Impacts of Sex Ratio Meiotic Drive on Genome Structure and Function in a Stalk-Eyed Fly

Josephine A. Reinhardt [1,*], Richard H. Baker[2], Aleksey V. Zimin[3], Chloe Ladias[1], Kimberly A. Paczolt[4], John H. Werren[5], Cheryl Y. Hayashi[2], and 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]Department of Biology, University of Rochester, Rochester, New York, USA

*Corresponding author: E-mail: reinhardt@geneseo.edu.

Accepted: 15 June 2023

Abstract

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 assemble and describe a chromosome-level genome assembly of the stalk-eyed fly, *Teleopsis dalmanni*, to elucidate patterns of divergence associated with SR. The genome contains tens of thousands of transposable element (TE) insertions and hundreds of transcriptionally and insertionally active TE families. By resequencing pools of SR and ST males using short and long reads, we find widespread differentiation and divergence between X^{SR} and X^{ST} associated with multiple nested inversions involving most of the *SR* haplotype. Examination of genomic coverage and gene expression data revealed seven X-linked genes with elevated expression and coverage in SR males. The most extreme and likely drive candidate involves an XSR-specific expansion of an array of partial copies of JASPer, a gene necessary for maintenance of euchromatin and associated with regulation of TE expression. In addition, we find evidence for rapid protein evolution between X^{SR} and X^{ST} for testis expressed and novel genes, that is, either recent duplicates or lacking a Dipteran ortholog, including an X-linked duplicate of *maelstrom*, which is also involved in TE silencing. Overall, the evidence suggests that this ancient X^{SR} polymorphism has had a variety of impacts on repetitive DNA and its regulation in this species.

Key words: selfish genetic elements, transposable elements, Diptera, JASPer.

Significance

Sex ratio meiotic drivers can have powerful impacts on evolution, impacting the genome structure and putting species in danger of extinction, but the genomic causes and consequences of drive are poorly understood. By assembling and analyzing the genomes of two stalk-eyed flies, we are able to document the dramatic consequences of meiotic drive, including a promising candidate involved in chromatin regulation, adding to a growing body of work connecting disruption of genome packaging to these selfish genetic elements.

Introduction

The genome was once thought to be little more than a blueprint needed to accomplish biological functions and reproduction of an organism. Yet research in the past few decades has demonstrated that genomes of most organisms are heavily 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). Meiotic drivers are a well-studied category of SGE, also known as segregation distorters. These elements spread by manipulating gametogenesis in their favor, leading to greater than 50% representation of the driver in mature gametes (Lyttle 1991). Gamete killers are a common type of meiotic driver (sperm and spore killers), which cause gametes not inheriting the driver to fail to develop. If drivers are on a sex chromosome, a skew in the sex ratio (SR) of offspring will result. Such SR distortion may cause population collapse or even extinction (Hamilton 1967) but can be maintained stably (reviewed in Curtsinger and Feldman 1980; Jaenike 2001; Lindholm et al. 2016), and production of excess females has been theorized to contribute to success in interspecific competition (Unckless and Clark 2014; Mackintosh et al. 2021).

The molecular mechanisms underlying meiotic drive have been elucidated in an increasing number of species in recent years (Grognet et al. 2014; Hu et al. 2017; Nuckolls et al. 2017), (t locus), and signs of common mechanisms have begun to emerge in *Drosophila* systems (Wu et al. 1988; Houtchens and Lyttle 2003; Montchamp-Moreau et al. 2006; Nagao et al. 2010; Gell and Reenan 2013; Larracuente 2014; Helleu et al. 2016; Lin et al. 2018; Courret et al. 2019; Muirhead and Presgraves 2021; Vedanayagam et al. 2021), including a role for repetitive DNA either in the driver or in the target of drive, disruptions in DNA packaging (histones or protamines), and a role for genome defense pathways (Courret et al. 2019). However, meiotic drivers are often associated with chromosomal inversions (Jaenike 2001), making them resistant to standard genetic analysis (Wu and Beckenbach 1983; Dyer et al. 2007; Paczolt et al. 2017; Fuller et al. 2020). Like any inversion, drive-associated inversions may exchange alleles within inversion type via recombination when homozygous, but this will be relatively rare if drive is at a low frequency. A reduction in recombination slows or prevents purging of new deleterious mutations (Muller 1964) and may also reduce nucleotide diversity (Smith and Haigh 1974; Charlesworth et al. 1993), presumably leading to reduced fitness to carriers over time. The role of such linked variation in the evolutionary trajectory of drive systems remains poorly understood.

Here, we analyze the impacts of SGEs within the genome of a stalk-eyed fly (*Teleopsis dalmanni*). In this species, 10–30% of males possess X-linked elements that prevent proper development of Y-bearing sperm and result in carrier males producing 90% or more daughters (Presgraves et al. 1997). This *SR* X chromosome has multiple impacts on individual 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. 2014). The SR X chromosome (X^{SR}) appears to have originated approximately 500 Kya (Paczolt et al. 2017), and hundreds of mostly X-linked genes are differentially expressed in the testes of SR males (Reinhardt et al. 2014). X^{SR} contains at least one large chromosomal inversion compared with the standard (ST) 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). Recombination occurs between X^{SR} haplotypes in homozygous females, but the rate of recombination is about half of that in X^{ST} females (Paczolt et al. 2017). Reduced recombination and effective population size have likely contributed to drastically reduced polymorphism on X^{SR} (Christianson et al. 2011).

But how much differentiation has occurred between X^{SR} and X^{ST} in this species? How many inversions are on the X? What impacts has long-term association with a meiotic drive element had on the landscape of genetic variation on X^{SR}? Can we identify likely candidate genes involved in establishing the meiotic drive phenotype? To answer these questions, we created a chromosome-level genome assembly for *T. dalmanni*, annotated transposable elements and genes, and then combined RNA sequencing (RNAseq), pooled short-read and long-read resequencing data from males exhibiting *SR* to identify sequence, copy number, and expression differences between the two types of X chromosomes.

Results

A Chromosome Length Assembly of the *T. dalmanni* Genome

A primary assembly of the genome of *T. dalmanni* (NLCU01000000), a stalk-eyed fly from southeast Asia, was created using MaSuRCA from hybrid sequencing data containing long-read and short-read sequences (supplementary table S1, Supplementary Material online). All data used in this assembly were generated using females from a standard SR inbred line (see methods). After haplotig filtering, the assembly was scaffolded using chromatin conformation information, producing three chromosome-length scaffolds with a total size of 438.2 Mbp (NLCU04000000) comprising 95.7% of the filtered MaSuRCA assembly. We validated the assembly by comparison with an independently generated linkage map produced using a backcross family from a prior QTL study (Wilkinson et al. 2014) (supplementary fig. S1, Supplementary Material online). Although the maps were largely concordant, a 13.3-Mbp region (61–74 Mbp) is inverted between the assembly and the linkage map, consistent with an inversion difference between the two populations used in the QTL study. BUSCO analysis confirmed the presence of 96.7% of 3,285 conserved Dipteran genes (supplementary table S2, Supplementary Material online), with 2.0% of BUSCO genes duplicated. Overall, 89.8% of 1-to-1 *Drosophila melanogaster* orthologs are located on the same Muller element in these two

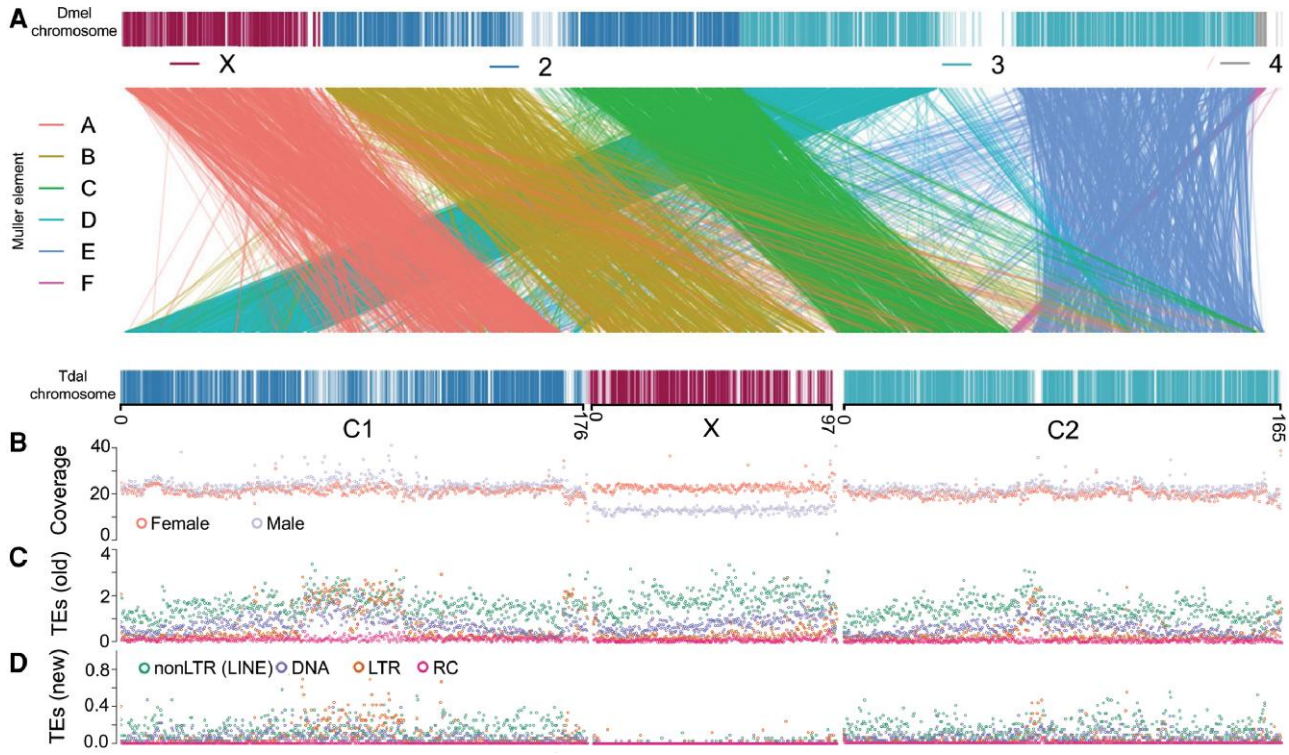

Fig. 1.—A chromosome length assembly of the stalk-eyed fly genome reveals gene synteny and movement compared with *D. melanogaster*, and unique patterns of sequence variation and differentiation on the X chromosome influenced by meiotic drive. (*A*) Slopegraph indicating the locations of 7,634 *T. dalmanni*—*D. melanogaster* 1-to-1 orthologs in each genome. Genes are ordered left to right by their Muller element (*A*–*F*) in *D. melanogaster*, with 89.8% found on the same Muller element in the two species. The X chromosome in *T. dalmanni* is Muller element B (chromosome 2*L* in *D. melanogaster*). C1 consists of Muller D and A (chromosomes 3*L* and X in *D. melanogaster*), and C2 contains Muller C, F, and E in that order (chromosomes 2*R*, 4, and 3*R* in *D. melanogaster*). (*B*) As expected, the X chromosome has reduced WGS coverage (reads per bp) in a male genomic DNA library compared with a female library. (*C*) Transposable element copies are more abundant on the *T. dalmanni* X than on the two autosomes. (*D*) New transposable element insertions are less common on the X.

Schizophoran fly species (fig. 1*A*). As previously reported (Baker and Wilkinson 2010), the 97.2-Mbp *Teleopsis* X chromosome is orthologous to chromosome 2L in *D. melanogaster* (Muller element B). Stalk-eyed flies are among only a few Dipterans in which Muller B is the X (Vicoso and Bachtrog 2015). The two autosomes, previously (Baker and Wilkinson 2010) referred to as "C1" and "C2" are similarly sized (176 and 165 Mbp). We also produced a draft assembly of a closely related cryptic species (Christianson et al. 2005; Paczolt et al. 2017), *T. dalmanni sp 2* (*Td2*) to polarize molecular evolutionary changes. This assembly contains 50,545 scaffolds and is less complete (Diptera BUSCO = 90.0%, N50 = 35,545) than the *T. dalmanni* sensu stricto assembly. Also as expected, genomic coverage in females was approximately twice that in males across the X, but not the autosomes (fig. 1*B*).

Compared with *D. melanogaster* (dm6 RepeatMasker open-4.0.6), repetitive sequences cover more of the *T. dalmanni* genome (~13.5% vs. ~35.8%). About 10.0% of the genome is comprised of unclassified interspersed elements, whereas the rest includes 716 classified families from 38 superfamilies of Class I (DNA) elements (4.8% of genome) and Class II elements including LINE (10.7%) and LTR (5.9%) elements, but no SINE elements (supplementary fig. S2*A*, Supplementary Material online). LINE elements were significantly ($\chi^2 = 3401$, $P < 0.001$) overrepresented in *T. dalmanni* compared with the other two species. The most abundant transposable element (TE) superfamilies in *T. dalmanni* include R1-LOA, Jockey, and RTE-BovB non-LTR (LINE) Class II elements (24,798, 15,719, and 11,712 copies, respectively), Gypsy LTR Class-II elements (14,484 copies), and TcMar-Mariner Class I (DNA) elements (13,210 copies). Overall, transposable elements are more abundant on the *T. dalmanni* X than on the two autosomes (293.9 TEs/Mbp on the X vs. 204.5 TEs/Mbp on the two autosomes, $\chi^2 = 364.46$, $P < 0.0001$) (fig. 1*C*). TE distribution also varies by element type, with more X-linked DNA elements (26.6%, $\chi^2 = 364$, $P < 0.0001$) and LINE elements (25.3%, $\chi^2 = 372$, $P < 0.0001$) but fewer X-linked LTR elements than expected (14.9%, $\chi^2 = 532$, $P < 0.0001$) given the X comprises 22.2% of the genome. The assembly is from female tissue, so we do not have a Y-chromosome

assembly; however, we found several transposable elements showing excess of male coverage (supplementary table S3, Supplementary Material online), suggesting an excess of insertions on the Y chromosome. For example, one Penelope element was present on male reads 8-fold more than on female reads. Many TE families are transcriptionally active (supplementary fig. S2*B*, Supplementary Material online) and are producing new insertions in the genome (fig. 1*D*).

The SR X Is Diverging from the ST X due to Multiple Overlapping Inversions

Using long reads generated from *SR* male siblings aligned to the reference genome, we identified six large X-chromosomal inversions that differ between SR males and the reference genome (fig. 2*A*), which we further validated by examining short-read pool-seq data from SR and ST males at breakpoint regions (supplementary table S4 and fig. S3, Supplementary Material online). These inversions spanned the entire 97-Mbp chromosome except for a small region at the proximal end (0–2.04 Mbp) and a region between 60.4 Mbp and 82.7 Mb. Many of the inversions overlap, particularly near the proximal end. Of the four inversions with outgroup data available from the *Td2* draft genome, three are derived in the SR lineage and one (inversion 2) is derived in the ST lineage (supplementary table S4, Supplementary Material online).

We compared patterns of genetic variation on the autosomes and the X^{SR} and X^{ST} chromosomes using the pool-seq short-read resequencing data from two pools of SR and four pools of ST males derived from two field sites (Gombak Field Studies Center and Kanching Forest Park) near Kuala Lumpur, Malaysia (fig. 2*B* and *C* and supplementary fig. S4, Supplementary Material online). As expected, the nucleotide diversity on autosomes in SR (0.020134 ± 0.0000839) and ST (0.021375 ± 0.001432) pools does not differ significantly (*t*-test SR vs. ST for C1, $P = 0.309$, for C2, $P = 0.3332$) from each other but we find that X^{SR} has significantly reduced diversity (0.0057492 ± 0.001775) compared with X^{ST} (0.015670 ± 0.001410 CI, *t*-test $P = 0.00122$, fig. 2*B*). All three chromosomes contain regions with reduced nucleotide diversity in all pools (supplementary fig. S4*A*, Supplementary Material online), which are presumably centromeric regions (Begun and Aquadro 1992; Begun et al. 2007), as they are either near the center of the metacentric autosomes or the end of the X chromosome. Genetic differentiation (F_{ST}) is strongly elevated across the X chromosome between *SR* types (X^{SR} vs. X^{ST}) and to a lesser degree was elevated between collection sites but within the *SR* type (fig. 2*C*) but is not elevated on the autosomes (supplementary fig. S4*B*, Supplementary Material online). Patterns of nucleotide diversity and differentiation appear to be influenced by proximity to the inversions on the X chromosome (fig. 2). F_{ST} was elevated, and polymorphism reduced where there is a higher density of overlapping inversions. In particular, the region between 17 Mbp and 20 Mb, where five of the six inversions overlap, has high F_{ST} between X^{SR} and X^{ST} and reduced diversity within all pools.

Gene Expression and Copy Number Variation Reveal Drive Candidates on the SR X

Using the pool-seq reads from X^{SR} and X^{ST} males above in combination with X^{SR} and X^{ST} RNAseq reads previously obtained from pools of mature testes (Reinhardt et al. 2014), we jointly evaluated differences in gene copy number and expression between X^{SR} and X^{ST}. We identified 596 DE genes (37.6% had higher expression in X^{SR}) and 120 genes with differential genomic coverage (DC genes, 62.5% with higher coverage on X^{SR}). Among the 48 genes that were *both* DE and DC, there was a highly significant and positive association in the direction of DE and differential genomic coverage (FET $P < 0.0001$) with only four genes having higher X^{SR} coverage but higher X^{ST} expression. This could be due to a direct effect of the copy number on expression but is not definitive of a causal relationship. As expected from prior work (Reinhardt et al. 2014; Paczolt et al. 2017), most SR–ST differences in gene expression (75.6%) and genomic coverage (89.2%) were confined to the X. Seven annotated protein-coding genes (*JASPer*, *Pbp95*, *Tetraspanin 29Fb*, *Minichromosome maintenance 10*, *isopeptidase-T-3*, *Cyclin K*, and *santa-maria*) exhibited both differential expression (DE) and differential genomic coverage (DC) between SR and ST males (supplementary table S5, Supplementary Material online). Strikingly, all seven were X linked and had a higher level of both expression and coverage in *SR* males. We also identified 11 transposable element families that differ in either expression or copy number between SR and ST, with roughly an equal number of families differentially expressed in each (supplementary table S3 and fig. S2*C*, Supplementary Material online). Maverick, Gypsy, and R1-LOA elements are overexpressed in SR males, whereas TcMariner, PIF-Harbinger, and two types of RTE elements have excess coverage on the SR X chromosome.

The highest level of differential expression and coverage was found for an X-linked paralog of *JASPer* (*Jil-1* Anchoring and Stabilizing Protein), a gene which normally regulates the maintenance of euchromatin and is involved in female fertility and X-chromosome dosage compensation (Albig et al. 2019; Dou et al. 2020). Examination of genomic coverage near the *JASPer* region (X:18.18 Mbp) shows a 20-fold increase in coverage over a 1.5-Kbp region containing a 1.1-Kbp gene (fig. 3*A*). Long- and short-read sequences from SR males showed supplemental alignment

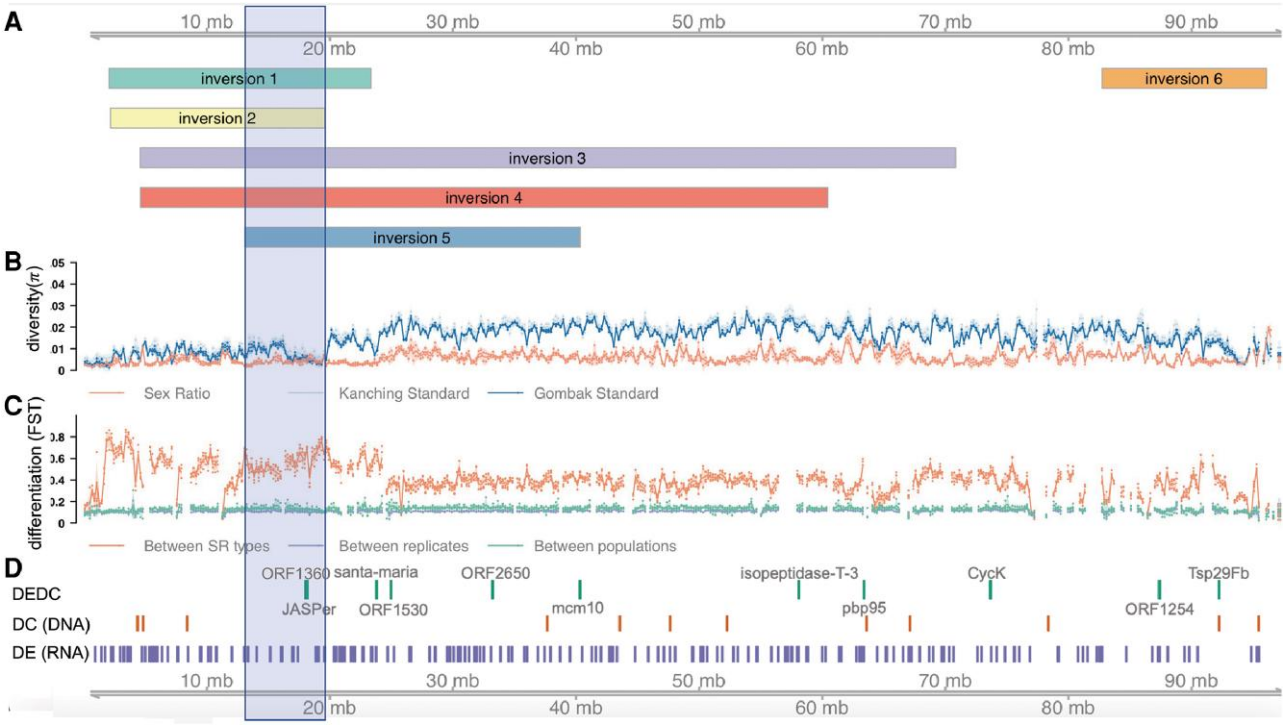

Fig. 2.—Comparison of the SR X chromosome (X^{SR}) with the ST X chromosome (X^{ST}) using long-read and short-read resequencing. (*A*) Six inversions were identified using alignment of PacBio long-read sequencing of males carrying the X^{SR} chromosome to the reference genome. The shaded area indicates a 7-Mbp region where five of the six inversions overlap. Two replicate WGS Illumina sequencing pools were constructed from screened males exhibiting standard SRs from each of two collection sites near Kuala Lumpur: Gombak Research station (GST) and Kanching Forest park (KST). One pool of males exhibiting female-biased SRs was also sequenced from each site. (*B*) Nucleotide diversity (pi) estimated from pool-seq libraries is lower across the X for the two SR pools compared with the four ST pools. (*C*) Compared with between-replicate comparisons (KST and GST), differentiation (F_{ST}) is only slightly elevated between populations with the same SR type (between populations) but F_{ST} is notably elevated chromosome wide in pairwise comparisons between ST and SR pool-seq (between SR types). (*D*) Locations of named genes that were differentially expressed (DE) in testes between SR and ST RNAseq samples or have differential coverage (DC) between SR and ST WGS pools or both (DCDE) are also shown.

to nearby regions of the X at 18.19 and 18.27 Mbp. These two regions also contain a ~1.5-Kbp region with 20-fold increased raw short-read coverage. At 18.19 Mbp, the gene exhibits the same two-exon structure and length as the copy at 18.16 Mbp but is positioned in the opposite orientation. At 18.27 Mbp, a smaller (~600 bp) single-exon region is transcribed. Translations of the transcripts produced by these three partial *JASPer* genes contain only one of the two major functional domains of *D. melanogaster JASPer*, PWWP, which normally interacts with activating chromatin marks (H3K36me3). There are also three full-length paralogs of *JASPer* on the C2 autosome and two additional X-linked copies that contain only the other functional domain, LEDGF, known to interact with *JASPer's* functional partner, *JIL-1* in *Drosophila* (fig. 3*B*). Although eight other *D. melanogaster* proteins contain PWWP domains, a maximum likelihood phylogeny shows that the *T. dalmanni* JASPer PWWP domains are closer orthologs to the PWWP domain of *D. melanogaster JASPer* than to other Dipteran PWWP domains (fig. 3*C*), confirming these are partial *JASPer* paralogs. Neither of the LEDGF-only X-linked copies exhibit differential expression or coverage between SR and ST males (fig. 3*A*). The *JASPer* paralogs are all single copy in X^{ST} but are amplified in copy number in X^{SR}. Using only uniquely mapping SR and ST short reads, we estimated the increase in copy number of each JASPer copy on X^{SR}. The fold increase rates in genomic coverage of SR libraries across the amplified regions are ~58.9-fold (at 18.181:Mbp), 3.4-fold (at 18.194:Mbp) and 6.5-fold (at 18.270:Mbp) compared with ST libraries (fig. 3*A*, **"unique"**). For the copy at 18.181:Mbp, we were also able to identify long reads which aligned uniquely to sequence matching both the left and right flanking regions and extended into the gene copy and these contained five or six tandem copies of the PWWP-only *JASPer* paralog (supplementary fig. S5, Supplementary Material online). We also searched for supplemental alignments of PWWP-JASPer aligning reads to other places on the X chromosome, as these could indicate additional places that a PWWP-JASPer might have

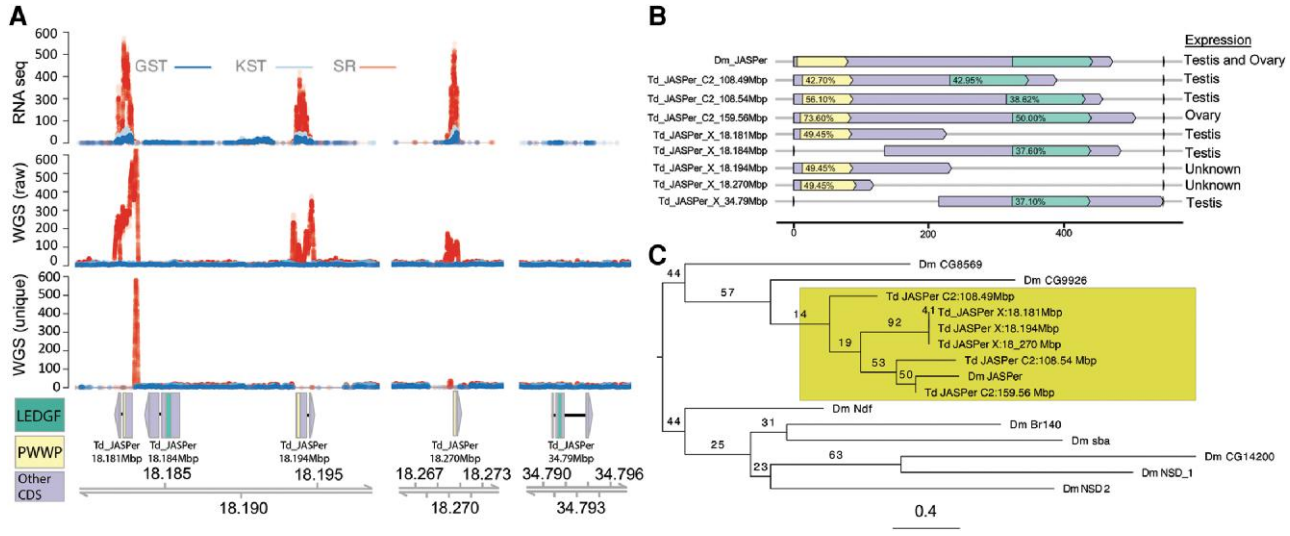

Fig. 3.—*JASPer*-PWWP has amplified in expression and coverage on the SR X. (*A*) Examination of mapped reads in X-linked gene regions with homology to *JASPer* shows elevated expression (RNAseq) and coverage in pool-seq WGS libraries (pool-seq raw) from SR males, but not ST males from the same collections (GST and KST). Only those copies of *JASPer* including the PWWP domain showed increases in expression and coverage. Examination of uniquely mapping pool-seq reads (pool-seq Unique) demonstrates that the excess SR coverage is largely limited to the *JASPer* copy at 18.181 Mbp. (*B*) *T. dalmanni* contains eight transcripts with detectable homology to *D. melanogaster JASPer*. Copies on the second autosome (C2) contain both canonical domains (PWWP and LEDGF) whereas the five X-linked copies contain only one of the two domains. Amino acid percent identity compared with *D. melanogaster JASPer* is shown for each domain. (*C*) *T. dalmanni JASPer* copies are orthologous to PWWP from *D. melanogaster JASPer* as confirmed by a maximum likelihood phylogeny of all PWWP domains in *D. melanogaster* and all PWWP domains from *T. dalmanni JASPer* copies. The three X linked *JASPer* PWWP domains from ~18 Mbp are identical in amino acid sequence.

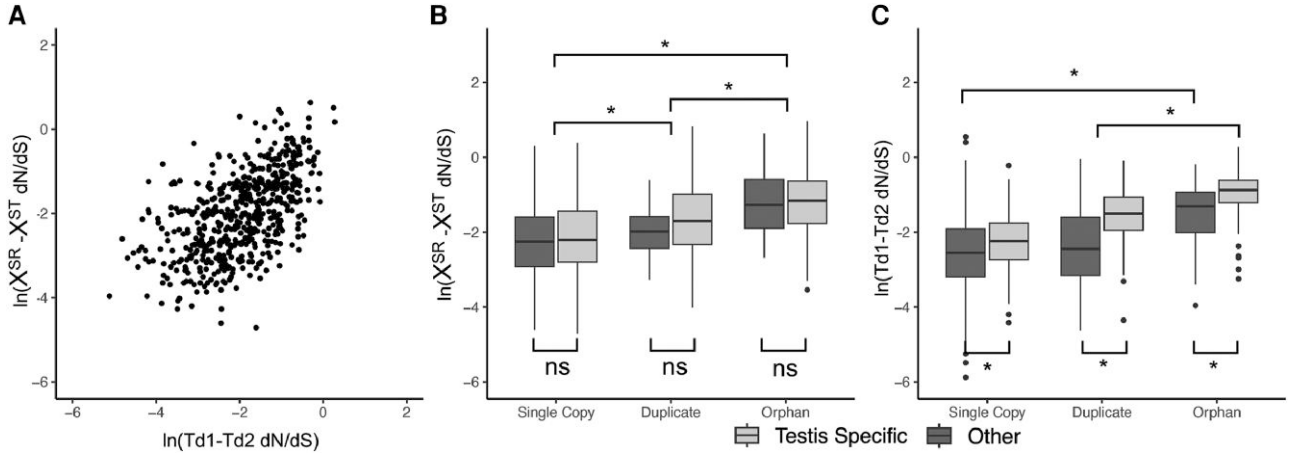

Fig. 4.—Rapid protein evolution on the SR X chromosome. *dN/dS* between *T. dalmanni* and *T. dalmanni* sp 2 (Td2) was assessed for 9,525 protein-coding genes, and *dN/dS* between the X^{SR} and X^{ST} chromosomes was assessed for 2,642 X-linked genes. Significance of comparisons was assessed using linear models on log-transformed *dN/dS* values and ANOVA (supplementary tables S5–S8, Supplementary Material online). The expression pattern of genes was annotated based on tissue-specific RNAseq in *T. dalmanni* (Reinhardt et al. 2014), and the age of genes was based on a prior comparative analysis (Baker et al. 2016). (*A*) *dN/dS* was significantly correlated in the two comparisons. (*B*) *dN/dS* between X^{SR} and X^{ST} copies of genes was elevated in testis-specific genes and newer categories of genes, with no significant interaction. (*C*) The same was true for the interspecific comparison (T2*dN/dS*).

inserted, providing an explanation of the excess coverage observed at 10.181 Mbp. We found a few reads mapping partly elsewhere on the X chromosome. However, none of these locations were represented by more than a single sequencing read.

Rapid Protein Evolution between X^{ST} and X^{SR}

Given most of the X chromosome shows elevated genetic differentiation (*FST*) between X^{ST} and X^{SR}, we calculated *dN/dS* to detect protein evolution occurring between these two chromosomal haplotypes, and also between

T. dalmanni and a cryptic sister species we described previously, *T. dalmanni sp 2* (Paczolt et al. 2017). For X-linked genes observed in both comparisons, *dN/dS* was positively correlated (Pearson's product–moment correlation = 0.57, $P < 2.2e^{-16}$, fig. 4*A*). Using prior annotations of gene novelty and expression pattern (Reinhardt et al. 2014; Baker et al. 2016) and a linear model to predict *dN/dS*, we find that novel protein-coding genes (duplicated genes and genes without an identifiable ortholog outside of *Diopsidae*) and testis-expressed genes have significantly higher *dN/dS* in both the interspecific and X^{SR}–X^{ST} comparisons (fig. 4*B* and *C* and supplementary tables S6–S9, Supplementary Material online). In the interspecific comparison, there is also a significant interaction effect between gene age and expression.

In the X^{SR}–X^{ST} comparison, 25 genes were putatively evolving under positive selection (*dN/dS* > 1, table 1). Of those with a *dN/dS* estimate in the interspecific comparison, most (11/13) were not under positive selection between species (*dN/dS* < 1). Among the positively selected genes, nine of 25 were *also* differentially expressed between SR and ST. These nine included three genes with *Drosophila* orthologs, Pol32 (a component of DNA polymerase involved in double-strand break repair), an uncharacterized transmembrane protein (Tetraspanin 29Fb), and Cyclin K, which phosphorylates RNA polymerase II and contributes to pre-mRNA processing, transcription, and chromatin structure.

Discussion

The presence of meiotic drivers can lead to dramatic genomic changes due to their conflicting interests with their hosts (Burt and Trivers 2006). Here, we analyzed the impacts on a stalk-eyed fly (*T. dalmanni*) genome of a long-term association with such an element. Prior work suggested widespread genetic differentiation between X^{SR} and X^{ST} chromosomes (Reinhardt et al. 2014) and identified at least one inversion distinguishing the two (Paczolt et al. 2017). By assembling the genome into three chromosome-length scaffolds, we find dramatic differentiation between X^{SR} and X^{ST} that extends across the entire X chromosome, a sign that genetic recombination between the two types of X chromosomes has been severely limited for a long time (figs. 1 and 2). By combining pool-seq short-read and long-read resequencing data, we located the breakpoints of six overlapping inversions that span most of the X (fig. 2 and supplementary table S3, Supplementary Material online) and have permitted expression, copy number, and sequence divergence to accumulate between the inversion haplotypes.

We identified several genes that have diverged dramatically in copy number, and expression between X^{SR} and X^{ST} including a paralog of *JASPer* (*JIL-1 Anchoring and Stabilizing Protein*, also known as *dP75*). *JASPer* has multiple new paralogs in the genus *Teleopsis*, and these have amplified in expression and copy number on X^{SR} relative to X^{ST} (fig. 3 and supplementary table S4, Supplementary Material online). In *D. melanogaster*, *JASPer* positively regulates the transition of heterochromatin to euchromatin with its partner JIL-1, is involved in X-chromosome dosage compensation (Albig et al. 2019), and is essential during oogenesis (Dou et al. 2020). A partial paralog of *JASPer* appears to have formed tandem arrays on X^{SR}, with only part of the gene—the chromatin-binding PWWP domain—being duplicated and upregulated in expression in X^{SR} (fig. 3). This amplified partial paralog (*Td-JASPer:X:18.181Mbp*) is found in a region overlapped by five of the six inversions we identified as distinguishing X^{SR} and X^{ST} (fig. 2). Interestingly, the excess coverage detected in the Illumina sequencing (~X) is much more than can be explained by the spanning PacBio reads at the copy inserted at18.181 Mbp and we were unable to identify additional copies inserted elsewhere on X^{SR}. It may be that additional copies are present on X^{SR}, but that the flanking sequences are also X^{SR} specific and/or were not assembled completely in the reference genome. *Td-JASPer:X:18.181Mbp* is also testis specific, whereas the full-length paralog closest in sequence to the *D. melanogaster* ortholog (*Td-JASPer:C2:156Mbp*) is primarily expressed in the ovary (fig. 3*B*).

PWWP domains, like those found in *JASPer*, primarily bind to H3K36me3 chromatin marks (Albig et al. 2019; Dou et al. 2020), which are associated with regions of active gene expression. Although we do not know the target of any *JASPer* paralog in *T. dalmanni*, it would seem likely this function is conserved given it occurs across *metazoa*. Plausibly, the PWWP-only *JASPer* duplicates expressed from X^{SR} might bind to their targets but are unable to recruit *JIL-1* and therefore fail to properly activate target genes, similar to a dominant negative allele. *JIL-1/JASPer* binding is also implicated in positive regulation of expression of *Gypsy5* retroelements, which are found in arrays in *Drosophila* telomeres (Albig et al. 2019), providing a plausible connection between this drive candidate and differences in expression of certain TEs. We found that Gypsy elements are generally less common on the X chromosome and several Gypsy families show a male bias in genomic coverage suggesting bias towards insertion on the Y (supplementary table S3, Supplementary Material online). The identification of a partial duplicate of a chromatin-binding protein bears a striking resemblance to another case of Dipteran meiotic drive, the so-called "Paris" SR drive system in *D. simulans* (Montchamp-Moreau et al. 2006). As in the present case, the SR copy of the causal gene—HP1D2—retains one domain—the histone-binding chromodomain, but not the protein binding chromoshadow (CSD) domain (Helleu et al. 2016). Knocking out the

Table 1
Genes Diverging under Positive Selection between *SR* and ST X Chromosomes

Name	Gene Age	Differential Coverage	Differential Expression	X^{SR}–X^{ST} *dN/dS*	Interspecific *dN/dS*
ORF2072	Single copy	No	No	2.632	NA
Barren	Duplicate	No	No	2.273	NA
ORF702	Orphan	No	No	2	NA
ORF1962	Orphan	No	No	1.887	0.73529412
ORF1691.1	Orphan	No	Up in X^{SR}	1.667	1.28205128
Maelstrom	Duplicate	No	No	1.587	0.34602076
La autoantigen-like	Single copy	No	No	1.471	0.36363636
ORF2361	Orphan	No	No	1.471	NA
Calmodulin	Single copy	No	No	1.429	NA
ORF2114	Orphan	No	Up in X^{ST}	1.37	NA
CG10195	Single copy	No	No	1.351	0.135318
ORF2326	Orphan	No	No	1.333	NA
ORF705	Orphan	No	Up in X^{ST}	1.299	0.70921986
CG34109	Single copy	No	No	1.205	0.31446541
ORF1530	Orphan	Up in X^{ST}	Up in X^{ST}	1.19	NA
ORF1691.2	Orphan	No	Up in X^{SR}	1.19	1.31578947
La autoantigen-like	Single copy	No	No	1.163	0.1980198
Pol32	Single copy	No	Up in X^{ST}	1.111	0.42194093
Heat shock protein 60	Duplicate	No	No	1.099	NA
Tetraspanin 29Fb	NA	Up in X^{SR}	Up in X^{SR}	1.087	NA
ORF739	Orphan	No	Up in X^{ST}	1.064	NA
ORF944	Orphan	No	No	1.064	No
Cyclin K	Duplicate	Up in X^{SR}	Up in X^{SR}	1.053	0.70921986
CG15435	Single copy	No	No	1.042	0.24630542
ORF1457	Orphan	No	No	1.02	0.36630037

standard copy of HP1D2 or deleting its CSD partly recapitulates drive, and HP1D2 binds specifically to Y-chromosomal heterochromatin. If the chromatin targets of *PWWP-JASPer* are similarly Y specific, development of Y-bearing, but not X^{SR}-bearing, sperm could be similarly disrupted.

The widespread divergence of X-linked sequences between X^{SR} and X^{ST} has had impacts beyond meiotic drive itself. Not only are hundreds of genes differentially expressed or varying in copy number, we further find that many X-linked genes are diverging rapidly (*dN/dS* > 1) between the chromosome types (fig. 4). Unsurprisingly, new genes and testis-specific genes evolved more quickly both between species and between the SR types, though the patterns of evolution are loosely correlated between the comparisons ($r = 0.56$). The model built to predict X^{SR}–X^{ST} *dN/dS* explains less (19%) of the total variation in *dN/dS* than the interspecific comparison (30%), indicating that drive-associated protein evolution may be more idiosyncratic and less driven by "typical" trends in molecular evolution (e.g., gene expression level and age). Whereas some of this difference is presumably due to stronger genetic drift occurring on X^{SR} due to a lack of recombination (Muller 1964), some genes are evolving under positive selection (table 1) and these are rarely (2 of 13 cases) the same genes as those evolving adaptively between species. Among those diverging adaptively between X^{SR} and X^{ST} but not between species are a number of novel and uncharacterized genes, the genome defense gene *maelstrom*, Pol32, a DNA polymerase subunit involved in DNA repair at chromosome fragile sites (Ji et al. 2019), and RNA-binding protein La autoantigen like. Given JASPer's putative role in chromatin regulation, the identification of *maelstrom* and *Pol32* is intriguing. *Maelstrom* acts to silence transposable elements by promoting the spread of heterochromatic states to nearby genes (Sienski et al. 2012), and chromatin states have been found to be tightly integrated with genome integrity via multiple DNA repair pathways (reviewed in Papamichos-Chronakis and Peterson 2013), including at chromosome fragile sites.

Overall, these results extend the work on the genomic impacts of SR meiotic drive (reviewed in Courret et al. 2019) to a charismatic organism—*Teleopsis* stalk-eyed flies—and suggest that although the specific mechanisms of drive are idiosyncratic, drivers may be converging repeatedly on similar genomic vulnerabilities.

Material and Methods

Genome Assembly of *T. dalmanni s.s.*

A draft genome assembly for *T. dalmanni*, NLCU01000000, was created with a combination of Roche 454, Illumina,

and Pacific Biosciences sequence data (supplementary table S1, Supplementary Material online) using MaSuRCa (Zimin et al. 2013) and is available on Ensembl metazoa, GCA002237135v2 (Kersey et al. 2018). A chromosome-level assembly (NLCU04000000) was then created by incorporating chromatin conformation information and validated with a high-density linkage map.

All DNA sequences used in the assembly were obtained from an inbred population (line "2A") of *T. 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 and 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 (Poynton et al. 2018). Male and female genomic short-read resequencing data were aligned to each scaffold using NextGenMap v0.5.5 (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. The NLCU01 assembly was then filtered for haplotigs and other redundant sequences using Purge Haplotigs v1.1.1 (Roach et al. 2018) prior to scaffolding to create the NLCU04000000 assembly. A prior scaffolding attempt was also done without first haplotig filtering resulting in an assembly not used in the present analysis (NLCU02000000).

For scaffolding, chromatin conformation capture data were 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 v. 0.7.15-r1144-dirty (Li and Durbin 2009) with the -5SP and -t 8 options specified, and all other options default. SAMBLASTER (commit 37142b37e4f0026e1b83ca3f1545d1807ef77617, 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 nonprimary and secondary alignments. Putative misjoined contigs were broken using Juicebox v1.11.08 (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 (commit aa3382b6e63f7b99e92a9d95c553ef1c6a5a6a38) 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 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 438.2 Mb, comprising 95.7% of the filtered MaSuRCa assembly. Finally, Juicebox was again used to correct the remaining scaffolding errors.

Linkage groups used to constrain and later validate the gene order in 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 non-inbred 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 that 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 for each scaffold by combining all loci present on a scaffold into a single "super locus." Reads were aligned using BWA

v0.7.17 (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 nonbackcross 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 nonbackcross 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 nonbackcross (Wilkinson et al. 2014) allele, which we defined as the less common allele across all female individuals. Markers were removed if they violated expected allele ratios for a backcross using a binomial test against the expectations described above, or if more than 20% of female individuals were found to carry **only** the nonbackcross 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 five 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 v4.1 (Stam 1993) to assign the top 1,000 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. We used regression mapping, rather than maximum likelihood, because it is less sensitive to missing genotype data (Van Ooijen 2006) as is typical for multiplex shotgun sequencing data sets. 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 position of scaffolds which were found in both maps. Chromosomal synteny of the assembly scaffolds with *D. melanogaster* chromosome arms was assessed by alignment of a set of 7,634 previously annotated (Baker et al. 2016) 1-to-1 *Drosophila* orthologs to the genome assembly using GMAP v2019-12-01 (Wu and Watanabe 2005) (–npaths=1 –format=gff3_gene –min-identity=0.9). Assemblies were assessed for completeness using BUSCOv5.0.0 (Seppey et al. 2019) comparing to pan-Dipteran (diptera_odb10 2020-08-05) and pan-eukaryotic (eukaryota_odb10 2020-09-10) genesets.

Draft Genome Assembly of *T. dalmanni* Species 2

We previously 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 #158667). 1 μg of DNA was sent to the New York Genome Center (NYGC) where it was prepped with the Chromium Genome-linked read kit (10× 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 (Weisenfeld et al. 2017). 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. Although incomplete, this genome was sufficient to determine inversion history and polarize molecular divergence for some genes.

Short-Read Resequencing of SR and ST Males

To identify sequence and structural variations specific to X^{SR}, we sequenced DNA from replicate pools of SR males (males with female-biased offspring SRs), 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) that were previously phenotyped for offspring SR (Paczolt et al. 2017). When an excess of individuals was available from a collection and *SR* category, genotype data from nine X-linked STR loci from the same analysis were used to avoid oversampling closely related individuals (e.g., male full siblings from the same brood). Haplotype diversity within pools was not significantly different to haplotype diversity among all candidate males for that pool (following Christianson et al. 2011, Dunnett's *t*-test, $P > 0.05$ for all comparisons, supplementary table S8, Supplementary Material online). DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen #69504) and quantified using the PicoGreen Quant-IT dsDNA quantification kit (Thermofisher Q33130). Pools were then assembled using an equimolar amount of DNA from each sample. Sample size for each pool ranged from 15 to 18 individuals

(supplementary table S8, Supplementary Material online). Six barcoded libraries were prepared and multiplexed on two lanes of a HiSeq1500 set to a RapidRun mode to generate 150-bp paired-end sequences. Bam-formatted alignments of these libraries to the genome were produced using NextGenMap v0.5.5 (Sedlazeck et al. 2013) with default parameters and used in subsequent analyses. Pairwise genetic diversity for each pool and F_{ST} between each pairwise combination of pool-seq data was calculated from the pooled resequencing alignments in 5-kb nonsliding windows using popoolation2 (Kofler et al. 2011). After pool-seq 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 supplementary table S8, Supplementary Material online, (Paczolt et al. 2017), and a previous work (Christianson et al. 2005; Swallow et al. 2005) had suggested that *Td2* would not be present in the collection sites we visited. Both SR pools and one ST pool (Kanching ST2) were entirely composed of *T. dalmanni* s.s. (supplementary Table S8, Supplementary Material online) so all analyses where excess polymorphism (potentially caused by 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 (supplementary table S9, Supplementary Material online).

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 an 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) to distinguish X^{SR} and X^{ST} sons (Paczolt et al. 2017). DNA from 46 X^{SR} sons was then extracted using the Gentra PureGene Tissue Kit (Qiagen #158667) 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 a 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 v0.2.7 with default parameters, and structural variants were called using Sniffles v1.0.12 (Sedlazeck et al. 2018), requiring at least two 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 with the read-pair orientation in the SR and ST pool-seq data at the breakpoint using IGV (Thorvaldsdottir et al. 2013). An inversion was considered validated if reads from the SR pool-seq samples but none of the ST pool-seq samples agreed with the sniffles call at that position (supplementary table S3, Supplementary Material online). 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) support another arrangement (scaffold present near breakpoint), or 4) are uninformative (no scaffolds map near the breakpoint) (supplementary table S3, Supplementary Material online).

Transposable Element Annotation

Transposable elements were annotated in the *T. dalmanni s.s.* assembly using RepeatModeler v. 1.0.4 (Smit and Hubley 2008) with default parameters and the NLCU04000000 assembly as input. RepeatModeler seed alignments and consensus sequences were submitted to dfam. Resulting consensus fasta formatted TE sequences were input into RepeatMasker open4.0.9 (Smit et al. 2013) with the assembly as the reference, producing a repeat-masked reference genome and repeat annotations. The tool One code to find them all v2014 (Bailly-Bechet et al. 2014) was used with the RepeatMasker output (.out) to count the numbers and locations of each type of insertion in the *T. dalmanni* genome. These were compared with the RepeatMasker annotations for two other Dipterans (*D. melanogaster* dm6 RepeatMasker open-4.0.6 and *Anopheles gambiae* anoGam1 RepeatMasker open-4.0.5) analyzed using the same procedure. Further classification of element superfamilies followed prior universal classification schemes (Wicker et al. 2007; Makałowski et al. 2019). X (or autosomal) chromosome bias in the distribution of elements of each annotated family was determined by comparing the observed number of intact elements of each element type on the X to the number expected assuming the X comprises 22.5% of the genome using a chi-squared goodness-of-fit test. To correct for multiple testing, we applied a Benjamini–Hochberg (BH) 5% false discovery rate (FDR).

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 consensus sequences. Sites were subsampled to 20× coverage and were discarded if coverage was less than 20× in any sample. 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 male testes was assessed using TEtools copyright (C) 2015 Laurent Modolo (Lerat et al. 2016), using the default settings and including alignment with bowtie2 v2.2.4 (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 (BioProject PRJNA240197). Unannotated repetitive elements ("Unknown" interspersed and simple repeats) were removed after normalization but prior to differential expression analysis with DESeq2 v1.32.0 (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 the TE family copy number within the SR and ST genomic resequencing pools and in male and female genomic libraries (SRS2309195–SRS2309198). We identified potentially Y-inserted elements by qualitative comparison DESeq2 normalized genomic read counts from one male and one female HiSeq library.

Annotation of Gene Duplication and Expression and Differential Coverage

A set of protein-coding genes annotated from a transcriptome assembly (BioProject PRJNA240197) was aligned to the three largest scaffolds using GMAP v2019-12-01 allowing for up to ten 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 RepeatMasker. 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) were aligned to the genome using HISAT2 (Kim et al. 2019) v2.0.1-beta (–dta -X 800). Genomic pool-seq data were aligned to the genome using NextGenMap v0.5.5 (Sedlazeck et al. 2013). Genomic expression and coverage in each library were estimated using Featurecounts v2.0.0 (Liao et al. 2014). Differential coverage and differential expression were each assessed from the Featurecounts read count matrix 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 with the *T. 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 v2019-12-01 (Wu and Watanabe 2005) (–nosplicing –format=samse). A *Td2* consensus was called from the GMAP alignment of *Td2* scaffolds to the *T. dalmanni* genome by sorting and indexing with samtools (Li et al. 2009) v1.3.1, then calling the consensus with bcftools v 1.9 (bcftools call –ploidy 1 -mA) (Li 2011). 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.

For molecular evolutionary comparisons, 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 one replicate from each collection site (Gombak ST1 and Kanching ST2) and consensus called as above. The best alignment of the coding regions of genes previously (Baker et al. 2016) assembled and annotated using data from a multitissue RNAseq experiment (BioProject PRJNA240197) was localized to the genome via alignment with GMAP v2019-12-01 (–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 GffRead v0.12.7 (Pertea and Pertea 2020) and the mRNA gff annotations from GMAP. Some genes contained in-frame stop codons in one or more libraries (e.g., stop codons were polymorphic). These genes were trimmed to the longest open reading frame present in all libraries, and if what remained was longer than 50 amino acids, they were retained. Genes were also excluded if they contained only ambiguity sequence ("N") in one or more of the consensus genomes or were less than 50 aa in length. After exclusions, we counted nonsynonymous divergent sites and calculated pairwise *dN/dS* of X^{SR} versus X^{ST} for 2,642 X-linked genes and for *T. dalmanni* versus Td2 for 9,525 genes using the SNAP utility v6/15/98 (Korber et al. 2000). For the linear models and ANOVA predicting ln(*dN/dS*), genes were identified as being testis specific based on prior analyses of a multitissue transcriptome (Reinhardt et al. 2014; Baker et al. 2016). Following Baker et al. 2016, gene duplication events occurring within *Diopsidae* were assigned based on a four-species transcriptomic analysis and protein coding genes without identifiable *Dipteran* orthologs were designated as orphan genes. In addition, genes that were previously designated as single copy were assessed as duplicates if they mapped (GMAP, see above) to multiple locations (paths) in the genome.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Melanie Kirk, Nathaniel Lowe, Wyatt Shell, George Ru, and Gabriel Welsh for the assistance with analysis, sample preparation, and fly rearing; Philip Johns and Max Brown for fly collections; Shawn Sullivan and Hayley Mangelson 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). We thank two anonymous reviewers for suggestions that improved the manuscript. Funding for this work was provided by the National Science Foundation grants DEB-0951816 to R.H.B., DEB-0952260 to G.S.W., and DEB1257053 and DEB1950078 to J.H.W; by the USDA National Institute of Food and Agriculture grant 2018-67015-28199 to A.V.Z; by the University of Maryland by the Research Foundation for the State University of New York; and by the Geneseo Foundation.

Author Contributions

J.A.R., R.H.B., and G.S.W. prepared the manuscript. J.A.R., A.V.Z., K.A.P., C.L., J.H.W., G.S.W., and R.H.B. analyzed the data. C.Y.H. and G.S.W. provided sequencing data.

Data Availability

Raw data and genome assemblies used in this project are available on NCBI BioProjects PRJNA655584 (sex ratio resequencing), PRJNA391339 (Teleopsis dalmanni s.s. genome assembly), PRJNA662429 (multiplexed shotgun genotyping), and PRJNA659474 (Teleopsis dalmanni sp2 aka isolate:KP12SP2M1 genome assembly). The Whole Genome Shotgun project for Teleopsis dalmanni s.s. has been deposited at GenBank under the accession NLCU00000000. The version described in this paper is version NLCU04000000. The following additional data sets are available on the digital repository at University of Maryland (DRUM): transposable element family consensus fasta sequences for *Teleopsis dalmanni* (ID 1903/26380), the MSG genotype matrix (ID fgxn-tuaf), and gff3 formatted gene annotations (ID gfqi-iktk).

Literature Cited

Albig C, et al. 2019. JASPer controls interphase histone H3S10 phosphorylation by chromosomal kinase JIL-1 in *Drosophila*. Nat Commun. 10:5343.

Andolfatto P, et al. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21:610–617.

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

Baker RH, et al. 2016. Spermatogenesis drives rapid gene creation and masculinization of the X chromosome in stalk-eyed flies (Diopsidae). Genome Biol Evol. 8:896–914.

Baker RH, Wilkinson GS. 2010. Comparative genomic hybridization (CGH) reveals a neo-X chromosome and biased gene movement in stalk-eyed flies (genus *Teleopsis*). PLoS Genet. 6:e1001121.

Begun DJ, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5: e310.

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

Bickhart DM, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet. 49:643–650.

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

Burton JN, et al. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 31:1119–1125.

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

Christianson SJ, Brand CL, Wilkinson GS. 2011. Reduced polymorphism associated with X chromosome meiotic drive in the stalk-eyed fly *Teleopsis dalmanni*. PLoS One. 6:e27254.

Christianson SJ, Swallow JG, Wilkinson GS. 2005. Rapid evolution of postzygotic reproductive isolation in stalk-eyed flies. Evolution 59:849–857.

Cotton AJ, Földvari M, Cotton S, Pomiankowski A. 2014. Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). Heredity (Edinb) 112:363–369.

Courret C, Chang CH, Wei KH, Montchamp-Moreau C, Larracuente AM. 2019. Meiotic drive mechanisms: lessons from *Drosophila*. Proc Biol Sci. 286:20191430.

Curtsinger JW, Feldman MW. 1980. Experimental and theoretical analysis of the "sex-ratio" polymorphism in *Drosophila pseudoobscura*. Genetics 94:445–466.

Dou K, et al. 2020. *Drosophila* P75 safeguards oogenesis by preventing H3K9me2 spreading. J Genet Genomics. 47:187–199.

Durand NC, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3:99–101.

Dyer KA, Charlesworth B, Jaenike J. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc Natl Acad Sci U S A. 104:1587–1592.

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

Finnegan SR, et al. 2019. Meiotic drive reduces egg-to-adult viability in stalk-eyed flies. Proc Biol Sci. 286:20191414.

Fuller ZL, et al. 2020. Extensive recombination suppression and epistatic selection causes chromosome-wide differentiation of a selfish sex chromosome in *Drosophila pseudoobscura*. Genetics 216: 205–226.

Gell SL, Reenan RA. 2013. Mutations to the piRNA pathway component *aubergine* enhance meiotic drive of segregation distorter in *Drosophila melanogaster*. Genetics 193:771–784.

Grognet P, Lalucque H, Malagnac F, Silar P. 2014. Genes that bias mendelian segregation. PLoS Genet. 10:e1004387.

Hamilton WD. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. Science 156:477–488.
Helleu Q, et al. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. Proc Natl Acad Sci U S A. 113:4110–4115.
Houtchens K, Lyttle TW. 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.
Hu W, et al. 2017. A large gene family in fission yeast encodes spore killers that subvert Mendel's Law. Elife 6:e26057.
Jaenike J. 2001. Sex chromosome meiotic drive. Annu Rev Ecol Syst. 32:25–49.
Ji J, Tang X, Hu W, Maggert KA, Rong YS. 2019. The processivity factor Pol32 mediates nuclear localization of DNA polymerase delta and prevents chromosomal fragile site formation in *Drosophila* development. PLoS Genet. 15:e1008169.
Johns PM, Wolfenbarger LL, Wilkinson GS. 2005. Genetic linkage between a sexually selected trait and X chromosome meiotic drive. Proc Biol Sci. 272:2097–2103.
Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res 12: 656–664.
Kersey PJ, et al. 2018. Ensembl genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 46: D802–D808.
Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37:907–915.
Kofler R, Gómez-Sánchez D, Schlötterer C. 2016. PoPoolationTE2: comparative population genomics of transposable elements using pool-seq. Mol Biol Evol. 33:2759–2764.
Kofler R, Pandey RV, Schlotterer C. 2011. Popoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.
Korber B, Rodrigo A, Learn G. 2000. HIV signature and sequence variation analysis. in Computational analysis of HIV molecular sequences. Dordrecht: Kluwer Academic Publishers. p. 55–72.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Larracuente AM. 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 Evol Biol. 14:233.
Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. 2016. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Res. 45:e17.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.
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, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
Liao Y, Smyth GK, Shi W. 2014. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.
Lieberman-Aiden E, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293.
Lin C-J, et al. 2018. The hpRNA/RNAi pathway is essential to resolve intragenomic conflict in the *Drosophila* male germline. Dev Cell. 46: 316–326.e5.
Lindholm AK, et al. 2016. The ecology and evolutionary dynamics of meiotic drive. Trends Ecol Evol. 31:315–326.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.
Lyttle TW. 1991. Segregation distorters. Ann Rev Genet. 25:511–557.
Mackintosh C, Pomiankowski A, Scott MF. 2021. X-linked meiotic drive can boost population size and persistence. Genetics 217:1–11.
Makałowski W, Gotea V, Pande A, Makałowska I, Anisimova M. 2019. Transposable elements: classification, identification, and their use as a tool for comparative genomics. Evolutionary genomics. New York (NY): Springer New York. p. 177–207.
McLaughlin RN Jr, Malik HS. 2017. Genetic conflicts: the usual suspects and beyond. J Exp Biol. 220:6–17.
Meade L, Finnegan SR, Kad R, Fowler K, Pomiankowski A. 2019. Maintenance of fertility in the face of meiotic drive. Am Nat. 195:743–751.
Montchamp-Moreau C, Ogereau D, Chaminade N, Colard A, Aulard S. 2006. Organization of the sex-ratio meiotic drive region in *Drosophila simulans*. Genetics 174:1365–1371.
Muirhead CA, Presgraves DC. 2021. Satellite DNA-mediated diversification of a sex-ratio meiotic drive gene family in *Drosophila*. Nat Ecol Evol. 5:1604–1612.
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 106:2–9.
Nagao A, et al. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. RNA 16:2503–2515.
Nuckolls NL, et al. 2017. Wtf genes are prolific dual poison-antidote meiotic drivers. Elife 6:e26033.
Paczolt KA, Reinhardt JA, Wilkinson GS. 2017. Contrasting patterns of X-chromosome divergence underlie multiple sex-ratio polymorphisms in stalk-eyed flies. J Evol Biol. 30:1772–1784.
Papamichos-Chronakis M, Peterson CL. 2013. Chromatin and the genome integrity network. Nat Rev Genet. 14:62–75.
Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. F1000Res 9:ISCB Comm J-304.
Poynton HC, et al. 2018. The toxicogenome of *Hyalella azteca*: a model for sediment ecotoxicology and evolutionary toxicology. Environ Sci Technol. 52:6009–6022.
Presgraves DC, Severance E, Wilkinson GS. 1997. Sex chromosome meiotic drive in stalk-eyed flies. Genetics 147:1169–1180.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.
Rao SSP, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–1680.
Reinhardt JA, et al. 2014. Meiotic drive impacts expression and evolution of X-linked genes in stalk-eyed flies. PLoS Genet. 10: e1004362.
Roach MJ, Schmidt SA, Borneman AR. 2018. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 19:460.
Sedlazeck FJ, et al. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 15:461–468.
Sedlazeck FJ, Rescheneder P, von Haeseler A. 2013. Nextgenmap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29:2790–2791.
Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227–245.
Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151:964–980.
Smit A, Hubley R. 2008. RepeatModeler Open-1.0.
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0.

Smith JM, Haigh J. 1974. The hitchhiking effect of a favourable gene. Genet Res. 23:23–25.
Stam P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. The Plant J. 3:739–744.
Swallow JG, Wallace LE, Christianson SJ, Johns PM, Wilkinson GS. 2005. Genetic divergence does not predict change in ornament expression among populations of stalk-eyed flies. Mol Ecol. 14: 3787–3800.
Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14:178–192.
Unckless RL, Clark AG. 2014. Sex-ratio meiotic drive and interspecific competition. J Evol Biol. 27:1513–1521.
Van Ooijen JW. 2006. Joinmap® 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV. Wageningen 33:10–1371.
Vedanayagam J, Lin C-J, Lai EC. 2021. Rapid evolutionary dynamics of an expanding family of meiotic drive factors and their hpRNA suppressors. Nat Ecol Evol. 5:1613–1623.
Vicoso B, Bachtrog D. 2015. Numerous transitions of sex chromosomes in Diptera. PLoS Biol. 13:e1002078.
Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid genome sequences. Genome Res. 27: 757–767.
Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc Natl Acad Sci U S A. 108(Suppl 2): 10863–10870.
Werren JH, Nur U, Wu CI. 1988. Selfish genetic elements. Trends Ecol Evol. 3:297–302.
Wheeler D, Redding AJ, Werren JH. 2013. Characterization of an ancient lepidopteran lateral gene transfer. PLoS One. 8:e59262.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Wilkinson GS. 1993. Artificial sexual selection alters allometry in the stalk-eyed fly *Cyrtodiopsis dalmanni* (Diptera: Diopsidae). Gene Res. 62:213–222.
Wilkinson GS, Christianson SJ, Brand CL, Ru G, Shell W. 2014. Haldane's rule is linked to extraordinary sex ratios and sperm length in stalk-eyed flies. Genetics 198:1167–1181.
Wilkinson GS, Johns PM, Kelleher ES, Muscedere ML, Lorsong A. 2006. Fitness effects of X chromosome drive in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. J Evol Biol. 19:1851–1860.
Wilkinson GS, Presgraves DC, Crymes L. 1998. Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. Nature 391:276–279.
Wolfenbarger LL, Wilkinson GS. 2001. Sex-linked expression of a sexually selected trait in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. Evolution 55:103–110.
Wu CI, Beckenbach AT. 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 CI, Lyttle TW, Wu ML, Lin GF. 1988. Association between a satellite DNA sequence and the responder of segregation distorter in *D. melanogaster*. Cell 54:179–189.
Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859–1875.Zimin AV, et al. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677.

Associate editor: Rebecca Zufall