artificial sexual selection alters allometry in the stalk-eyed fly Cyrtodiopsis dalmanni (*Diptera*: *Diopsidae*). Genetical Research 62:213–222.

Wilkinson, G. S., S. J. Christianson, C. L. Brand, G. Ru, and W. Shell. 2014. Haldane's Rule Is Linked to Extraordinary Sex Ratios and Sperm Length in Stalk-Eyed Flies. Genetics 198:1167–1181.

Wilkinson, G. S., P. M. Johns, E. S. Kelleher, M. L. Muscedere, and A. Lorsong. 2006. Fitness effects of X chromosome drive in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. Journal of Evolutionary Biology 19:1851–1860.

Wilkinson, G. S., D. C. Presgraves, and L. Crymes. 1998. Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. Nature 391:276–279.

Wolf, D., and S. P. Goff. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature 458:1201–1204.

Wolfenbarger, L. L. & Wilkinson, G. S. 2001. Sex-linked expression of a sexually selected trait in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. Evolution 55:103–110.

Wu, C. I., T. W. Lyttle, M. L. Wu, and G. F. Lin. 1988. Association between a satellite DNA sequence and the *Responder* of *Segregation Distorter* in *D. melanogaster*. Cell 54:179–189.

Wu, C.-I. & Beckenbach, A. T. 1983. Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *D. persimilis* and identification of hybrid sterility factors. Genetics 105:71–86.

Wu, T. D., and C. K. Watanabe. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875.

Zhang, S., B. Pointer, and E. S. Kelleher. 2020. Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant *de novo* mutations. Genome Research. 30:566–575.

Zimin, A. V., G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, and J. A. Yorke. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677.