Defective splicing of Y-chromosome-linked gigantic genes underlies hybrid male sterility in Drosophila

Adrienne Fontan[1, 2, *], Romain Lannes[1, *], Jaclyn M. Fingerhut[1,3], Jullien M. Flynn[1,3, #], and Yukiko M. Yamashita[1,2,3, **]

[1]Whitehead Institute for Biomedical Research

[2]Massachusetts Institute of Technology, Department of Biology

[3]Howard Hughes Medical Institute

* these authors equally contributed to this work

** corresponding author: yukikomy@wi.mit.edu

[#] present address: Université Laval, Institute of Integrative and Systems Biology, Department of Biology, Québec, Canada

Abstract

The Y chromosome evolves rapidly, often differing dramatically even between closely related species. While such divergence has long been suspected to contribute to hybrid male sterility, leading to reproductive isolation and thus speciation, the underlying mechanisms remain elusive. Here, we identify a molecular basis linking Y chromosome divergence to reproductive isolation in *Drosophila*. We show that male hybrids between *D. simulans* and *D. mauritiana* fail to properly express key Y-linked fertility genes. These genes contain unusually large introns, exceeding megabases and show substantial sequence divergence between species. In the hybrids, these gigantic introns are misprocessed, resulting in widespread splicing defects, including aberrant “back-splicing” events that join later exons to earlier ones. Our findings suggest that sequence divergence within introns can disrupt essential gene expression through defective splicing, providing a mechanistic link between rapid Y chromosome evolution and hybrid sterility. This work highlights the underappreciated role of intronic divergence in speciation.

Introduction

Extensive studies have provided insights into the process of reproductive isolation, which prevents gene flow between two populations, leading to speciation [1]. Although there are multiple causes of reproductive isolation, one that often appears early in the course of speciation is hybrid sterility [2]. The heterogametic sex (XY males and ZW females) suffers hybrid sterility and inviability more often than the homogametic sex (XX females and ZZ males), an observation known as Haldane's rule [3]. Multiple hypotheses have been put forward to explain Haldane's rule [4, 5]. For example, the dominance theory proposes that the heterogametic sex, which has only one X chromosome from one species, may suffer incompatibility with the autosomes from the other species, leading to inviability or sterility in hybrids [6]. It has also been hypothesized that incompatible interactions between two sex chromosomes (X vs. Y or Z vs. W) from different species may explain Haldane's rule [4]. It is generally thought that Haldane's rule likely represents a composite phenomenon resulting from multiple causes [5, 7]. Accordingly, much remains to be understood about how individual factors contribute to hybrid sterility in a manner that preferentially impacts the heterogametic sex.

While many studies have documented hybrid sterility across species including mammals, insects, and plants, Drosophila species have served as a leading model of hybrid sterility for a century [8-10], including the three *D. simulans* clade species (*D. simulans, D. mauritiana*, and *D. sechellia)* that diverged ~250 thousand years ago (Fig 1A) [2, 11, 12]. These species exhibit hybrid male sterility in reciprocal crosses [11], whereas hybrid females are fertile, conforming to Haldane's rule [3]. Mapping of the genes that cause sterility of *D. simulans/D. mauritiana* hybrid males has revealed the presence of multiple factors on the X chromosome and the autosomes [13-17]. One of such genes, the X-linked *OdsH* gene*,* is sufficient to cause male sterility in a heterospecific context: *D. simulans* males with the *D. mauritiana OdsH* gene are sterile [18, 19]. Cytologically, *D. simulans/D. mauritiana* hybrid males exhibit abnormalities during spermatid differentiation, such as defects in axoneme elongation and sperm individualization [11, 20, 21]. Several autosomal genes required for spermatogenesis (e.g., *sa, dj, Mst84Dc,* and *Mst98Ca*) were downregulated in *D. simulans/D. mauritiana* hybrids [22-26], potentially explaining the cytological phenotype observed in the sterile hybrids.

In parallel, the Y chromosome, essential for male fertility in Drosophila [27], has been speculated to contribute to hybrid male sterility. The Drosophila Y chromosome has diverged

rapidly, as is the case for many species, including mammals [28, 29], leading to a model that Y chromosome divergence may contribute to the hybrid male sterility observed in Drosophila species. Indeed, introgression of the Y chromosome into other species (e.g., the *D. simulans* Y chromosome in a *D. mauritiana* background and vice versa) causes sterility, suggesting that the Y chromosome is not compatible between these species [12, 30, 31]. However, the underlying molecular mechanisms that render the Y chromosome nonfunctional in hybrids remain poorly understood.

In recent years, Y chromosome divergence between Drosophila species has been characterized in more detail, revealing the rearrangement of gene locations on the Y chromosomes and considerable changes in repetitive DNA sequences, such as satellite DNA [32, 33]. Satellite DNA divergence has been speculated to underlie hybrid incompatibility, although the molecular mechanisms behind it are still not fully understood [2, 34]. Curiously, satellite DNA is found in the introns of several Y-linked fertility genes. Satellite DNA-containing introns sometimes exceed megabases in size while only containing a small fraction of protein-coding sequence (typically up to ~15kb), a phenomenon called intron gigantism [35-37]. Intron gigantism is observed in a handful of genes within the genomes of a broad range of species, from flies to humans: a well-known example is the mammalian dystrophin gene, which spans over 2 megabases but encodes only 11kb of the protein-coding sequence [38, 39]. Although the functional significance of intron gigantism remains elusive, previous studies on Y-linked fertility genes in *D. melanogaster* revealed that gigantic introns likely impose considerable challenges for gene expression, including transcription and splicing [40, 41]. For example, the transcription of gigantic genes attenuated within gigantic introns upon perturbation of splicing [41].

Here, we show that hybrid males produced by the mating of *D. simulans* females and *D. mauritiana* males exhibit unique defects in the expression of Y-linked gigantic genes, likely contributing to hybrid sterility. Our analyses reveal that the compromised expression of fertility genes in hybrids results, at least in part, from splicing errors particularly at gigantic introns. DNA sequences of gigantic introns are poorly conserved between *D. simulans* and *D. mauritiana*. Taken together, we propose that the defective splicing of gigantic introns due to poorly conserved intronic sequences contributes to hybrid male sterility. Our results may provide a mechanistic link between the rapid divergence of the Y chromosome and hybrid male sterility, providing a potential explanation for Haldane's rule.

Results

Defective spermatogenesis of *D. simulans/D. mauritiana* hybrids is associated with downregulation of Y-chromosome-linked fertility genes

Interspecies crosses between *D. simulans* females and *D. mauritiana* males have long been known to produce sterile hybrid males due to defects in post-meiotic processes, including sperm axoneme development and sperm DNA packaging [11, 20, 21] (Fig 1). Spermatogenesis of the parental species, *D. simulans* and *D. mauritiana,* is highly similar in gross anatomy and cellular features to the well-characterized *D. melanogaster* spermatogenesis, as described previously [42, 43] (Fig 1B-F, Supplementary Fig S1). The apical tip of the testis contains mitotically proliferating cells (germline stem cells, gonialblasts, and spermatogonia), which undergo four rounds of mitosis with incomplete cytokinesis to yield a cluster of 16 spermatogonia. These cells then enter the spermatocyte stage, progressively growing in size (Fig 1C, Supplementary Fig S1A). Mature spermatocytes undergo meiosis (Fig 1D, Supplementary Fig S1B), and the resulting 64 interconnected haploid spermatids morphologically transform in a cyst (Fig 1E, Supplementary Fig S1C): the nuclei undergo stereotypical morphological changes through leaf, canoe, and needle stages, resulting in highly condensed sperm DNA [44, 45](Fig 1E, F, Supplementary Fig S1C, D). The entire cellular morphology also undergoes striking morphological changes, elongating the sperm axoneme and surrounding mitochondria to form 1.9mm sperm tails (Fig 1B). An actin-based structure called the individualization complex forms around needle-stage spermatid nuclei (Fig 1F, Supplementary Fig S1D), which subsequently progresses downwards toward sperm tails to eliminate excess cytoplasm and produce individual sperm [44, 46]. Mature sperm are then released into the seminal vesicle.

In contrast to the parental species, *D. simulans/D. mauritiana* hybrid males exhibited various defects in post-meiotic stages, as described previously [11, 20, 21]. Although they possessed normal gross anatomy with the stereotypical coiled tubular structure (Fig 1G), and underwent apparently normal germ cell development until meiotic divisions (Fig 1H), the hybrids began to exhibit clear defects after meiosis. First, in contrast to the parental species, where all 64 nuclei within a cyst polarize uniformly towards the distal end of the testis (Fig 1E, Supplementary Fig S1C), the hybrid cyst displayed depolarization with some nuclei incorrectly polarized towards the opposite end (Fig 1I, arrowheads). Some nuclei were hyper-condensed, indicating cell death (Fig 1I, arrows). Spermatids ultimately failed to condense their nuclear DNA properly (Fig 1J), never forming the needle-shaped nuclei of mature sperm observed in parental

species (Fig 1F, Supplementary Fig S1D). Furthermore, the individualization complex (Fig 1F, Supplementary Fig S1D) did not form in hybrids (Fig 1J). These cytological defects (described previously and further detailed in this study) explain the well-known sterility of *D. simulans/D. mauritiana* hybrids.

RNA sequencing on the testes of hybrids compared to those of the parental species revealed the downregulation of several Y-linked genes (Table 1). We conducted RNA sequencing using total RNA to capture transcription intermediates in addition to mature mRNAs. The results confirmed the previously reported downregulation of several autosomal genes in hybrids (see Methods). We then focused on Y-linked genes, because the Y chromosome harbors genes required for post-meiotic stages of spermatogenesis [47], and it has been shown that introgression of heterospecific Y chromosomes into other species' backgrounds led to sterility [12, 30]. Because the hybrid males studied here carry the *D. mauritiana* Y chromosome, the hybrids' RNA sequence reads were aligned to the *D. mauritiana* Y chromosome assembly [32] and were compared to the transcriptome of *D. mauritiana* testes by using a manually curated annotation of *D. mauritiana* Y-linked genes (Supplementary_file_1, Methods). Differential expression analysis showed that a subset of Y-linked genes, most notably *kl-3* and *kl-5*, were downregulated in hybrids (Table 1). We observed a slight but significant reduction in the expression of *kl-2*, *WDY*, *ppr-Y*, and *CCY* as well.

The *kl-2*, *kl-3*, and *kl-5* genes encode axonemal dynein subunits, which are required for sperm tail development and thus fertility [35]. Using single molecule RNA FISH (smRNA FISH) with probes for *kl-3* exons 1 and 14, we monitored the expression of *kl-3* in *D. simulans/D. mauritiana* hybrids over the course of spermatocyte development. Because the *kl-2*, *kl-3*, and *kl-5* genes contain gigantic introns that exceed megabases in length, their expression takes the entirety of spermatocyte development (80-90 hours) [40, 48], where the transcription of early exons begins in early spermatocytes, followed by the transcription of gigantic introns, then concludes with transcription of later exons in mature spermatocytes [40]. Expression of *kl-3* in *D. simulans* and *D. mauritiana* exhibited a spatiotemporal pattern similar to that observed in *D. melanogaster* (Fig 2B, D) [40]: the transcription of exon 1 was observed in early spermatocytes, whereas exon 14 was transcribed much later. Mature *kl-3* mRNA was observed in cytoplasmic granules, termed kl-granules (Fig 2C, E), which also contained a protein component Pontin (Fig 2H, I).

In contrast to *D. simulans* and *D. mauritiana*, we found that *D. simulans/D. mauritiana* hybrids exhibit striking defects in *kl-3* expression. Although exon 1 transcripts appeared in the early spermatocyte stage, exon 14 was rarely visible, and cytoplasmic kl-granules were absent, suggesting functional *kl-3* mRNA was not produced (Fig 2F, G). Similar results were obtained for the *kl-2* and *kl-5* genes, where the parental species exhibited the same expression patterns as *D. melanogaster* (Supplementary Fig S2B, C, E, F), but the hybrids did not (Supplementary Fig S2D, G). As expected from the lack of cytoplasmic *kl-2*, *kl-3*, and *kl-5* transcripts, immunofluorescence staining for Pontin showed that kl-granules were indeed absent in the hybrids (Fig 2J). We conclude that *D. simulans/D. mauritiana* hybrid males fail to express a subset of Y-linked axonemal dynein genes, which may contribute to hybrid sterility.

Transcription of *kl-2*, *kl-3,* and *kl-5* attenuates within gigantic introns in hybrids

More detailed analysis of the RNA sequencing reads revealed that the expression of *kl-2*, *kl-3*, and *kl-5* was specifically decreased toward the 3' side of the gene. For example, the read depth of *kl-3* dropped dramatically between exons 5 and 6 in hybrids relative *to D. mauritiana* (Fig 3A). Notably, the intron between exons 5 and 6 is a gigantic intron, containing 207kb of assembled sequence and a gap in the genome assembly due to the presence of repetitive DNA (Supplementary_file_1)[32]. A similar drop was observed between exons 13 and 14 of *kl-5,* and between exons 6 and 7 of *kl-2*, which are both gigantic introns (35kb plus a gap and 432kb plus a gap, respectively) (Fig 3A, Supplementary_file_1). These results are consistent with the absence of late exon of *kl-2*, *kl-3*, and *kl-5* visualized by smRNA FISH (Fig 2F, G, Supplementary Fig S2D, G). Together, these results revealed that *kl-2*, *kl-3*, and *kl-5* become downregulated in hybrid testes due to transcription downregulation within gigantic introns.

The sudden drop in read depth within gigantic introns is reminiscent of a phenotype we recently discovered in *D. melanogaster* following depletion of splicing factors [41]. We have shown that these gigantic introns are co-transcriptionally spliced, and co-transcriptional splicing is critical for the continuation of transcription [41]. We postulated that transcription may attenuate within gigantic introns upon perturbation of splicing, perhaps because excessively long unspliced transcripts attached to RNA polymerase II hinder the progression of transcription (Fig 3B, see also an alternative interpretation below) [41]. The dependence of transcription on splicing was specific to genes containing gigantic introns [41], suggesting large introns are more

sensitive to splicing perturbation. Considering these similarities, we hypothesized that Y-linked genes with gigantic introns may be downregulated in hybrids due to splicing defects.

D. simulans/D. mauritiana hybrids are defective in the splicing of gigantic introns

To test whether the read depth drop observed in the *D. simulans/D. mauritiana* hybrids may be associated with splicing defects, we analyzed the splicing of *kl-2*, *kl-3*, and *kl-5* in *D. simulans/D. mauritiana* hybrids compared to *D. mauritiana*. To this end, we developed a bioinformatics pipeline that examines not only the efficiency of splicing (i.e., the number of correctly spliced reads divided by the total number of spliced and nascent reads) but also potentially incorrect splicing events in which the end of an exon is connected to an unpredicted sequence (see Methods). Using this method, all reads containing the 3' end of a given exon were identified, and the identity of the adjoining sequence was analyzed. These reads were categorized as follows (Fig 4A):

i. Spliced: the 3' end of the exon was correctly joined to the 5' end of the next exon.
ii. Unspliced: the junction between the 3' end of the exon and the juxtaposing intron remained intact, indicative of a nascent RNA.
iii. Exon-other: the 3' end of the exon was connected to a downstream sequence that was not the predicted splice acceptor site. These reads may represent erroneous splicing events or products of recursive splicing. (Note that the reads that correspond to known alternative splice isoforms or overlapping gene were excluded from this category).
iv. Exon-clipped: the read was partially mapped to the 3' end of the exon, but the adjoining sequence could not be mapped linearly with the genome, and was marked as clipped by the aligner. This category of reads is indicative of chimeric fragments between 3' end of an exon and an RNA fragment that is not normally expected to be joined to the given exon. Such chimera may be biological, or also arise from issues with the library quality, such as the artifactual ligation of two unrelated DNA molecules during library preparation. Therefore, these reads require additional analysis for validation. Note that RNA sequence reads were considered in this category only if the 'clipped' portion was greater than 10 nucleotides.

We performed this analysis for each exon of *kl-2, kl-3,* and *kl-5* (Fig 4B). Several locations displayed significant difference between *D. mauritiana* and *D. simulans/D. mauritiana* hybrids (Fig 4B, introns marked with asterisks), including *kl-3* exons 5 and 15, and *kl-5* exons 13 and 19, all of which immediately precede gigantic introns. The sequence reads across these

junctions exhibited several notable features. First, the proportion of correctly spliced products decreased considerably in the hybrids, compared to *D. mauritiana*. Correspondingly, the proportion of unspliced reads often increased in the hybrids, suggesting that hybrids indeed fail to splice gigantic introns. It should be noted that even pure *D. mauritiana* often showed a large fraction of unspliced reads, likely reflecting the early spermatocyte population that had not transcribed the next exon or splice acceptor site.

Most notably, we found a considerable increase in exon-clipped reads in hybrids, where the 3' end of an exon (particularly, *kl-3* exon 5, *kl-5* exon 14, and *kl-5* exon 19, all of which immediately precede gigantic introns) was joined to 'unexpected' sequence (Fig 4A, B). Because exon-clipped reads are normally assumed to be a result of experimental artifacts, they are frequently excluded from analyses. However, given the possibility of splicing defects in hybrids, we further analyzed the identity of the clipped portion of the sequence following the 3' end of the exon. To our surprise, most of the clipped sequences were identical to the 5' end of an earlier exon: e.g., 24.3% of the reads containing the 3' end of *kl-3* exon 5 were exon-clipped, and of these, 66% were joined to the 5' end of either exons 2 or 4, suggestive of erroneous back-splicing (Fig 4C, D). Back-splicing is a phenomenon in which the 3' end of an exon is spliced to the 5' end of an earlier exon, likely forming a circular RNA molecule [49]. Such back-spliced products have been identified in a broad range of conditions from healthy tissues to cancer [50]. We speculate that such back-splicing may occur if the 3' end of an exon (at the splice donor site) is unable to locate the correct downstream splice acceptor site (Fig 4C). Similar exon-clipped reads indicative of back-splicing events were observed for *kl-5* exon 19, in which nearly all of the clipped fragments mapped to exons 15 or 17, and *kl-2* exon 6 reads, which mapped to exons 3 and 4 (Fig 4D, Supplementary_file_2).

The presence of exon-clipped reads (instead of experimental artifacts during library preparation) was validated by RT-PCR designed to specifically amplify the correctly spliced products vs. back-spliced products (Fig 5A), which was confirmed by sequencing of the PCR products (Supplementary Fig S3A-E). By further using RT-qPCR with primers to detect correctly spliced vs. back-spliced products, we found that the hybrids had much lower amounts of correctly spliced products (*kl-3* exon 5-exon 6 junction) than the wild type *D. mauritiana*, (Fig 5B, C, Supplementary_file_3). Correspondingly, the hybrids displayed increased levels of back-splicing (the *kl-3* exon 5-exons 4/2 junctions) (Fig 5B, C, Supplementary_file_3), confirming the RNA sequencing results (Fig 4B, D). It is important to note that exon-clipped reads and back-

spliced products were also present in wild type *D. mauritiana* at junctions where the read depth drops, implying that the splicing of gigantic introns is challenging even in the pure species. Importantly, however, mature mRNA was observed in the pure species, suggesting that a sufficient amount of RNAs undergo correct splicing.

Back-splicing may also provide an alternative explanation to the drop in RNA sequencing reads observed in the hybrids (Fig 3A): although we have originally speculated that the unspliced gigantic RNA attached to RNA polymerase II may hinder the progression of transcription (Fig 3B)[41], back-splicing will result in uncapped RNA attached to RNA polymerase II, which in turn may cause degradation of nascent RNA and termination of transcription (Fig 5D). Taken together, we conclude that the hybrid males fail to correctly splice gigantic introns, explaining the lack of mature mRNAs of *kl-2*, *kl-3*, and *kl-5* in hybrids (Fig 2). These results indicate that defective splicing of Y-linked fertility genes may contribute to hybrid sterility.

Splicing defects are a general feature of Y-linked genes with gigantic introns

kl-2, kl-3, and *kl-5* are not the only Y-linked genes with gigantic introns. Although other Y-linked gigantic genes did not exhibit clear downregulation in the hybrids (Table 1), several other Y-linked genes displayed drops in read depth within gigantic introns (e.g., *wdy* intron 4, *ppr-Y* intron 1, Fig 6A), suggesting that gigantic introns indeed tend to cause attenuation of transcription. In these cases, the degree of read depth drop was considerably muted compared to *kl-2*, *kl-3*, and *kl-5*, and/or occurred toward the end of the gene (Fig 6A), explaining why the defects were not detected as downregulation of gene expression (Table 1). Nonetheless, in many cases (*WDY, Ppr-y, ORY, CCY, PRY*), gigantic introns were associated with defective splicing in hybrids, encompassing defects similar to *kl-2*, *kl-3*, and *kl-5*, including back-splicing, which was validated by RT-PCR and qPCR (Fig 6B, Supplementary Fig S3F, G, H, Supplementary_file_3). These results imply that splicing failure is a common feature for many gigantic introns.

Size and sequence conservation correlate with splicing errors in hybrids

We found that there is a general correlation between the size of an intron and the degree of its splicing defects (Fig 7, note that for this analysis, we only considered the known length of an intron, ignoring genome assembly gaps). First, we examined the levels of defective splicing for introns of Y-linked genes, comparing the hybrids to the parental species *D. mauritiana*, which

shares the Y chromosomes with the hybrids (Fig 7A). This analysis revealed the clear trend that larger introns have a higher likelihood of defective splicing, both in the pure species and hybrids [a positive Spearman correlation of 0.48 (p-value $5.62*10^{-6}$) in the pure species, and 0.48 (p-value $8.68*10^{-6}$) in the hybrids]. The hybrids exhibited a higher likelihood of defective splicing than the pure species (Fig 7A, Supplementary_file_4). Importantly, the splicing defects in hybrids became more profound as intron size increased, while the pure species showed a lesser increase in such defects. These results imply that although large introns are generally challenging to splice even in pure species, the pure species are able to handle them relatively well. In contrast, the hybrids may be defective in such a mechanism, failing in the splicing of gigantic introns.

We applied the same analysis to the X-linked genes, comparing the hybrids with the *D. simulans* data (Fig. 7B, Supplementary_file_4). We chose to analyze X-linked genes because the X chromosome in these hybrids comes solely from *D. simulans*, allowing each read to be mapped unambiguously to the *D. simulans* genome. Although X-linked genes also exhibited an intron size-dependent increase in splicing defects, and the hybrids exhibited more splicing defects than *D. simulans*, the degree of splicing defects in the hybrids was more profound with the Y-linked gigantic genes than X-linked genes. These results indicate that introns of Y-linked genes may also have other features, in addition to size, that make them sensitive to splicing defects. Indeed, analysis of DNA sequence homology of the *kl-3* gene revealed that its introns are highly divergent between *D. simulans* and *D. mauritiana*, whereas the exon sequences are well conserved (Fig. 7C). In contrast, both exons and introns of *Flotillin-2*, an example of an X-linked gene, are highly conserved between two species (Fig. 7D). The genome-wide analysis of sequence conservation further revealed that the introns of Y-linked genes exhibit marked divergence between *D. simulans* and *D. mauritiana,* whereas exons of Y-linked genes, as well as both introns and exons of the X chromosome/autosomes, are highly conserved (Supplementary Figure S4, Methods). In summary, we propose that the divergence of intronic sequences associated with Y-linked fertility genes may explain the difficulty in splicing, leading to hybrid sterility.

Discussion

In this study, we show that the splicing of Y-linked gigantic genes is severely impacted in male hybrids between *D. simulans* and *D. mauritiana*, leading to the downregulation of multiple fertility genes. We propose that such downregulation of fertility genes contributes to hybrid

sterility, together with other proposed causes of hybrid male sterility. Our work provides a potential explanation as to why the heterospecific Y chromosome does not support male fertility. The Y chromosome is the most rapidly diverging chromosome across species [28, 29, 51, 52], thus our finding may provide a generalizable principle of Y chromosome-mediated hybrid sterility.

Previous studies have provided significant insights into the mechanisms of hybrid male sterility, and the presence of multiple hybrid incompatibility genes was shown (see Introduction). More broadly, one of the leading explanations for hybrid sterility is that genomic conflicts caused by independently evolved male meiotic drivers result in hybrid sterility [53]. Indeed, there are several examples where male sterility and meiotic drive are caused by the same genetic elements, including *tmy* in *D. simulans/D. mauritiana* [54], *Overdrive* in *D. pseudoobscura* [55], and *Sex Ratio (SR)* in *D. pseudoobscura* [56]. However, these drivers and their suppressors are mostly mapped to X chromosomes and autosomes [54, 57-59]. Thus, they do not explain why the Y chromosome is nonfunctional when introgressed into other species' genomic context. Another proposed cause of hybrid sterility is transposon derepression [60-64], although there are cases where transposons are not dramatically derepressed in sterile hybrids (e.g., hybrids between *D. arizonae* and *D. mojavensis*) [65]. Our RNA sequencing analysis revealed some degrees of transposon derepression in the *D. simulans/D. mauritiana* hybrid males (Supplementary Table S1), which may also contribute to hybrid dysfunction.

The present study demonstrates the defective splicing of gigantic introns of Y-linked fertility genes in *D. simulans/D. mauritiana* hybrids, potentially providing a link between rapid divergence of the Y chromosome and hybrid male sterility. Previous studies have shown that gigantic introns of Y-linked fertility genes pose challenges in the processes of gene expression, such as transcription and/or splicing, even in wild type flies, requiring additional factors to aid in the process: multiple proteins have been identified to bind to intronic transcripts (e.g., Blanks, Heph, and Maca) and are required for the production of mature mRNA from fertility genes with gigantic introns [40, 66]. Additionally, the transcription of gigantic genes is sensitive to the perturbation of splicing [41], further illuminating the challenging nature of producing mRNA from gigantic genes. Because intronic sequences are highly divergent (Supplementary Figure S4), it is conceivable that RNA-binding proteins, which bind gigantic introns to facilitate their splicing, become incompatible between species (Fig 7E). It is tempting to speculate that incompatibility of this nature contributes to Haldane's rule [3], specifically impacting male fertility.

In summary, the present study provides a new model on a cause of hybrid male sterility, providing a potential molecular explanation as to why the Y chromosomes are incompatible between species.

Methods

Fly husbandry and strains used

All *Drosophila melanogaster* strains were raised on standard Bloomington medium at 25 °C. The following stocks were used: *D. simulans* wXD1 and *D. mauritiana* w12 (obtained from Dr. Ching-Ho Chang).

Immunofluorescence staining

Testes from 1- to 3-day-old males were dissected in 1x PBS and fixed in 4% formaldehyde in 1x PBS for 30 minutes. Fixed testes were then washed in 1x PBST (PBS containing 0.1% Triton X-100) for at least 1.5 hours, followed by incubation with primary antibodies diluted in 1x PBST containing 3% BSA at 4 °C overnight. Samples were washed three times in 1x PBST for 30 minutes each and then incubated with secondary antibodies in 1x PBST with 3% BSA at 4 °C overnight. After a similar washing procedure, samples were mounted in VECTASHIELD with DAPI (Vector Labs). Images were acquired using a Leica Stellaris 8 confocal microscope with a 63x oil immersion objective lens (numerical aperture 1.4) and processed with Fiji (ImageJ) software. The primary antibodies used were: anti-Pontin (1:200; guinea pig) [67], anti-pH3 ser10 (1:200, rabbit), and anti-ATP5a (1:1000; mouse; Abcam, ab14748). Phalloidin-Alexa Fluor 488 (1:200; Thermo Fisher Scientific, A12379) was used to stain F-actin. Alexa Fluor-conjugated secondary antibodies (Life Technologies) were used at a 1:200 dilution.

Single-molecule RNA Fluorescent *in situ* hybridization

RNA FISH was performed as previously described [68]. All solutions used were RNase free. Testes from 1–3 day old flies were dissected in 1X PBS and fixed in 4% formaldehyde in 1X PBS for 30 minutes. Testes were washed briefly in 1X PBS and permeabilized in 70% ethanol overnight at 4°C. Testes were briefly rinsed with wash buffer (2X saline-sodium citrate (SSC), 10% formamide) and then hybridized overnight at 37°C in hybridization buffer (2X SSC, 10% dextran sulfate (sigma, D8906), 1mg/mL E. coli tRNA (sigma, R8759), 2mM Vanadyl Ribonucleoside complex (NEB S142), 0.5% BSA (Ambion, AM2618), 10% formamide). Following hybridization, samples were washed three times in wash buffer for 20 minutes each at

37°C and mounted in VECTASHIELD with DAPI (Vector Labs). Images were acquired using a Leica Stellaris8 confocal microscope with a 63X oil immersion objective lens (NA = 1.4) and processed using Fiji (ImageJ) software.

Fluorescently labeled probes were added to the hybridization buffer to a final concentration 100nM. Probes against *kl-3*, *kl-5*, and *kl-2* exons were designed using the Stellaris RNA FISH Probe Designer (Biosearch Technologies, Inc.) available online at www.biosearchtech.com/stellarisdesigner. Each set of custom Stellaris RNA FISH probes was labeled with Quasar 670, Quasar 570 or Fluorescein-C3. Probe information can be found in Supplementary_file_5 [40].

RNA isolation and sequencing

Total RNA was purified from 2-5 day old adult testes (100 pairs/sample) by TRIzol (Invitrogen) extraction according to the manufacturer's instructions. Libraries were prepared for RNA sequencing using the KAPA Biosystems RNA HyperPrep Kit with RiboErase according to manufacturer's directions with some modifications. Briefly, 500 ng of total RNA was ribo-depleted by hybridization of complementary DNA oligonucleotides. The set of complementary oligonucleotides was a custom panel designed for *Drosophila*. This was followed by treatment with RNase H and DNase to remove rRNA duplexed to DNA and original DNA oligonucleotides. The enriched fraction was then fragmented with heat and magnesium, and first-strand cDNA was generated using random primers. Strand specificity was achieved during second-strand cDNA synthesis by replacing dTTP with dUTP, which quenches the second strand during amplification, and the cDNA is then A-Tailed. The final double strand cDNA was then ligated with indexed adapters. Finally, the library was amplified using a DNA Polymerase that cannot incorporate past dUTPs, effectively quenching the second strand during PCR. Libraries were enriched for fragments between 500 – 1000bp with two additional cycles of PCR followed by a size selection using a 1.5% gel on a Pippin Prep (Sage Science) electrophoresis instrument. Final libraries were quantified by qPCR and Fragment Analyzer. Samples were sequenced on a NOVASEQ 6000, producing 250 × 250bp paired-end reads.

RT-PCR, RT-qPCR, and sequencing

Total RNA was purified from 2-5 day old adult testes (50 pairs/sample) by TRIzol (Invitrogen) extraction according to the manufacturer's instructions. 1 µg total RNA was reverse transcribed with SuperScript III First-Strand Synthesis Supermix for qRT-PCR (Invitrogen), followed by PCR using Phusion High Fidelity DNA Polymerase (New England Biolabs) with

DMSO according to the manufacturer's instructions. PCR products were purified via gel extraction using a QIAquick Gel Extraction Kit (Qiagen). All reactions were done in technical duplicates with two biological replicates. Exon junctions were chosen based on the most highly abundant back-spliced transcripts from our RNAseq data set. PCR primers were designed to span exon junctions such that the product would only be detected if the 3' end of the first exon was spliced directly to the 5' end of the second exon. Primer sequences are listed in Supplementary_file_S5. DNA concentration was measured by a Qubit 3.0 Fluorometer (Invitrogen). One replicate of *D. mauritiana* and one replicate of *D. simulans/D. mauritiana* hybrid PCR products were sequenced for each PCR target. Premium PCR Sequencing was performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

For RT-qPCR, total RNA was isolated and reverse transcribed as described above. qPCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) on a QuantStudio 6 Flex Real-Time PCR system (Applied Biosystems). Relative expression levels were normalized to GAPDH and *D. mauritiana* controls. All reactions were done in technical triplicates with two biological replicates. Graphical representation is inclusive of all replicates. Primer sequences are listed in Supplementary_file_S5.

Bioinformatics analysis

Read alignment: Paired-end reads (251 x 251 bp) were quality trimmed using fastp with the following option "--length_required 100 --disable_adapter_trimming --trim_poly_g --cut_tail -M 20". We used *D. simulans* refseq genome (GCF_016746395.2) and the long reads assembly of *D. mauritiana* by (https://doi.org/10.5061/dryad.280gb5mr6). Alignment was performed using STAR (2.7.1a) with default parameters.

Y gene annotation: MMseqs2 was used to align the translated product of *D. melanogaster* Y genes to both the *D. mauritiana* and *D. simulans* genome (https://doi.org/10.5061/dryad.280gb5mr6). We manually parsed the alignment file to annotate the gene positions. We further manually refined the annotation using our RNA-seq data.

Transposable element: RepeatMasker (http://www.repeatmasker.org) was used to extract the transposable element using One Code To find Them All, from both *D. simulans* and *D. mauritiana* (https://doi.org/10.5061/dryad.280gb5mr6). We merged the sequences of the transposable elements from both species. Then we used TEtools [69] to count the reads

matching to transposable element. Finally, we used DESeq2 [70] to test from differential expression of transposable element.

Splicing defect: We Used OmniSplice [71] (https://github.com/rLannes/Fontan_2025) to identify splicing defect.

Genome wide intron and exon sequence identity estimates: Because the sequences of gigantic were poorly conserved and so large between *D. simulans* and *D. mauritiana*, it was not possible to calculate the homology of intronic sequences. Thus, using bowtie2, DNA sequencing of *D. simulans* (SRR22548176) was aligned to both the genomes of *D. simulans* and *D. mauritiana* (long reads genome assembly (https://doi.org/10.5061/dryad.280gb5mr6)). The proportion of base covered (by at least one read) was used as a proxy of sequence conservation (because poorly conserved sequence will not be mapped). Comparison between *D. simulans* short read sequencing vs. *D. simulans* long-read assembled genome served as the control, where all sequence reads must be perfectly conserved (however, note that some sequences showed low coverage, likely because of genome assembly/strain difference). For the X genes, we used the gtf file of the genome assembly (https://doi.org/10.5061/dryad.280gb5mr6) and for the Y linked genes we used our annotation. Then using those gtf files, we made two bed files one describing all exons and the other all introns. Finally, using bedtools coverage (version v2.29.2)[72], we computed the percentage of base covered by at least one read for both all introns, and all exons. (see GitHub repository to reproduce the code: https://github.com/rLannes/Fontan_2025)

Differential expression: We aligned both *D. simulans* and *D. simulans/D. mauritiana* hybrids to *D. simulans* refseq genome. Then using features count and Deseq2, we performed differential gene expression analyses. Analysis of the RNA sequencing results confirmed downregulation of several autosomal genes in hybrids, consistent with previous studies [22-26]: *dj, Mst98Ca*, and *Mst84Dc* (which is annotated as a lncRNA LOC27208299 in the recent *D. simulans* genome annotation (GCF_016746395.2)) (estimated log2 fold change -1.37, -0.81, -1.03, and adjusted p-value $7.8*10^{-51}$, $1.54*10^{-14}$, 0.07 for *dj*, *Mst98Ca*, and *LOC27208299,* respectively).

Plotting: All figures were generated using python (3.8.10), matplotlib (3.7.1). The dot plot was generated using a custom code available on GitHub. Briefly, we cut all sequence in smaller chunks. Using the Needleman-Wunsch algorithm we aligned all the chunks of a sequence to all the chunks of the other sequence. We used the ECDNA matrix for scoring with the following

parameters (gap opening: -10, gap extension: -0.5). From those alignments we reconstructed the dot plot. For the coverage plot, we designed an in-house code available at https://github.com/rLannes/Fontan_2025.

Data availability

All data are provided in the manuscript. RNA sequencing data is deposited to SRA under the bioproject accession number: PRJNA1248387.

Code Availability

Bioinformatics code used in this study is available at https://github.com/rLannes/Fontan_2025.

Acknowledgements

We thank the members of the Yamashita lab for discussions and comments on the manuscript, and Phil Sharp for helpful suggestions. We thank Flybase and the National Drosophila Species Stock Center for reagents and critical information. We thank the Genome Technology Core at the Whitehead Institute for their consultation and aid in designing and performing RNA sequencing experiments.

Funding: Howard Hughes Medical Institute (Y.M.Y)

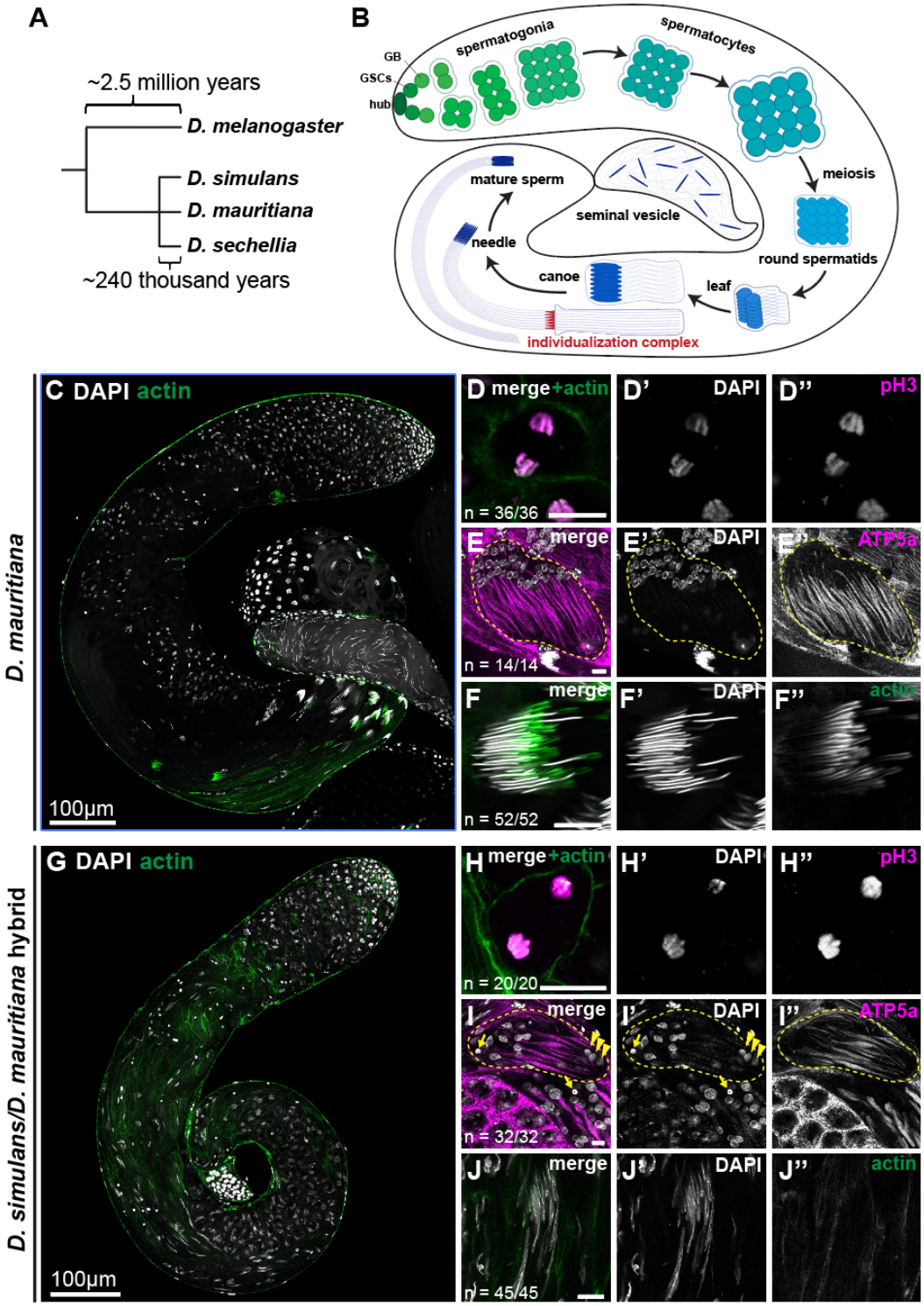
Author contributions:

Conceptualization: AF, RL, YY
Methodology: AF, RL, JFi, JFl, YY
Investigation: AF, RL, JFi, JFl, YY
Funding acquisition: YY
Supervision: YY
Writing and editing: AF, RL, JFi, YY

Competing interests

Authors declare no competing interests

1 **Figures**



2

3 **Figure 1: *D. simulans/D. mauritiana* hybrids display cytological defects in post-meiotic**
4 **spermatids.**
5 A. Phylogeny of the *Drosophila melanogaster* complex.

B. Schematic of *Drosophila* spermatogenesis. Germline stem cells reside in the niche (formed by hub cells) at the apical tip of the testis. Differentiating progeny of germline stem cells undergo transit-amplifying divisions as gonialblasts (GB) and then as spermatogonia with incomplete cytokinesis, yielding a cyst of 16 interconnected germ cells. These cells enter the spermatocyte stage, growing in size and expressing genes required for meiosis and post-meiotic development. Spermatocytes then undergo meiosis, and haploid spermatids undergo morphological transformation to become functional sperm.

C-F. *D. mauritiana* testis and germ cells. Low magnification view of the whole testis (C) stained for F-actin (green, stained with phalloidin-Alexa 488) and DAPI (gray). Meiotic anaphase (D) stained for phospho-H3 ser10 (magenta) and DAPI (gray). Early elongating spermatids (E) stained for ATP5a (magenta, mitochondria) and DAPI (gray). Needle stage spermatids (F) stained for F-actin (green, stained with phalloidin-Alexa 488, individualization complex). Scale bar: 10μm, unless otherwise noted.

G-J. *D. simulans/D. mauritiana* hybrid testis and germ cells. Low magnification view of the whole testis (G), meiotic anaphase (H), early elongating spermatids (I), and canoe stage spermatids (J), with the same staining as C-F. Scale bar: 10μm, unless otherwise noted. Arrows indicate hyper-condensed (dying) nuclei; arrowheads indicate depolarized spermatid nuclei.

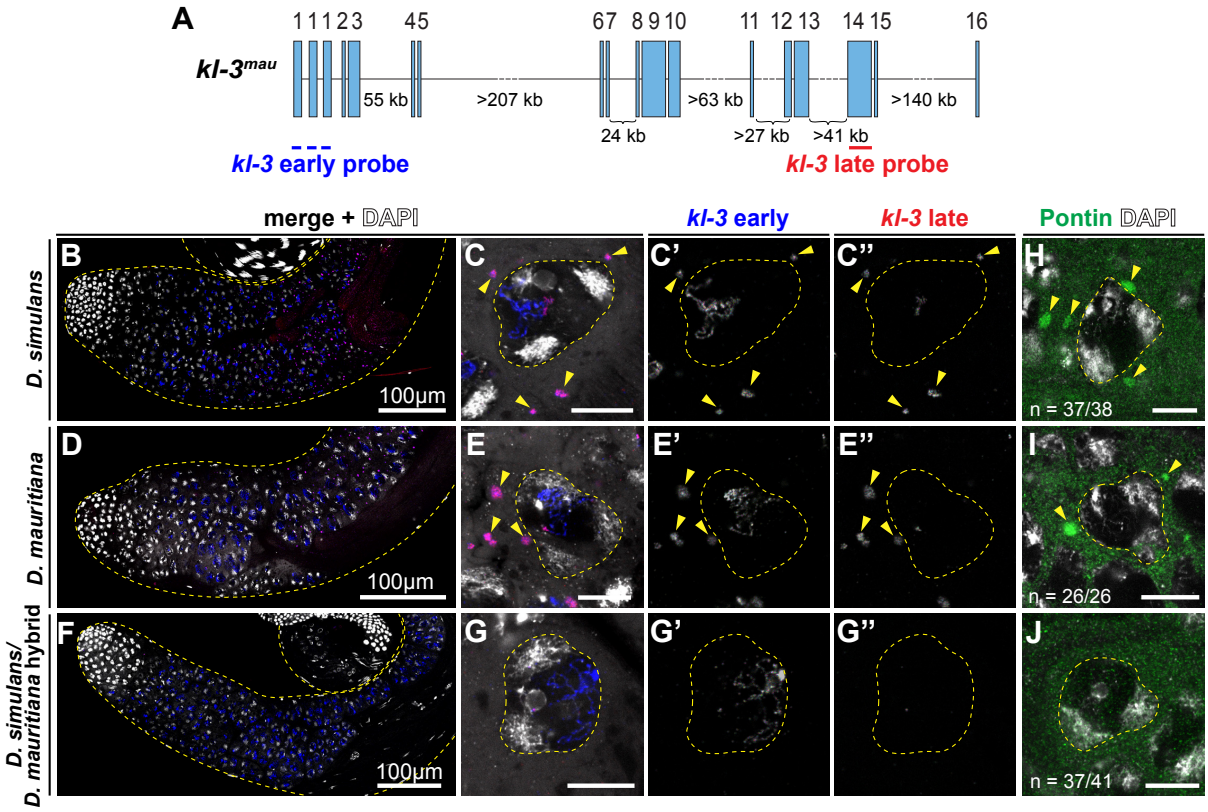


Figure 2: kl-granules are absent from *D. simulans/D. mauritiana* hybrid spermatocytes.

A. *D. mauritiana kl-3* (*kl-3*^{mau}) gene structure. The positions of smRNA FISH probes are indicated.

B-G. RNA in situ hybridization for *kl-3* transcripts in *D. simulans* (B, C), *D. mauritiana* (D, E) and *D. simulans/D. mauritiana* hybrids (F, G). Apical third of the testis (B, D, F) and individual late spermatocyte nuclei (C, E, G) are shown. Spermatocyte nuclei are indicated by dashed lines. Arrowheads indicate cytoplasmic kl-granule containing mature *kl-3* mRNA. Blue: *kl-3* early exons. Red: *kl-3* late exons. Gray: DAPI. Scale bar: 10µm, unless otherwise noted.

H-J. Individual late spermatocyte nucleus from *D. simulans* (H), *D. mauritiana* (I) and *D. simulans/D. mauritiana* hybrids (J). Spermatocyte nuclei are indicated by dashed lines. Arrowheads indicate cytoplasmic kl-granule stained for anti-Pontin antibody (green).

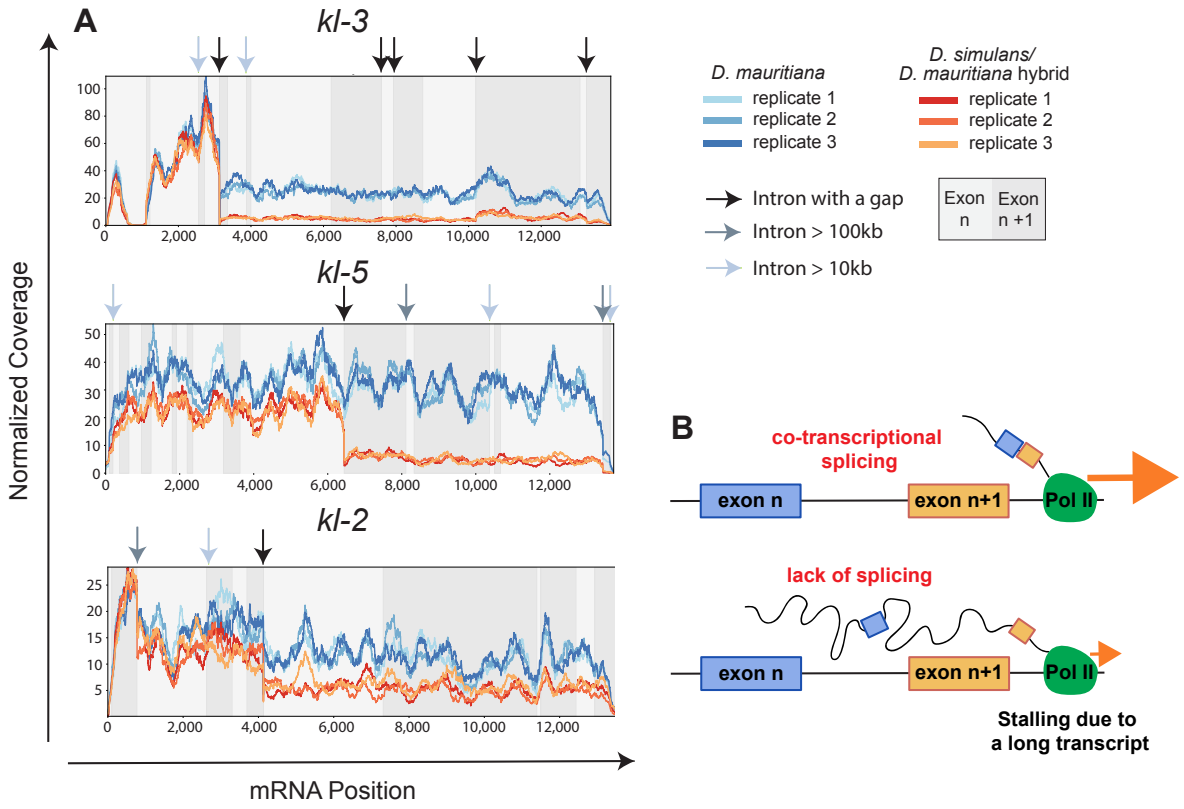


Figure 3: Transcription of *kl-2, kl-3,* and *kl-5* attenuates in *D. simulans/D. mauritiana* hybrids relative to *D. mauritiana* within gap-containing gigantic introns.

A. Normalized read depth coverage of the *kl-2, kl-3,* and *kl-5* genes along the coding sequence in *D. mauritiana* vs. *D. simulans/D. mauritiana* hybrids. Alternating shades of gray in the background indicate different exons. Arrows point to the location of large or gap-containing introns.

B. A proposed model of stalled transcription due to defective splicing (after [41]).

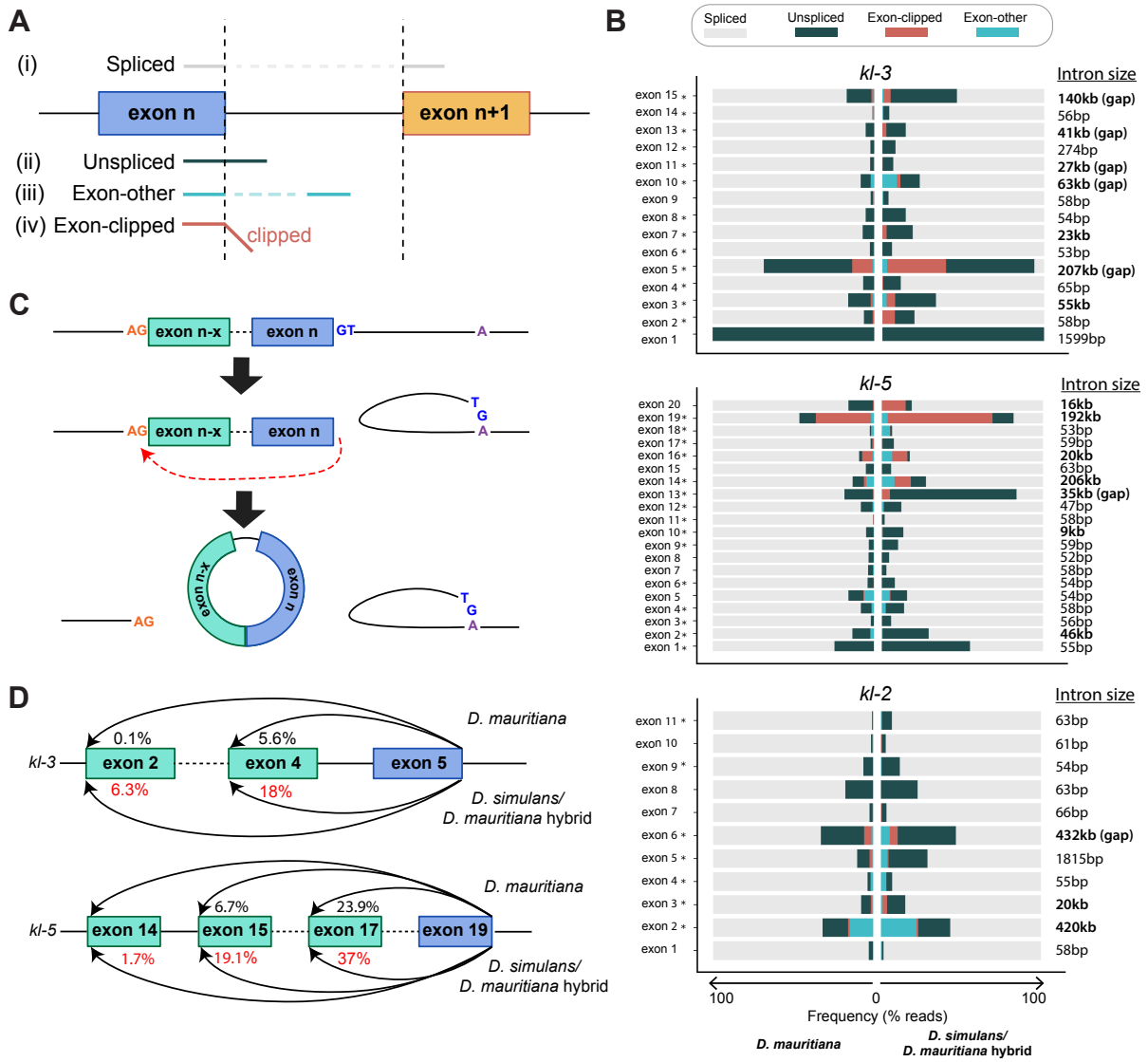


Figure 4. *D. simulans/D. mauritiana* hybrids exhibit splicing defects in *kl-2, kl-3,* and *kl-5*.

A. Categories of sequence reads that overlap with the end of exons.

B. Sequence read analysis for each exon end of the *kl-2, kl-3,* and *kl-5* genes in *D. mauritiana* vs. *D. simulans/D. mauritiana* hybrids. Asterisks(*) indicate significant difference in the proportion of correctly spliced reads between *D. mauritiana* and *D. simulans/D. mauritiana* hybrids.

C. Schematic of back-splicing producing circular RNA.

D. Frequency of back-spliced reads for *kl-3* exon 5 ends and *kl-5* exon 19 ends in *D. mauritiana* vs. *D. simulans/D. mauritiana* hybrids. Percentages indicate the frequency within the total reads aligning to the end of *kl-3* exon 5 and *kl-5* exon 19.

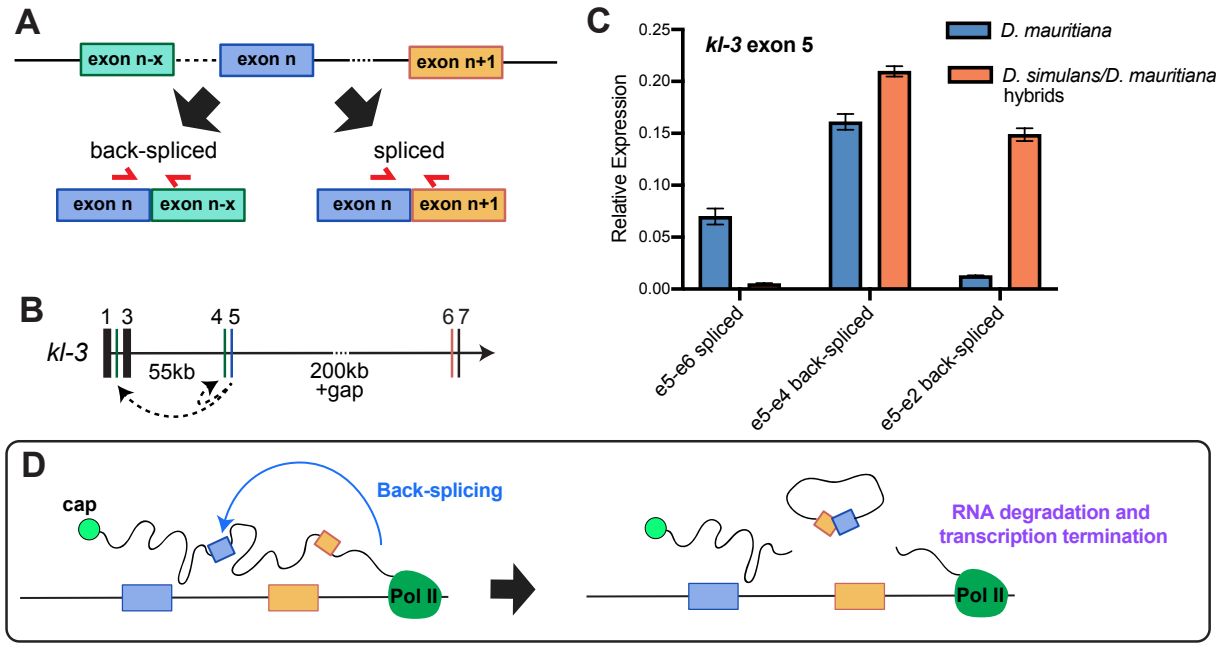


Figure 5. Back-splicing is increased in *D. simulans/D. mauritiana* hybrids.

A. RT-qPCR primer designs to detect spliced vs. back-spliced products
B. Schematic of the exon-intron structure surrounding kl-3 exon 5 back-splicing junctions. Exons are represented by vertical lines and labeled by exon number. Introns are represented by horizontal lines; a dotted line indicates a gap in the genome assembly
C. RT-qPCR to detect the presence of correctly spliced and back-spliced transcripts of kl-3 exon 5, normalized to GAPDH. Error bars represent SEM across three technical replicates. (See Supplementary_file_3 for raw CT values and biological replicates.)
D. A model of how back-splicing may lead to transcriptional attenuation (alternative model to the one depicted in Fig. 3b).

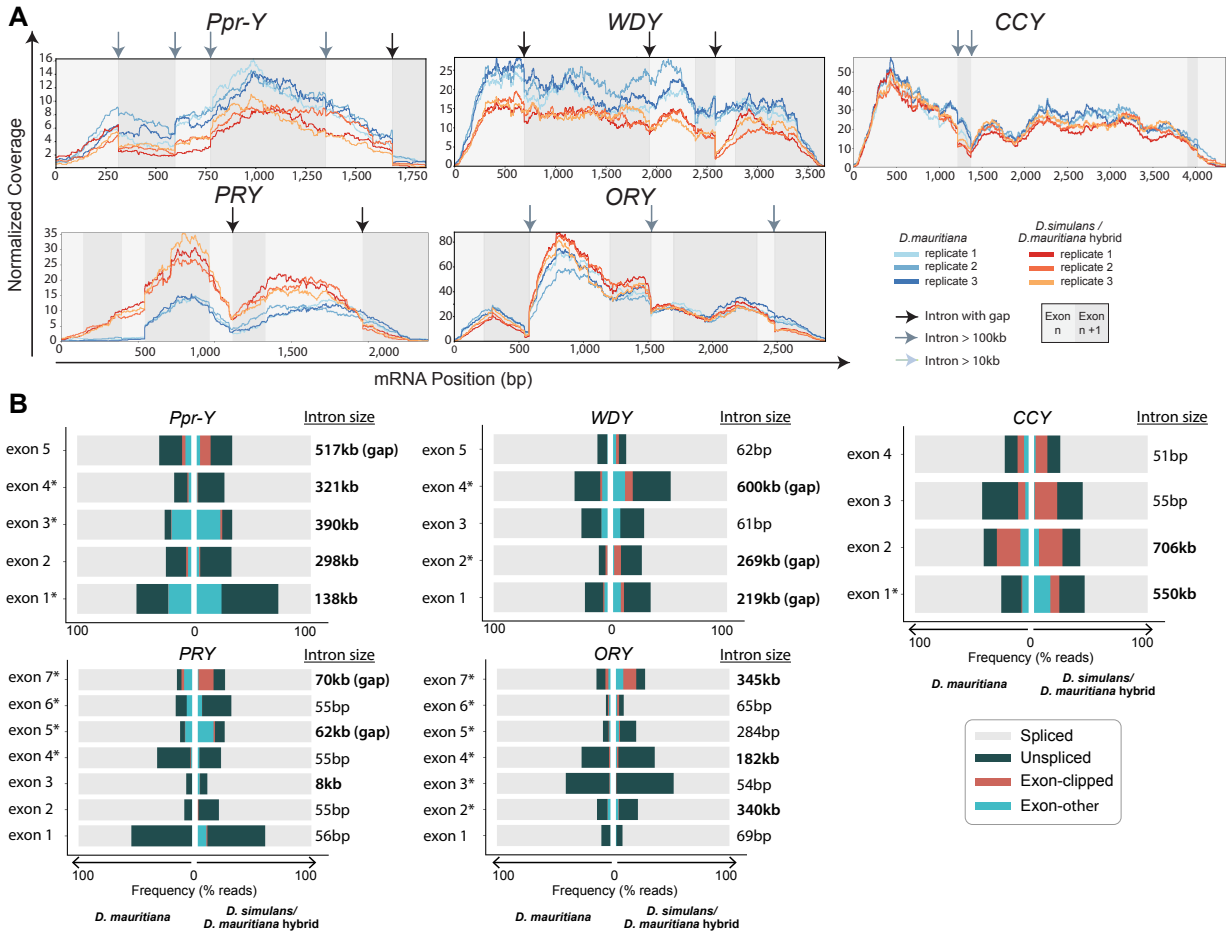


Figure 6. *D. simulans/D. mauritiana* hybrids exhibit splicing defects in Y-linked gigantic genes.

A. Normalized read depth coverage of the Y-linked fertility genes (*Ppr-Y, WDY, PRY, ORY, CCY*) in *D. mauritiana* vs. *D. simulans/D. mauritiana* hybrids.
B. Sequence read analysis for each exon end of the Y-linked fertility genes (*Ppr-Y, WDY, PRY, ORY, CCY*) in *D. mauritiana* vs. *D. simulans/D. mauritiana* hybrids. Asterisks(*) indicate significant difference in the proportion of correctly spliced reads between *D. mauritiana* and *D. simulans/D. mauritiana* hybrids.

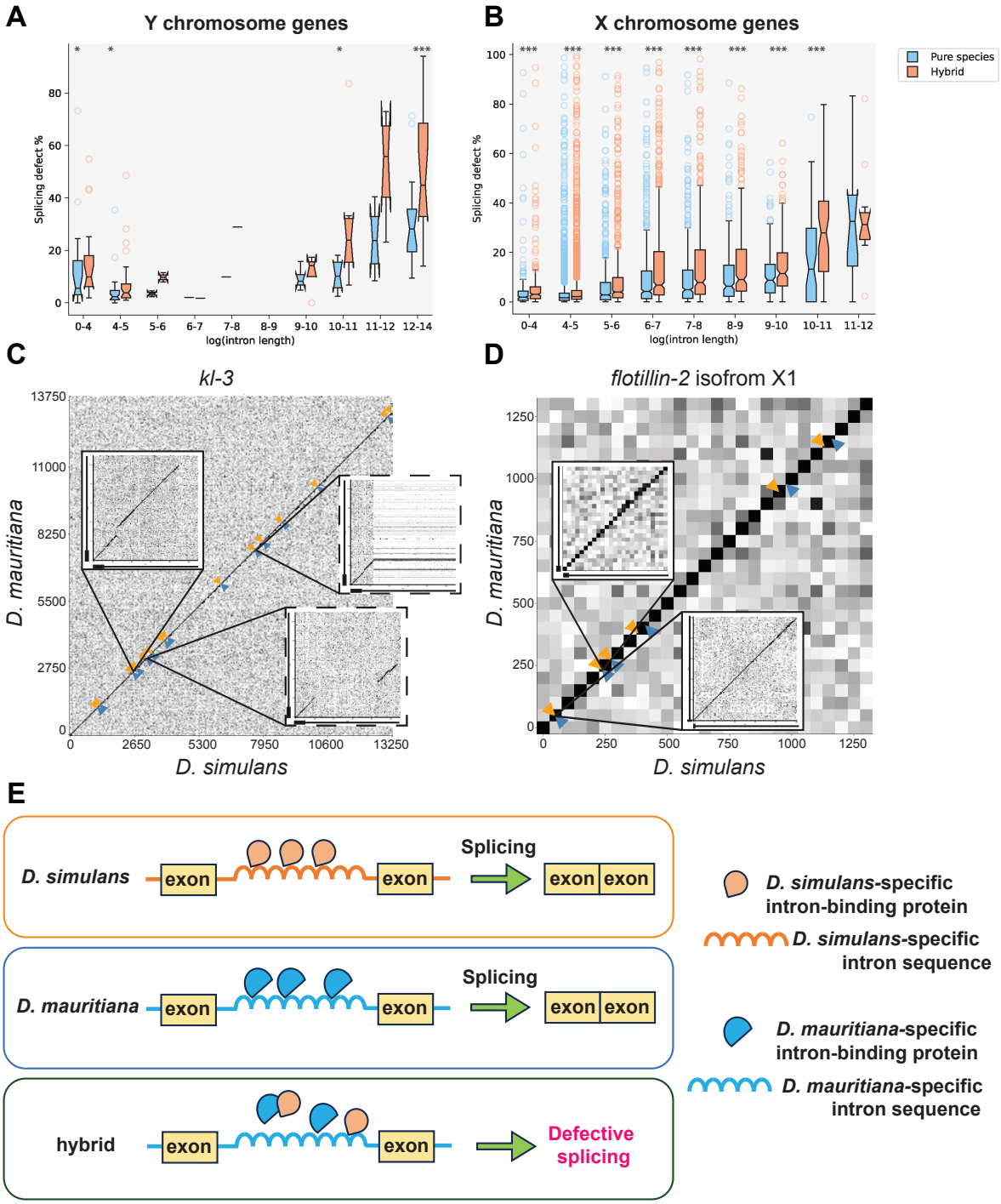


Figure 7. Intron size correlates with splicing defects of Y-linked fertility genes in *D. simulans/D. mauritiana* hybrids.

A, B. Frequency of splicing defects (100% - correctly spliced reads/all reads) for Y-linked genes (A) and X-linked genes (B) based on the size of introns. Introns were binned by their size (logarithmic scale with the base of e). Stars indicate statistical significance

between *D. simulans/D. mauritiana* hybrids and *D. mauritiana* (Supplementary_file_4 for statistic details.)

C, D. Dot plot of sequence conservation between *D. simulans* (X-axis) vs *D. mauritiana* (Y-axis) for kl-3 gene (Y-linked, C) and *flotillin-2* isoforms X1 gene (X-linked, D). Arrows indicate the position of introns. Zoomed-in squares cover the areas containing the end of the preceding exon (up to 500pb, represented by a bold line on the X-Y axes) and the first 5 kb of the intron (represented by a thin line on the X-Y axes). The solid boxes indicate fully assembled introns, and the dotted boxes indicate gap-containing introns. The first intron of *flotillin-2* is ~90kb.

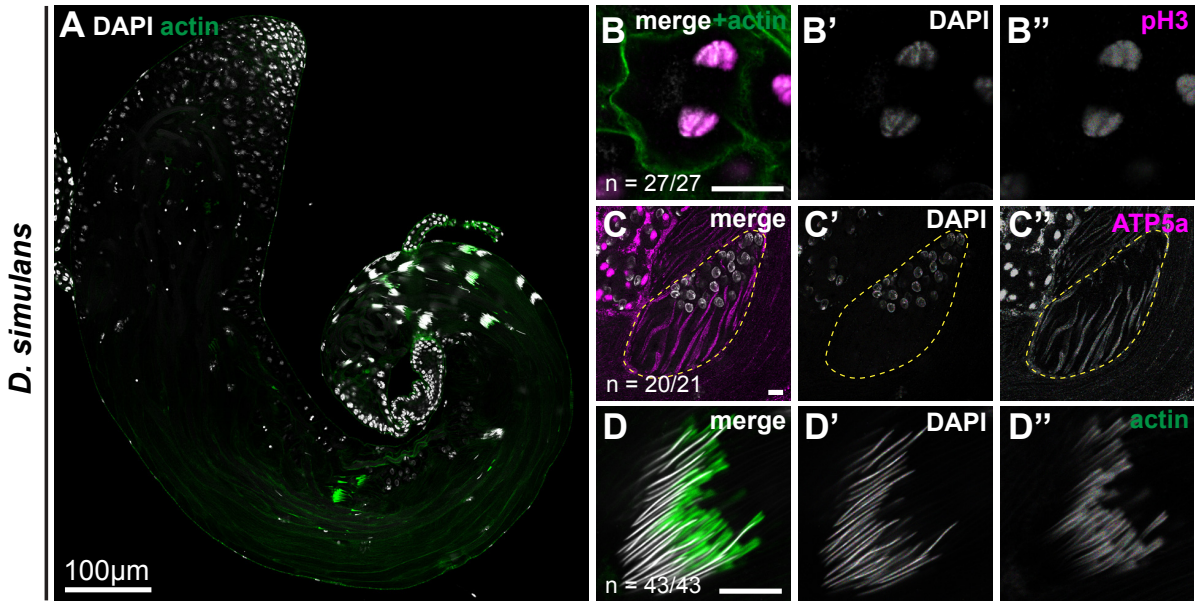
E. Model of how distinct intronic RNA sequences may cause hybrid sterility. RNA-binding proteins that bind to intronic RNA and aid in the process of splicing are divergent between two species due to distinct intronic sequences. When RNA-binding proteins and intronic RNA sequence are brought together in the context of hybrid, incompatibility between RNA-binding proteins and intronic sequence may cause splicing defects.

	baseMean	log2FoldChange	pvalue	padj	symbol
FBgn0267433	8429.782124	-1.34968047	3.15E-41	2.60E-39	**kl-5**
FBgn0267432	7865.223966	-1.21942604	2.86E-46	3.08E-44	**kl-3**
FBgn0001313	3951.734923	-0.80832938	2.39E-16	3.77E-15	**kl-2**
FBgn0046697	353.783317	-0.74483373	2.08E-05	8.12E-05	**Ppr-Y**
FBgn0267449	1449.592824	-0.71955795	8.67E-08	4.94E-07	**WDY**
FBgn0261399	34.476346	-0.44963307	2.67E-01	3.60E-01	**Pp1-Y1**
FBgn0267592	2728.81308	-0.31097907	6.01E-03	1.40E-02	**CCY**
FBgn0265047	5976.340682	-0.19448045	3.97E-02	7.21E-02	**FDY**
FBgn0046323	2412.807673	-0.04123393	6.60E-01	7.34E-01	**ORY**
FBgn0267490	59004.19059	-0.01515801	8.61E-01	8.97E-01	**Mst77Y-4**
FBgn0058064	1420.364084	0.29391915	7.25E-03	1.65E-02	**ARY**
FBgn0046698	444.653182	0.81141702	1.16E-08	7.54E-08	**Pp1-Y2**
FBgn0267489_alt	630.185477	0.8587783	2.03E-11	1.91E-10	**PRY**

Table 1: Differential gene expression analysis of Y-linked genes in *D. simulans/D. mauritiana* hybrids

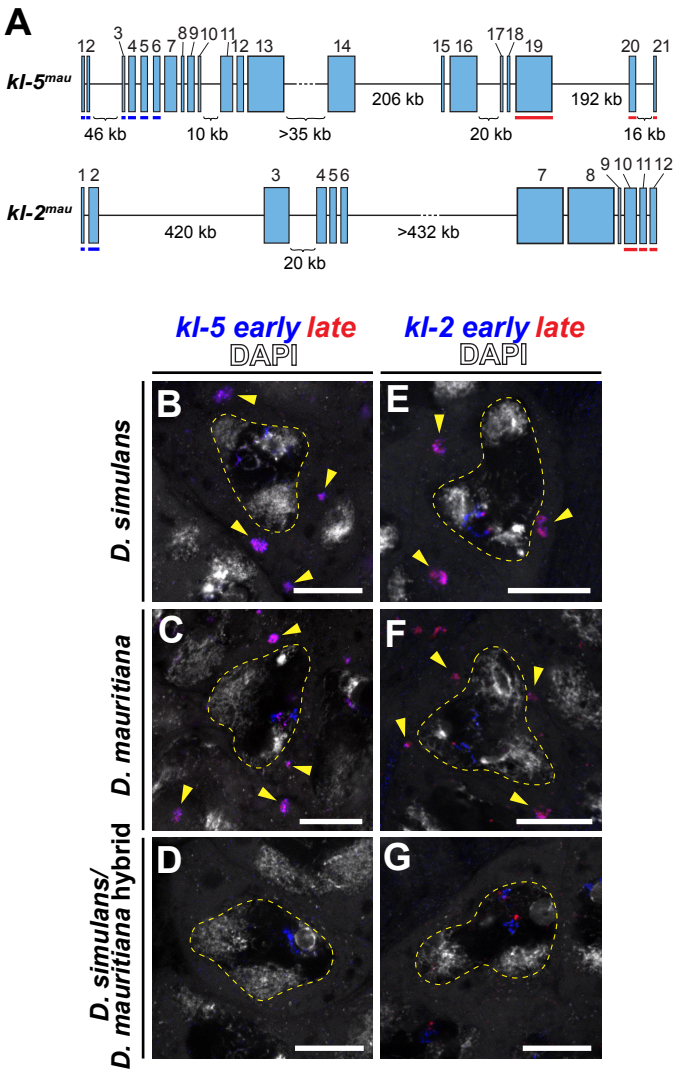
D. simulans/D. mauritiana hybrids vs. *D. mauritiana* differential Y genes expression results. Blue indicates statistically overexpressed genes in the *D. simulans/D. mauritiana* hybrids. Red indicates genes statistically under expressed in the *D. simulans/D. mauritiana* hybrids. Darker red highlight genes statistically under expressed in the *D. simulans/D. mauritiana* hybrids with log2 fold change < -1.

Supplementary Figures and Table



Supplementary Figure S1: *D. simulans* spermatogenesis

A-D. *D. simulans* testis and germ cells. Low magnification view of the whole testis (A) stained for F-actin (green, stained with phalloidin-Alexa 488) and DAPI (gray). Meiotic anaphase (B) stained for phospho-H3 ser10 (magenta) and DAPI (gray). Early elongating spermatids (C) stained for ATP5a (magenta, mitochondria) and DAPI (gray). Needle stage spermatids (D) stained for F-actin (green, stained with phalloidin-Alexa 488, individualization complex). Scale bar: 10µm, unless otherwise noted.

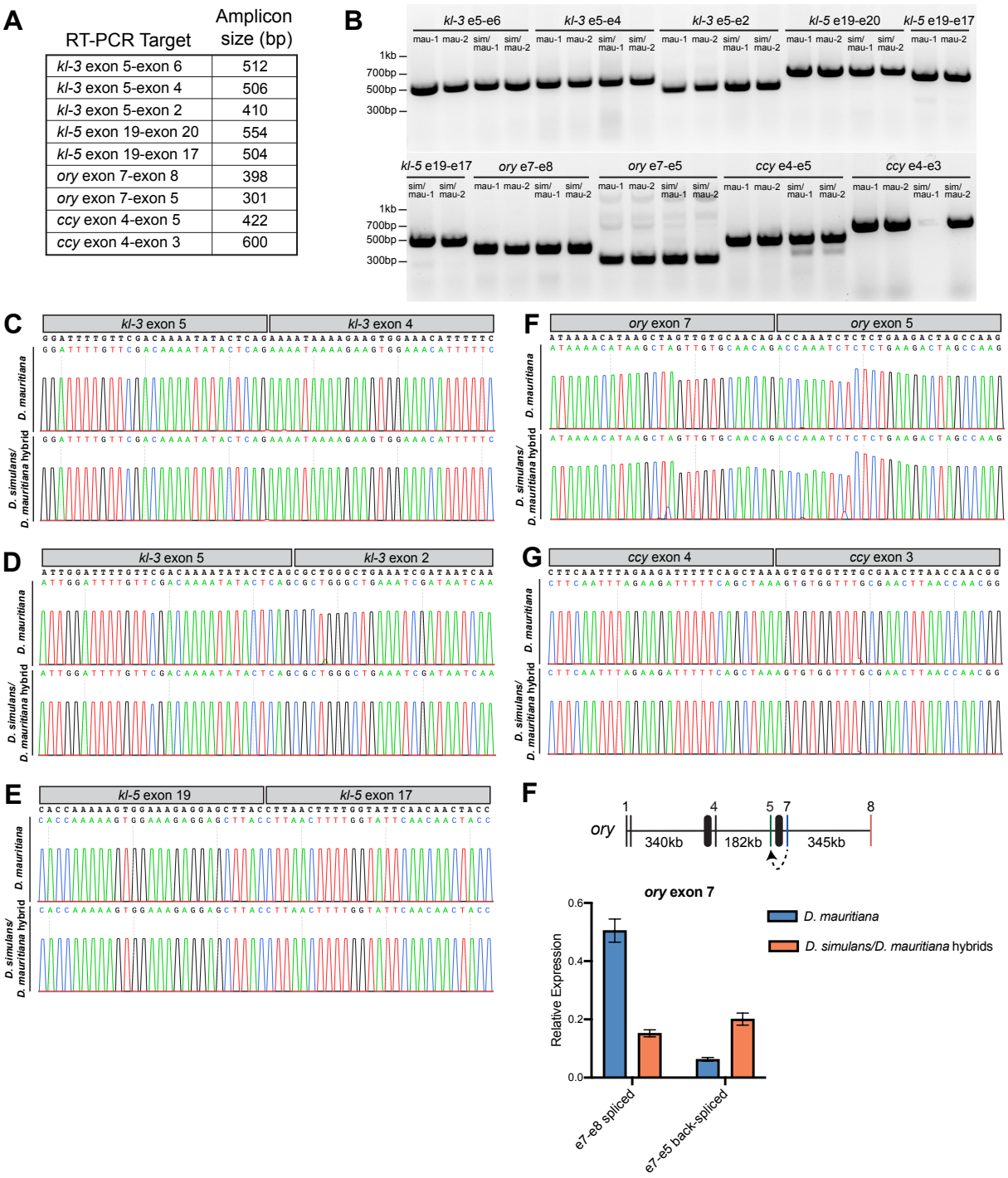


Supplementary Figure S2: Expression of *kl-2* and *kl-5* in the spermatocytes of the hybrids and parental species.

A. *D. mauritiana kl-5* (*kl-5*^{mau}) and *kl-2* (*kl-2*^{mau}) gene structure. The positions of smRNA FISH probes are indicated.

B-F. RNA in situ hybridization for *kl-5* transcripts in late-stage spermatocyte nuclei of *D. simulans* (B), *D. mauritiana* (C) and *D. simulans/D. mauritiana* hybrids (D). Spermatocyte nuclei are indicated by dashed lines. Arrowheads indicate cytoplasmic kl-granule containing mature *kl-5* mRNA. Blue: *kl-5* early exons. Red: *kl-5* late exons. Gray: DAPI. Scale bar: 10μm.

E-G. RNA in situ hybridization for *kl-2* transcripts in late-stage spermatocyte nuclei of *D. simulans* (E), *D. mauritiana* (F) and *D. simulans/D. mauritiana* hybrids (G). Spermatocyte nuclei are indicated by dashed lines. Arrowheads indicate cytoplasmic kl-granule containing mature *kl-2* mRNA. Blue: *kl-2* early exons. Red: *kl-2* late exons. Gray: DAPI. Scale bar: 10μm.

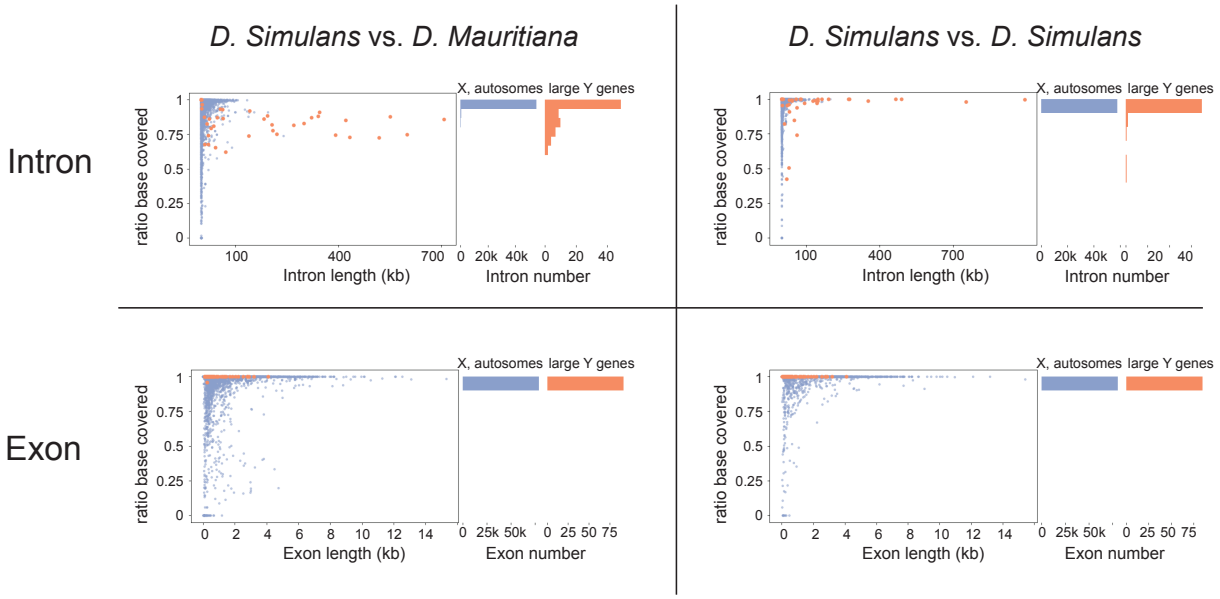


Supplementary Figure S3: Validation of back-splicing products.

A-B. RT-PCR for back-spliced transcripts. (A) List of expected PCR amplicon sizes. Primer sequences can be found in Supplementary_file_3. (B) RT-PCR products for two biological replicates of each *D. mauritiana* and *D. simulans/D. mauritiana* hybrids. *CCY* e4-e3 failed to amplify in one of the hybrid replicates.

C-G. Oxford Nanopore sequencing of the PCR products generated in (A-B) aligned to the corresponding exon-exon junction. Only the sequence immediately around the junction is shown.

H. RT-qPCR results of spliced vs. back-spliced products involving *ORY* exon 7. Schematic of the exon-intron structure surrounding *ory* exon 7 is shown in the upper panel. Exons are represented by vertical lines and labeled by exon number. Introns are represented by horizontal lines; a dotted line indicates a gap in the genome assembly. RT-qPCR results were normalized to GAPDH. Error bars represent SEM across three technical replicates. (See Supplementary_file_3 for raw CT values and biological replicates.)



Supplementary Figure S4: estimation of DNA sequence conservation between *D. simulans* and *D. mauritiana* for exons and introns

Ratio of covered bases for genomic features (intron: top row, exon: bottom row) by short read DNA sequencing from *D. simulans* aligned against either *D. mauritiana* genome (left) or *D. simulans* genome (right). Features from X and autosomes are shown in blue, and large Y genes in orange. Histograms (right to each scatter plot) indicate the distribution of introns based on the conservation (Y axis).

References

1. Peichel CL, Bolnick DI, Brannstrom A, Dieckmann U, Safran RJ. Speciation. Cold Spring Harb Perspect Biol. 2024. Epub 20240610. doi: 10.1101/cshperspect.a041735. PubMed PMID: 38858075.
2. Presgraves DC, Meiklejohn CD. Hybrid Sterility, Genetic Conflict and Complex Speciation: Lessons From the Drosophila simulans Clade Species. Front Genet. 2021;12:669045. Epub 20210623. doi: 10.3389/fgene.2021.669045. PubMed PMID: 34249091; PubMed Central PMCID: PMCPMC8261240.
3. Haldane JBS. Sex ratio and unisexual sterility in hybrid animals. Journal of genetics. 1922;12(2):101-9. doi: 10.1007/bf02983075.
4. Delph LF, Demuth JP. Haldane's Rule: Genetic Bases and Their Empirical Support. J Hered. 2016;107(5):383-91. Epub 20160527. doi: 10.1093/jhered/esw026. PubMed PMID: 27233288.
5. Orr HA. Haldane’s Rule. Annual Review of Ecology, Evolution, and Systematics. 1997;28(Volume 28, 1997):195-218. doi: https://doi.org/10.1146/annurev.ecolsys.28.1.195.
6. Turelli M, Orr HA. The dominance theory of Haldane's rule. Genetics. 1995;140(1):389-402. doi: 10.1093/genetics/140.1.389. PubMed PMID: 7635302; PubMed Central PMCID: PMCPMC1206564.
7. Cutter AD. Beyond Haldane's rule: Sex-biased hybrid dysfunction for all modes of sex determination. Elife. 2024;13. Epub 20240819. doi: 10.7554/eLife.96652. PubMed PMID: 39158559; PubMed Central PMCID: PMCPMC11333046.
8. Presgraves DC. Sex chromosomes and speciation in Drosophila. Trends Genet. 2008;24(7):336-43. doi: 10.1016/j.tig.2008.04.007. PubMed PMID: 18514967; PubMed Central PMCID: PMCPMC2819171.
9. Searle JB, Pardo-Manuel de Villena F. Meiotic Drive and Speciation. Annu Rev Genet. 2024;58(1):341-63. doi: 10.1146/annurev-genet-111523-102603. PubMed PMID: 39585909.
10. Coughlan JM, Matute DR. The importance of intrinsic postzygotic barriers throughout the speciation process. Philos Trans R Soc Lond B Biol Sci. 2020;375(1806):20190533. Epub 20200713. doi: 10.1098/rstb.2019.0533. PubMed PMID: 32654642; PubMed Central PMCID: PMCPMC7423280.
11. Lachaise D, David JR, Lemeunier F, Tsacas L, Ashburner M. The reproductive relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D. melanogaster from the afrotropical region. Evolution; international journal of organic evolution. 1986;40(2):262-71.
12. Johnson NA, Hollocher H, Noonburg E, Wu CI. The effects of interspecific Y chromosome replacements on hybrid sterility within the Drosophila simulans clade. Genetics. 1993;135(2):443-53. Epub 1993/10/01. PubMed PMID: 8244006; PubMed Central PMCID: PMCPMC1205647.
13. Cabot EL, Davis AW, Johnson NA, Wu CI. Genetics of reproductive isolation in the Drosophila simulans clade: complex epistasis underlying hybrid male sterility. Genetics. 1994;137(1):175-89. doi: 10.1093/genetics/137.1.175. PubMed PMID: 8056308; PubMed Central PMCID: PMCPMC1205934.
14. Palopoli MF, Wu CI. Genetics of hybrid male sterility between drosophila sibling species: a complex web of epistasis is revealed in interspecific studies. Genetics.

1994;138(2):329-41. doi: 10.1093/genetics/138.2.329. PubMed PMID: 7828817; PubMed Central PMCID: PMCPMC1206152.
15. Hollocher H, Wu CI. The genetics of reproductive isolation in the Drosophila simulans clade: X vs. autosomal effects and male vs. female effects. Genetics. 1996;143(3):1243-55. Epub 1996/07/01. PubMed PMID: 8807297; PubMed Central PMCID: PMCPMC1207394.
16. True JR, Weir BS, Laurie CC. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of Drosophila mauritiana chromosomes into Drosophila simulans. Genetics. 1996;142(3):819-37. doi: 10.1093/genetics/142.3.819. PubMed PMID: 8849890; PubMed Central PMCID: PMCPMC1207021.
17. Coyne JA, Charlesworth B. Location of an X-linked factor causing sterility in male hybrids of Drosophila simulans and D. mauritiana. Heredity (Edinb). 1986;57 (Pt 2):243-6. doi: 10.1038/hdy.1986.114. PubMed PMID: 3781872.
18. Perez DE, Wu CI, Johnson NA, Wu ML. Genetics of reproductive isolation in the Drosophila simulans clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, Odysseus (Ods). Genetics. 1993;134(1):261-75. doi: 10.1093/genetics/134.1.261. PubMed PMID: 8514135; PubMed Central PMCID: PMCPMC1205429.
19. Perez DE, Wu CI. Further characterization of the Odysseus locus of hybrid sterility in Drosophila: one gene is not enough. Genetics. 1995;140(1):201-6. doi: 10.1093/genetics/140.1.201. PubMed PMID: 7635285; PubMed Central PMCID: PMCPMC1206547.
20. Kulathinal R, Singh RS. Cytological Characterization of Premeiotic Versus Postmeiotic Defects Producing Hybrid Male Sterility among Sibling Species of the Drosophila Melanogaster Complex. Evolution; international journal of organic evolution. 1998;52(4):1067-79. doi: 10.1111/j.1558-5646.1998.tb01834.x. PubMed PMID: 28565214.
21. Kanippayoor RL, Alpern JHM, Moehring AJ. A common suite of cellular abnormalities and spermatogenetic errors in sterile hybrid males in Drosophila. Proceedings Biological sciences / The Royal Society. 2020;287(1919):20192291. Epub 2020/01/23. doi: 10.1098/rspb.2019.2291. PubMed PMID: 31964309; PubMed Central PMCID: PMCPMC7015338.
22. Catron DJ, Noor MA. Gene expression disruptions of organism versus organ in Drosophila species hybrids. PLoS ONE. 2008;3(8):e3009. Epub 20080820. doi: 10.1371/journal.pone.0003009. PubMed PMID: 18714377; PubMed Central PMCID: PMCPMC2500191.
23. Michalak P, Noor MA. Genome-wide patterns of expression in Drosophila pure species and hybrid males. Molecular biology and evolution. 2003;20(7):1070-6. Epub 20030530. doi: 10.1093/molbev/msg119. PubMed PMID: 12777520.
24. Michalak P, Noor MA. Association of misexpression with sterility in hybrids of Drosophila simulansand D. mauritiana. J Mol Evol. 2004;59(2):277-82. doi: 10.1007/s00239-004-2622-y. PubMed PMID: 15486701.
25. Moehring AJ, Teeter KC, Noor MA. Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Molecular biology and evolution.

2007;24(1):137-45. Epub 20061010. doi: 10.1093/molbev/msl142. PubMed PMID: 17032727.
26. Sundararajan V, Civetta A. Male sex interspecies divergence and down regulation of expression of spermatogenesis genes in Drosophila sterile hybrids. J Mol Evol. 2011;72(1):80-9. Epub 20101116. doi: 10.1007/s00239-010-9404-5. PubMed PMID: 21079940.
27. Bridges CB. NON-DISJUNCTION AS PROOF OF THE CHROMOSOME THEORY OF HEREDITY. Genetics. 1916;1(1):1.
28. Hughes JF, Page DC. The Biology and Evolution of Mammalian Y Chromosomes. Annu Rev Genet. 2015;49:507-27. Epub 20151006. doi: 10.1146/annurev-genet-112414-055311. PubMed PMID: 26442847.
29. Kotov AA, Bazylev SS, Adashev VE, Shatskikh AS, Olenina LV. Drosophila as a Model System for Studying of the Evolution and Functional Specialization of the Y Chromosome. Int J Mol Sci. 2022;23(8). Epub 20220410. doi: 10.3390/ijms23084184. PubMed PMID: 35457001; PubMed Central PMCID: PMCPMC9031259.
30. Zeng LW, Singh RS. The genetic basis of Haldane's rule and the nature of asymmetric hybrid male sterility among Drosophila simulans, Drosophila mauritiana and Drosophila sechellia. Genetics. 1993;134(1):251-60. Epub 1993/05/01. PubMed PMID: 8514134; PubMed Central PMCID: PMCPMC1205428.
31. Araripe LO, Tao Y, Lemos B. Interspecific Y chromosome variation is sufficient to rescue hybrid male sterility and is influenced by the grandparental origin of the chromosomes. Heredity (Edinb). 2016;116(6):516-22. Epub 20160316. doi: 10.1038/hdy.2016.11. PubMed PMID: 26980343; PubMed Central PMCID: PMCPMC4868264.
32. Chang CH, Gregory LE, Gordon KE, Meiklejohn CD, Larracuente AM. Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. Elife. 2022;11. Epub 2022/01/07. doi: 10.7554/eLife.75795. PubMed PMID: 34989337.
33. Jagannathan M, Warsinger-Pepe N, Watase GJ, Yamashita YM. Comparative Analysis of Satellite DNA in the Drosophila melanogaster Species Complex. G3 (Bethesda). 2017;7(2):693-704. Epub 2016/12/23. doi: 10.1534/g3.116.035352. PubMed PMID: 28007840; PubMed Central PMCID: PMCPMC5295612.
34. Yunis JJ, Yasmineh WG. Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science. 1971;174(4015):1200-9. PubMed PMID: 4943851.
35. Carvalho AB, Lazzaro BP, Clark AG. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(24):13239-44. Epub 2000/11/09. doi: 10.1073/pnas.230438397. PubMed PMID: 11069293; PubMed Central PMCID: PMC27209.
36. Kurek R, Trapitz P, Bunemann H. Strukturdifferenzierungen in Y-chromosom von Drosophila hydei: the unique morphology of the Y chromosomal lampbrush loops Threads results from 'coaxial shells' formed by different satellite-specific subregions within megabase-sized transcripts. Chromosome Res. 1996;4(2):87-102. PubMed PMID: 8785614.

37. Kurek R, Reugels AM, Lammermann U, Bunemann H. Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of Drosophila sp. Genetica. 2000;109(1-2):113-23. PubMed PMID: 11293786.
38. Pozzoli U, Sironi M, Cagliani R, Comi GP, Bardoni A, Bresolin N. Comparative analysis of the human dystrophin and utrophin gene structures. Genetics. 2002;160(2):793-8. doi: 10.1093/genetics/160.2.793. PubMed PMID: 11861579; PubMed Central PMCID: PMCPMC1461978.
39. Pozzoli U, Elgar G, Cagliani R, Riva L, Comi GP, Bresolin N, et al. Comparative analysis of vertebrate dystrophin loci indicate intron gigantism as a common feature. Genome Res. 2003;13(5):764-72. doi: 10.1101/gr.776503. PubMed PMID: 12727896; PubMed Central PMCID: PMCPMC430921.
40. Fingerhut JM, Moran JV, Yamashita YM. Satellite DNA-containing gigantic introns in a unique gene expression program during Drosophila spermatogenesis. PLoS genetics. 2019;15(5):e1008028. doi: 10.1371/journal.pgen.1008028.
41. Fingerhut JM, Lannes R, Whitfield TW, Thiru P, Yamashita YM. Co-transcriptional splicing facilitates transcription of gigantic genes. PLoS genetics. 2024;20(6):e1011241. Epub 20240613. doi: 10.1371/journal.pgen.1011241. PubMed PMID: 38870220; PubMed Central PMCID: PMCPMC11207136.
42. Fuller MT. Spermatogenesis. In The Development of Drosophila melanogaster Martinez-Arias MBaA, editor: New York: Cold Spring Harbor Laboratory Press.; 1993.
43. Yamashita YM. Subcellular Specialization and Organelle Behavior in Germ Cells. Genetics. 2018;208(1):19-51. Epub 2018/01/06. doi: 10.1534/genetics.117.300184. PubMed PMID: 29301947; PubMed Central PMCID: PMCPMC5753857.
44. Fabian L, Brill JA. Drosophila spermiogenesis: Big things come from little packages. Spermatogenesis. 2012;2(3):197-212. Epub 2012/10/23. doi: 10.4161/spmg.21798. PubMed PMID: 23087837; PubMed Central PMCID: PMC3469442.
45. Tokuyasu KT. Dynamics of spermiogenesis in Drosophila melanogaster. IV. Nuclear transformation. J Ultrastruct Res. 1974;48(2):284-303. Epub 1974/08/01. doi: 10.1016/s0022-5320(74)80083-3. PubMed PMID: 4210599.
46. Tokuyasu KT, Peacock WJ, Hardy RW. Dynamics of spermiogenesis in Drosophila melanogaster. I. Individualization process. Z Zellforsch Mikrosk Anat. 1972;124(4):479-506. Epub 1972/01/01. doi: 10.1007/BF00335253. PubMed PMID: 4622067.
47. Gatti M, Pimpinelli S. Functional elements in Drosophila melanogaster heterochromatin. Annual review of genetics. 1992;26:239-75. Epub 1992/01/01. doi: 10.1146/annurev.ge.26.120192.001323. PubMed PMID: 1482113.
48. Chandley AC, Bateman AJ. Timing of spermatogenesis in Drosophila melanogaster using tritiated thymidine. Nature. 1962;193:299-300. doi: 10.1038/193299b0. PubMed PMID: 13878048.
49. Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. The biogenesis, biology and characterization of circular RNAs. Nat Rev Genet. 2019;20(11):675-91. Epub 20190808. doi: 10.1038/s41576-019-0158-7. PubMed PMID: 31395983.
50. Conn VM, Chinnaiyan AM, Conn SJ. Circular RNA in cancer. Nat Rev Cancer. 2024;24(9):597-613. Epub 20240729. doi: 10.1038/s41568-024-00721-7. PubMed PMID: 39075222.

51. Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature. 2012;483(7387):82-6. Epub 20120222. doi: 10.1038/nature10843. PubMed PMID: 22367542; PubMed Central PMCID: PMCPMC3292678.
52. Soh YQ, Alfoldi 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(4):800-13. Epub 20141030. doi: 10.1016/j.cell.2014.09.052. PubMed PMID: 25417157; PubMed Central PMCID: PMCPMC4260969.
53. McDermott SR, Noor MA. The role of meiotic drive in hybrid male sterility. Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1265-72. doi: 10.1098/rstb.2009.0264. PubMed PMID: 20308102; PubMed Central PMCID: PMC2871811.
54. Tao Y, Hartl DL, Laurie CC. Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proc Natl Acad Sci U S A. 2001;98(23):13183-8. doi: 10.1073/pnas.231478798. PubMed PMID: 11687638; PubMed Central PMCID: PMC60845.
55. Phadnis N, Orr HA. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science. 2009;323(5912):376-9. doi: 10.1126/science.1163934. PubMed PMID: 19074311; PubMed Central PMCID: PMC2628965.
56. Bladen J, Nam HJ, Phadnis N. Transformation of meiotic drive into hybrid sterility in Drosophila. Genetics. 2024;228(2). doi: 10.1093/genetics/iyae133. PubMed PMID: 39196789.
57. 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. PLoS Biol. 2007;5(11):e293. doi: 10.1371/journal.pbio.0050293. PubMed PMID: 17988173; PubMed Central PMCID: PMC2062476.
58. Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL. A sex-ratio meiotic drive system in Drosophila simulans. I: an autosomal suppressor. PLoS Biol. 2007;5(11):e292. doi: 10.1371/journal.pbio.0050292. PubMed PMID: 17988172; PubMed Central PMCID: PMC2062475.
59. Cazemajor M, Landre C, Montchamp-Moreau C. The sex-ratio trait in Drosophila simulans: genetic analysis of distortion and suppression. Genetics. 1997;147(2):635-42. doi: 10.1093/genetics/147.2.635. PubMed PMID: 9335600; PubMed Central PMCID: PMCPMC1208185.
60. Kelleher ES, Edelman NB, Barbash DA. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 2012;10(11):e1001428. doi: 10.1371/journal.pbio.1001428. PubMed PMID: 23189033; PubMed Central PMCID: PMCPMC3506263.
61. Castillo DM, Moyle LC. Hybrid incompatibility between Drosophila virilis and D. lummei is stronger in the presence of transposable elements. Journal of evolutionary biology. 2022;35(10):1319-34. Epub 20220821. doi: 10.1111/jeb.14079. PubMed PMID: 35988129; PubMed Central PMCID: PMCPMC9805144.
62. Dion-Cote AM, Renaut S, Normandeau E, Bernatchez L. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake

whitefish species. Molecular biology and evolution. 2014;31(5):1188-99. Epub 20140206. doi: 10.1093/molbev/msu069. PubMed PMID: 24505119.
63. Kotov AA, Adashev VE, Kombarov IA, Bazylev SS, Shatskikh AS, Olenina LV. Molecular Insights into Female Hybrid Sterility in Interspecific Crosses between Drosophila melanogaster and Drosophila simulans. Int J Mol Sci. 2024;25(11). Epub 20240523. doi: 10.3390/ijms25115681. PubMed PMID: 38891872; PubMed Central PMCID: PMCPMC11172174.
64. Parhad SS, Tu S, Weng Z, Theurkauf WE. Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev Cell. 2017;43(1):60-70 e5. doi: 10.1016/j.devcel.2017.08.012. PubMed PMID: 28919205.
65. Banho CA, Oliveira DS, Haudry A, Fablet M, Vieira C, Carareto CMA. Transposable Element Expression and Regulation Profile in Gonads of Interspecific Hybrids of Drosophila arizonae and Drosophila mojavensis wrigleyi. Cells. 2021;10(12). Epub 20211218. doi: 10.3390/cells10123574. PubMed PMID: 34944084; PubMed Central PMCID: PMCPMC8700503.
66. Zhu L, Fukunaga R. RNA-binding protein Maca is crucial for gigantic male fertility factor gene expression, spermatogenesis, and male fertility, in Drosophila. PLoS genetics. 2021;17(6):e1009655. Epub 20210628. doi: 10.1371/journal.pgen.1009655. PubMed PMID: 34181646; PubMed Central PMCID: PMCPMC8248703.
67. Fingerhut JM, Yamashita YM. mRNA localization mediates maturation of cytoplasmic cilia in Drosophila spermatogenesis. J Cell Biol. 2020;219(9). Epub 2020/07/25. doi: 10.1083/jcb.202003084. PubMed PMID: 32706373; PubMed Central PMCID: PMCPMC7480094.
68. Fingerhut JM, Yamashita YM. Analysis of Gene Expression Patterns and RNA Localization by Fluorescence in Situ Hybridization in Whole Mount Drosophila Testes. Methods Mol Biol. 2023;2666:15-28. doi: 10.1007/978-1-0716-3191-1_2. PubMed PMID: 37166654.
69. Lerat E, Fablet M, Modolo L, Lopez-Maestre H, Vieira C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Res. 2017;45(4):e17. doi: 10.1093/nar/gkw953. PubMed PMID: 28204592; PubMed Central PMCID: PMCPMC5389681.
70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.
71. Lannes R, Fingerhut JM, Yamashita YMM. OmniSplice: a framework-free splicing event reporter. biorxiv. 2025. doi: 10.1101/2025.04.06.647416.
72. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. Epub 20100128. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278; PubMed Central PMCID: PMCPMC2832824.