Recurrent gene amplification on Drosophila Y chromosomes suggests cryptic sex chromosome drive is common on young sex chromosomes

Chris Ellison and Doris Bachtrog

University of California Berkeley, Berkeley, CA 94720

*corresponding author: dbachtrog@berkeley.edu

Theory predicts that selfish genetic elements that increase their transmission are prone to originate on sex chromosomes but create strong selective pressure to evolve suppressors due to reduced fertility and distorted population sex ratios. Here we show that recurrent genetic conflict over sex chromosome transmission appears to be an important evolutionary force that has shaped gene content evolution of sex chromosomes in Drosophila. We demonstrate that convergent acquisition and amplification of spermatid expressed gene families are common on Drosophila sex chromosomes, and especially on recently formed ones, and harbor characteristics typical of meiotic drivers. We carefully characterize one putative novel cryptic sex chromosome distortion system that arose independently several times in members of the *Drosophila obscura* group. Co-amplification of the *S-Lap1/GAPsec* gene pair on both the X and the Y chromosome occurred independently several times in members of the *D. obscura* group, where this normally autosomal gene pair is sex-linked due to a sex chromosome - autosome fusion. Investigation of gene expression and short RNA profiles at the *S-Lap1/GAPsec* system suggest that meiotic drive and suppression likely involves RNAi mechanisms. Our finding suggests that recurrent conflict over sex chromosome transmission has shaped widespread genomic and evolutionary patterns, including the epigenetic regulation of sex chromosomes, the distribution of sex-biased genes, and the evolution of hybrid sterility.

Introduction

Selfish genetic elements whose evolutionary trajectories are in conflict with those of their host were first described almost 100 years ago [1]. However, only in the recent decade has it become apparent that the antagonistic coevolution resulting from genetic conflict has shaped genome content and structure across the tree of life, from bacteria to plants and animals [2]. Antagonistic coevolution can occur between organisms, as in the evolutionary "arms race" experienced between pathogens and their hosts, or within genomes, where genetic elements can manipulate meiosis (or gametogenesis) so that they are transmitted to more than 50% of offspring (so called "segregation distorters" or "meiotic drivers")[3]. These processes can leave behind a variety of distinct genetic signatures. For example, genes involved in pathogen virulence and host resistance consistently show

elevated levels of amino acid substitutions (i.e. dN/dS) whereas genes involved in intra-genomic conflict, such as segregation distorters and their suppressors, often have high rates of lineage-specific duplications and gene amplifications (see below).

While segregation distorters can arise on all chromosomes, they are particularly prone to evolve on sex chromosomes [4]. Most characterized segregation distortion systems involve a distorting locus that targets a sensitive responder locus on its homolog, so that the chromosome where the driver resides is transmitted to more than 50% of the offspring of a heterozygous carrier [5]. A distorter must be able to discriminate its host chromosome from its homolog, and the distorter and responder loci must be in strong linkage to avoid the generation of suicide chromosomes that carry the distorter and a sensitive responder [4]; both of these conditions are met on differentiated, non-recombining X and Y chromosomes. The propensity of sex chromosomes to evolve distorting alleles could be strong enough to drive several unique characteristics of sex chromosomes, and it has been suggested that heterochromatinization of the Y and meiotic inactivation of the X during spermatogenesis evolved in part as protective measures to silence such alleles [5,6]. If this is the case, young sex chromosomes may be particularly susceptible to evolving sex ratio distortion because they often lack these features.

Disruption of equal transmission of the X and Y chromosomes will skew the population sex ratio and thus create strong selective pressure to evolve loci that repress sex chromosome drive [7,8]. Sex chromosome drive may thus be prevalent in evolution but is often cryptic, since bursts of meiotic drive elements arising on sex chromosomes should be followed by the quick invasion of suppressor alleles that restore equal sex ratios [4,7]. Cryptic drive systems are thus typically only revealed through detailed genetic manipulations or crosses between populations or species.

While cryptic drive systems cannot be observed directly, recurrent bursts of invading drivers and repressors may leave common footprints at patterns of genome evolution that are indicative of past meiotic drive. In particular, molecular characterization of meiotic drive systems in both mammals and fruit flies has found the following commonalities: In several instances, suppressors were found to share common sequences with the distorter itself [9-12]. This is the case for the Winters sex-ratio system within *D. simulans*, for instance. Here, two X-linked distorters, *Dox* and *MDox* (with *Dox* being derived from *MDox*), are both necessary for drive [9,10], and two autosomal genes (called *Nmy* and *Tmy*) that both originated from *Dox*/*MDox* suppress this drive [13]. Often, both the distorter and suppressor alleles are not only homologous to each other, but also amplified on both the X and Y chromosome [12,14,15]. A cryptic sex chromosome drive in mouse, for example, involves the convergent acquisition and amplification of the same gene families (*Slx*/*Sly*) on both the X and Y chromosome, and careful experimentation has shown that the co-amplified genes are in a co-evolutionary battle over sex chromosome transmission, whereby the X-and Y-linked copies of a gene family directly silence each other [11,12]. *Sly* knockdowns show female-biased sex ratios, while *Slx* deficiency causes a sex ratio distortion towards males. A

similar mechanisms of cryptic segregation distortion has been implicated in the *Stellate*/ *Suppressor of Stellate* (*Ste*/*Su*(*Ste*) system in *D. melanogaster*, where the expression of the X-linked gene *Ste* leads to the production of defective sperm, and *Su*(*Ste*), which is a multi-gene copy of *Ste* that moved to the Y-chromosome, silences *Ste* [16,17]. Thus, cryptic sex chromosome drive has repeatedly led to characteristic patterns of gene amplification of homologous genes on both the X and the Y chromosomes (**Figure 1A**).

In several cases, suppression is likely mediated by sequence homology between the suppressor and distorter, through RNAi mechanisms resulting in posttranscriptional gene silencing. For example, piRNA pathway mutants enhance the drive caused by segregation distorter in *D. melanogaster* (the *Sd*/*Rsp* system) [18]. In addition, the sex-ratio driver Winters in *D. simulans* (*Dox*/*Nmy*) is suppressed through small RNA-based silencing [9,10,13], where *Nmy* and *Tmy* encode related hairpin RNAs that generate siRNA that repress their paralogous distorters [13]; knockout of RNAi pathway genes derepresses these sex-ratio drivers and disrupts spermatogenesis [13]. Similarly, the Y-linked *Su*(*Ste*) in *D. melanogaster* silences *Ste* through anti-sense expression and generation of small RNA's [16]. This shows that the emergence of homologous repeated sequences can resolve an intragenomic conflict, through invoking the RNAi pathway.

Thus, cryptic sex chromosome drive has repeatedly led to characteristic patterns of gene amplification of homologous genes on both the X and the Y chromosomes that are targeted by short RNAs. Sex chromosomes also provide opportunity for other types of intragenomic conflict to arise. Sex chromosomes are hotspots for sexually antagonistic genes (i.e. genes that are beneficial to one sex but detrimental to the other [19,20]), and Y chromosomes often contain genes involved in male fertility that can undergo antagonistic coevolution due to sperm competition [21,22]. Recent studies have shown that the Y chromosomes of many organisms contain testes-specific genes that have amplified in copy number [23-25]. Some of these Y-linked gene families, such as those in mice, have been shown to be involved in sex chromosome drive, whereas for other gene families, the extra copies may either act to increase gene dosage or prevent degeneration by providing a substrate for non-allelic gene conversion [26]. Alternatively, the extra copies may be neutral or even slightly deleterious, yet they remain on the Y due to the reduced efficiency of selection on this non-recombining chromosome [27]. One key signature that appears to be unique to Y-amplified genes involved in sex chromosome drive is that their X-linked homologs have duplicated as well. This pattern is consistent with antagonistic coevolution resulting from repeated bouts of sex ratio distortion and suppression.

Here, we use bioinformatics and functional genomic analyses to assess the importance of sex chromosome drive across Drosophila species. Consistent with a role for Y-amplified genes unrelated to sex chromosome drive (see discussion), we find hundreds of genes that appear to be present in multiple copies on the Y chromosomes of various Drosophila species. However, we also find a second category of Y-amplified genes whose X homolog has been duplicated as well. We show that species with young sex chromosomes have repeatedly evolved genes that

have co-amplified on the X and the Y, and show functions and expression patterns that are consistent with widespread meiotic conflict. Detailed investigation of one of these putative drive systems in the *pseudoobscura* group suggests that the putative driver is targeted by RNAi, and reveal that the same genes appear to have become involved in a meiotic conflict independently among multiple species of this group.

Results

Bioinformatic inference of co-amplified X and Y genes across Drosophila. To analyze gene content evolution and identify amplified X- and Y-linked genes, we sequenced both male and female genomic DNA in 26 Drosophila species from across the Drosophila phylogeny (**Table S1**). Roughly half of the species considered (11 out of 26) harbor the typical sex chromosome complement of Drosophila (that is, a single pair of ancient sex chromosomes, shared by all members of Drosophila). In addition to the ancestral pair of sex chromosomes, the other 15 species have a younger pair of "neo-sex" chromosomes, which formed when an autosome became fused to one or both of the ancient X and Y chromosomes (**Figure S1**). These younger "neo-sex" chromosomes are at various stages of evolving the typical properties of ancestral sex chromosomes, with neo-Y chromosomes losing their original genes and acquiring a genetically inert heterochromatic appearance, and neo-X chromosomes acquiring their unique gene content and sex-specific expression patterns [27,28]. We identified putative Y-amplified genes based on male and female gene coverage without relying on a genome assembly (**Figure 1B**, see Methods) and validated our approach using a high-quality genome assembly from *D. pseudoobscura* (**Figure S2**). Using this approach, we identify (depending on our cutoffs) 100s of genes that have multiple copies on the Y across the 26 species investigated (**Table S2, S3**, **Figure S3**). Genes might amplify on the Y for a variety of reasons, but co-amplification of testis genes may be a defining feature of genes evolving under meiotic conflict (see Discussion).

Among these multi-copy gene families on the Y, we found 46 amplified Y-linked genes with co-amplified X homologs in 10 species (**Table 1, Table S4**). We infer that the copy number of these co-amplified X/Y genes ranges from 8 copies on the Y up to 297 Y-linked copies (for an uncharacterized testis gene in *D. melanogaster* that amplified on the Y of *D. robusta*), with a mean copy number of 58 (see **Table S5**). We detect between 2-4 X-linked copies in our assemblies for these co-amplified X/Y genes (**Table S5**). However, the number of assembled X copies is likely an underestimate, since recent gene duplicates are typically collapsed in assemblies derived from short read sequencing data (see also Discussion). Several features of these co-amplified genes are typical of meiotic drive systems identified in flies and mammals. First, consistent with their proposed action of cheating fair meiosis, many are highly expressed in reproductive tissues in *D. melanogaster* (**Table S4**), and several of the genes have meiosis-related functions that may be exploited for sex chromosome drive (**Table 1, Table S4**). For example, we identify genes that are associated with spindle assembly involved in male meiosis (*fest*), chromosome segregation (*mars*), or male meiosis cytokinesis (*scra*), amongst others (**Table 1**).

Amplified Y genes were detected in each species investigated (**Table S2**). Interestingly, however, co-amplification of genes is much more common in species with recently added neo-sex chromosomes: of the 10 species where we found co-amplified genes, nine harbor neo-sex chromosomes (**Table 1**). Young sex chromosomes have many more genes that can evolve to cheat meiosis, and in the vast majority of cases the amplified genes were ancestrally present on the chromosome that formed the neo-sex chromosomes (**Table 1**). Also, young sex chromosomes may not yet be transcriptionally down-regulated during spermatogenesis, and sex-linked drivers are thus less likely to be silenced [5].

Characterization of *S-Lap1* / *GAPsec* gene family in *D. pseudoobscura*. We decided to more carefully characterize a putative cryptic drive system in *D. pseudoobscura,* a species with a high quality PacBio-based genome assembly. *D. pseudoobscura* currently lacks an assembled Y chromosome, but we inferred Y-linkage of contigs based on male and female read coverage using Illumina data (see Methods). We identified two adjacent genes that exist in multiple copies on the X and Y chromosome of *D. pseudoobscura*: *S-Lap1* (Dpse\GA19547) and *GAPsec* (Dpse\GA28668). *S-Lap1* is a member of a leucyl aminopeptidase gene family that encodes the major protein constituents of Drosophila sperm [29], while *GAPsec* is a GTPase activating protein. This situation is reminiscent of the *Segregation distorter* meiotic drive system in *D. melanogaster*, where the distorter is a tandem duplication of RanGAP, which is also a GTPase activator [30]. Both *S-Lap1* and *GAPsec* show partial tandem duplications on the X (**Figure 2A**), and we detect roughly 100 (partial and full-length) copies of both *S-Lap1* and *GAPsec* on the Y chromosome (the Y-linked contigs contain 127 copies of *S-Lap1* and 91 copies of *GAPsec*; **Figure 2B, Figure 3**). *S-Lap1* appears to have independently duplicated in different parts of the Drosophila phylogeny (**Figure 2C**), but phylogenetic clustering of *S-Lap1* and *S-Lap2* in certain species groups could also result from gene conversion homogenizing gene duplicates within a clade [31]. In most species of the Drosophila clade, *S-Lap1* and *S-Lap2* are similar in size; in *D. pseudoobscura* and its sister *D. persimilis*, however, *S-Lap2* has acquired a large deletion, removing more than half of the 3' end of the gene (**Figure 2A**; **Figure S4**). The partial duplication of *GAPsec*, on the other hand, is only found in *D. pseudoobscura* and its close relative *D. persimilis* (**Figure 2A, C**). *S-Lap1* and *GAPsec* probably dispersed onto the Y chromosome simultaneously, as there are multiple locations on the Y that preserve their X orientation (**Figure 2B**); note however, that the amplified copies on the Y do not include the tandemly duplicated copies. For both *S-Lap1* and *GAPsec*, the X- and Y-linked copies are highly expressed in testis of *D. pseudoobscura* (**Table S4, Figure 4, Table S6**), as expected for genes that try to cheat fair male meiosis.

In several cases, meiotic drive and suppression involves RNAi mechanisms [13]. We used stranded RNA-seq and small RNA profiles from wildtype *D. pseudoobscura* testes, to obtain insights into the molecular mechanism of the putative meiotic drive system involving *S-Lap1* and *GAPsec*. Interestingly, we detect both sense and antisense transcripts and short RNA's derived from *S-Lap1* (**Figure 4A, Table S6;** see **Figure S5, S6** for the size distribution of short RNAs mapping and **Figure S7, S8** for cross-mapping of RNA-seq and short RNA reads between different gene copies). In particular, stranded RNA-seq data reveal that the X-linked copy of *S-Lap1*-duplicate produces both

sense and anti-sense transcripts, resulting in the production of short RNA's (see **Figure 4A**, **Table S6**). Close inspection of this genomic region in *D. pseudoobscura* shows that the duplicated *GAPsec* gene is directly adjacent to where the *S-Lap1*-duplicate antisense transcript begins (**Figure 4A**). Intriguingly, this segment scores highly as a potential promoter sequence when using the Berkeley Drosophila Genome Project (BDGP) neural network promoter prediction algorithm [32] (score=0.89, highest possible score=1). Thus, this suggests that the partial duplication of *GAPsec* provided a promoter-like sequence in *D. pseudoobscura* for antisense transcription of *S-Lap1*-duplicate. Note that this putative promoter sequence is not part of the Y copies of *S-Lap1* (which lack the *GAPsec* duplicate), and we detect virtually no antisense transcripts that originate from the Y-amplified copies of *S-Lap1* (**Table S6**). CAGE-seq data support that the *GAPsec* duplicate generated a new TSS resulting in antisense transcript of *S-Lap1*-dup (**Figure 4A**).

These data suggest a possible model of cryptic sex chromosome drive in *D. pseudoobscura* (**Figure 4B**). *S-Lap1* is the most abundant sperm protein in *D. melanogaster* [29] but its function is unknown. If this protein is crucial for generating Y-bearing sperm, depletion of *S-Lap1* during spermatogenesis would result in drive. *S-Lap1* was duplicated in an ancestor of *D. pseudoobscura*, and a partial duplication of *GAPsec* (and truncation of *S-Lap1*-duplicate) created a TSS for anti-sense transcription of *S-Lap1* duplicate. Anti-sense production of *S-Lap1*-dup transcript may trigger siRNA production and silencing of *S-Lap1*, which could result in elimination of Y-bearing sperm. Acquisition of multiple copies of S-*Lap1* on the Y chromosome could restore *S-Lap1* function, and create a cryptic drive system in *D. pseudoobscura* (**Figure 4B**). Detailed molecular testing will be necessary to characterize the wildtype function of *S-Lap1*, and the cellular basis of the drive phenotype and its suppression.

Independent co-amplification of *S-Lap1* and *GAPsec* in other flies of the *Drosophila obscura* group. In most Drosophila species, *S-Lap1* and *GAPsec* are located on an autosome (chromosome 3L in *D. melanogaster*). In the *D. pseudoobscura* and *affinis* group, however, this chromosome arm fused with the sex chromosomes about 15MY ago, causing *S-Lap1* to become sex-linked. Intriguingly, patterns of molecular evolution at *S-Lap1* and *GAPsec* suggest that they may be involved in meiotic conflicts in several members of the *pseudoobscura* species group. We used high-quality PacBio genome assemblies for two additional members of that species group [33], *D. miranda*, which diverged form *D. pseudoobscura* about 2-4 MY ago, and *D. athabasca*, which diverged 10-15 MY ago [34]. While our Illumina sequencing-based approach failed to detect co-amplified X and Y genes in these species (see **Table 1**), examination of the assembled PacBio genomes revealed that both gene pairs independently amplified on the sex chromosomes of both *D. miranda* and *D. athabasca* (see **Figure 5**). We identify tandem duplications of the entire genomic region containing a total of 11 copies of *S-Lap1* and 6 copies of *GAPsec* on chromosome XR in *D. miranda*, and these two genes have amplified 5 and 4 times, respectively, on the neo-Y chromosome of *D. miranda* (**Figure 5B**). Both the nature of the duplication event and patterns of sequence evolution suggest that co-amplification of *S-Lap1* and *GAPsec* occurred independently in *D. miranda*. Here, the XR copies arose from individual duplications of these two genes followed by three tandem duplications of the entire

genomic region encompassing *S-Lap1* and *GAPsec*, producing a total of 11 copies of *S-Lap1* and 6 copies of *GAPsec*. All six of the X-linked copies of *GAPsec* are highly similar to each other (>99% identical), and more similar to their Y-linked paralogs than they are to *D. pseudoobscura* **(Figure 5A**). Also, *S-Lap1* and *GAPsec* appear to have moved only to a single location on the neo-Y of *D. miranda*, instead of being dispersed all across the Y, as in *D. pseudoobscura* (**Figure 5B**). Patterns of gene expression and short RNA production in *D. miranda* mimic that of *D. pseudoobscura*, with *SLap-1* (and *GAPsec*) transcripts being produced from both strands, and short RNA's are generated across that genomic region (**Figure 6**). The mechanism of antisense production appears to differ from *D. pseudoobscura* (**Figure 6**). In particular, transcriptional read-through at both *S-Lap1* and *GAPsec* appear to generate anti-sense transcripts of both genes. Again, this would provide dsRNA targets for both *S-Lap1* and *GAPsec* in *D. miranda* that trigger the RNAi response, and select for copy number increases of those genes on the neo-Y, to silence the driving X. Close inspection of this genomic region in *D. miranda* reveals sequence differences between the X and Y copies that may account for antisense production at X-linked gene copies. In particular, we detect a polyadenylation signal (AATAAA) for *GAPsec* that is present in most (3 of the 4) Y copies, and in the homologous copy on the X in *D. pseudoobscura*, but which is missing in the *D. miranda* X-linked copies of *GAPsec*. This mutational event could account for the production of read-through transcripts on the X of *D. miranda*, leading to production of antisense transcript for *S-Lap1* and initiation of RNAi, analogous to the model proposed for *D. pseudoobscura*..

Likewise, we infer independent amplifications of *S-Lap1* and *GAPsec* on the X chromosome of *D. athabasca*, another member of the *pseudoobscura* group for which we have generated a high-quality female genome assembly (**Figure 5C**). We detect 6 copies of *GAPsec* and 5 copies of *S-Lap1* on the X chromosome of *D. athabasca*, and genomic coverage analysis suggests a similar number of copies on the Y chromosome. This indicates that *S-Lap1* and *GAPsec* are involved in sex chromosomes drive several times in species where this locus has become sex-linked.

Discussion

Theory predicts that selfish genetic elements that cheat fair meiosis may be common on sex chromosomes [4]. Sex chromosome drive has extreme evolutionary consequences, and distorted sex ratios can lead to population collapse and extinction, unless suppressors evolve rapidly to silence drivers on sex chromosomes [7]. Recurrent bursts of invasion of sex chromosome drivers followed by the quick fixation of repressors may thus be common, and may be a main evolutionary force driving the rapid evolution of sex determination mechanisms [35], shape the distribution of genes and patterns of gene expression on sex chromosomes [5], and lead to the evolution of hybrid sterility and account for the special role the X plays in the evolution of reproductive isolation [36,37]. However, given the short-lived nature of sex chromosome drive, the prevalence and the molecular bases of such

conflicts remains poorly understood.

Co-amplification of sex-linked genes is common in Drosophila. Here, we identify cryptic sex chromosome drive systems based on the footprints they leave behind in patterns of genome evolution: co-amplification of genes on both the X and the Y chromosome. Genes may amplify on the Y chromosome for a variety of reasons, and our current data do not allow us to evaluate their relative importance. In particular, multi-copy genes may simply arise on the Y at a higher rate, since the high repeat content on the Y facilitates structural re-arrangements that can promote gene family expansion [38]. Additionally, the efficacy of natural selection is reduced on the non-recombining Y, and Y chromosomes across diverse taxa accumulate functionless and deleterious repetitive DNA [27]. Amplified Y genes thus may either provide no benefit for their carriers, or could in fact be slightly deleterious, yet natural selection is unable to remove them [39]. Heterochromatin formation on the Y may further dampen any functional consequences of gene family expansion, and multi-copy Y genes may simply be more tolerated on the silenced Y. Finally, some multi-copy Y genes may actually contribute to male fitness and fertility [23-25,40]. Y chromosomes are transmitted from father to son, and are thus a perfect genomic location for genes that specifically enhance male fitness [41]. Y chromosomes of several species, including mammals and Drosophila, have been shown to contain multi-copy gene families that are expressed in testis and contribute to male fertility [23-25]. Our analysis shows that multi-copy Y genes are common across flies, and it will be of great interest to identify the diverse evolutionary processes driving their amplification.

Co-amplification of X/Y genes, on the other hand, is difficult to explain under scenarios that do not involve meiotic conflict: The repeat content of X chromosomes is comparable to that of autosomes [33]; natural selection efficiently purges deleterious mutations from the recombining X; and transcription of the X chromosome in Drosophila males is increased, rather than reduced [42]. Most importantly, co-amplified X and Y genes are enriched for meiosis functions (see also [43]), and the X-linked copies of co-amplified genes are highly expressed in testis [43]. Testis expression of co-amplified X-linked genes is unusual, as the X chromosome of Drosophila is normally devoid of testis-expressed genes [44], but can be understood under intragenomic conflict models invoking cryptic sex chromosome drive [5,9,10,45,46].

Our comparative analysis shows that co-amplification of genes on sex chromosomes is common in Drosophila and we find evidence of cryptic sex ratio distortion systems across distantly related species. Note, however, that our method for identifying co-amplified X/Y genes is conservative, and we might greatly underestimate the true magnitude of cryptic sex chromosome drive. On one hand, our approach for detecting amplified Y-genes requires them to have much higher coverage in male than female genomic reads (i.e. 4-fold higher coverage), and can thus only detect genes that have acquired considerably more copies on the Y chromosome relative to the X. Indeed, our recent careful examination of gene family evolution on the fully sequenced and assembled neo-Y of *D. miranda* confirms that the true number of co-amplified X/Y gene families is much higher than what we can detect

here: Direct sequence inspection revealed that at least 94 genes co-amplified on the X and Y of *D. miranda* [43], while we could only detect 16 genes with our methodology. In addition, we only probed for genes that are present in the *D. melanogaster* annotation. Most of the species that we surveyed here are only distantly related to *D. melanogaster*, and genes that may have evolved novel drive-related functions may simply not have a homolog in *D. melanogaster.* Indeed, about 1/3 of the co-amplified X/Y genes that we identified in *D. miranda* did not have an ortholog in *D. melanogaster* [43]. Finally, we required X-linked co-amplified gene copies to be present in our Illumina assemblies; however, recent gene duplicates are often collapsed is such assemblies [33].

Co-amplification of sex-linked genes and meiotic drive. Our data suggest that several putative drive genes have well-characterized functions in meiosis ensuring precise chromosome segregation, and selfish elements have recurrently succeeded to manipulate these normally tightly regulated cellular processes. How does co-amplification of meiosis-related genes on the X and Y cause meiotic drive and its suppression? If amplified Y genes are involved in a battle with the X over fair transmission, changes in gene copy number may tip the balance over inclusion into functional sperm, and could result in repeated co-amplification of distorters and suppressors on the sex chromosomes (**Figure 1A**). In particular, an X-linked gene involved in chromosome segregation may evolve a duplicate that acquires the ability to incapacitate Y-bearing sperm (**Figure 1A**). Invasion of this sex-ratio distorter skews the population sex ratio and creates a selective advantage to evolve a Y-linked suppressor that is resistant to the distorter. Suppression may be achieved at the molecular level by increased copy number of the wildtype function or by inactivation of X-linked drivers using RNAi [9,10,13]. If both driver and suppressor are dosage sensitive, they would undergo iterated cycles of expansion, resulting in rapid co-amplification of both driver and suppressor on the X and Y chromosome [4].

Co-amplification of X/Y genes is thus a characteristic signature of recurrent drive, but does not tell us which chromosome started the battle over inclusion into the sperm. That is, who drives whom? While we cannot determine the sequence of evolutionary events with certainty, the X chromosome is more likely to acquire segregation distorters, for several reasons. X chromosomes typically harbor many more genes that could acquire a driving function. Y chromosomes have a lower effective population size, and natural selection is impaired on the non-recombining Y [27], making drivers less likely to originate and spread on the Y. In addition, old Y chromosomes in many species are highly heterochromatic, and defects in heterochromatin formation can lead to chromosome mis-segregation [47]. This makes the Y chromosome especially vulnerable for mis-segregation, and the X chromosome—via DNA-binding proteins or other proteins involved in the regulation of the chromatin state—can take advantage of the heterochromatin state of the Y chromosome to induce meiotic drive. Indeed, most of the cellular meiotic drive phenotypes described in Drosophila species suggest a failure in chromatin state regulation [4,9,10,48-51]. We therefore expect *a priori* that most drivers originate on the X, and create strong selective pressure for the Y to evolve suppressors.

Most of the species where we identify co-amplified X/Y genes harbor neo-sex chromosomes. This suggests that sex ratio distorters have repeatedly evolved to exploit genomic vulnerabilities associated with the formation of new sex chromosomes. Different features of young vs. old sex chromosome create different susceptibilities to sex chromosome drive. Old Y chromosomes are typically highly repetitive and heterochromatic, a feature that may easily be exploited by a driver on the X. Also, old sex chromosomes show much higher levels of sequence divergence, which makes identification and targeting of the homolog by a driver easier. Yet, young Y chromosomes typically contain many more genes that can evolve to cheat meiosis, thereby increasing the chances of a Y-linked driver. Finally, young X chromosomes may not yet be transcriptionally inactive during spermatogenesis and thus express more drivers. In many species, including Drosophila, expression from the X chromosome is reduced during spermatogenesis [52]. As discussed above, the low gene number yet high repeat content makes Y chromosomes especially vulnerable to meiotic drive, and silencing of the X during spermatogenesis may have evolved as a genome defense against driving X's [5]. Suppression of transcription during spermatogenesis may not yet have fully evolved on young X chromosomes, allowing the expression of more X-linked drivers.

RNAi and drive. Our data further support growing evidence that the production of antisense transcripts, hairpin RNAs and small RNAs may be a common feature of meiotic drive elements [13,43]. RNA interference (RNAi) related pathways provide defense against viruses and transposable elements, and have been implicated in the suppression of meiotic drive elements [13]. Intriguingly, genes in these pathways often evolve rapidly, and show frequent gene duplication and loss over long evolutionary time periods. *Argonaute 2* (*Ago2*), for example, is one of the key RNAi genes in insects, and has repeatedly formed new testis-specific duplicates in the recent history of the *Drosophila obscura* group [53]. Analysis of additional RNAi-pathway genes confirms that they undergo frequent independent duplications and that their history has been particularly labile within the *Drosophila obscura* group [54]. Our finding suggests that the presence of young sex chromosomes in this species group makes them especially vulnerable to the invasion of meiotic drive elements, and may thus drive the rapid evolution of RNAi genes in this clade. It will be of interest to study the dynamics of RNAi genes in other species groups that have gained novel sex chromosomes, to see if diversification of RNAi genes is correlated with the emergence of new sex chromosomes.

In order to trigger the RNAi response, the production of dsRNA is required. This can be achieved in multiple ways. In the *D. simulans* Winters system, the two suppressor genes both encode related long inverted repeats that can form hairpin RNAs (hpRNAs), which are then processed by the RNAi machinery to generate siRNAs that repress the paralogous distorters [13]. Alternatively, the production of dsRNA can occur through anti-sense transcription of the target genes, and this mechanism is creating siRNAs in the putative drive involving *GAPsec* and *S-Lap1*, both in *D. pseudoobscura* and *D. miranda*.

The RNAi pathway can be utilized in different ways to either create a meiotic driver, or to suppress it. For example, a gene on the X (or it's duplicate) may gain a novel function that disrupts segregation of the Y chromosome. The homologous Y gene (or duplicates of it) may then silence the driving X by producing anti-sense transcripts that generate dsRNAs and launch the RNAi response, to silence the X-linked driver. This scenario resembles the Winters sex ratio system, even though the suppressors of X-linked drive are autosomal and not Y-linked. The RNAi machinery can also be hijacked to create a driving X. In particular, if an X-linked gene is required for producing Y-bearing sperm, an X chromosome that silences this gene could evolve a drive phenotype. It could do so by antisense RNA production of this X-linked gene (or duplicates of it), thereby triggering the RNAi response to inactivate the gene. The organism could restore the wildtype function of this gene by increasing its dose through its amplification on the Y, and even non-functional copies may act as a decoy to soak up endo-siRNA that are targeting this locus. This pathway may underlay the putative *GAPsec /S-Lap1* drive system in *D. pseudoobscura*, where the X-linked duplicates of *S-Lap1* produce the vast majority of antisense transcripts (roughly 95%, see **Table S6**), while the Y-linked *S-Lap1* copies predominantly generate sense RNA (>99.9%). Most of the short RNA are produced from *S-Lap1-dup* (the putative driver) and the Y-linked copies of *S-Lap1* (about 96% in total), consistent with the idea that amplification of this spermatogenesis gene allows restoration of wildype function, possibly by acting as a decoy to dilute RNAi induced silencing triggered by antisense transcripts of *S-Lap1-dup.* Under either scenario, if both the driver and suppressor are dosage-sensitive, this can lead to the repeated invasion of driving and suppressing chromosomes through co-amplification of genes on the X and Y chromosome.

Conclusion

To conclude, our comparative analysis suggests that cryptic sex chromosome drive may be common, especially on young sex chromosomes. The prevalence of cryptic sex ratio drive systems in insects and animals may account for several evolutionary and molecular phenomena [5]. Sex ratio distorters can fuel the rapid turn-over of sex determination mechanisms across species [35], and the transcriptional inactivation of sex chromosomes during spermatogenesis may have evolved as a defense against meiotic drive elements. The recurrent fixation of cryptic drive systems on sex chromosomes might explain the prominent role of the X chromosome in the evolution of hybrid sterility in a wide range of species [36,37,55-57], and contribute to genomic biases in the location of sex-biased genes or gene duplicates [58,59]. Our work suggests that the short RNA pathway has an important role not only in controlling genomic parasites such as transposable elements, but also selfish genetic elements that try to exploit highly regulated cellular proceeses such as chromosome segregation [60]. Future characterization of the putative drive systems identified here will provide a full picture of how distorting elements manipulate and cheat meiosis, what molecular pathways or developmental processes are particularly vulnerable, and how the genome has launched evolutionary responses to counter distortion**.**

Methods

Genome sequencing & assembly

Strains were acquired from the Drosophila Species Stock Center (UC San Diego) or the EHIME stock center (Ehime University, Japan) as indicated in **Table S1**. For each strain, DNA was extracted from a single male and a single female, using the Qiagen Gentra Puregene cell kit. The Illumina TruSeq Nano DNA library preparation kit was used to prepare 100 bp paired-end sequencing libraries for all species except *D. robusta*, *D. melanica*, and *D. willistoni*. For these species, the Illumina Nextera DNA library preparation kit was used to prepare 150 bp paired-end sequencing libraries. The genome assemblies produced for this study are noted in **Table S1**. Assemblies were produced from the female data: reads were error-corrected using BFC [61] and assembled using IDBA-UD [62] with default parameters.

Identification of X-A fusions

X chromosome/autosome fusions were identified in two steps [63]. For each species, genomic scaffolds were assigned to Muller elements based on their gene content, inferred from the results of a translated BLAST search of *D. melanogaster* peptides to the assembly of interest. Scaffolds smaller than 5kb were excluded. Next, the male and female Illumina data were separately mapped to the female assembly using Bowtie2 [64] and excluding alignments with mapping quality less than 20. The coverage ratio (M/F) was calculated for each scaffold that was assigned to a Muller element. The distribution of coverage ratios for each Muller element (**Figure S1**) was then examined to determine if any of the ancestral autosomes had become X-linked.

Identification of co-amplified genes on the X- and Y-chromosome

To characterize co-amplified genes on the sex chromosomes, we first identify genes amplified on the Y. For each species, male and female Illumina reads were separately aligned to a filtered version of the *D. melanogaster* peptide set, where only the longest isoform of each gene was retained. To generate these alignments, the DIAMOND software package [65] was used to perform a translated search of each Illumina read to the peptide set. Read coverage for each peptide sequence was calculated in 30 amino acid non-overlapping windows and normalized by dividing by the total number of mapped reads. The M/F coverage ratio was computed by dividing the median male coverage by the median female coverage, for each peptide. We required that potentially Y-amplified genes have a normalized M/F coverage ratio of at least 2.5 and only retained genes whose parent copy was X-linked in the species of interest. We searched for X-linked duplicates in the female genome assemblies by first using Exonerate [66] to extract the coding sequence of the best hit between the *D. melanogaster* peptide and the female assembly. We then used BLASTN [67] to obtain a stringent (E-value threshold = 1e-20) list of all non-overlapping hits between each exon of the coding sequence and the genome assembly. We considered a gene to

be duplicated in females if at least 25% of the parent coding sequence aligned to more than one location in the genome assembly.

S-Lap1 and _GAPsec_ gene trees and Y chromosome gene copies

The Muller-D copies of *S-Lap1* and *GAPsec* were identified in the 12 Drosophila genomes [68] by synteny with *D. melanogaster* and their coding sequences were downloaded from FlyBase [69]. The PRANK software package [70] was used to generate codon-aware alignments of coding sequences for each gene. The resulting alignment was trimmed using trimAl [71] and RaxML [72] was used to infer a maximum likelihood phylogeny (100 bootstrap replicates). *D. pseudoobscura* Y-linked contigs were identified using read coverage information from male versus female genomic sequencing data. Exonerate [66] was used to determine the location of the amplified copies of *S-Lap1* and *GAPsec* on these scaffolds with the *D. pseudoobscura S-Lap1* (*FBpp0285960*) and *GAPsec* (*FBpp0308917*) peptide sequences as queries. The *D. pseudoobscura* Y copies of each gene were aligned using MAFFT [73], trimmed with trimAl [71], and a Y consensus sequence for each gene was generated using PILER [74].

RNA libraries and mapping

We dissected testes from 3-8 day old virgin males of *D. pseudoobscura* (strain MV25, transgenic *P{w$^+$:sh-S-Lap1-Y}* and *white* control males) reared at 18°C on Bloomington food. We used Trizol (Invitrogen) and GlycoBlue (Invitrogen) to extract and isolate total RNA. *D. pseudoobscura* CAGE-seq data were obtained from the ModEncode project [75]. We resolved 20 μg of total RNA on a 15% TBE-Urea gel (Invitrogen) and size selected 19-29 nt long RNA, and used Illumina's TruSeq Small RNA Library Preparation Kit to prepare small RNA libraries, which were sequenced on an Illumina HiSeq 4000 at 50 nt read length (single-end)., We used to Ribo-Zero to deplete ribosomal RNA from total RNA, and used Illumina's TruSeq Stranded Total RNA Library Preparation Kit to prepare stranded testis RNA libraries, which were sequenced on an Illumina HiSeq 4000 at 100 nt read length (paired-end). Total RNA data were aligned to the *D. pseudoobscura* reference genome using HISAT2 [76], whereas Bowtie2 [64] (seed length: 18) was used to align small RNA and CAGE-seq data. In all cases, alignments with mapping quality less than 20 were discarded.

References:

1. Gershenson S. A New Sex-Ratio Abnormality in DROSOPHILA OBSCURA. Genetics. Genetics Society of America; 1928;13: 488–507.
2. Rice WR. Nothing in Genetics Makes Sense Except in Light of Genomic Conflict. Annual review of Ecology, Evolution and Systematics. 2013;44: 217–237.
3. McLaughlin RN, Malik HS. Genetic conflicts: the usual suspects and beyond. J Exp Biol. The Company of Biologists Ltd; 2017;220: 6–17. doi:10.1242/jeb.148148
4. Jaenike J. Sex Chromosome Meiotic Drive. http://dxdoiorg/101146/annurevecolsys32081501113958. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA; 2003;32: 25–49. doi:10.1146/annurev.ecolsys.32.081501.113958

5. Meiklejohn CD, Tao Y. Genetic conflict and sex chromosome evolution. Trends Ecol Evol (Amst). 2010;25: 215–223. doi:10.1016/j.tree.2009.10.005
6. Hurst LD, Pomiankowski A. Maintaining Mendelism: might prevention be better than cure? Bioessays. John Wiley & Sons, Ltd; 1991;13: 489–490. doi:10.1002/bies.950130910
7. Lindholm AK, Dyer KA, Firman RC, Fishman L, Forstmeier W, Holman L, et al. The Ecology and Evolutionary Dynamics of Meiotic Drive. Trends Ecol Evol (Amst). 2016;31: 315–326. doi:10.1016/j.tree.2016.02.001
8. Hurst GD, Werren JH. The role of selfish genetic elements in eukaryotic evolution. Nat Rev Genet. 2001;2: 597–606. doi:10.1038/35084545
9. Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL. A sex-ratio meiotic drive system in Drosophila simulans. I: an autosomal suppressor. Barbash D, editor. PLoS Biol. Public Library of Science; 2007;5: e292. doi:10.1371/journal.pbio.0050292
10. Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, Hartl DL. A sex-ratio Meiotic Drive System in Drosophila simulans. II: An X-linked Distorter. Barbash D, editor. PLoS Biol. Public Library of Science; 2007;5: e293. doi:10.1371/journal.pbio.0050293
11. Cocquet J, Ellis PJI, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS. A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. Nachman MW, editor. PLoS Genet. Public Library of Science; 2012;8: e1002900. doi:10.1371/journal.pgen.1002900
12. Cocquet J, Ellis PJI, Yamauchi Y, Mahadevaiah SK, Affara NA, Ward MA, et al. The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. Hastie N, editor. PLoS Biol. 2009;7: e1000244. doi:10.1371/journal.pbio.1000244
13. Lin C-J, Hu F, Dubruille R, Vedanayagam J, Wen J, Smibert P, et al. The hpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the Drosophila Male Germline. Dev Cell. 2018;46: 316–326.e5. doi:10.1016/j.devcel.2018.07.004
14. Soh YQS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, et al. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell. 2014;159: 800–813. doi:10.1016/j.cell.2014.09.052
15. Balakireva MD, Shevelyov YuYa, Nurminsky DI, Livak KJ, Gvozdev VA. Structural organization and diversification of Y-linked sequences comprising Su(Ste) genes in Drosophila melanogaster. Nucleic Acids Res. Oxford University Press; 1992;20: 3731–3736.
16. Malone CD, Lehmann R, Teixeira FK. The cellular basis of hybrid dysgenesis and Stellate regulation in Drosophila. Curr Opin Genet Dev. 2015;34: 88–94. doi:10.1016/j.gde.2015.09.003
17. Hurst LD. Is Stellate a relict meiotic driver? Genetics. Genetics Society of America; 1992;130: 229–230.
18. Gell SL, Reenan RA. Mutations to the piRNA pathway component aubergine enhance meiotic drive of segregation distorter in Drosophila melanogaster. Genetics. 2013;193: 771–784. doi:10.1534/genetics.112.147561
19. Gibson JR, Chippindale AK, Rice WR. The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc Biol Sci. The Royal Society; 2002;269: 499–505. doi:10.1098/rspb.2001.1863
20. Innocenti P, Morrow EH. The sexually antagonistic genes of Drosophila melanogaster. Hurst LD, editor. PLoS Biol. Public Library of Science; 2010;8: e1000335. doi:10.1371/journal.pbio.1000335
21. Roldan E, Gomendio M. The Y chromosome as a battle ground for sexual selection. Trends Ecol Evol (Amst). 1999;14: 58–62.
22. Chippindale AK, Rice WR. Y chromosome polymorphism is a strong determinant of male fitness in Drosophila melanogaster. Proc Natl Acad Sci USA. National Academy of Sciences; 2001;98: 5677–5682. doi:10.1073/pnas.101456898
23. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. Nature Publishing Group; 2003;423: 825–837. doi:10.1038/nature01722
24. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 2014;508: 494–499. doi:10.1038/nature13206
25. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional

evolution of Y chromosomes across mammals. Nature. Nature Publishing Group; 2014;508: 488–493. doi:10.1038/nature13151
26. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature. Nature Publishing Group; 2003;423: 873–876. doi:10.1038/nature01723
27. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. Nature Publishing Group; 2013;14: 113–124. doi:10.1038/nrg3366
28. Lucchesi JC. Gene dosage compensation and the evolution of sex chromosomes. Science. 1978;202: 711–716.
29. Dorus S, Wilkin EC, Karr TL. Expansion and functional diversification of a leucyl aminopeptidase family that encodes the major protein constituents of Drosophila sperm. BMC Genomics. BioMed Central; 2011;12: 177. doi:10.1186/1471-2164-12-177
30. Merrill C. Truncated RanGAP Encoded by the Segregation Distorter Locus of Drosophila. Science. 1999;283: 1742–1745. doi:10.1126/science.283.5408.1742
31. Casola C, Ganote CL, Hahn MW. Nonallelic gene conversion in the genus Drosophila. Genetics. Genetics; 2010;185: 95–103. doi:10.1534/genetics.110.115444
32. Reese MG. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem. 2001;26: 51–56.
33. Mahajan S, Wei KH-C, Nalley MJ, Gibilisco L, Bachtrog D. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. Tyler-Smith C, editor. PLoS Biol. Public Library of Science; 2018;16: e2006348. doi:10.1371/journal.pbio.2006348
34. Gao J-J, Watabe H-A, Aotsuka T, Pang J-F, Zhang Y-P. Molecular phylogeny of the Drosophila obscura species group, with emphasis on the Old World species. BMC Evol Biol. BioMed Central; 2007;7: 87. doi:10.1186/1471-2148-7-87
35. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, et al. Sex determination: why so many ways of doing it? PLoS Biol. Public Library of Science; 2014;12: e1001899. doi:10.1371/journal.pbio.1001899
36. Hurst LD, Pomiankowski A. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics. Genetics Society of America; 1991;128: 841–858.
37. Frank SA. Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution. Wiley/Blackwell (10.1111); 1991;45: 262–267. doi:10.1111/j.1558-5646.1991.tb04401.x
38. Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol. 2010;20: 211–221. doi:10.1016/j.semcancer.2010.03.001
39. Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philos Trans R Soc Lond, B, Biol Sci. The Royal Society; 2000;355: 1563–1572. doi:10.1098/rstb.2000.0717
40. Piergentili R. Multiple roles of the Y chromosome in the biology of Drosophila melanogaster. ScientificWorldJournal. Hindawi; 2010;10: 1749–1767. doi:10.1100/tsw.2010.168
41. Rice WR. Sex chromosomes and the evolution of sexual dimorphism. Evolution. Wiley/Blackwell (10.1111); 1984;38: 735–742. doi:10.1111/j.1558-5646.1984.tb00346.x
42. Vicoso B, Bachtrog D. Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosome Res. 2009;17: 585–602. doi:10.1007/s10577-009-9053-y
43. Mahajan S, Bachtrog D. *Massive Gene Amplification on a Recently Formed Drosophila Y Chromosome*.
44. Sturgill D, Zhang Y, Parisi M, Oliver B. Demasculinization of X chromosomes in the Drosophila genus. Nature. 2007;450: 238–241. doi:10.1038/nature06330
45. Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JMA. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat Genet. Nature Publishing Group; 2008;40: 794–799. doi:10.1038/ng.126
46. Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, et al. Independent specialization of the human and mouse X chromosomes for the male germ line. Nat Genet. Nature Publishing Group; 2013;45:

1083–1087. doi:10.1038/ng.2705
47. Kellum R, Alberts BM. Heterochromatin protein 1 is required for correct chromosome segregation in Drosophila embryos. J Cell Sci. 1995;108 (Pt 4): 1419–1431.
48. Policansky D, Ellison J. “Sex ratio” in Drosophila pseudoobscura: spermiogenic failure. Science. 1970;169: 888–889.
49. Montchamp-Moreau C. Sex-ratio meiotic drive in Drosophila simulans: cellular mechanism, candidate genes and evolution. Biochem Soc Trans. 2006;34: 562–565. doi:10.1042/BST0340562
50. Cazemajor M, Joly D, Montchamp-Moreau C. Sex-ratio meiotic drive in Drosophila simulans is related to equational nondisjunction of the Y chromosome. Genetics. Genetics Society of America; 2000;154: 229–236.
51. Helleu Q, Gérard PR, Dubruille R, Ogereau D, Prud'homme B, Loppin B, et al. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. Proc Natl Acad Sci USA. 2016;113: 4110–4115. doi:10.1073/pnas.1519332113
52. Landeen EL, Muirhead CA, Wright L, Meiklejohn CD, Presgraves DC. Sex Chromosome-wide Transcriptional Suppression and Compensatory Cis-Regulatory Evolution Mediate Gene Expression in the Drosophila Male Germline. Becker PB, editor. PLoS Biol. 2016;14: e1002499. doi:10.1371/journal.pbio.1002499
53. Lewis SH, Webster CL, Salmela H, Obbard DJ. Repeated Duplication of Argonaute2 Is Associated with Strong Selection and Testis Specialization in Drosophila. Genetics. Genetics; 2016;204: 757–769. doi:10.1534/genetics.116.192336
54. Crysnanto D, Obbard DJ. Widespread gene duplication and adaptive evolution in the RNA interference pathways of the *Drosophila obscura* group.
55. Zhang L, Sun T, Woldesellassie F, Xiao H, Tao Y. Sex ratio meiotic drive as a plausible evolutionary mechanism for hybrid male sterility. Begun DJ, editor. PLoS Genet. Public Library of Science; 2015;11: e1005073. doi:10.1371/journal.pgen.1005073
56. Presgraves DC. The molecular evolutionary basis of species formation. Nat Rev Genet. Nature Publishing Group; 2010;11: 175–180. doi:10.1038/nrg2718
57. McDermott SR, Noor MAF. The role of meiotic drive in hybrid male sterility. Philos Trans R Soc Lond, B, Biol Sci. The Royal Society; 2010;365: 1265–1272. doi:10.1098/rstb.2009.0264
58. Vibranovski MD, Zhang Y, Long M. General gene movement off the X chromosome in the Drosophila genus. Genome Res. 2009;19: 897–903. doi:10.1101/gr.088609.108
59. Emerson JJ, Kaessmann H, Betrán E, Long M. Extensive gene traffic on the mammalian X chromosome. Science. American Association for the Advancement of Science; 2004;303: 537–540. doi:10.1126/science.1090042
60. Wen J, Duan H, Bejarano F, Okamura K, Fabian L, Brill JA, et al. Adaptive regulation of testis gene expression and control of male fertility by the Drosophila hairpin RNA pathway. [Corrected]. Mol Cell. 2015;57: 165–178. doi:10.1016/j.molcel.2014.11.025
61. Li H. BFC: correcting Illumina sequencing errors. Bioinformatics. 2015;31: 2885–2887. doi:10.1093/bioinformatics/btv290
62. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28: 1420–1428. doi:10.1093/bioinformatics/bts174
63. Vicoso B, Bachtrog D. Reversal of an ancient sex chromosome to an autosome in Drosophila. Nature. Nature Publishing Group; 2013;499: 332–335. doi:10.1038/nature12235
64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. Nature Publishing Group; 2012;9: 357–359. doi:10.1038/nmeth.1923
65. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12: 59–60. doi:10.1038/nmeth.3176
66. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. BioMed Central; 2005;6: 31. doi:10.1186/1471-2105-6-31
67. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. BioMed Central; 2009;10: 421. doi:10.1186/1471-2105-10-421

68. Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007;450: 203–218. doi:10.1038/nature06341
69. Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, et al. FlyBase at 25: looking to the future. Nucleic Acids Res. 2017;45: D663–D671. doi:10.1093/nar/gkw1016
70. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. Totowa, NJ: Humana Press; 2014;1079: 155–170. doi:10.1007/978-1-62703-646-7_10
71. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973. doi:10.1093/bioinformatics/btp348
72. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033
73. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010
74. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21 Suppl 1: i152–8. doi:10.1093/bioinformatics/bti1003
75. Chen Z-X, Sturgill D, Qu J, Jiang H, Park S, Boley N, et al. Comparative validation of the D. melanogaster modENCODE transcriptome annotation. Genome Res. Cold Spring Harbor Lab; 2014;24: 1209–1223. doi:10.1101/gr.159384.113
76. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. Nature Publishing Group; 2015;12: 357–360. doi:10.1038/nmeth.3317

Acknowledgements Funded by NIH grants (R01GM076007, GM101255 and R01GM093182) to DB. We thank Lauren Gibilisco for generating short RNA libraries and testis RNA-seq libraries.

Author contributions. C.E. generated the genome data and performed the bioinformatics analysis and D.B. oversaw the project and wrote the manuscript with input from all authors.

Competing interests: The authors declare that no competing interests exist.

Figure Legends

Figure 1. Meiotic conflict can fuel co-amplification of X and Y genes. A. An X-linked gene duplicate can evolve a novel function that eliminates Y bearing sperm. Amplification of the homologous Y gene, and production of antisense transcript may trigger the RNAi response, and silence the distorter. Repeated cycles of amplification of dosage-sensitive distorters and suppressors can result in the co-amplification of X/Y genes that are targeted by short RNAs. **B.** Bioinformatic identification of co-amplified X/Y genes. Multicopy Y genes are identified based on mapping of male and female genomic reads to D. melanogaster proteins using translated BLAST searches. Multi-copy X-linked homologs are identified for multi-copy Y genes, based on genome assemblies. Sex-linkage of contigs is inferred based on male and female read coverage of contigs, or based on published genome assemblies for a subset of species.

Figure 2. **Molecular evolution of the *S-Lap1* and *GAPsec* genes in Drosophila**. **A.** Organization of *S-Lap1* and *GAPsec* genes across the Drosophila genus. *S-Lap1* is duplicated in all Drosophila species investigated, and *GAPsec* shows a partial duplication in *D. pseudoobscura* and its sister species *D. persimilis*. **B.** Amplification of *S-Lap1* and *GAPsec* on Y-linked scaffolds of *D. pseudoobscura*. **C.** Gene trees of *S-Lap1* and *GAPsec* copies. *S-Lap1* independently duplicated multiple times across the Drosophila phylogeny. Inferred duplication events are shown by the yellow circle.

Figure 3. **Phylogeny of *S-Lap1* and *GAPsec* gene families in *D. pseudoobscura***. Y-linked copies are shown by blue branches, and X-linked copies are shown in red.

Figure 4. Short RNA's may be involved in suppression of the cryptic *S-Lap1* drive system. A. Expression and short RNA profiles from wildtype *D. pseudoobscura* testis. Stranded RNA-seq (red tracks) reveals that the X-linked copy of *S-Lap1*-duplicate produces both sense and anti-sense transcripts, resulting in the production of short RNA's (blue tracks). CAGE-seq data (grey tracks) support that the *GAPsec* duplicate generated a new TSS resulting in antisense transcript of *S-Lap1*-dup. **B.** A hypothetical evolutionary model of the cryptic *S-Lap1* drive system. *S-Lap1* was duplicated in an ancestor of *D. pseudoobscura*, and a partial duplication of *GAPsec* created a TSS for anti-sense transcription of *S-Lap1* duplicate. Production of small RNA's may deplete *S-Lap1* transcripts, which may result in elimination of Y-bearing sperm, and could be compensated by amplification of *S-Lap1* on the Y chromosome.

Figure 5. Independent amplification of S-Lap1 / GAPsec in different species of the *obscura* group.

Figure 6. Short RNAs may be involved in suppression of the cryptic *S-Lap1* drive system in *D. miranda.* For legend, see Figure 4A.

Table 1. Multi-copy Y-linked genes across Drosophila species. Shown is which chromosome arms form the neo-sex chromosomes (based on synteny in *D. melanogaster*). Genes in bold have functions related to chromosome segregation.

Figure S1. Identification of newly formed sex chromosomes across Drosophila. Sex chromosomes are inferred using male and female coverage data. Plotted is the normalized male / female genomic read coverage for scaffolds mapped to the *D. melanogaster* genome, to infer the location of Muller elements.

Figure S2. **Validation of bioinformatics pipeline to infer multi-copy Y genes in *D. pseudoobscura***. Shown is the predicted coverage based on mapping of Illumina reads on the x-axis, versus the number of Y-linked copies of a gene found in the genome assembly. Note that our bioinformatics pipeline is conservative and underestimates the number of Y-linked copies found in the assembly, presumably due to many multi-copy genes being fragmented in the assembly.

Figure S3. Chromosomal location of multi-copy Y genes.

Figure S4. Alignment of X-linked and Y-linked copies of A. *S-Lap1* and B. *GAPsec*.

Figure S5. Size distribution of short RNAs mapping to X and Y linked copies of *S-Lap1*.

Figure S6. Size distribution of short RNAs mapping to X and Y linked copies of *S-Lap1*.

FigureS7. Mappability of RNA-seq and short RNA data to X-linked and Y-linked copies of *S-Lap1*.

FigureS8. Mappability of RNA-seq and short RNA data to X-linked and Y-linked copies of *GAPsec.*

Table S1. Species used in this study. Shown are species and stock numbers, total assembly size, and sex chromosome karyotype (see Figure S1).

Table S2. Number of amplified Y genes. Shown are the number of inferred amplified Y-linked genes found in each species, for a cut-off of male/female coverage ratio (M/F) >= 2.5.

Table S3. Number of amplified Y genes identified, for different cut-offs of male/female coverage ratio (M/F from 2.5 to 10).

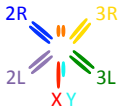
Table S4. Multi-copy Y-linked genes across Drosophila species. Shown are the orthologous location of multi-copy Y genes in *D. melanogaster*, and their inferred molecular function and gene expression pattern in *D. melanogaster* (data from flybase.org).

Table S5. Inferred copy numbers for co-amplified X and Y genes.

Table S6. Mapping of total RNA and short RNA for X-linked and Y-linked copies of *GAPsec* and *S-Lap1*.

Table 1. Multi-copy Y-linked genes across Drosophila species. Shown is which chromosome arms form the neo-sex chromosomes (based on synteny in *D. melanogaster*). Genes in bold have functions related to chromosome segregation.

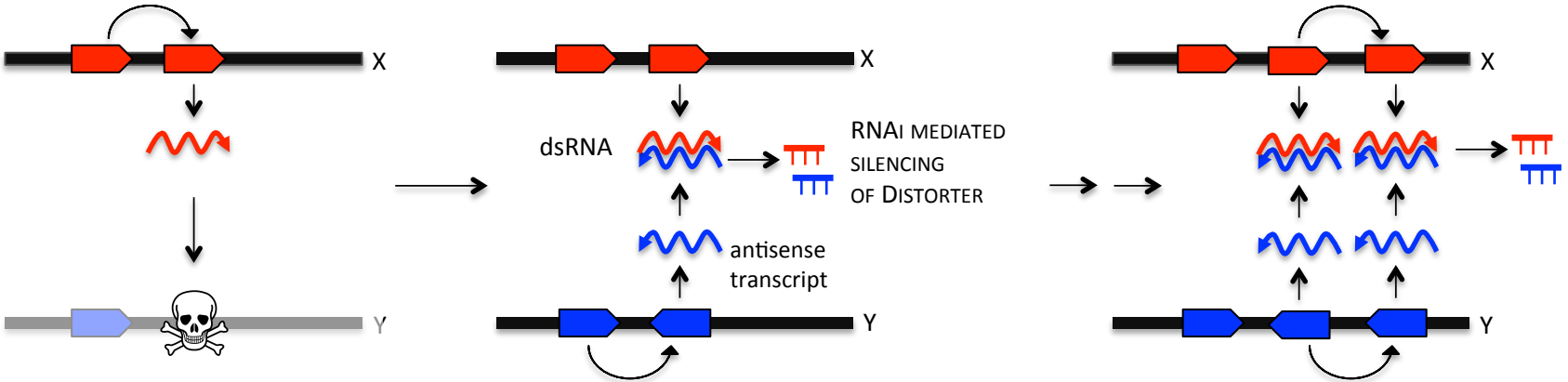
Species	Neo-sex chromosome	Amplified Y genes
D. albomicans	3L, 2R	FBgn0053725*
D. americana	2L	FBgn0032715*
D. athabasca	3L, 2R	FBgn0045483*, FBgn0034429*, FBgn0034328*, FBgn0036640, FBgn0266456
D. lummei	no	**dhd**#
D. melanica	3L	FBgn0028668*, FBgn0010424*, FBgn0053796*
D. miranda	3L, 2R	FBgn0026582*, FBgn0060296*, **Klp61F**, FBgn0034491*, **PCNA***, **fest***, FBgn0034324*, **thr***, S-Lap5*, **mars***, FBgn0033788*, FBgn0033354*, **scra***, FBgn0033216*, FBgn0054045*, **Drak**#
D. nannoptera	3L	FBgn0031103#
D. nigromelanica	2R	FBgn0063491*, **Iswi***
D. pseudoobscura	3L	S-Lap1*, S-Lap2*, GAPsec*, FBgn0035690*
D. robusta	3L	FBgn00103178, FBgn0046301*, FBgn0053795*



Muller element's

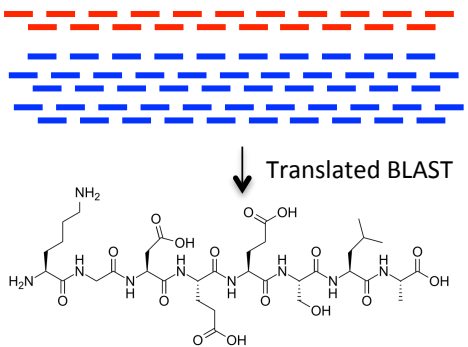
located on ancestral X
* located on neo-sex chromosome

A



B

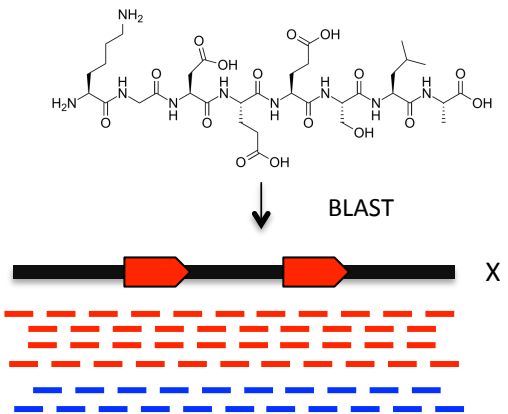
I. Infer Y-amplified genes based on Genomic read coverage



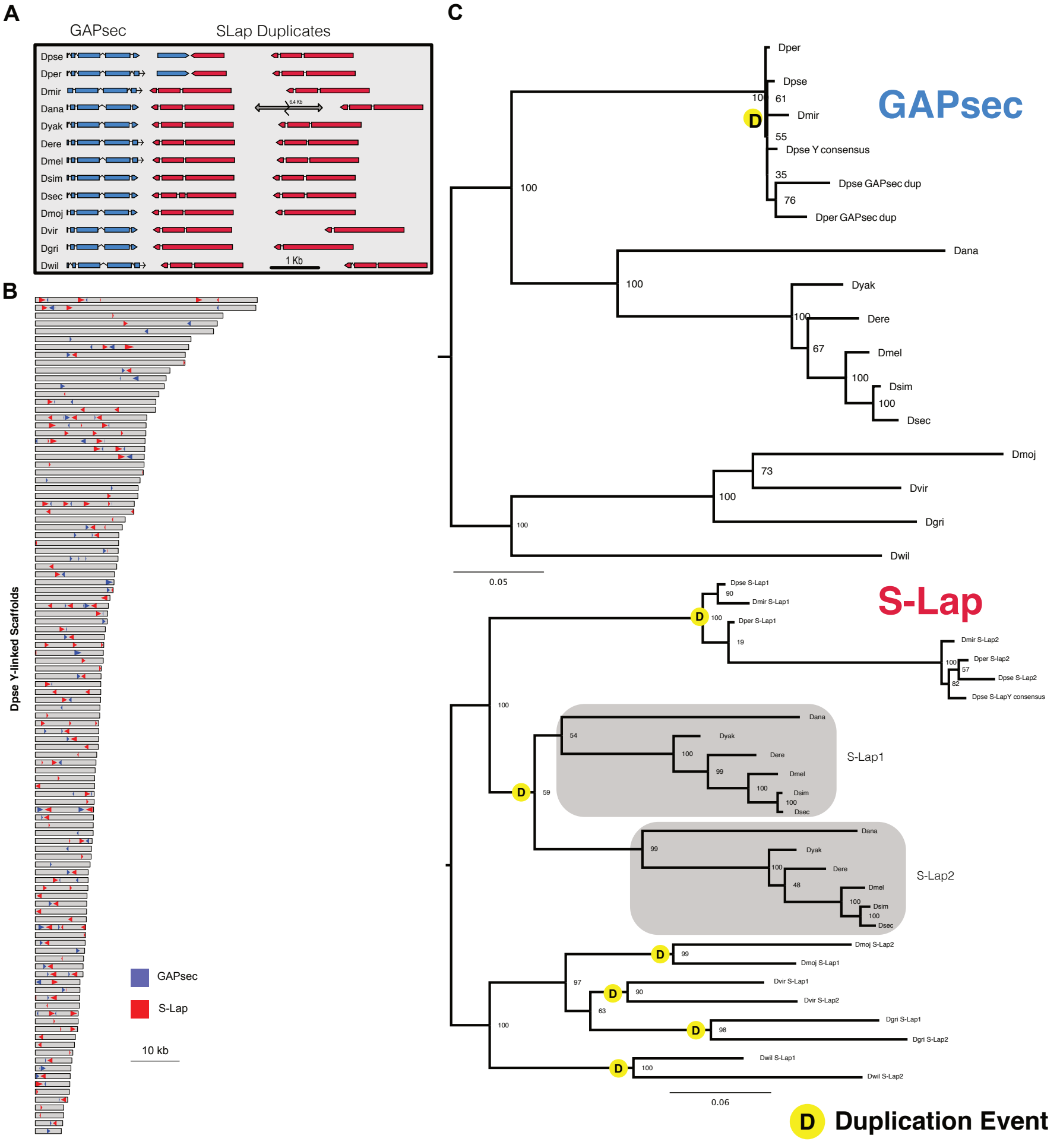
Map Male & Female Genomic reads to *D. melanogaster* peptides

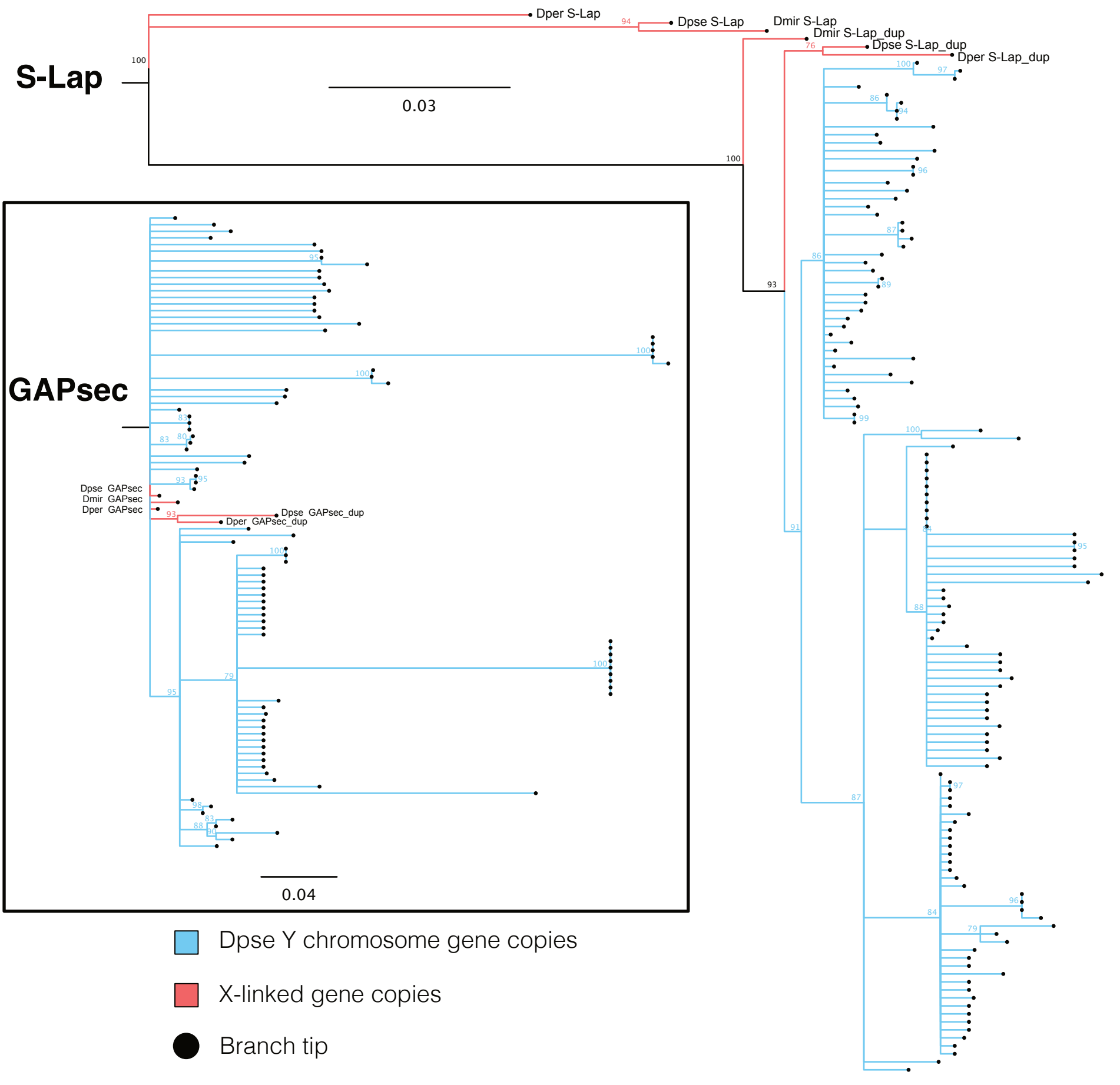
Peptide *D. melanogaster*

II. Identify X-linked contigs and Multi-copy X genes in Female Assembly

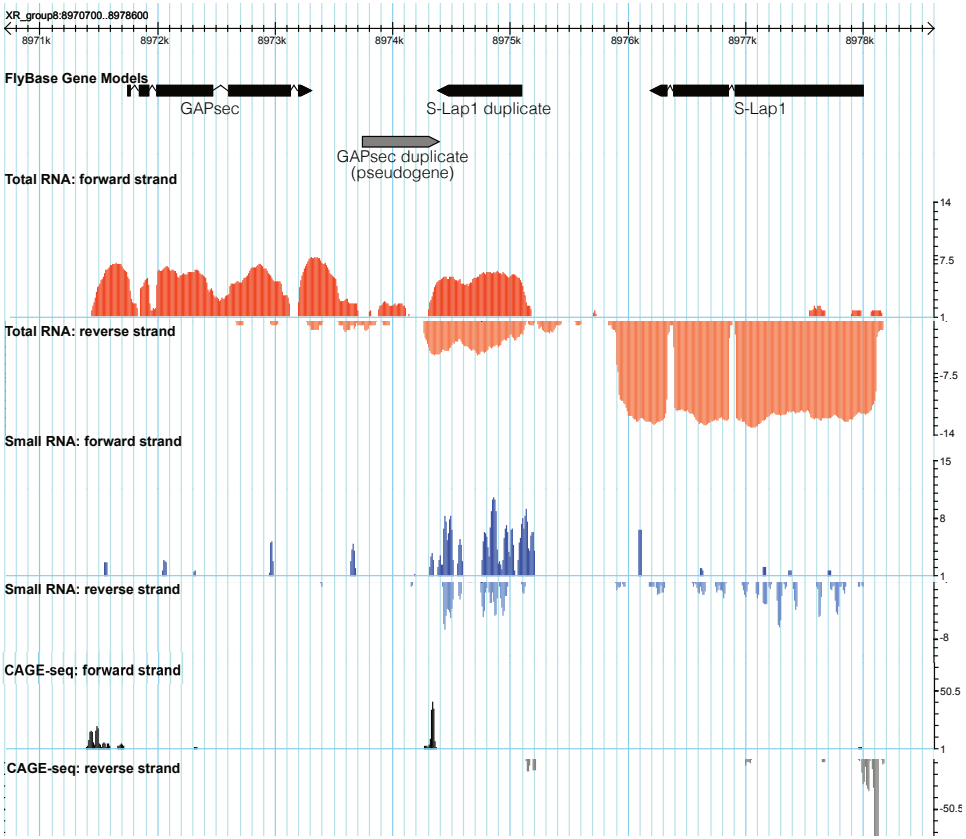


Map Male & Female Genomic reads to Female Assembly

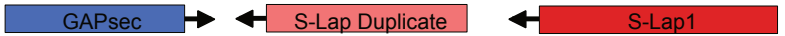

A
GAPsec
SLap Duplicates
Dpse
Dper
Dmir
Dana
Dyak
Dere
Dmel
Dsim
Dsec
Dmoj
Dvir
Dgri
Dwil
6.4 Kb
1 Kb
B
Dpse Y-linked Scaffolds
GAPsec
S-Lap
10 kb
C
GAPsec
Dper
Dpse
Dmir
Dpse Y consensus
Dpse GAPsec dup
Dper GAPsec dup
Dana
Dyak
Dere
Dmel
Dsim
Dsec
Dmoj
Dvir
Dgri
Dwil
0.05
S-Lap
Dpse S-Lap1
Dmir S-Lap1
Dper S-Lap1
Dmir S-Lap2
Dper S-lap2
Dpse S-Lap2
Dpse S-LapY consensus
S-Lap1
S-Lap2
Dmoj S-Lap2
Dmoj S-Lap1
Dvir S-Lap1
Dvir S-Lap2
Dgri S-Lap1
Dgri S-Lap2
Dwil S-Lap1
Dwil S-Lap2
0.06
D Duplication Event


S-Lap
Dper S-Lap
Dpse S-Lap
Dmir S-Lap
Dmir S-Lap_dup
Dpse S-Lap_dup
Dper S-Lap_dup
0.03
GAPsec
Dpse GAPsec
Dmir GAPsec
Dper GAPsec
Dpse GAPsec_dup
Dper GAPsec_dup
0.04
Dpse Y chromosome gene copies
X-linked gene copies
Branch tip

A



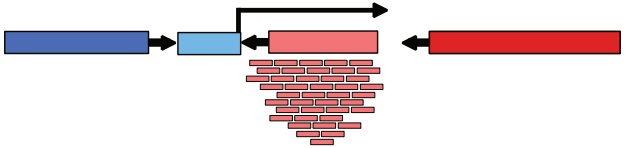
B



D. pseudoobscura

Partial duplication of GAPsec and truncation of the S-Lap Duplicate.

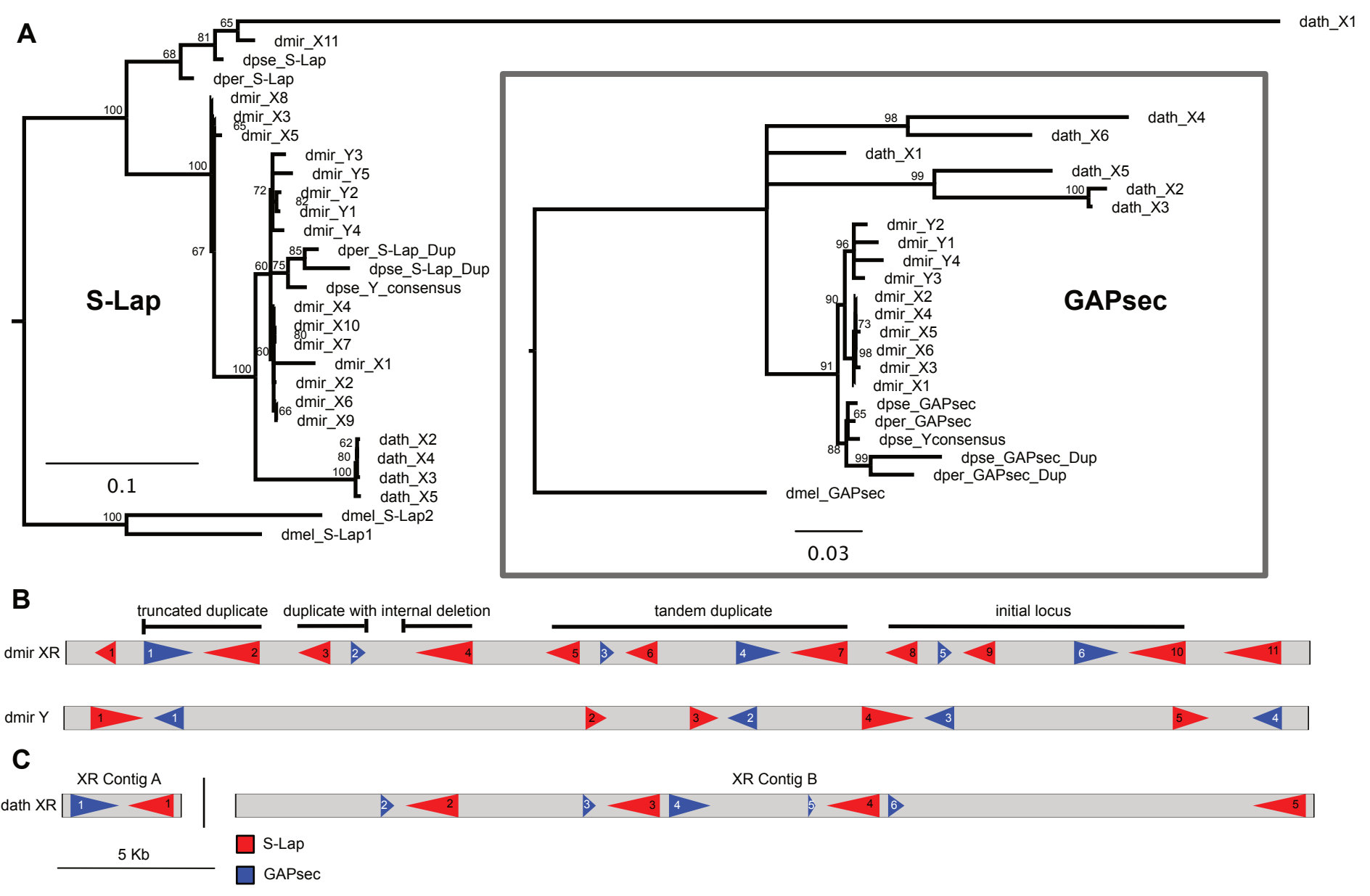
A transcript antisense to the S-Lap duplicate is expressed with its TSS originating from the GAPsec duplicate

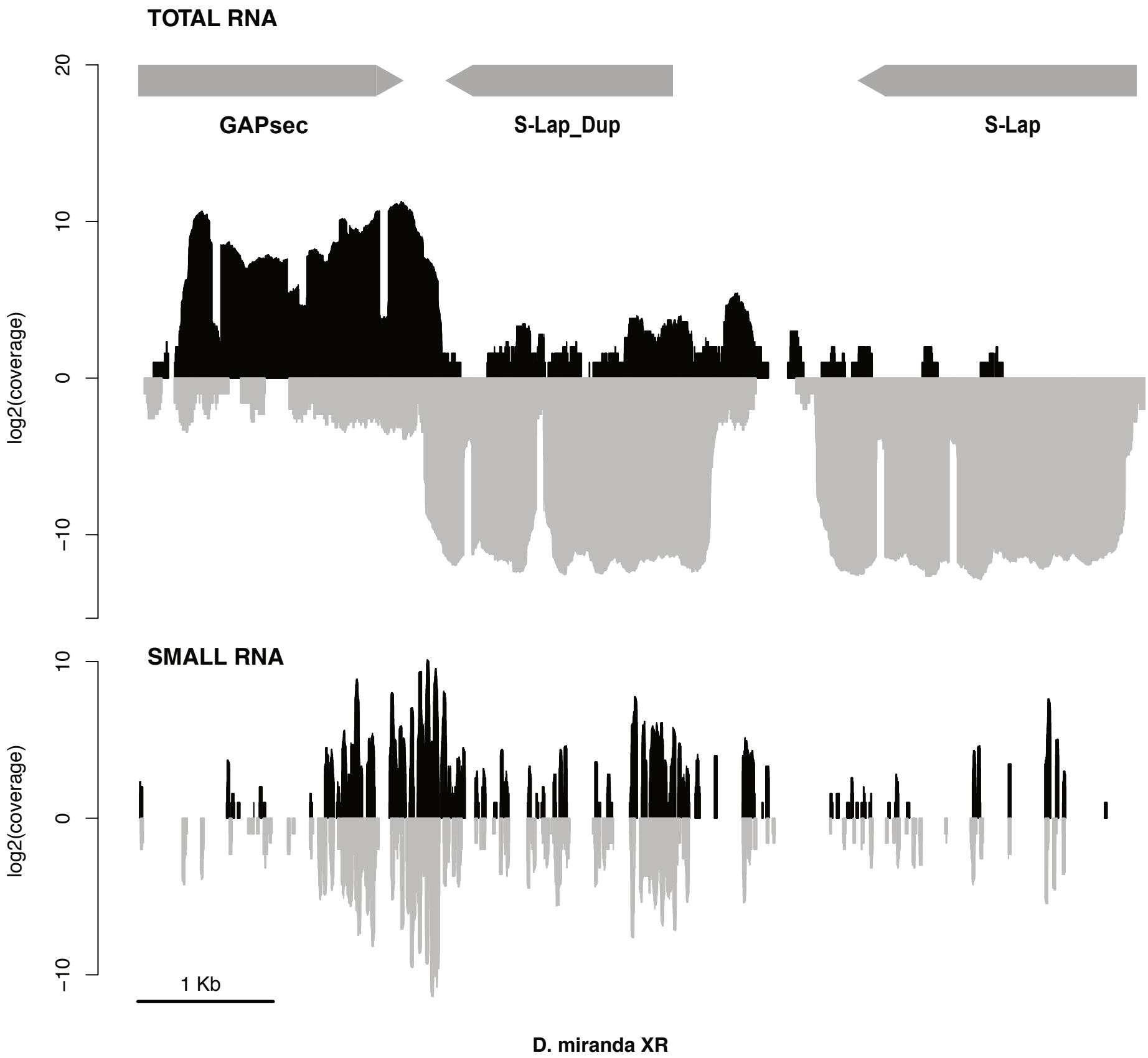


The overlapping transcripts from the S-Lap duplicate produce small RNAs that downregulate expression of the sense transcript.

The reduction in protein product from the S-Lap duplicate results in a defect in the maturation of Y-bearing sperm, causing sex-ratio distortion

Full-length versions of GAPsec and the S-Lap duplicate amplify on the Y chromosome. Expression from the Y-linked copies compensate for the downregulation of the X-linked copy and suppress sex-ratio distortion.


A
S-Lap
dath_X1
dmir_X11
dpse_S-Lap
dper_S-Lap
dmir_X8
dmir_X3
dmir_X5
dmir_Y3
dmir_Y5
dmir_Y2
dmir_Y1
dmir_Y4
dper_S-Lap_Dup
dpse_S-Lap_Dup
dpse_Y_consensus
dmir_X4
dmir_X10
dmir_X7
dmir_X1
dmir_X2
dmir_X6
dmir_X9
dath_X2
dath_X4
dath_X3
dath_X5
dmel_S-Lap2
dmel_S-Lap1
0.1
GAPsec
dath_X4
dath_X6
dath_X1
dath_X5
dath_X2
dath_X3
dmir_Y2
dmir_Y1
dmir_Y4
dmir_Y3
dmir_X2
dmir_X4
dmir_X5
dmir_X6
dmir_X3
dmir_X1
dpse_GAPsec
dper_GAPsec
dpse_Yconsensus
dpse_GAPsec_Dup
dper_GAPsec_Dup
dmel_GAPsec
0.03
B
truncated duplicate
duplicate with internal deletion
tandem duplicate
initial locus
dmir XR
dmir Y
C
XR Contig A
XR Contig B
dath XR
5 Kb
S-Lap
GAPsec


TOTAL RNA
GAPsec
S-Lap_Dup
S-Lap
log2(coverage)
20
10
0
−10
SMALL RNA
log2(coverage)
10
0
−10
1 Kb
D. miranda XR