

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

The Epigenetics and Evolution of Y Chromosomes in Drosophila

Permalink

<https://escholarship.org/uc/item/94r143z7>

Author

Nguyen, Alison Hanh

Publication Date

2021

Peer reviewed|Thesis/dissertation

The Epigenetics and Evolution of
Y Chromosomes in *Drosophila*

by

Alison Hanh Nguyen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Doris Bachtrog, Chair
Professor Noah K. Whiteman
Professor Donald C. Rio

Summer 2021

Abstract

The Epigenetics and Evolution of Y Chromosomes in *Drosophila*

by

Alison Hanh Nguyen

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Doris Bachtrog, Chair

Y chromosomes are fascinating parts of the genome. In XY determination species, they specify the male sex and to do so, they influence a wide variety of molecular and cellular processes. Furthermore, since the Y is restricted to the male lineage, they undergo unique evolutionary trajectories resulting in gene-poor and repeat-rich chromosomes. This begs the question of how Y chromosomes can elicit drastic biological differences such as differences between sexes and differences between individuals. In my dissertation, I leveraged the Y chromosomes of two *Drosophila* species to address this fundamental question.

Males often die earlier than females in XY species and recent studies have shown that the Y chromosome may play a role in this sex-specific phenomenon. Intriguingly, repetitive DNA such as transposable elements (TEs) show increased expression over aging, suggesting that the Y's repetitive nature can contribute to this observation. I study sex-specific aging by leveraging the well-assembled genome of *D. miranda*, whose males have a very large young Y chromosome (neo-Y). The neo-Y was recently formed through a fusion between an autosome and a Y chromosome and therefore contains both genes and TEs. Using chromatin immunoprecipitation followed by sequencing (ChIP-Seq) and RNA-Seq, I showed that *D. miranda* males age faster than females possibly because of gene activity on the Y that also increases overall TE expression. Therefore, ineffective repression of deleterious elements both on and off the Y can explain the differences in aging.

While many Y's are degenerate, they can still vary in size within a population. The *D. pseudoobscura* Y especially has long been known to harbor dramatic size variation in the wild. I show that the 24Mb size difference between a large Y chromosome variant and a small variant is due to a handful of recently active TEs using cytological and sequencing approaches. These same TEs are also implicated in the size variation of a related species, *D. affinis*. Consistent with previous work, these Y chromosomes produce subtle changes in male lifespan and gene expression. Interestingly, I find that males bearing the large Y variant have higher expression of

large Y-abundant TEs compared to males bearing the small Y. Thus, the very large Y chromosome variant of *D. pseudoobscura* was recently formed through highly active TEs and that these particular TEs have the propensity to also amplify in other species.

Gene content on the Y was previously believed to be sparse but advances in sequencing have countered these old notions. In particular, Y chromosomes have been implicated in intragenomic conflicts with the X chromosome via gene amplification. I inspect the genomes and gene annotations of three *D. pseudoobscura* Y variants to investigate gene content evolution within a population. I find that the *D. pseudoobscura* Y is not a gene desert. In fact, it contains many gene families with hundreds of gene copies and not all variants contain the same families or copy numbers. Expression of these Y chromosome genes reveal that select genes are testis-expressed with sperm-related functions while others are involved in chromatin regulation. Therefore, Y chromosome gene content is highly dynamic and Y chromosome genes may play a role in wide-reaching biological processes.

DEDICATION

To my family,
near and far, literal and figurative.

Acknowledgements

I thank my advisor Doris Bachtrog for advising and funding my graduate program. I also thank Professors Noah Whiteman and Donald Rio for serving on my dissertation committee. Together, their insights made challenges surmountable and I deeply appreciate their guidance. I also thank Professors Rasmus Nielsen, Nipam Patel, Donald Rio, and Noah Whiteman for being on my qualifying exam committee. I learned a lot from all of them over these past six years.

My lab mates, current and former, never hesitated to give me honest advice and reassurance. I thank Dat Mai and Kevin Wei for their instantaneous, candid feedback and Carolus Chan for helping me troubleshoot many protocols. Reema Aldemailani and Kamalakar Chatla reliably supported me in the lab, especially in the last chapter, and Silu Wang always had many ideas and suggestions. Emily Brown and Lauren Gibilisco gave me a ton of “veteran” graduate student advice. My undergraduates provided a lot of assistance: Emily Chong, Daisy Fang, Jasmine Tatak, Amy Weixiang Wang, and Lydia Yu. In particular, Emily and Amy helped kick start the Y chromosome project. Without them, it would have taken much longer.

Many people taught me various skills. Hector Nolla and Carl Hjelman taught me flow cytometry, Lydia Smith trained me on qPCR equipment, and Holly Aaron taught me confocal microscopy. Alison Killilea also made fly food for all the experiments, Karen Lundy and Shana McDevitt facilitated the sequencing, and Oanh Nguyen and Jessie Li ran the Bionano project.

My graduate social life was filled with wonderful conversations from members of the Whiteman lab, CTEG, and IB - especially Monica Albe, Debora Brandt, Ana Lyons, Dat Mai, and Nina Pak. Doing research is difficult, but it feels less so when surrounded by kindness and generosity.

Previous mentors were instrumental in my academic career. Professors Alan Bergland and Dmitri Petrov advised my undergraduate research, introducing me to the fun of fly genetics and the fundamentals of molecular evolution. I also had many conversations about personal and academic growth with Professors Susan McConnell and Michael Simon. My former high school teachers Kristen Kirby and Brian Morrison encouraged me to apply to my first research opportunity. Howard Rafal and Megan Rokop mentored me in that internship and the experience inspired me to pursue a degree in biology.

Staying in the Bay Area was a tough call, but I am glad I did thanks to my Stanford family. Han Altae-Tran, Emmerich Anklam, Arthur Brant, Christopher Kremer, Eric Lopez, Clara Luu, Arjun Mathur, Aran Nayebi, Julia Quintero, Laza Upatasing, Matt Vitelli, Joshua Wilbur, and Clara Wong-Fillinger recharged my batteries many times. I also sincerely thank Gene and Vicki Becker for welcoming me into their home figuratively and, during the pandemic, literally.

Finally, I thank my family. My mom and dad knew little about college and research but they have always trusted my choices and judgment. They also made sure I did not worry about them so I could have the luxury to focus on my intellectual pursuits. My brother Anton has provided me with empathy and encouragement, the latter often delivered in a comically nagging manner. Last, my partner Dominic Becker has brought me joy and perspective throughout this entire endeavor. His patience with me is unrivaled and I am forever grateful for his companionship.

Chapter 1: Toxic Y chromosome: increased repeat expression and age-associated heterochromatin loss in male *Drosophila* with a young Y chromosome

Originally published with slight modifications as Nguyen, Alison H., and Doris Bachtrog. “Toxic Y Chromosome: Increased Repeat Expression and Age-Associated Heterochromatin Loss in Male *Drosophila* with a Young Y Chromosome.” *PLOS Genetics* 17, no. 4 (April 22, 2021): e1009438. <https://doi.org/10.1371/journal.pgen.1009438>.

Abstract

Sex-specific differences in lifespan are prevalent across the tree of life and influenced by heteromorphic sex chromosomes. In species with XY sex chromosomes, females often outlive males. Males and females can differ in their overall repeat content due to the repetitive Y chromosome, and repeats on the Y might lower survival of the heterogametic sex (toxic Y effect). Here, we take advantage of the well-assembled young Y chromosome of *Drosophila miranda* to study the sex-specific dynamics of chromatin structure and repeat expression during aging in male and female flies. Male *D. miranda* have about twice as much repetitive DNA compared to females, and live shorter than females. Heterochromatin is crucial for silencing of repetitive elements, yet old *D. miranda* flies lose H3K9me3 modifications in their pericentromere, with heterochromatin loss being more severe during aging in males than females. Satellite DNA becomes de-repressed more rapidly in old vs. young male flies relative to females. In contrast to what is observed in *D. melanogaster*, we find that transposable elements (TEs) are expressed at higher levels in male *D. miranda* throughout their life. We show that epigenetic silencing via heterochromatin formation is ineffective on the TE-rich neo-Y chromosome, presumably due to active transcription of a large number of neo-Y linked genes, resulting in up-regulation of Y-linked TEs already in young males. This is consistent with an interaction between the evolutionary age of the Y chromosome and the genomic effects of aging. Our data support growing evidence that “toxic Y chromosomes” can diminish male fitness and a reduction in heterochromatin can contribute to sex-specific aging.

Summary

Y chromosomes can be toxic. The Y chromosome of many species contains a large number of transposable elements (TEs), which are transcriptionally constrained by repressive chromatin marks. When relieved of these epigenetic modifications, many TEs can readily move from one genomic location to another. We show that TEs located on the Y chromosome are less effectively silenced in male *Drosophila*, and the toxic Y effect appears more pronounced in a species that contains a larger Y chromosome with more repeats and more actively transcribed genes. Our data demonstrate that repeat-rich Y chromosomes are a genomic liability for males.

Introduction

Males and females differ in many life history traits, and sexual dimorphism in fitness-related traits is often driven by natural selection (Tarka et al. 2018). However, some traits could vary between the sexes in a non-adaptive manner. For example, males and females often differ in their lifespan (Maklakov and Lummaa 2013), and the chromosomal sex determination system has been shown to influence sex-specific longevity. In particular, several studies have suggested that the sex with the heteromorphic sex chromosomes (males in XY species; females in ZW species)

has a shorter lifespan on average (Xirocostas, Everingham, and Moles 2020; Pipoly et al. 2015). The aging process is associated with an overall loss of heterochromatin in many species (Larson et al. 2012; Haithcock et al. 2005), and differences in the heterochromatin content between sexes could in principle contribute to sex-specific mortality (the toxic Y effect; (Marais et al. 2018)). In species with heteromorphic sex chromosomes, the amount of repetitive DNA and thus heterochromatin can vary dramatically between males and females, due to the presence of a large, repetitive Y (or W) chromosome in the heterogametic sex. Heterochromatin loss can result in de-repression and mobilization of silenced transposable elements (TEs) (Wood et al. 2016; 2010; De Cecco et al. 2013; W. Li et al. 2013), and might disproportionally affect the sex with the higher heterochromatin content.

We previously showed that heterochromatin loss differs between the sexes in *D. melanogaster* (Brown, Nguyen, and Bachtrog 2020a). TEs are up-regulated in old male flies, and Y-linked repeats are especially prone to become de-repressed during aging in *D. melanogaster*. Genetic manipulations of sex chromosome karyotypes suggested that the Y chromosome directly contributes to sex-specific aging in flies: females containing a Y chromosome (XXY females) lived shorter than wildtype XX females, males lacking a Y (X0 males) outlived wildtype males, and males containing two Y chromosomes (XYY males) showed a drastically reduced lifespan compared to XY males (Brown, Nguyen, and Bachtrog 2020a). However, repeat-rich Y chromosomes are typically poorly assembled (only 14.6 Mb of the roughly 40-Mb large Y chromosome of *D. melanogaster* is present in the latest assembly; (Chang and Larracunte 2019)), making it impossible to directly observe chromatin changes on the heterochromatic Y chromosome and link them to expression changes of repeats using genomic approaches.

Here, we test for an association between heteromorphic sex chromosomes and sex-specific heterochromatin formation and silencing of repetitive DNA in a species with a well-assembled Y chromosome (Mahajan et al. 2018). We assay chromatin and gene expression profiles in young and aged males and females of *D. miranda* (Fig 1A), a model system for studying the molecular basis of sex chromosome differentiation (Mahajan et al. 2018; Steinemann and Steinemann 1992; Bachtrog 2006; Bachtrog et al. 2008; Zhou and Bachtrog 2012; C. E. Ellison and Bachtrog 2013). *D. miranda* has a recently evolved neo-sex chromosome that was formed only 1.5MY ago from an ordinary autosome that fused to the ancestral Y (Bachtrog and Charlesworth 2002). The neo-X and neo-Y are in the process of differentiating into heteromorphic sex chromosomes, and the neo-Y has dramatically increased in size mainly through an accumulation of TEs (Mahajan et al. 2018; Steinemann and Steinemann 1992; Bachtrog et al. 2008; Bachtrog 2003). Specifically, a high-quality genome assembly that includes large fractions of heterochromatin contained >90 Mb of neo-Y-linked sequence (compared to ~25 Mb of neo-X linked sequence), and 65 Mb are derived from repetitive elements (compared to 5 Mb on the neo-X) (Mahajan et al. 2018). TEs are uniformly enriched along the neo-Y chromosome and a main contributor to its dramatically increased genome size [(Mahajan et al. 2018; Steinemann and Steinemann 1992; Bachtrog et al. 2008; Bachtrog 2003). For example, the most abundant repeat on the neo-Y is the *ISY* element, a helitron transposon that is inserted about 22,000 times on the neo-Y/Y chromosome and amounts to more than 16 Mb of DNA (i.e., 16% of neo-Y sequence), while its former homolog, the neo-X chromosome only harbors about 1,500 copies (3% of the neo-X; less than 1 Mb; (Mahajan et al. 2018)). The second most common repeat class that has amplified on the neo-Y are *gypsy* TEs; roughly 14,300 insertions were found on the neo-Y (15% of neo-Y sequence, 15 Mb) and less

than 1 Mb on the neo-X (about 800 insertions, i.e., 3% of its sequence; (Mahajan et al. 2018)). Thus, a huge number of TEs has accumulated on the neo-Y chromosome since its origination 1.5 MY ago. Yet this chromosome still contains thousands of protein-coding genes that are embedded within these islands of TEs (Mahajan et al. 2018), and many of the neo-Y linked protein-coding genes are still actively transcribed (Zhou and Bachtrog 2012; Bachtrog, Mahajan, and Bracewell 2019). Thus, unlike the Y chromosome of *D. melanogaster*, the neo-Y of *D. miranda* contains thousands of transcribed, euchromatic genes, which are intermingled with repetitive heterochromatin.

We use chromatin and transcription profiling to show that heterochromatin formation is compromised on the repeat-rich neo-Y of *D. miranda* already in young males, resulting in global up-regulation of TEs in males. Age-associated heterochromatin loss is more pronounced in male flies compared to females, which also live shorter, and accompanied by de-repression of satellite repeats in old males. Our data provide empirical support that toxic Y chromosomes can diminish male fitness, and contribute to sex-specific aging in species with heteromorphic sex chromosomes.

Results

***Drosophila miranda* males have more repeats and live shorter.**

We chose the standard lab strain MSH22 of *D. miranda*, a fly that has served as a model for Y chromosome evolution (**Fig 1A**) (Mahajan et al. 2018; Zhou and Bachtrog 2012; Bachtrog, Mahajan, and Bracewell 2019) to investigate how the ‘toxic’ Y chromosome may influence sex-specific behaviors of repetitive sequences, such as TE expression, the dynamics of heterochromatin during the lifespan of a fly, and sex-specific aging. In *D. miranda*, an autosome fused to the ancestral sex chromosome about 1.5MY ago (Bachtrog and Charlesworth 2002), creating a neo-Y chromosome which is at an intermediate stage of degeneration. The neo-Y still contains many of its ancestral genes, but has expanded in size by about 3-fold, mostly due to the accumulation of TEs (Mahajan et al. 2018; Bachtrog et al. 2008). This increase in repeat content is accompanied by a change in its chromatin structure with many regions along the neo-Y becoming heterochromatic (**Fig 1B**) (Zhou et al. 2013). *D. miranda* has a high-quality genome assembly that comprise large stretches of repetitive heterochromatin, including major parts of the pericentromeres and the Y chromosome (Mahajan et al. 2018) (**Fig 1C**). In particular, *D. miranda* males have a enormous, ~92 Mb large recently formed and highly repetitive and heterochromatic neo-Y chromosome (Mahajan et al. 2018), while the pericentromeric heterochromatin on the X/neo-X only amounts to ~24 Mb (**Table 1**) (Mahajan et al. 2018; Bracewell et al. 2019). Thus, males contain significantly more repetitive DNA than females, and flow cytometry estimates agree with our assembled genome sizes (**Table 1**).

In most *Drosophila* species male flies live significantly shorter than female flies (Yoon, Gagen, and Zhu 1990; Tower and Arbeitman 2009). We determined longevity for males and females for the reference lab strain from *D. miranda* (MSH22) at 18°C (from 3 biological replicates, see Materials & Methods). Median survivorship for males is 78 days, and >98 days for females (**Fig 1D**). Thus, lifespan assays confirm that males live significantly shorter than females, consistent with multiple studies on sex-specific lifespan in *Drosophila* (Yoon, Gagen, and Zhu 1990; Tower and Arbeitman 2009). Longevity patterns are thus consistent with the notion that the repeat-rich

Y chromosome reduces survivorship in the heterogametic sex (Xirocostas, Everingham, and Moles 2020; Pipoly et al. 2015), but note that other factors besides sex (such as male-male aggression) may contribute to differences in longevity between males and females (Brenman-Suttner et al. 2020).

Heterochromatin loss is more pronounced in males.

We collected replicate ChIP-seq data (4 replica for each sex and age) for a histone modification typical of repressive heterochromatin (H3K9me3) from young (9-11 days) and old (80-90-day) *D. miranda* male and female brains (MSH22) (**S1 Table, S2 Table**). In order to increase mappability of our ChIP-seq data (both to repetitive regions, and to distinguish between reads derived from the neo-X and neo-Y), we collected 100-bp paired-end read data. We spiked-in a fixed amount of chromatin from *D. melanogaster* to each *D. miranda* chromatin sample prior to ChIP and sequencing, to compare the genomic distribution of chromatin marks across samples using a ‘spike in’ normalization method (Brown, Nguyen, and Bachtrog 2020a; X.-Y. Li et al. 2014). This species pair is sufficiently diverged from each other that there is very little ambiguity in the assignment of reads to the correct species, especially for fast-evolving repetitive DNA (**Supplementary Material 1**), and this strategy allows us to compare levels of H3K9me3 enrichment across sexes and time.

Fig 2 shows the genomic distribution of the repressive histone modification H3K9me3 for young and old male and female flies. As expected, heterochromatin is enriched at repetitive regions, including pericentromeres, the small dot chromosome and the repeat-rich Y chromosome (**Fig 2A**). Overall heterochromatin enrichment is similar for repetitive regions at the X and autosomes in young males and females (**Fig 2B**). Consistent with heterochromatin being lost during aging, both males and females show a reduction in H3K9me3 enrichment at their pericentromeres in old compared to young flies (**Figs 2 and 3**). Strikingly, heterochromatin loss is more pronounced in old male flies than in old females (**Figs 2A and 2B**). This finding is reproducible across individual replicates (**S1-S6 Figs and S3 Table**), and robust with regards to normalization strategy (**S1 Fig and S4 Table**) and using only uniquely mapping reads (**S3 Fig and S5 Table**). *D. miranda* males show significantly more regions that lose H3K9me3 signal (1.5-fold or more in 5-kb windows) during aging compared to females (2936 vs. 316, $p < 0.00001$; Fisher’s exact test; **S7 Fig**). Almost all regions (78% of windows) that lose heterochromatin in males are located within the pericentromeres of the autosome/X (1963 windows) and the neo-Y (328 windows) (**S7-S9 Figs**). In females, 25.6% of windows (81) losing heterochromatin occur within the pericentromeres. Many fewer regions gain H3K9me3 signal (1.5-fold or more) during aging (947 in males and 155 in females, $p < 0.00001$, Fisher’s exact test; **S7 Fig**). Almost all regions that gain H3K9me3 signal are found along the chromosome arms (**S7 Fig**), consistent with the idea that some heterochromatin becomes re-distributed in old flies (Sun, Yu, and Dang 2018). Faster heterochromatin loss in old males is in agreement with recent findings in *D. melanogaster* (Brown, Nguyen, and Bachtrog 2020a). Note that genes involved in heterochromatin formation and organization show little change in their expression during aging in both males and females, suggesting that heterochromatin loss in males is not due to a male-specific decline in proteins required for heterochromatin formation (**S6 Table**). These data suggest that increased deterioration of heterochromatin during aging may be a general feature of the sex with more repetitive DNA.

Heterochromatin content and loss on the neo-Y chromosome.

The neo-Y is highly repetitive (**Fig 1 and Table 1**) and enriched for heterochromatin (**Fig 2**), yet to a lesser extent than pericentromeric regions (**Figs 2A and 2B**). Note that while the neo-X and neo-Y still show considerable homology at protein-coding genes, there is little mis-mapping between reads derived from the neo-X or neo-Y chromosome (**Supplementary Material 2**), indicating that lower H3K9me3 enrichment on the neo-Y is not an artifact due to mis-mapping. The neo-Y harbors several thousand protein-coding genes embedded in highly repeat-rich regions (Mahajan et al. 2018; Bracewell et al. 2019), and overall repeat content on the neo-Y is slightly less than of pericentromeres on the X or autosomes (72% of bp are repeat-masked on the neo-Y chromosome, compared to 82-89% of the pericentromeres on X and autosomes).

To evaluate if lower repeat content on the neo-Y can account for lower levels of heterochromatin, we compared H3K9me3 enrichment versus repeat content for 5-kb windows across chromosomes (**Fig 3A**). As expected, genomic regions with a higher repeat density show more heterochromatin, and absolute levels of H3K9me3 enrichment are very similar between the X and autosomes, and young males and females. However, for repeat-rich windows (>70% repeat masked), the neo-Y chromosome shows considerably lower levels of H3K9me3 enrichment than other chromosomes, even in young males (**Fig 3A**). In particular, average H3K9me3 enrichment in highly repetitive windows (>70% of bases repeat masked) is about 30% lower on the neo-Y compared to X-linked or autosomal regions in young males, and 24% lower in old males. The same conclusion holds if the top repeat-rich regions (>99.5% repeat masked) are considered (**Fig 3A**). Thus, this suggests that repetitive DNA on the neo-Y shows less heterochromatin-induced epigenetic silencing than other repeat-rich regions of the genome. Immunofluorescence staining of H3K9me3 in mitotic chromosomes confirms much weaker labeling of the neo-Y chromosome relative to the large pericentromeric region of the X chromosome in male larvae (**Figs 3B and S10**). While we observe consistent and robust H3K9me3 staining along the repetitive pericentromere on the X chromosome and neo-X/autosomes, only the pericentromeric region of the neo-Y shows intense H3K9me3 staining; the highly repetitive chromosome arms of the neo-Y show considerable weaker H3K9me3 signal (**Fig 3B**), consistent with our ChIP-seq results (**Fig 2**). Thus, Immunofluorescence staining independently confirms that the neo-Y chromosome contains lower levels of heterochromatin, compared to other repetitive regions in the genome.

Multiple factors could contribute to lower levels of H3K9me3 enrichment on the neo-Y. The neo-Y chromosome still contains many unique regions with transcribed protein-coding genes (Mahajan et al. 2018; Zhou and Bachtrog 2012; Bachtrog, Mahajan, and Bracewell 2019) and thus encompasses a patchwork of heterochromatic repeats interspersed with unique and genetically active euchromatic regions. In *Drosophila*, the boundary between heterochromatin and euchromatin is not fixed but determined by the local balance between factors that promote either heterochromatin or euchromatin (Elgin and Reuter 2013). Active transcription could limit spreading of heterochromatin on the neo-Y and reduce overall H3K9me3 enrichment (Allshire and Madhani 2018). Indeed, the gene density on the neo-Y is substantially higher than in the pericentromeres. The *D. miranda* annotation contains 255 genes located in the pericentromeres (that is, roughly 9 genes/Mb) and 5625 genes on the Y/neo-Y (61 genes/Mb) (Mahajan et al. 2018). Most genes are transcribed in adult brain samples (246 of all the pericentromeric genes and 4428 of the neo-Y genes; see below), consistent with the idea that active transcription of

neo-Y genes could impede heterochromatin formation on this chromosome. Indeed, we find that H3K9me3 enrichment is lower around transcription start sites (TSS) for genes located on the neo-Y, similar to genes located in pericentromeric regions (**Fig 4A**), supporting the idea that active transcription counteracts local heterochromatin formation. Importantly, if we group neo-Y linked genes into those that are highly expressed from the neo-Y (top 20%) versus those lowly expressed (bottom 40%), we find that H3K9me3 enrichment is substantially lower neighboring genes that are highly expressed (**Fig 4B**). Lower H3K9me3 levels near highly transcribed genes supports the notion that active transcription interferes with heterochromatin formation on the neo-Y.

In addition, heterochromatin on the neo-Y is evolutionarily young, and its overall repeat structure may differ from pericentromeres. The neo-Y was an ordinary autosome until 1.5 MY ago and only acquired its high repeat content and heterochromatin structure after it became fused to the ancestral Y (Steinemann and Steinemann 1992; Bachtrog 2003). The repeats present on the neo-Y should therefore encompass mostly recently active TEs, and it is possible that these repeats may initiate less heterochromatin (Molaro and Malik 2016). Indeed, the relative abundance of different TE families differs between the neo-Y chromosome and the rest of the genome (**S11 Fig**; (Mahajan et al. 2018)); we find the neo-Y contains more LTR transposons and helitrons but less simple repeats in comparison to other chromosomes (**S11 Fig**).

While overall TE composition differs between the neo-Y and other chromosomes, transposons belonging to the same family may also show lower levels of H3K9me3 enrichment if present on the neo-Y. As expected, H3K9me3 profiles at TE families show that there is a general enrichment of this repressive mark in both male and female flies (**Fig 5A**). To test if neo-Y linked TEs show less H3K9me3 enrichment, we calculated average H3K9me3 enrichment for each TE family based on genome-wide mapping of H3K9me3 separately for copies located on the neo-Y versus other genomic locations. Indeed, copies of TEs inserted onto the neo-Y typically show lower levels of H3K9me3 enrichment than copies on other chromosomes. Global H3K9me3 enrichment is about 30% lower for neo-Y TEs compared to their X- and autosomal homologs in young flies (and 27% lower in old flies), and 92/16 show significantly less/more H3K9me3 if located on the neo-Y relative to copies on the X/autosome in young males (and 25/6 in old ones; **Fig 5B**). Thus, this suggests that most TEs experience less epigenetic silencing on the neo-Y relative to other regions of the genome.

Like other repeat-rich regions, overall levels of heterochromatin on the neo-Y decrease during aging; however, the absolute amount of H3K9me3 loss appears less on the Y in old males, compared to pericentric regions, or the dot chromosome (**Figs 2 and S12**). Overall, our results show that male *D. miranda* lose heterochromatin marks more rapidly than female, consistent with our previous findings in *D. melanogaster* (Brown, Nguyen, and Bachtrog 2020a). However, we also find that the neo-Y shows considerably lower levels of heterochromatin than what we might expect given its repeat content. Indeed, levels of H3K9me3 enrichment on the neo-Y in young males resemble reduced levels of heterochromatin at X-linked and autosomal pericentromeres in old males.

Repeat de-repression in young male flies

Heterochromatin silences repetitive DNA (Allshire and Madhani 2018), and reduced levels of H3K9me3 along the large repeat-rich neo-Y chromosome could have profound consequences on TE silencing in male *D. miranda* even in young flies. To study TE expression changes between sexes and during aging, we gathered replicated stranded RNA-seq data from young and old *D. miranda* (3 biological replicates, 100-bp paired-end reads to increase mappability; **S7-S9 Tables**). Indeed, comparison of TE expression between young and old male and female *D. miranda* reveals that repeat expression is strikingly increased in male flies compared to females, irrespective of age (**Figs 6A and 6B and S13**). In particular, 135 TE families are significantly upregulated, while 1 is expressed at lower levels in male *D. miranda* compared to females already in 9-day old flies, and the fraction of transcripts derived from TEs is more than twice as high in males compared to females (the fraction of repetitive reads in all RNA-seq reads is on average 8.0% at young males and 3.7% at young females, **S8-S10 Tables**).

It is difficult to identify the genomic copy from which a TE transcript originated, especially for recently active families with highly similar copies. We used several approaches to test if increased TE expression in males is primarily driven by repeats located on the neo-Y. First, we mapped RNA seq reads to the *D. miranda* genome, and estimated expression of neo-Y linked copies versus copies on other chromosomes for each TE family (Jin et al. 2015). We find that 53% of all TE transcripts in young males are derived from the autosomes and X and 47% are from the neo-Y, suggesting that most of the excess TE transcription in males could be attributed to expression from the neo-Y chromosome (**Figs S14 and S15**).

We also quantify expression of TEs enriched on the neo-Y independent of our assembly (**Figs 6A and 6B**). TE families with more copies on the neo-Y chromosome will show higher coverage in male DNA sequencing libraries compared to female libraries (Brown, Nguyen, and Bachtrog 2020a). We ranked TE families based on their male/female genomic coverage and classified 81 TEs as Y-enriched (based on $\log_2(\text{female/male gDNA})$ coverage; (Wei, Gibilisco, and Bachtrog 2020)). Indeed, TE families that are enriched on the neo-Y show higher expression in both males and females than the remaining TE families that are not enriched on the neo-Y (**Fig 6D**). In addition, most of the neo-Y enriched TE families are significantly more expressed in males, with 85% of the neo-Y enriched TEs (69 families) being more highly expressed in young males compared to females. Thus, this confirms the expectation that it is more recently active TEs that invaded the neo-Y chromosome since it stopped recombining with the neo-X 1.5MY ago. Increased copy number of these TEs and lower levels of heterochromatin on the neo-Y might drive their higher expression in males. Note that increased TE expression in male holds even if we account for higher copy number of repeats in males by normalizing by genomic DNA read counts or by estimated copy numbers in the genome assembly (**S16 Fig**). Thus, higher repeat expression in males is not simply a product of higher TE copy number in males.

No change in TE expression in old flies

A loss of heterochromatin in old individuals can lead to a re-activation of previously silenced TEs (Wood et al. 2016; 2010; De Cecco et al. 2013; W. Li et al. 2013). In *D. melanogaster*, we found that TEs are efficiently repressed in young male and female brain tissue (Brown, Nguyen, and Bachtrog 2020a). While old females maintained heterochromatin and silencing of repeats, males aged for a similar amount of time showed a dramatic reduction in heterochromatin at pericentric regions and on their Y chromosome. Heterochromatin loss in males was accompanied

by de-repression of TEs, and repeats from the Y chromosome were found to be disproportionately up-regulated in old *D. melanogaster* males (Brown, Nguyen, and Bachtrog 2020a).

Female *D. miranda* efficiently suppress TEs throughout their life, similar to *D. melanogaster*. In females, 4 TEs showed increased and 36 decreased expression during aging (**Fig 6A**, using $\text{abs}(\log_2(\text{fold-change})) = 0.58$), and the total fraction of transcripts derived from repeats decreases during aging (the fraction of repetitive reads in all RNA-seq reads is on average 3.7% at 9 days, vs. 2.9% at 80 days, **S3 Table**). In contrast to repeat de-repression found in old *D. melanogaster* males, however, expression of TEs does not change during aging in male *D. miranda* (**Figs 6C and S13**). In particular, 6/8 TEs show a in/decrease in expression during aging in *D. miranda* males (**Figs 6C and S13**; same fold-change as females), and the total fraction of transcripts derived from repeats slightly decreases during aging (from 8.0% at 9 days to 7.6% at 80 days, **S3 Table**). Thus, while old *D. miranda* males show a drastic loss in heterochromatin at pericentromeres on the X and autosomes, this does not result in global expression changes at TEs. However, unlike in *D. melanogaster*, overall levels of TE expression are substantially increased in male *D. miranda* relative to females, already in young flies (**Figs 6B and S13**).

Satellite mis-expression in old male flies

Another type of repetitive DNA that is targeted by heterochromatin is satellite DNA repeats (Allshire and Madhani 2018). Satellites are often transcribed, and satellite transcripts may have important biological roles, including heterochromatin formation or centromere assembly (Kuhn 2015; Usakin et al. 2007; Topp, Zhong, and Dawe 2004; Mills et al. 2019). Unlike *D. melanogaster*, which has large arrays of satellite DNA composed of simple repeats (sometimes several megabases in size), the genomes of *D. miranda* and its close relatives contain many fewer, often shorter satellites (Bracewell et al. 2019; Wei et al. 2018). Of the 780 different satellite repeat units that were identified in the *D. miranda* genome assembly (Mahajan et al. 2018), only 9 are enriched on the neo-Y (based on male/female genomic coverage), compared to 81 neo-Y/Y enriched TEs (out of 303 TE families total; see above). Thus, unlike TE's, satellites are not preferentially found on the neo-Y. Satellite DNA is highly enriched for H3K9me3 in young flies (**Fig 7A**), in both males and females. Mimicking overall patterns of heterochromatin enrichment, we see a clear drop in H3K9me3 levels in old male flies at satellites, but not females (**Fig 7A**).

Many satellite repeats are transcribed, and at similar levels in both sexes; 21/16 satellites show higher/lower expression in young males compared to young females (**Figs 7B and S17**). Thus, unlike TEs that are highly enriched on the neo-Y and inefficiently silenced in males (**Fig 6**), no such differences in satellite expression are seen in young males and females. Old males show a dramatic loss in repressive heterochromatin at satellites, and a simultaneous up-regulation at satellite DNA expression. In males, 21 satellites show increased expression during aging compared to 1 satellite showing decreased expression, while expression at 2/1 satellites in/decreased in old females (**Figs 7C and S17**). Thus, a loss of silencing heterochromatin at satellites in old male *D. miranda* is accompanied by a de-repression of satellite transcripts.

Mis-expression of neo-Y genes and genes embedded in heterochromatin in old flies

Changes in heterochromatin during aging could also affect protein-coding genes; while most genes are located in repeat-poor euchromatin, some are embedded in heterochromatin. In

addition to protein-coding genes located on the neo-Y (5586 in our combined ChIP and RNA analysis), *D. miranda* also has a total of 376 genes in pericentromeres and the dot. Genes that are naturally located in heterochromatin, such as pericentric genes in *Drosophila* or genes on the dot chromosome, have evolved mechanisms to assure gene transcription in this normally suppressive chromatin environment, and often are dependent on it (Riddle, Shaffer, and Elgin 2009; Yasuhara, DeCrease, and Wakimoto 2005; Yasuhara and Wakimoto 2008). Indeed, genes located on the pericentromeres appear to be expressed at slightly lower levels in old flies, compared to young ones, in both sexes (6/18 significantly up/downregulated in aged males and 3/29 in aged females; **S18 Fig**). Neo-Y genes, on the other hand, only became embedded in their heterochromatic environment within the past 1.5MY. Indeed, as flies age and lose H3K9me3, neo-Y genes become slightly upregulated (249/141 are significantly up/downregulated; **S18 Fig**). Thus, these results are consistent with a dynamic view of co-evolution of heterochromatin and its host genome; genes embedded in old heterochromatin, such as the pericentromere, had millions of years of evolution to properly function in this typically repressive environment, and depend on it. A loss of heterochromatin during aging thus results in downregulation of pericentric genes. In contrast, genes that only became invaded by repeats and thus heterochromatin more recently, such as neo-Y genes in *D. miranda*, have not yet had enough time to adapt to this novel environment; losing heterochromatin, therefore, results in slight upregulation of neo-Y genes.

Discussion

Repetitive DNA is composed of TEs and satellite DNA, and their misregulation can induce genomic instability (Bourque et al. 2018). Transposable elements (TEs) are mobile DNA sequences that can change their position and comprise a substantial fraction of eukaryotic genomes (Werren 2011). Satellite DNA is often found near centromeres, and incorrect packaging or expression of some satellite DNA can have profound negative fitness consequences (Kuhn 2015; Usakin et al. 2007; Topp, Zhong, and Dawe 2004; Mills et al. 2019). Repetitive elements and their host genome are in a constant arms race, which need to co-evolve to evade each other's response (Werren 2011; Hurst and Werren 2001). Mobilization of TEs can cause insertional mutations (Bourque et al. 2018) and natural selection acts to remove deleterious insertions (B. Charlesworth and Langley 1986). Several host-encoded mechanisms have emerged during eukaryotic evolution to suppress TE activity (Bourque et al. 2018; Cosby, Chang, and Feschotte 2019). One defense mechanism widely employed across eukaryotes is transcriptional silencing of repetitive DNA through heterochromatin formation (Allshire and Madhani 2018). A host-encoded defense mechanism that prevents movement of a TE family, however, creates selective pressure on these elements to evolve escape mechanisms to avoid the host defense, forcing the host to evolve further mechanisms to silence TEs. Thus, repetitive elements and their host genome are in a constant arms race, where they need to co-evolve to evade each other's response (Werren 2011; Hurst and Werren 2001).

The effectiveness of host defense mechanisms to eliminate or silence repetitive elements can vary in space and time. Epigenetic defense mechanisms can vary both on evolutionary time scales, and throughout an individual's life. Significant chromatin structural changes occur during aging, with a loss of heterochromatin commonly observed in old individuals (Larson et al. 2012; Sun, Yu, and Dang 2018). Disruption of heterochromatin during an individual's life can result in mobilization of previously silenced TEs (Wood et al. 2010; O'Sullivan and Karlseder 2012). The effectiveness of epigenetic silencing mechanisms can also vary across evolutionary time scales.

In particular, a novel TE that is invading a naïve host genome can escape host surveillance mechanisms and propagate across the genome, until silencing mechanisms evolve to target that TE (Molaro and Malik 2016; Bousios et al. 2020).

Natural selection to remove TE insertions can also be compromised. In the absence of recombination, natural selection is less efficient, and can result in fixation of deleterious mutations, including the accumulation of TEs (B. Charlesworth and Charlesworth 2000; W. G. Hill and Robertson 1966). Low or non-recombining regions of eukaryotic genomes such as pericentromeres are often an agglomerate of TEs (Bracewell et al. 2019; B. Charlesworth and Langley 1986; Bachtrog 2013), and Y chromosomes of many species, including *Drosophila*, are comprised almost entirely of repetitive DNA (Bachtrog 2013; Pimpinelli et al. 1995). In fact, the neo-Y of *D. miranda* has accumulated tens of thousands of TE insertions since it stopped recombining 1.5MY ago (Mahajan et al. 2018; Bachtrog 2003), demonstrating just how quickly TEs can invade a non-recombining chromosome. Non-recombining Y chromosomes of many species can thus act as reservoirs for TEs (Pimpinelli et al. 1995; Gatti and Pimpinelli 1992), and dramatically increase the repeat content in males compared to females (Brown, Nguyen, and Bachtrog 2020a).

The accumulation of deleterious mutations and repetitive elements on the Y chromosome might lower the survival of the heterogametic sex (“toxic Y” hypothesis) (Xirocostas, Everingham, and Moles 2020; Pipoly et al. 2015). While TEs may be efficiently silenced in young individuals, deteriorating heterochromatin can affect males disproportionately, and result in shorter lifespans in males (Brown, Nguyen, and Bachtrog 2020a). Under the toxic Y model, we expect males to show higher levels of TE expression than females, at least during some stages in their life (for example, in old males when heterochromatin is known to be compromised). Secondly, we would also expect that larger Y chromosomes that contain more repetitive elements are more ‘toxic’ to males than Y chromosomes with fewer or less active repeats.

Our data support both predictions of the toxic Y effect. We show that repeats from the Y are mis-expressed in male *D. miranda*. We find that the overall number of TE-derived transcripts is about twice as much in males than it is in females, and that TEs located on the Y are primarily responsible for increased TE transcript abundance in males. Secondly, comparison of our data to a recent analysis in *D. melanogaster* suggests that the larger neo-Y of *D. miranda* is a much bigger burden for males than the ancient Y of *D. melanogaster* (Brown, Nguyen, and Bachtrog 2020a). Young *D. melanogaster* males are effectively suppressing their Y-linked repeats and epigenetic defense mechanisms are only breaking down in old males (Brown, Nguyen, and Bachtrog 2020a). In contrast, TE suppression is compromised even in young *D. miranda* males. The Y chromosome of *D. melanogaster* is substantially smaller than the neo-Y of *D. miranda* and fully heterochromatic (Pimpinelli et al. 1995; Gatti and Pimpinelli 1992). About ¼ of the 40 Mb large Y chromosome of *D. melanogaster* is composed of TEs (that is, about 10 Mb), while the rest of the Y mainly consists of satellite DNA (Chang and Larracuente 2019). In contrast, almost 75 Mb of Y/neo-Y derived sequence in *D. miranda* is classified as TE-derived.

In addition to having substantially more repeats, the neo-Y shows lower levels of heterochromatin enrichment compared to repetitive regions on the X or autosomes. The neo-Y of *D. miranda* contains 1000s of expressed genes, and we find that active transcription of these

genes appears to impede the formation of silencing heterochromatin at adjacent TEs on the neo-Y. Thus, not only do male *D. miranda* have more repeats than females and their Y-chromosome has more TEs than other species (such as *D. melanogaster*), the TEs present on the neo-Y are particularly poorly silenced, due to the large number of active genes present. Therefore, while Y chromosomes can generally be toxic, the recently formed neo-Y of *D. miranda* is especially harmful (Wei, Gibilisco, and Bachtrog 2020). This increased toxicity of the *D. miranda* neo-Y can explain why Y-linked TEs are ineffectively silenced throughout the life of a fly, and not only in old males. Note, however, that our comparative study does not allow us to directly demonstrate that the neo-Y of *D. miranda* reduces male lifespan. Careful genetic manipulations of the sex chromosomes, as done in *D. melanogaster* (Brown, Nguyen, and Bachtrog 2020a), are necessary to link the absence or presence of the neo-Y to changes in lifespan.

Overall, we detect higher levels of expression of TEs in male flies relative to female flies, both for young and old flies. Our analysis suggests that low levels of heterochromatin on the neo-Y may be inefficient to fully suppress repetitive elements on this chromosome even in young males. Genome-wide deterioration of heterochromatin in old males may further decrease TE silencing, but may not substantially contribute to overall TE transcript abundance in old males that already have a large, inefficiently silenced highly repeat-rich neo-Y. Indeed, this resembles our findings in *D. melanogaster* flies with additional Y chromosomes (Brown, Nguyen, and Bachtrog 2020a). While wildtype *D. melanogaster* males efficiently suppress their Y-linked repeats when young, flies with additional Y chromosomes (XXY females or XYY males) show inefficient silencing of Y-linked repeats already in young flies (Brown, Nguyen, and Bachtrog 2020a). These patterns are consistent with a dynamic view of co-evolution between selfish TEs and host defense mechanisms. Flies typically can suppress wildtype levels of TE activity when young, but these defense mechanisms may fail as flies age (especially in males which have more repetitive DNA) or when they are being challenged with unusually high numbers of active TEs (extra Y chromosomes in XXY and XYY *D. melanogaster*, or the recently formed, highly repetitive neo-Y of *D. miranda*).

The molecular mechanisms underlying increased TE expression during aging, and the contribution of TE activity to aging are only beginning to be understood. A direct link between neurodegeneration and heterochromatin was found in *Drosophila*: The microtubule-associated protein tau is involved in a number of neurodegenerative disorders, including Alzheimer's disease. Wide-spread heterochromatin loss was found in tau transgenic *Drosophila* and mice, and oxidative stress and subsequent DNA damage were identified as mechanistic links between transgenic tau expression and heterochromatin deterioration (Frost et al. 2014). Several TEs become highly active in *Drosophila* brain during normal aging, and exacerbated TE expression in brain results in age-dependent memory impairment and shortened lifespan (W. Li et al. 2013). Movements of TEs result in DNA damage, and may provide the mechanistic link between TE activity and physiological aging. In mouse, L1 retrotransposons become active during the course of aging; SIRT6 - a powerful repressor of L1 activity - binds to the 5'UTR of L1 elements, and facilitates heterochromatin formation (Van Meter et al. 2014). During aging, and also in response to DNA damage, SIRT6 becomes depleted from L1, allowing the activation of these previously silenced TEs (Van Meter et al. 2014). De-repression of L1 elements results in activation of the type-I interferon response and age-associated inflammation, a phenotype of late senescence (De

Cecco et al. 2019). A direct link between systemic inflammation and heterochromatin was also observed in *Drosophila*: genes involved in immune response are enriched in lamin-associated heterochromatin in both flies and mammals, and fat body immunosenescence is caused by age-associated lamin-B reduction, contributing to loss of heterochromatin and de-repression of genes involved in immune responses (Chen, Zheng, and Zheng 2014). Future research will further establish the mechanistic connections between repetitive DNA, TE activity, heterochromatin loss, and sex-specific aging.

Materials & Methods

Drosophila strains. We used the reference genome strain MSH22 for analysis (Mahajan et al. 2018). For chromatin and gene expression analyses, flies were grown in incubators at 18°C, 60% relative humidity, and 12h light for the indicated number of days following eclosion, and were then flash-frozen in dry ice and stored at -80°C. Flies were reared on standard molasses fly food: 0.68% agar (w/v), 2.7% yeast (w/v), 6.67% cornmeal (w/v), 0.456% propionic acid (v/v), 1.6% sucrose (w/v), 0.76% of 95% ethanol (v/v), 0.09% Tegosept (w/v), 8.2% molasses (v/v), 0.0625% CaCl₂ (w/v), 0.75% Na Tartrate (w/v).

Lifespan assays and tissue preparation. Lifespan assays were conducted following (Linford et al. 2013) using the same rearing conditions as described above. Briefly, we collected synchronized embryos on a molasses plate with yeast paste for 16-20 hours. We washed embryos with 1x PBS pH 7.4 three times and dispensed 10µl of embryos per culture vial by pipette. To obtain synchronized adult flies, we collected emerging adults over a 3 day window and aged the adults for at least 7 days to mature and copulate. Afterwards, we separated males and females into separate vials, placing 20 individuals per vial, maintaining them at 18°C with a light:dark setting of 12h:12h. We moved flies to new vials every 3-5 days, without using CO₂, and fly deaths were recorded. Flies that were observed escaping the vial were censored. To collect samples for the RNA-seq and ChIP-seq experiment, we censored the entire lifespan experiment once it reached less than 50% survivorship for one sex and flash-froze the remaining flies in liquid nitrogen. In total, 1765 female flies in 107 vials and 1304 male flies in 75 vials were counted for the lifespan assays reported here. We dissected brains from males and females aged 5-9 days (young flies) and 80-98 days old in 1X PBS pH 7.4 (1 brain/sample). We then snap-froze and stored the tissue at -80°C prior to use in the ChIP or RNA-seq assay. For the spike-in, we dissected brains from *D. melanogaster yw* strain females and pooled 2 brains/sample.

Chromatin Immunoprecipitation and Sequencing. We modified the native and ultra-low input ChIP protocol (Brind'Amour et al. 2015) for single-brain ChIP-Seq. First, we added 40µL EZ nuclei lysis buffer to frozen tissues and homogenized them with a pestle grinder on ice. We spun cells down at 1000xg for approximately 10 minutes and decanted 20µL of the supernatant. We froze the cell pellet in the lysis buffer at -80°C. We then followed the 100,000 cell count digestion protocol from (Brind'Amour et al. 2015) with a few modifications. For each aged *D. miranda* sample and *D. melanogaster* spike-in sample, we added 4.4µL of 1% DOC and 1% Triton-X100 to resuspended cells and mixed thoroughly. Briefly, we digested each sample at 37°C for 04:58 (mm:ss) with MNase and quenched the reaction with 4.9µL of 100mM EDTA pH 8.0. We then added the fragmented spike-in samples to each *D. miranda* sample to account for

approximately 20% of the final pooled sample volume (i.e. final sample consisted of 80% *D. miranda*, 20% *D. melanogaster*). We reserved 10% of the pooled fragmented sample for the input (fragment control) and used the remaining 90% to perform the chromatin pull-down (ChIP sample, target antibody: H3K9me3 polyclonal classic Diagenode C15410056) according to the 100,000 cell specifications. We used the Rubicon Genomics ThruPlex kit to prepare ChIP-Seq libraries for sequencing with 10 PCR amplification cycles for the input samples and 12 cycles for the ChIP pull-down samples. We sequenced libraries on the Illumina HiSeq 4000 Platform at the Vincent J. Coates Genomics Sequencing Center (Berkeley, CA). In total, we made 4 biological replicates: 2 replicates from the first aging trial, 1 replicate from the second, and 1 replicate from the third.

RNA extraction and RNA-seq. We extracted total RNA from 20 brains per sex per age (3 biological replicates, one from each aging trial) following a standard TRIzol protocol. Briefly, we homogenized tissues first in 100µl TRIzol and then added 900µl TRIzol to proceed with the full-volume protocol. We generated libraries from total RNA using Illumina’s TruSeq Stranded Total RNA Library Prep kit with Ribo-Zero ribosomal RNA reduction chemistry, which depletes the highly abundant ribosomal RNA transcripts (Illumina RS-122-2201). Libraries were run on a Bioanalyzer for fragment traces and sequenced on the Illumina HiSeq 4000 at the Vincent J Coates Genomics Sequencing Center.

Mapping of sequencing reads and data normalization. We mapped all ChIP-seq datasets to the *D. miranda* genome assembly (Mahajan et al. 2018) to the *D. melanogaster* Release 6 of the genome assembly and annotation (Hoskins et al. 2015). For all ChIP-seq datasets, we used Bowtie2 (Langmead and Salzberg 2012) to map reads to both genomes separately, using the parameters “-D 15 -R 2 -N 0 -L 22 -i S,1,0.50 --no-1-mm-upfront”, which allowed us to reduce cross-mapping to the *D. melanogaster* genome to approximately 1% of 100bp-paired end reads (Brown, Nguyen, and Bachtrog 2020b). All ChIP and input samples were processed with bedtools coverage -counts (Quinlan and Hall 2010). To calculate the ChIP signal without the *D. melanogaster* spike-in normalization, we normalized all counts files by library size, and calculated the coverage across 5kb windows for both the ChIP and the input *D. miranda* alignments. These values were then normalized by the median autosome coverage of euchromatic arms (Muller B and Muller E) to obtain ChIP signal (ChIP/Input) for each sample. To calculate signals normalized by the *D. melanogaster* spike-in, we used three different approaches (see **S1 Fig**). First, following the approach from (Brown, Nguyen, and Bachtrog 2020b), we took the ChIP enrichment from the no-spike normalization method and multiplied these values by a scale factor. The scale factor c is calculated as follows with normal script indicating the number of reads mapped for the given sequencing sample and subscripts indicating the specific species to which reads were mapped:

$$c = 1 + \text{ChIP}_{D.miranda} / (\text{Input}_{D.melanogaster} / \text{Input}_{D.miranda}) / (\text{Input}_{D.melanogaster} + \text{ChIP}_{D.melanogaster}).$$

In this respect, we calculated normalization factor c based on the number of mapped reads and this is irrespective of mapping location. We also used a regression-based method from (Bonhoure et al. 2014). Briefly, we calculated the median ChIP signal for each sample’s *D. melanogaster* spike-ins. Then, the ChIP signal in each *D. miranda* sample was adjusted by a logistic regression based on each sample’s genome-wide *D. melanogaster* spike-in signal. In a separate spike-in

normalization approach, we used a quantile-quantile method (Wei, Gibilisco, and Bachtrog 2020). A reference ChIP signal for all *D. melanogaster* spike-ins was made by aggregating all *D. melanogaster* ChIP and input coverages, respectively. Then, all individual *D. melanogaster* spike-ins were adjusted to this reference and the corresponding *D. miranda* ChIP signals were adjusted accordingly.

We emphasize that all spike-in normalization methods do not use specific regions of the genome and are based only on the mapping and/or ChIP enrichment of the *D. melanogaster* spike-in samples. All three spike-in normalization methods as well as the no-spike normalization procedure yielded quantitatively similar results: young males and females show similar enrichment of H3K9me3 along their pericentromeres; both old males and females show less H3K9me3 enrichment than young flies, and the loss of heterochromatin is more pronounced in males than females, and finally, the neo-Y chromosome generally shows less heterochromatin than pericentromeres (**S1 Fig**). We further confirmed the consistency of these normalizations by showing that 5kb windows in euchromatin with elevated H3K9me3 enrichment are found consistently in all normalization methods (**S10 Table**). We used the no-spike normalization procedure for our analysis, and one sample with low ChIP efficacy (old male 1; see **S1 Fig**) was excluded from the analysis.

The boundaries of euchromatic chromosome arms versus pericentromeric heterochromatin was defined using repeat-enrichment, using a cut-off of 40% repeat-masked DNA across 100 kb sliding windows away from the centromere (from (Bracewell et al. 2019)). For the neo-Y chromosome, we identified the pericentromeric region based on increased H3K9me3 signal surrounding the centromeric satellite sequences (see (Mahajan et al. 2018)), which roughly corresponds to 9.8Mb of sequences at the end of neo-Y scaffold 1 (**S19 Fig**).

Gene expression analysis. For each replicate of RNA-seq data, we first mapped RNA-seq reads to a repository of ribosomal DNA scaffolds from NCBI and removed all reads that mapped to this scaffold. Although total RNA library preparation aims to remove the bulk of rRNA transcript, any additional differences in rRNA abundance in sequenced samples are likely to be technical artifacts. We then mapped the remaining reads to the *D. miranda* genome and to the repeat libraries separately using HISAT2 (Kim, Langmead, and Salzberg 2015) with default parameters. We then used Subread FeatureCounts (Liao, Smyth, and Shi 2014) to calculate gene counts and repeat counts and used DESeq2 (Love, Huber, and Anders 2014) to normalize the libraries and perform differential expression analysis between male/female flies of different ages. In an analogous approach, we used TECount from TEtranscripts (Jin et al. 2015) to estimate gene and TE expression counts across samples. We used TECounts with default parameters on RNA-seq reads mapped to the *D. miranda* genome. We also used DESeq2 (Love, Huber, and Anders 2014) to perform the differential expression analysis with these counts tables. Both counting methods for repeats yielded similar results for each sample. For a non-reference based approach, we used dnaPipeTE (Goubert et al. 2015) on RNA-seq reads using the following parameters: -coverage 0.5 -genome-size 200000000 -sample-number 2. We then calculated the proportion of TEs and repeats identified in each sample.

Repeat libraries. We used two repeat libraries for the ChIP-seq and RNA-seq analyses. Our first library was the consensus sequences of known repetitive elements identified in the *D.*

pseudoobscura group (T. Hill and Betancourt 2018) and the second was based on the consensus sequences of known satellites identified in *D. miranda* (Mahajan et al. 2018). The following methods were employed for both repeat libraries.

ChIP enrichment/RNA expression of specific repeats (and by chromosome). To assess H3K9me2 signal in repetitive elements, we took a similar approach as we did for calculating ChIP enrichment profiles across the genome. First, we mapped both ChIP and input sequencing reads to the *D. pseudoobscura* family TE library/satellite DNA library using Bowtie2 (Langmead and Salzberg 2012) and the parameters “-D 15 -R 2 -N 0 -L 22 -i S,1,0.50 --no-1-mm-upfront”. We then calculated the mean coverage across each repetitive element using Bedtools (Quinlan and Hall 2010) coverage, and normalized the coverage by the median autosome coverage. We then calculated the ChIP signal at each repeat by dividing the ChIP coverage by the corresponding input coverage. This method accounts for differences in copy number of the repetitive elements by dividing the ChIP coverage by each repeat’s coverage in the input. In a complementary approach, we took ChIP mappings to the entire genome and called enrichment at repetitive elements through TECounts (part of TETranscripts) (Jin et al. 2015). To call for repeat enrichment on a per chromosome basis, we split mappings by chromosome and ran TECounts separately for each split mapping. We then normalized counts at repeats in both the ChIP and input by mean autosomal coverage and calculated ChIP enrichment as ChIP/Input.

ChIP profiling at transcription start sites (TSS). To calculate H3K9me3 enrichment around genes on the neo-Y or pericentromeric regions, we followed the BedTools pipeline for estimating transcription factor occupancy at transcription start sites (Quinlan and Hall 2010). Briefly, we calculated genome-wide ChIP enrichment at 100bp resolution as described previously. We then generated intervals flanking transcription start sites of genes on the neo-Y and pericentromeric regions using BedTools slop. We then used Bedtools map to map our ChIP enrichments to these flanking regions and Bedtools groupby to aggregate enrichment profiles across all TSS.

Immunofluorescence staining in mitotic chromosome spreads. We followed an immunofluorescence protocol from (Fanti et al. 2008). We dissected brains from male 3rd instar larvae in 0.7% NaCl. We incubated brains first in 0.5% sodium citrate for 8 minutes and then in 2% paraformaldehyde in 45% acetic acid fixative solution for 8 minutes. We transferred 4 brains to small drops of fixative solution on a 18mmx18mm siliconized coverslip and placed a clean glass slide on top. The slides were pressed with a thumb for 1.5 minutes and then frozen in liquid nitrogen until bubbling stopped. We used a razor blade to pop off the coverslip and immediately immersed slides in PBS (pH 7.4). The slides were then treated with 100uL of 20units/mL DNase1 (Thermofisher #89836) in 50mM Tris-HCl pH 8.0, 5mM MgCl₂, and 0.05mg/mL BSA for 4.5 minutes. Treatment with DNase1 increases antibody binding to the H3K9me3 histone mark as reported in (Fanti et al. 2008). The slides were then washed 3x in PBS for 5 minutes/wash and incubated in 1% Triton X-100 in PBS for 10 minutes. Slides were blocked in 5% nonfat dry milk in PBS for 30 minutes, washed in PBS for 5 minutes, and then incubated with primary antibody (1:300 dilution, Abcam ab8898) in 3% BSA in PBS for 1 hour first at room temperature and later overnight at 4°C in a box lined with wet paper towels. We washed slides 3x in PBS for 5 minutes/wash and then incubated them with secondary antibody Goat anti-rabbit conjugated with Alexa Fluor-568 in 3% BSA in PBS (1:1000 dilution, Thermofisher #A-11011) for 1 hour at room temperature in a box lined with wet paper towels and covered

with aluminum foil. Slides were washed 3x in PBS for 5 minutes/wash and dried for 10 minutes in the dark. We mounted slides using Vectashield antifade mounting medium with DAPI (Vector Laboratories H-1200-10) on a coverslip and then sealed the edges with nail polish. Slides were imaged using a Nikon Spinning Disc Confocal microscope and the Nikon Elements AR software at the UC Berkeley Cancer Research Laboratory Molecular Imaging Center. We used Fiji/ImageJ to adjust image brightness and contrast.

Figure Legends

Fig 1. *Drosophila miranda* karyotype and lifespan. **A.** Karyotype of *D. miranda* males and females. Chromosome arms (Muller elements) are color-coded. **B.** Polytene chromosome squash of *D. miranda* male stained with *HPI* (magenta) marking heterochromatin. **C.** Genome-wide repetitive content per chromosome. Shown is repeat content (% repeat masked bp) in 50kb windows. **D.** Kaplan-Meier survivorship curves (Kaplan and Meier 1958) for MSH22 males (blue) and females (red), with the shaded region indicating the upper and lower 95% confidence interval calculated from the Kaplan-Meier curves. The number of flies counted for each sex (n) to obtain the survivorship curves is indicated.

Fig 2. Heterochromatin loss and aging. **A.** Genome-wide enrichment of H3K9me3 for young and old *D. miranda* males and females along the different chromosome arms. Values represent the mean of 4 biological replicates for Young/Old Females and Young Males and the mean of 3 biological replicates for Old Males in 25kb windows. Subtraction plots show the absolute difference in signal of 25kb windows between young and aged flies along the chromosome arms, with females in red and males in blue. One tick mark is equal to 1 Mb on the x-axis. **B.** Boxplot showing ChIP enrichment differences between biological samples at euchromatin and pericentromeres. Enrichment resolution is in 5kb windows. Significance values calculated (* $p < 0.05$, ** $p < 1e-6$, *** $p < 1e-12$, Wilcoxon test) for males (blue, cyan) and females (red, pink), with pericentromere boundaries defined by (Bracewell et al. 2019).

Fig 3. Heterochromatin enrichment across chromosomes. **A.** ChIP enrichment across samples grouped by chromosome and binned by amount of repeats (%) per 5kb-window. Values represent the mean of 4 replicates with standard error bars with the exception of the “Old Males” where we omit 1 failed ChIP replicate. **B.** Immunofluorescence staining for H3K9me3 in male mitotic chromosomes. Arrowhead denotes neo-Y, arrow denotes MullerA-AD (old-X), scale bar is 50µm.

Fig 4. Heterochromatin enrichment at transcription start sites (TSS) on the neo-Y and the pericentromeres. **A.** The average young male H3K9me3 ChIP enrichment profile +/- 5kb of TSS in 100bp resolution grouped by gene location (neo-Y or pericentromere). **B.** The average young male enrichment profile at TSS neo-Y genes grouped and plotted by expression level. The resolution is also 100bp.

Fig 5. Heterochromatin enrichment at TEs across sex, time and chromosomes. Shown is enrichment of H3K9me3 at different repeat families for young and old females and males. **A.** Log2 H3K9me3 enrichment of all transposable elements (from (Wei, Gibilisco, and Bachtrog 2020)) in young (5-9 day) and old (80-98 day) male and female MSH22, averaged across 4

replicates for Young/Old Females and Young Males and across 3 replicates for Old Males. H3K9me3 enrichment in males is separated by autosomes/X and neo-Y. Boxplots below with significance values calculated using the Wilcoxon test (* $p < 0.05$, ** $p < 0.01$, *** $p < 1e-5$). **B.** Log2(H3K9me3 enrichment) differences between TEs found on the Autosomes/X and the neo-Y for young (top) and old (bottom) males. Data represents averages of 4 replicates for Young/Old Females and Young Males and of 3 replicates for Old Males. The bars denote standard error. Data highlighted in yellow are TEs that show significant differences in H3K9me3 enrichment on autosomes/X and neo-Y ($p < 0.05$, two-tailed Student's t-test).

Fig 6. TEs are upregulated in male *D. miranda*. Shown is differential TE expression within and between sexes. **A.** Overall log2 expression by sex and age with rows sorted by log2 female/male TE coverage (left histogram). Boxplots below with significance values calculated using the Wilcoxon test (* $p < 0.05$, ** $p < 0.01$, *** $p < 1e-5$). Data represents averaged expression across samples. **B.** Log2 expression differences of TEs between young (left) and old (right) males versus females. Data indicates mean value of 3 biological replicates and bars denote standard error. Color-code indicates significant comparisons (50% significant higher or lower expression; Wald test, $p < 0.05$) and Y-enriched repeats (based on genomic coverage): cyan: Y-enriched, $p < 0.05$; yellow: unbiased, $p < 0.05$; blue: Y-enriched, n.s.; black: unbiased; n.s. Grey line indicates 1:1 expression levels. **C.** Same as B but comparison between young vs. old males (left) and females (right). **D.** Log2 TE expression by sex and age grouped by TEs enriched on the neo-Y (blue) and TEs not enriched on the neo-Y (yellow). Significance values calculated using the Wilcoxon test (* $p < 0.05$, ** $p < 0.01$, *** $p < 1e-5$).

Fig 7. Heterochromatin and expression at satellite DNA. **A.** Average log2 H3K9me3 enrichment at different satellite DNA for young and old females and males. Boxplots with significance values calculated using the Wilcoxon test (* $p < 0.05$, ** $p < 0.01$, *** $p < 1e-5$). **B.** Log2 expression differences between young (left) and old (right) males and females. Data represents mean value of 3 replicates with bars denoting standard error, data in yellow denote significant differentially expressed satellites (50% higher or lower expression, Wald test, $p < 0.05$). Grey line denotes 1:1 expression levels between the samples. **C.** Log2 expression differences between young vs. old males (left) and females (right). Plots made in similar fashion as in **B**.

S1 Fig. ChIP enrichment across samples using four different normalization strategies (5kb windows; see Materials & Methods for details). From top to bottom: No spike in normalization: ChIP enrichment normalized by coverage of the autosomes for all biological samples and replicates, ignoring the spike-in. Brown et al. (2020) normalization: ChIP enrichment normalized by assessing proportion of reads derived from sample and spike-in both the ChIP and Input. Bonhoure et al. (2014) normalization: ChIP enrichment normalized by employing a regression-based method using the *D. melanogaster* spike-in. Wei et al. (2020) normalization: ChIP enrichment normalized by employing a quantile-quantile method using the *D. melanogaster* spike-in. **A.** Enrichment for each sample replicate.

S1 Fig (continued) B. Average enrichment across samples. We omitted the first Old Male sample for all averaged Old Male values because it had lower enrichment signal compared to other samples in both the sample and the spike-in, indicative of inefficient ChIP pull-down.

S2 Fig. ChIP signal after subsampling reads to 37 million reads for each replicate. **A.** Enrichment for each sample replicate. **B.** Average enrichment across samples. Enrichment is normalized without spike-in and omits the failed Old Male replicate.

S3 Fig. ChIP enrichment across samples (average across replicates) with only uniquely mapped reads (5kb windows) using ChIP signal normalized by autosome coverage (3 replicates used for the Old Male category).

S4 Fig. Input coverage across sexes and ages. (Top) Coverage of the input DNA across samples normalized by the median autosomal coverage. (Bottom) Mean coverage of the input DNA across samples.

S5 Fig. Genome-wide ChIP enrichment with ± 1 standard error (denoted in yellow). 1 tick mark = 1Mb on the x-axis.

S6 Fig. Differential ChIP enrichment between the sexes. Each data point represents enrichment of old/young male (left) or female (right) samples at a 5-kb window. Data highlighted in yellow denote windows with 50% higher/lower enrichment between samples ($p < 0.05$, two-tailed Student's t-test).

S7 Fig. Regions with at least 1.5x H3K9me3 up-/down-regulation within sexes over aging in 5kb-window resolution. Data values represent the difference of means from old and young samples. One tick mark on the x-axis is equal to 2Mb.

S8 Fig. Enrichment of H3K9me3 (in 5kb windows) at the pericentromeres and 2Mb downstream for **A.** males and **B.** females. Boundaries are indicated by darker regions on the chromosome diagrams of the major chromosome arms. Subtraction plots show higher H3K9me2 signal in young (blue) or old (red) flies. Data represents the mean of 4 biological replicates (3 for old males; in black) or the difference of means (subtraction plots, blue or red).

S9 Fig. Sex-specific H3K9me3 gain/loss by compartment (euchromatin/heterochromatin/neo-Y) over aging.

S10 Fig. Immunofluorescence staining for H3K9me3 in males (A-I, replicates from different slides) and females (J). For all figures, arrowhead denotes neo-Y, arrows denote MullerA-AD, DAPI/DNA is blue channel, H3K9me3 is orange channel, and scale bar is 50 μ m.

S11 Fig. Repeat composition per chromosome. Plotted for each Muller element is the total length of repeats (bp) divided by the entire chromosome length (bp).

S12 Fig. ChIP enrichment differences between young and old females and males. Values represent the difference of means between young and old samples, grouped by chromosome and binned by amount of repeats (%) per 5kb-window.

S13 Fig. Volcano plots for TE expression by sex and age. Left: (Top) TEs in young females v. young males, (bottom) TEs in old females vs old males. Fold change is indicative of

male/female. Right: (Top) TEs in young males vs old males, (bottom) TEs in young females vs old females. Fold change is indicative of young/old.

S14 Fig. TE expression by sex/age and by chromosome group (Autosomes/X and Neo-Y).

S15 Fig. Male TE expression from the Autosomes/X and the Neo-Y in young (left) and old males (right). Diagonal line denotes 1:1 equal expression for TEs from both chromosome groups.

S16 Fig. Sex-differences in TE expression normalized by sex-differences in TE copy number. **A-C.** Shown are expression values for each biological sample accounting for TE copy number differences in the males and females when copies are calculated from **A.** adult gDNA, **B.** embryo gDNA, and **C.** genome repeat annotation. Expression values for each TE were normalized by their respective copy number in the male or female genome. Data plotted represents the averaged values of old and young female and male samples (3 replicates). Significance values calculated (n.s. not significant, * $p < 0.05$, Wilcoxon test). **D-F.** Shown are $\text{Log}_2(\text{Male/Female Expression})$ values for each TE along with the corresponding $\text{Log}_2(\text{Male/Female DNA})$ values calculated from **D.** adult gDNA, **E.** embryo gDNA, and **F.** genome repeat annotation. Line indicates male/female expression is proportional to the male/female copy number of a specific TE.

S17 Fig. Volcano plots for satellite DNA expression by sex and age. Left: (Top) Satellites in young females v. young males, (bottom) satellites in old females vs old males. Fold change is indicative of male/female. Right: (Top) Satellites in young males vs old males, (bottom) satellites young females vs old females. Fold change is indicative of young/old.

S18 Fig. Heterochromatin and expression of pericentromeric and neo-Y genes. **A.** Log_2 expression of genes located in the pericentromeres within sexes. Data represents mean values of 3 replicates with bars denoting standard error. Data highlighted in yellow denotes genes with significantly 50% higher or lower expression when comparing fold-change between samples (Wald test, $p < 0.05$). Grey line indicates 1:1 expression between samples. Boxplot values represent mean values of 3 replicates with significance values calculated using the Wilcoxon test (* $p < 0.05$, ** $p < 0.01$, *** $p < 1e-5$). **B.** Log_2 expression of genes located on the neo-Y between young and old males. Data represented in a similar manner as in **A** and data highlighted shows genes that are up-regulated (top-left) and/or down-regulated (bottom-right) by 50% during aging (Wald test, $p < 0.05$). **C.** Age-related differences in log_2 H3K9me3 enrichment and log_2 expression of neo-Y genes. Each data represents the difference of mean values between young and old males. Color-code indicates differentially expressed/enriched comparisons: black: absolute $\text{log}_2(\text{fold-change}) < 1$ in both H3K9me3 and expression; yellow: absolute $\text{log}_2(\text{fold-change}) \geq 1$ in both H3K9me3 and expression. **D.** Log_2 expression differences in genes on the neo-Y and their respective gene copy number. Values denote the difference of mean values between young and old males. Grey line denotes 0 expression difference between the ages.

S19 Fig. Heterochromatin cutoffs (dashed orange line) by % repeats overlap with regions of elevated H3K9me3 enrichment on (A) Muller B, (B) Muller E, (C) Muller C, and (D) Muller AD.

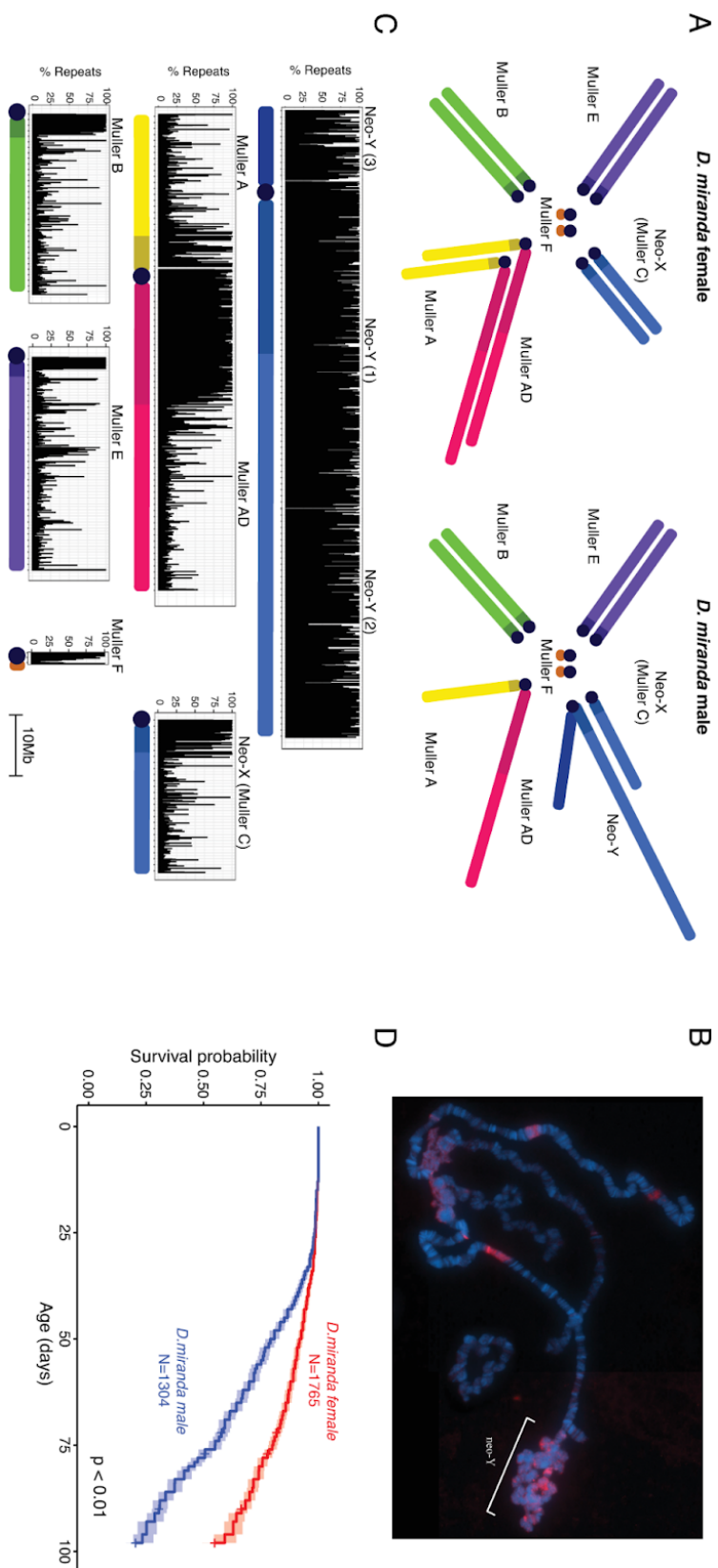
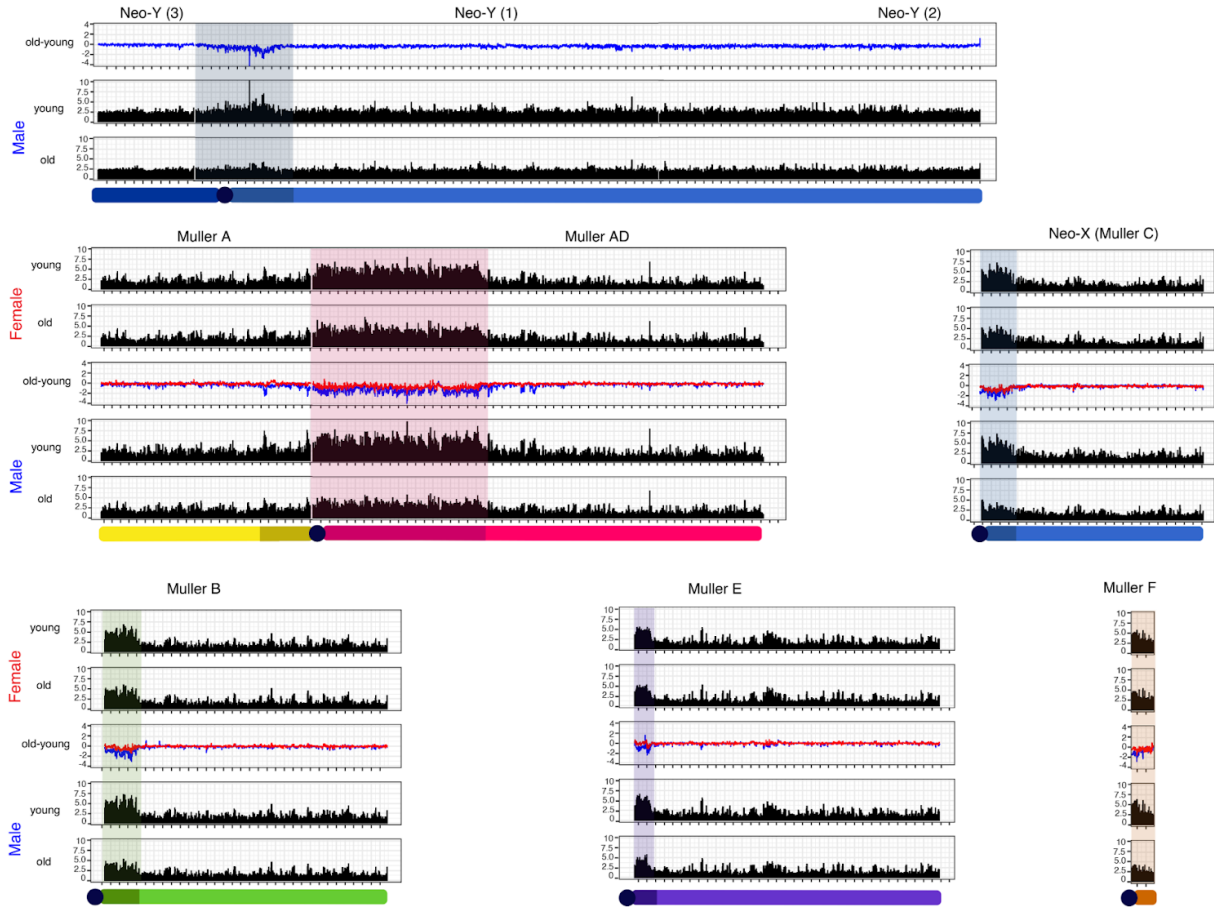


Figure 1.

A



B

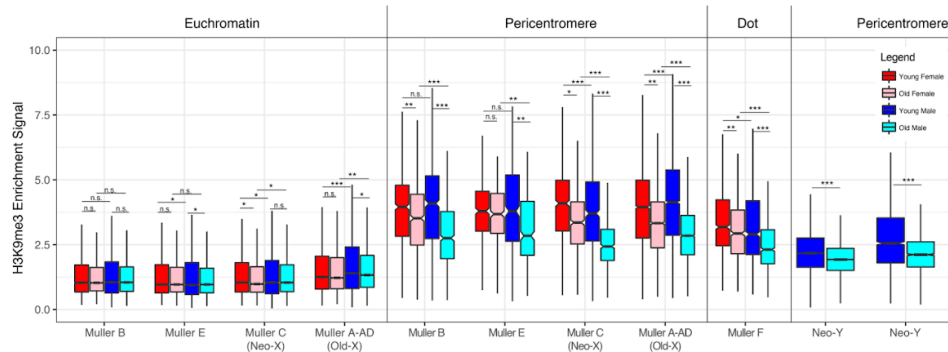
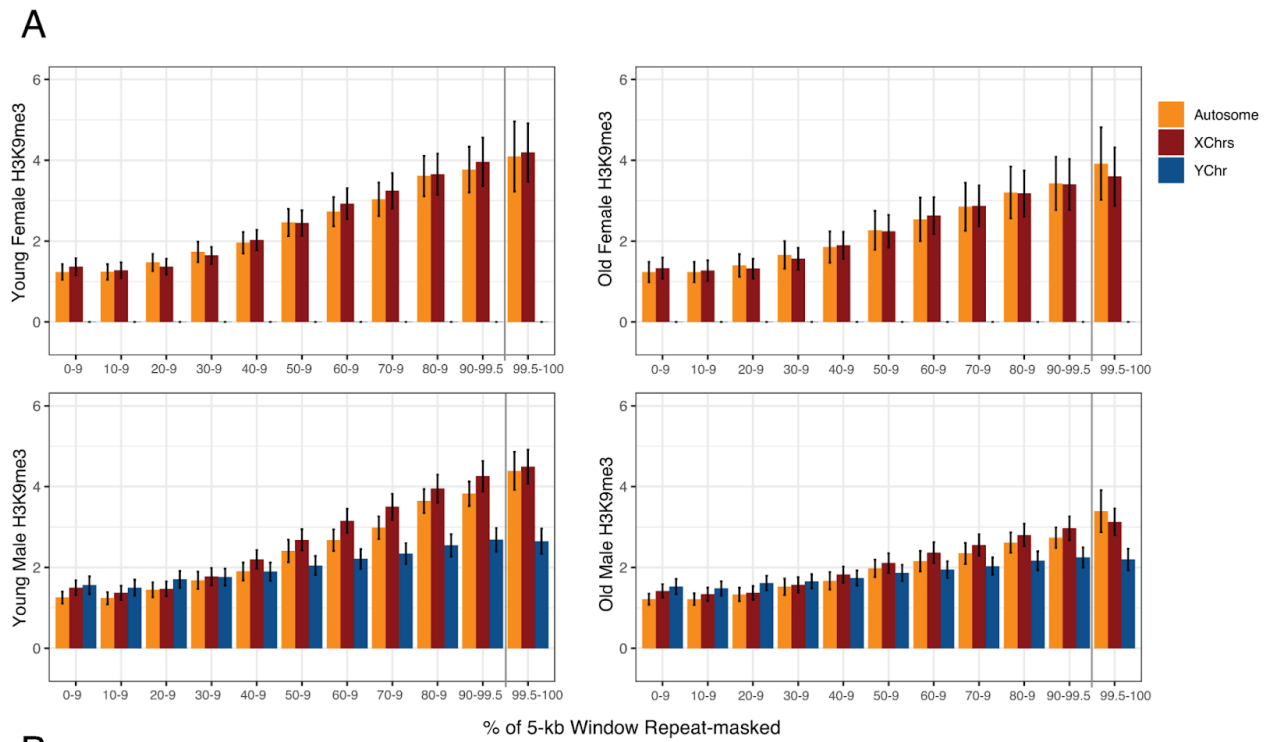


Figure 2.



B

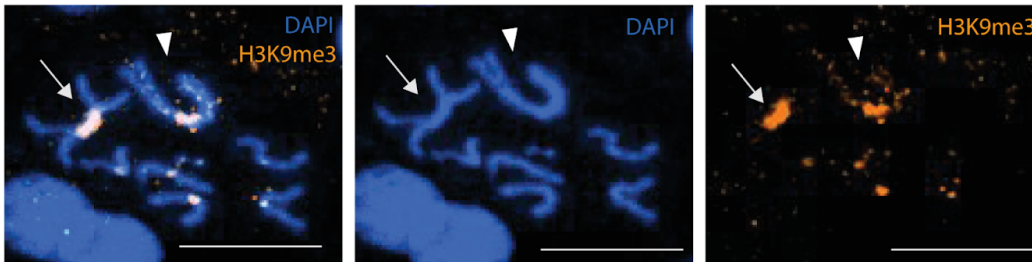


Figure 3.

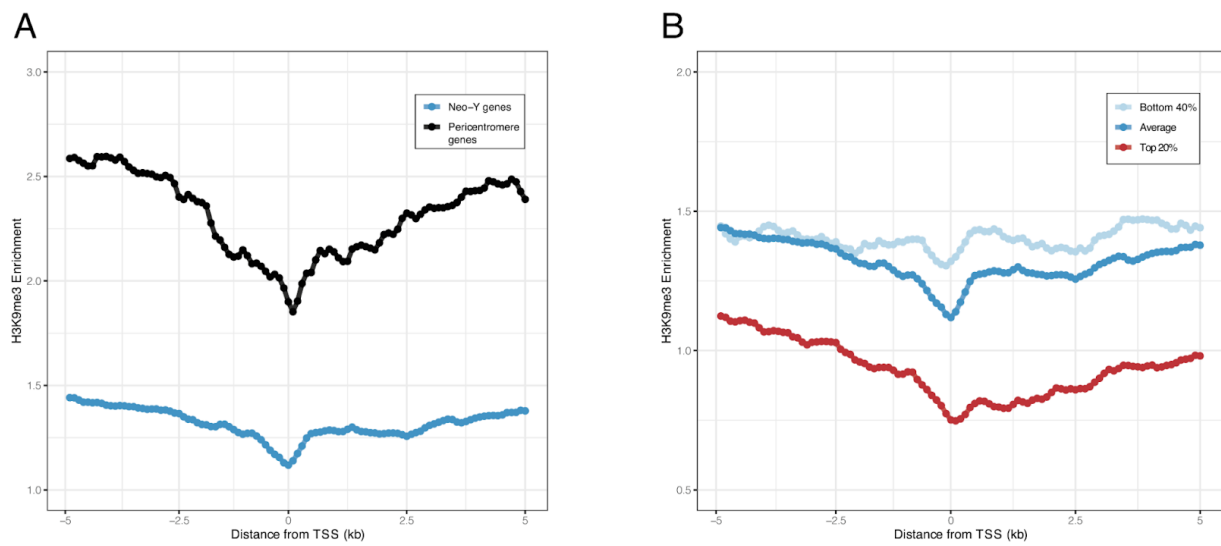


Figure 4.

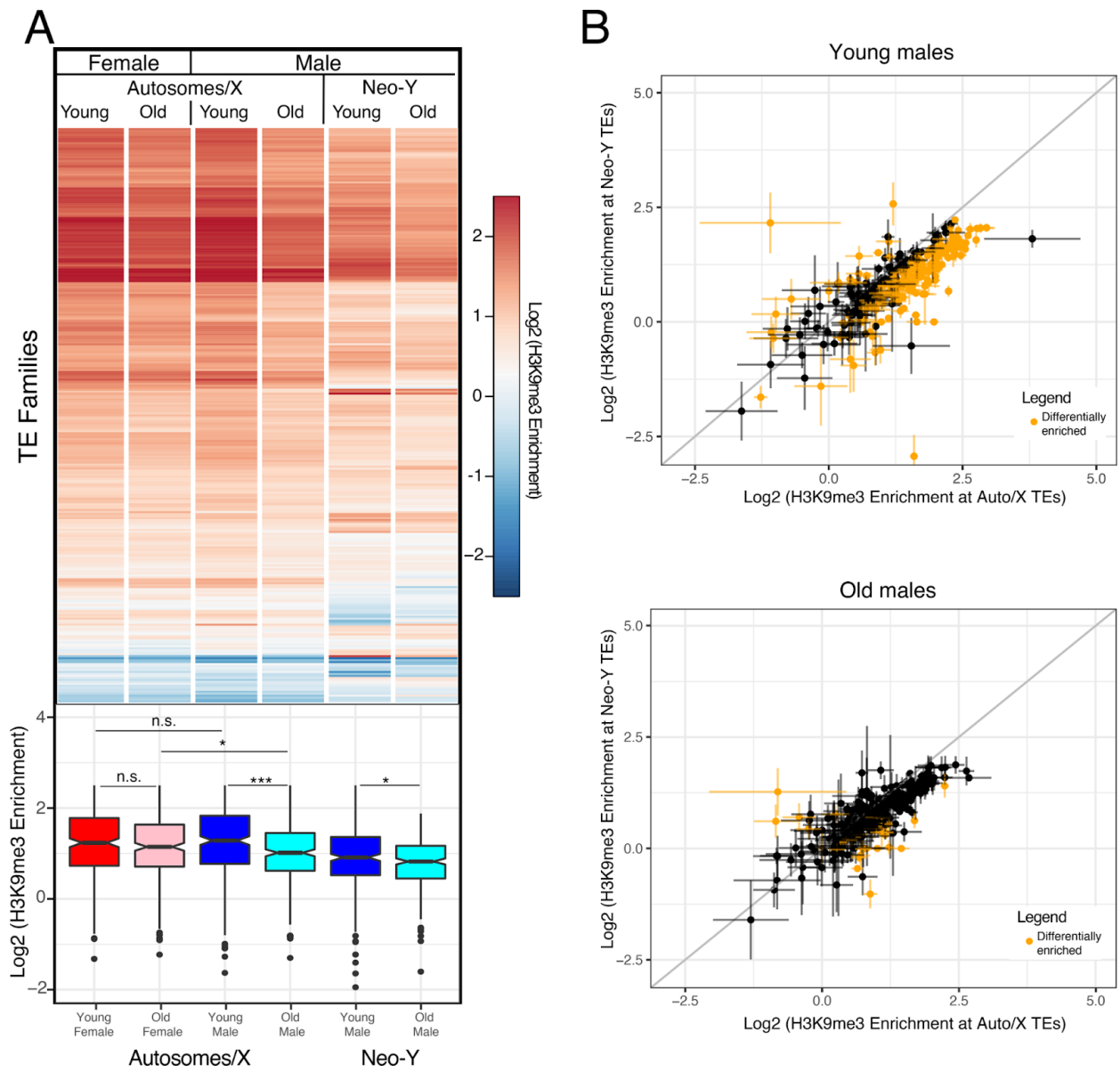


Figure 5.

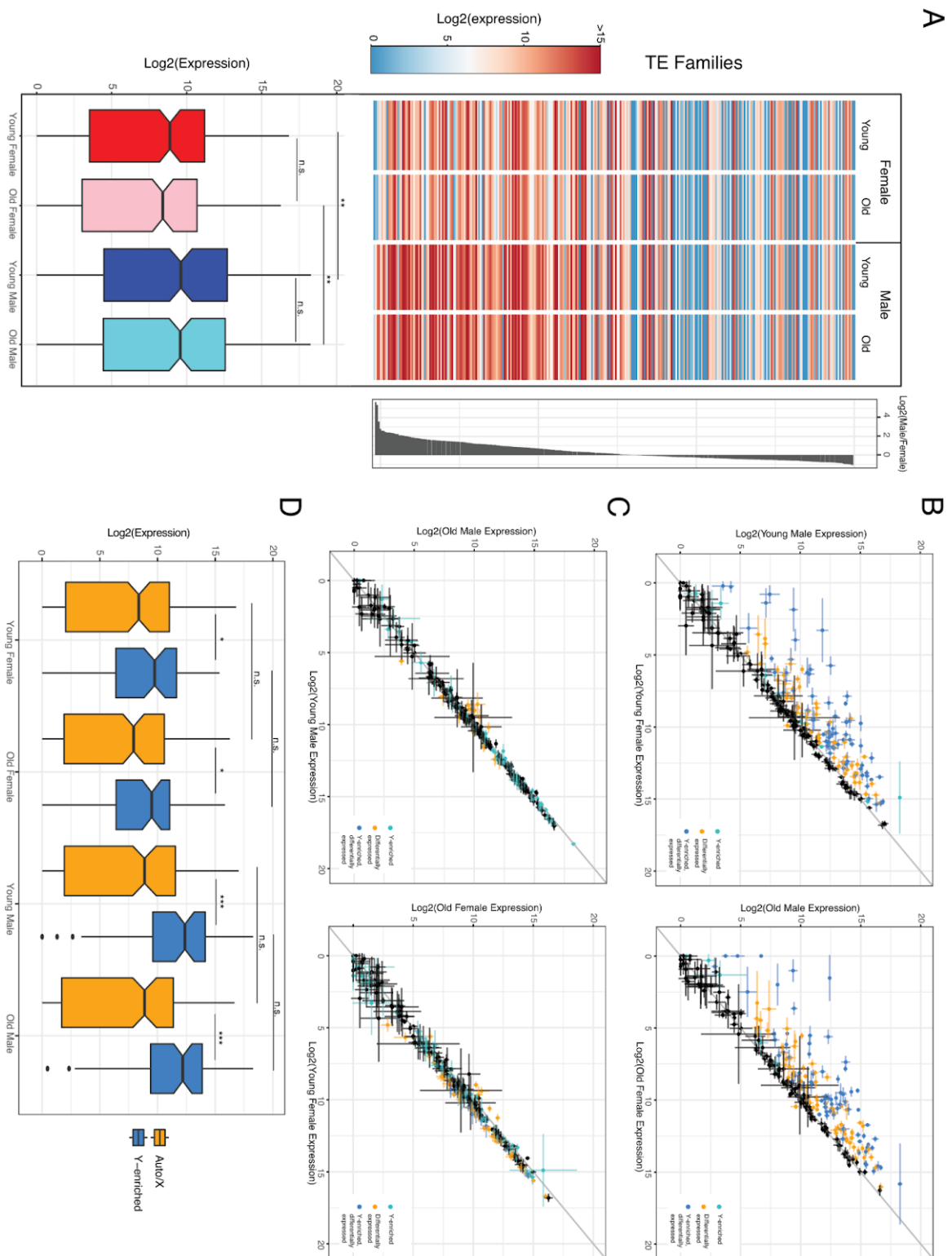


Figure 6.

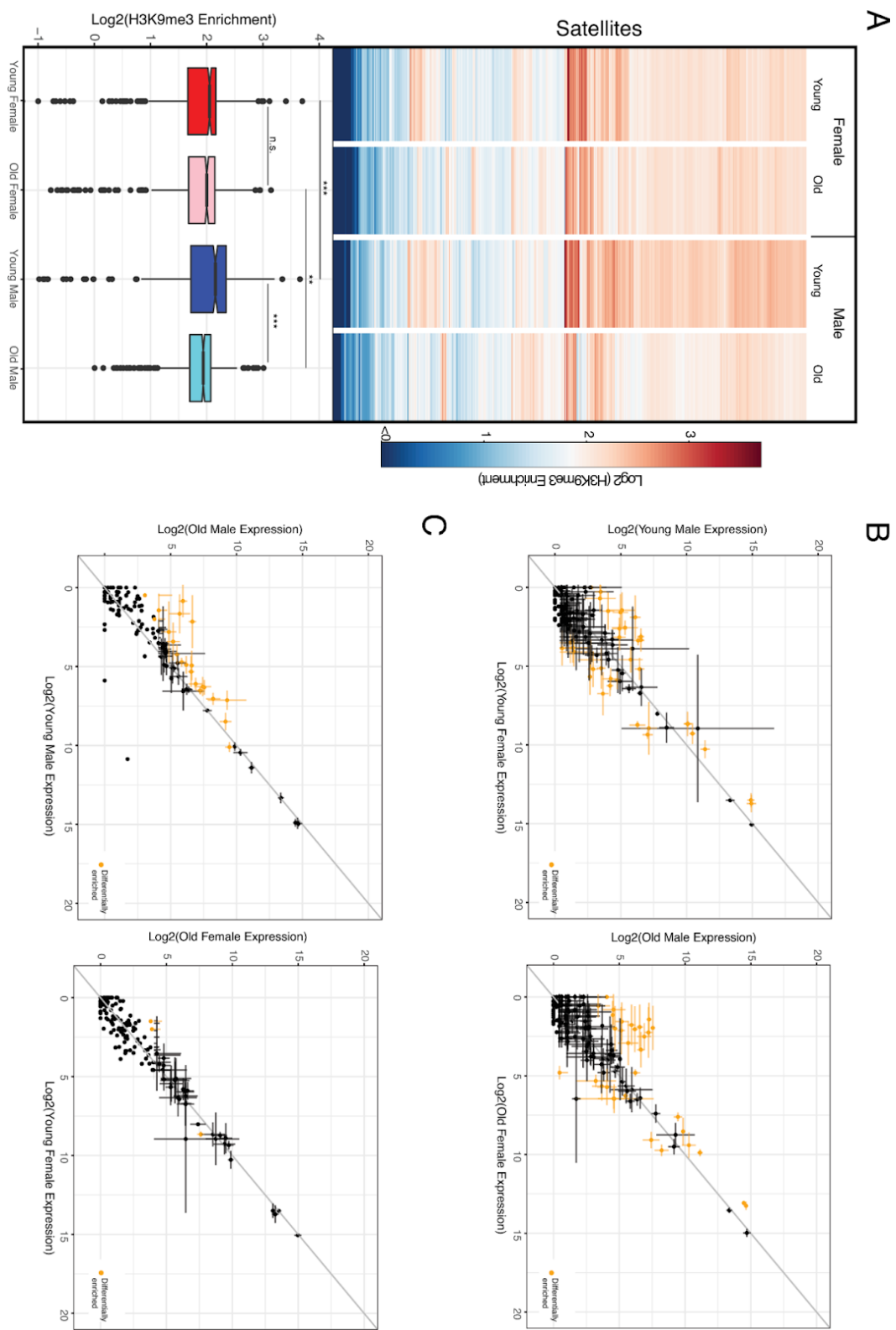


Figure 7.

Table 1. Assembled length for each chromosome, size of chromosome arms and pericentromere, and repeat-masked bases (in Mb). The total genome size inferred from flow cytometry is given in brackets.

	Total Assembly (Mb)	Euchromatic Arms (Mb)	Pericentric Heterochromatin	Repeatmasked (Mb)
X chr (A/D)	77.6	57.1	20.5	28.9
neo-X (C)	25.3	21.9	3.4	6.1
Y/neo-Y chr	92.1	n/a	n/a	76.8
chr4 (B)	32.5	29.1	3.4	6.6
chr2 (E)	35.3	33.3	2	6.2
dot	2.4	0	2.4	1.3
total	173.1	141.4	31.7	49.1
male (2n)	335.4 (354)	203.8	131.6	140.0
female (2n)	346.3 (351.2)	282.9	63.4	98.2

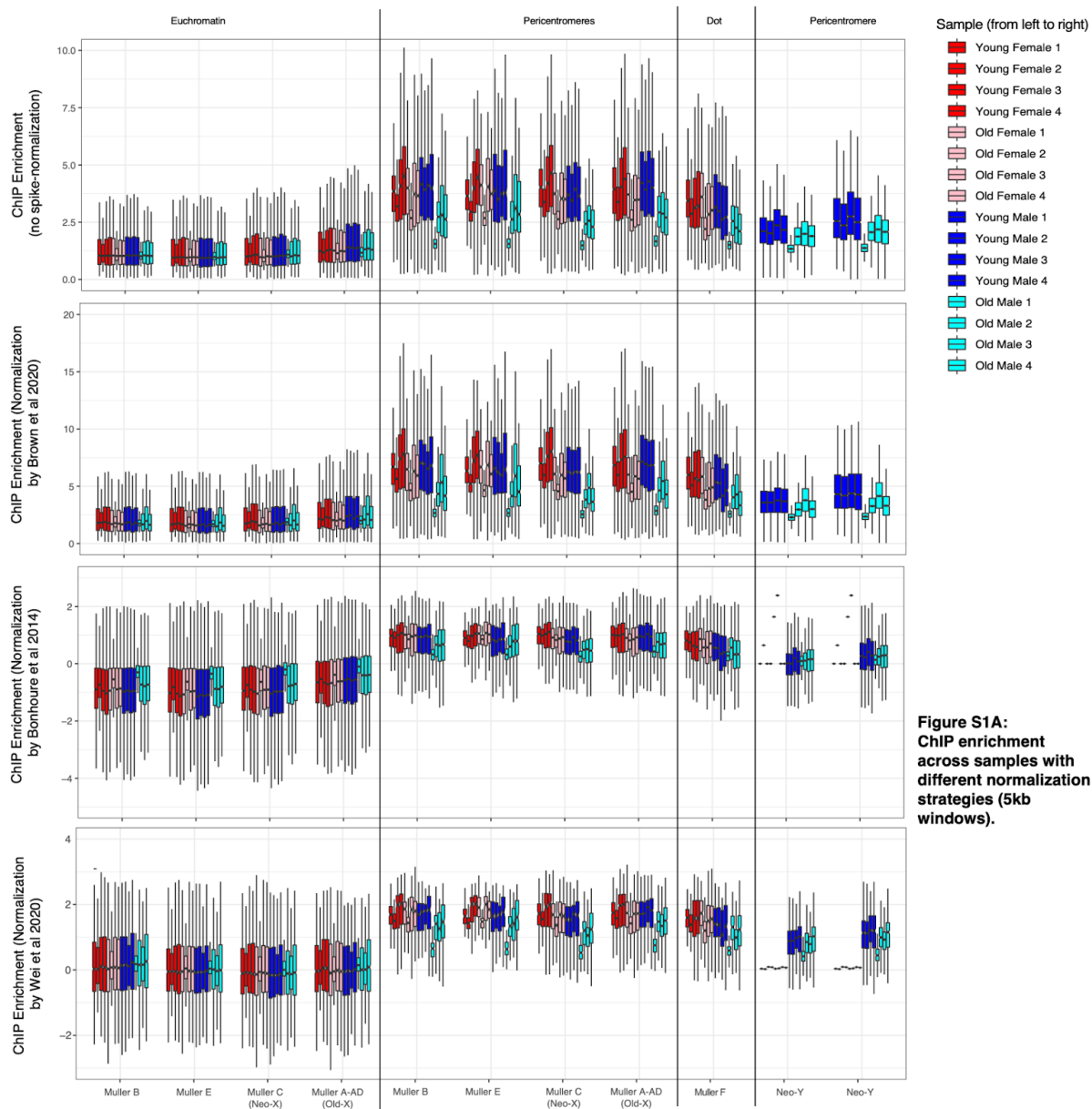


Figure S1A.

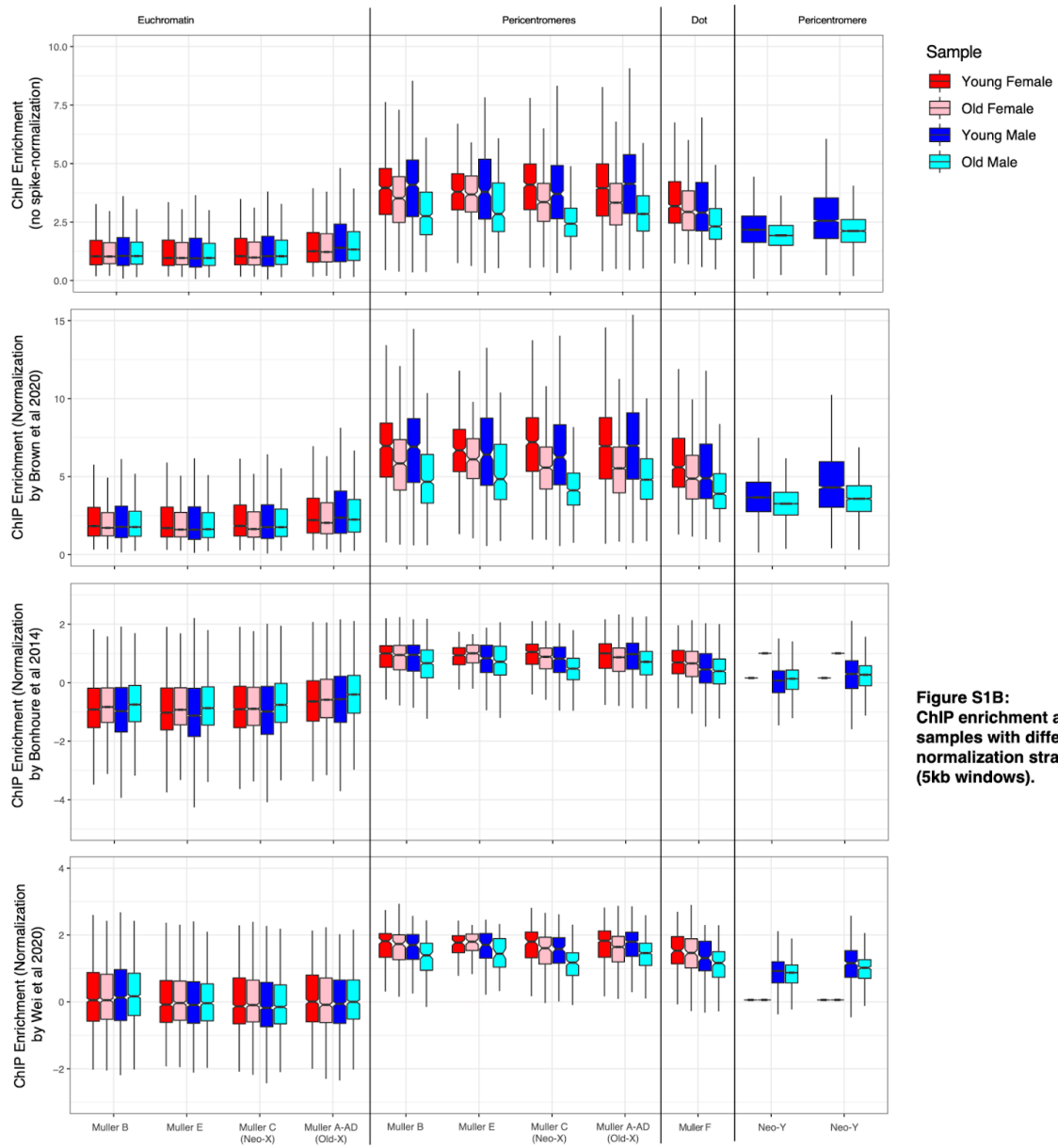


Figure S1B:
ChIP enrichment across
samples with different
normalization strategies
(5kb windows).

Figure S1B.

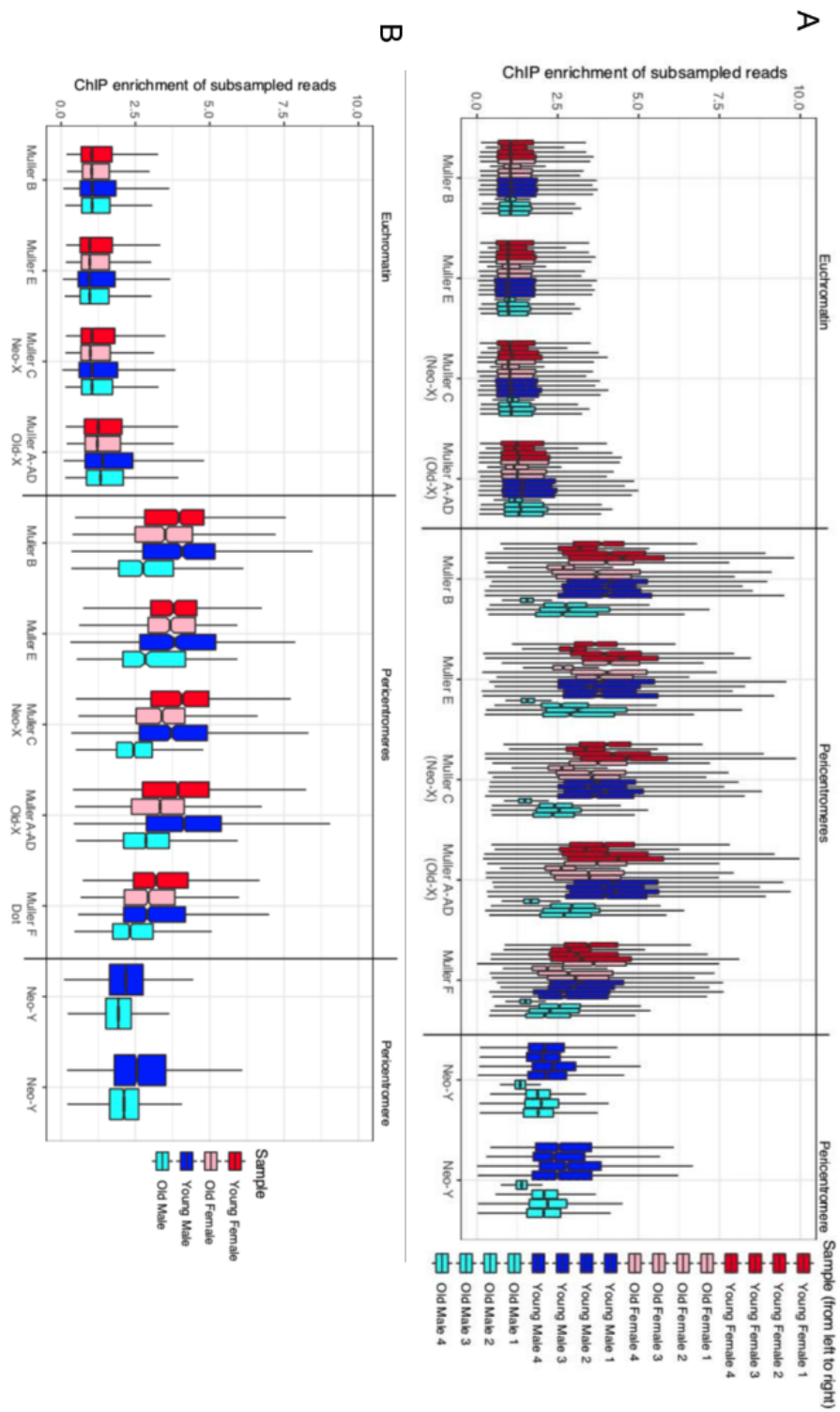


Figure S2.

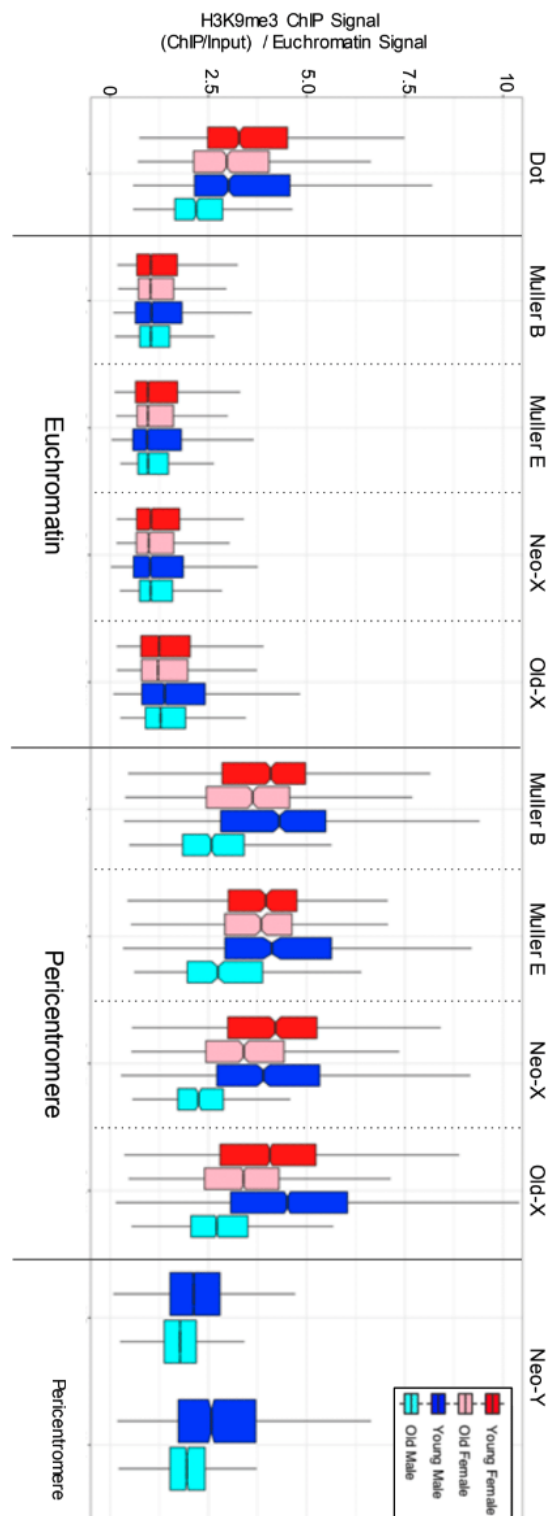


Figure S3.

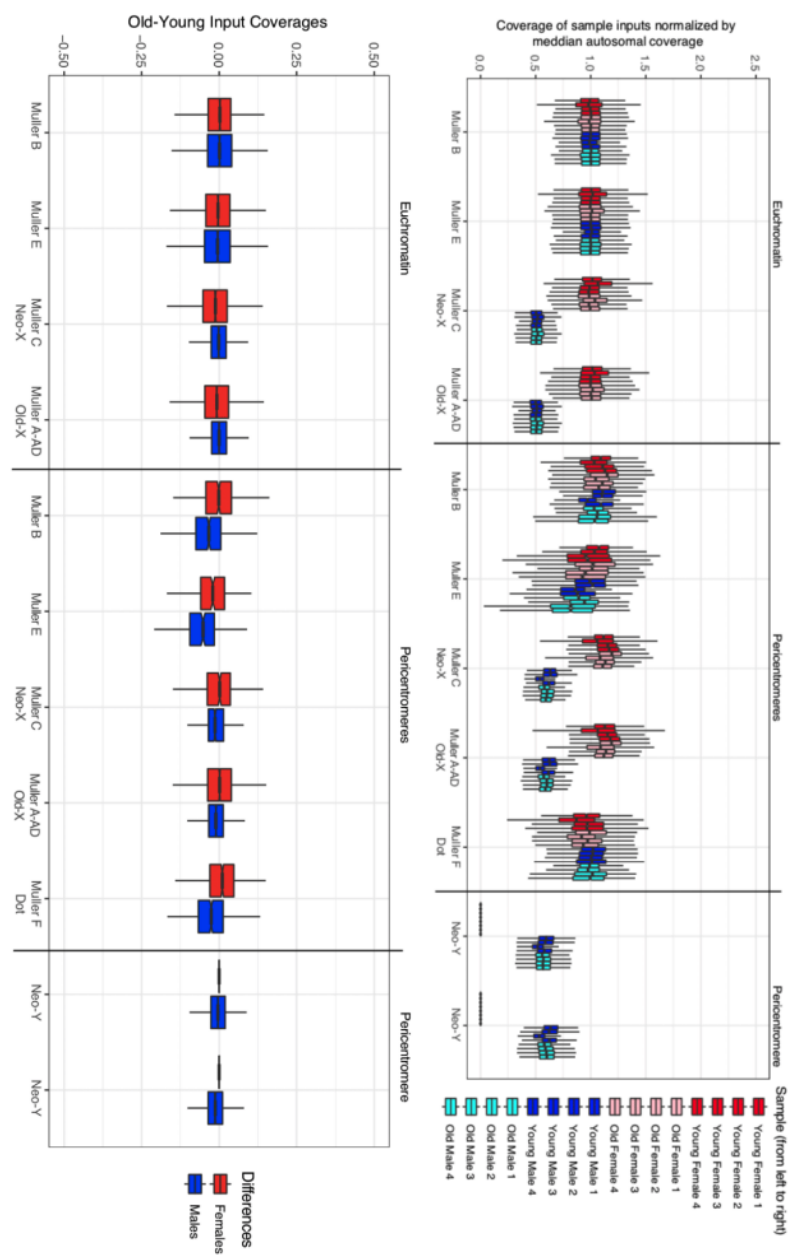
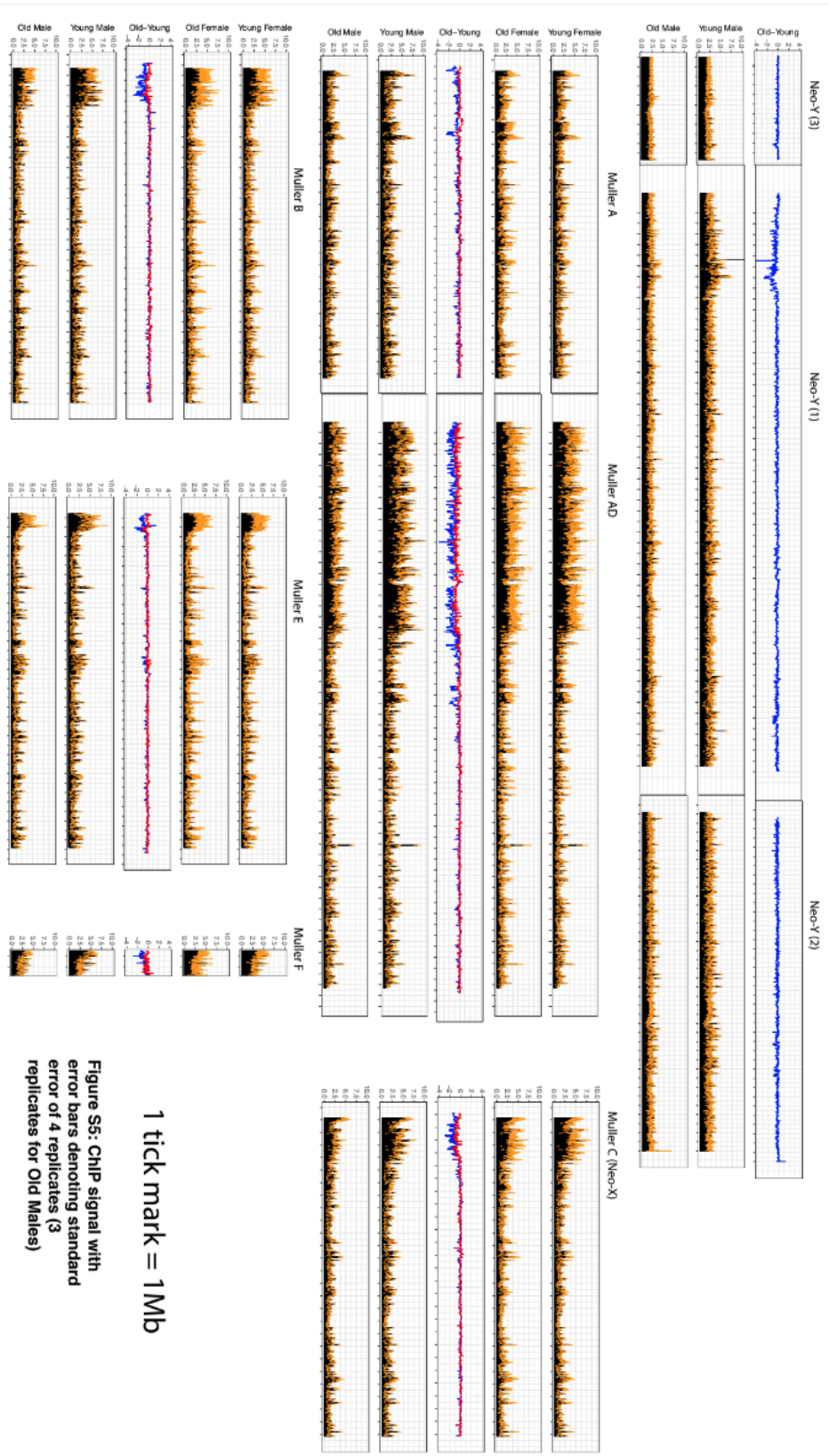


Figure S4.



1 tick mark = 1Mb

Figure S5: Chip signal with error bars denoting standard error of 4 replicates (3 replicates for Old Males)

Figure S5.

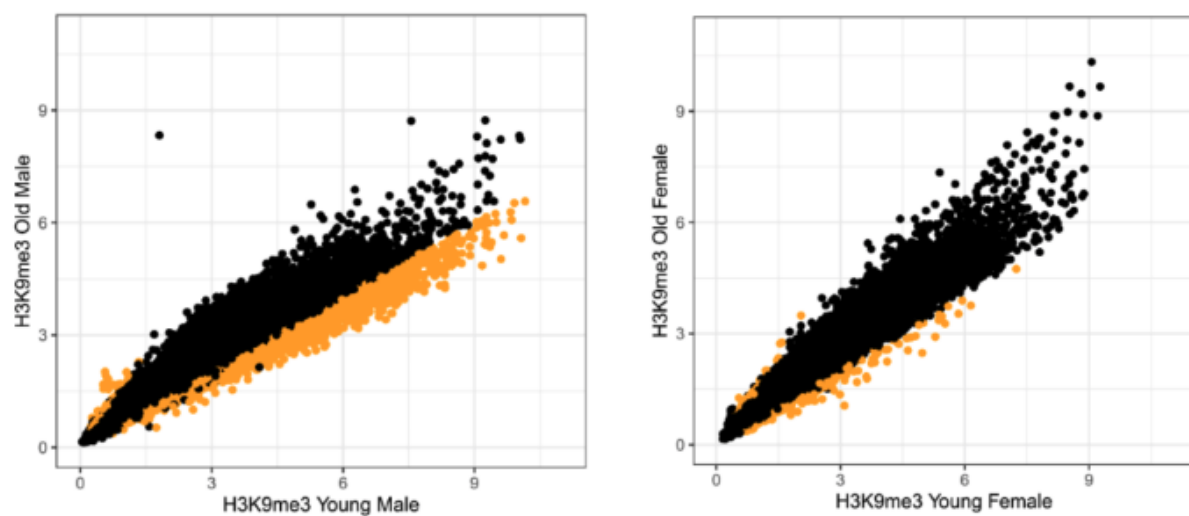


Figure S6.

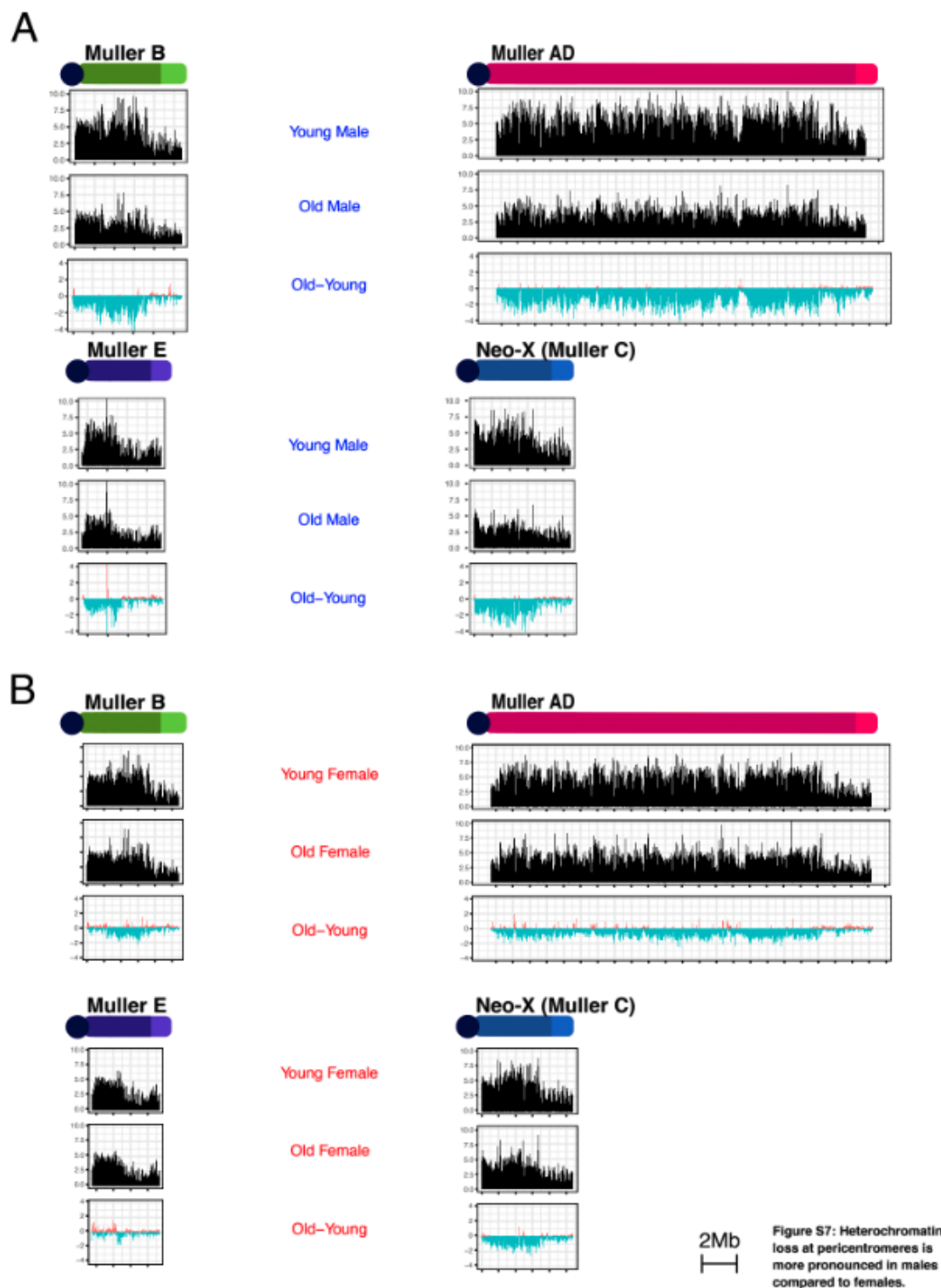


Figure S7.

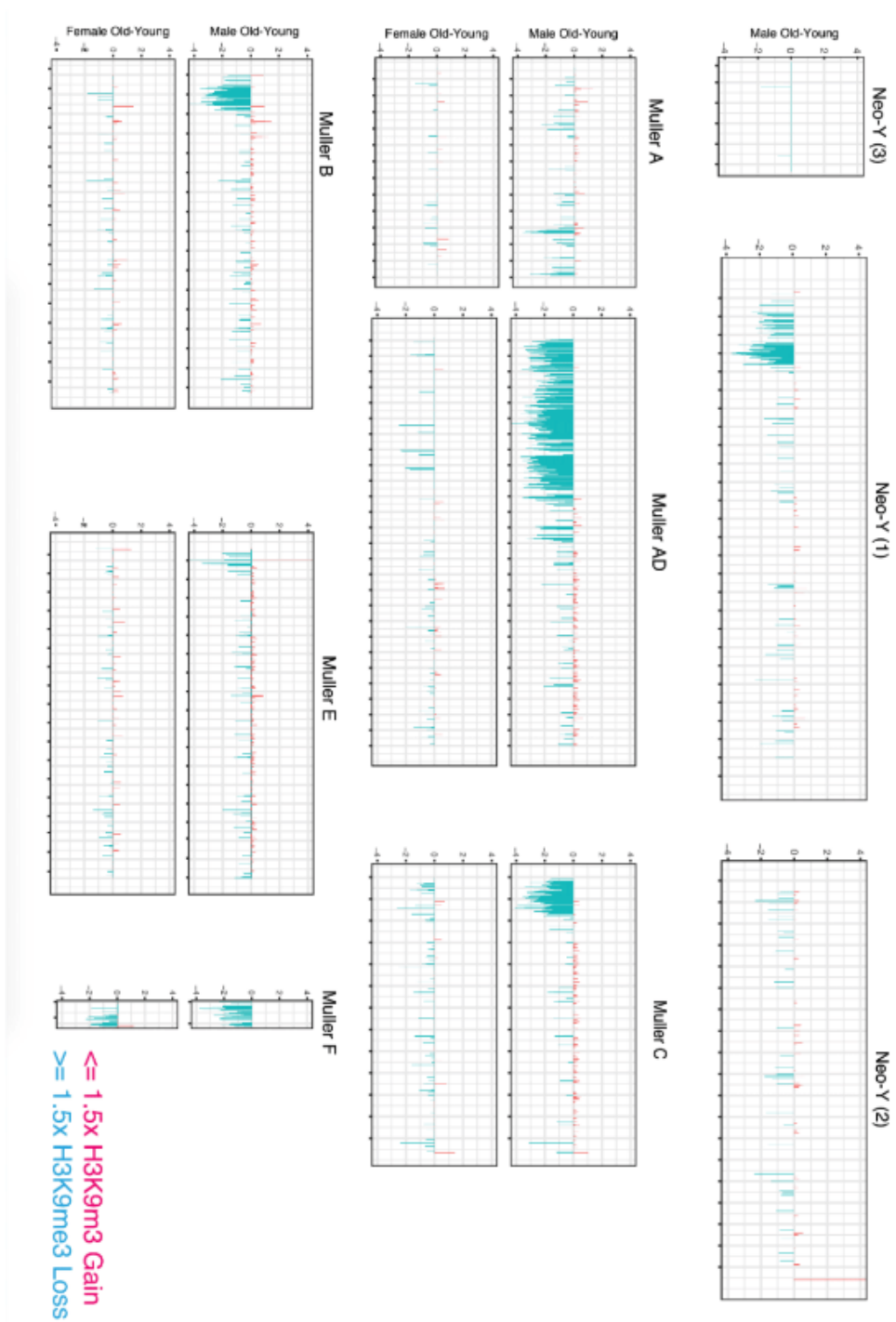


Figure S8.

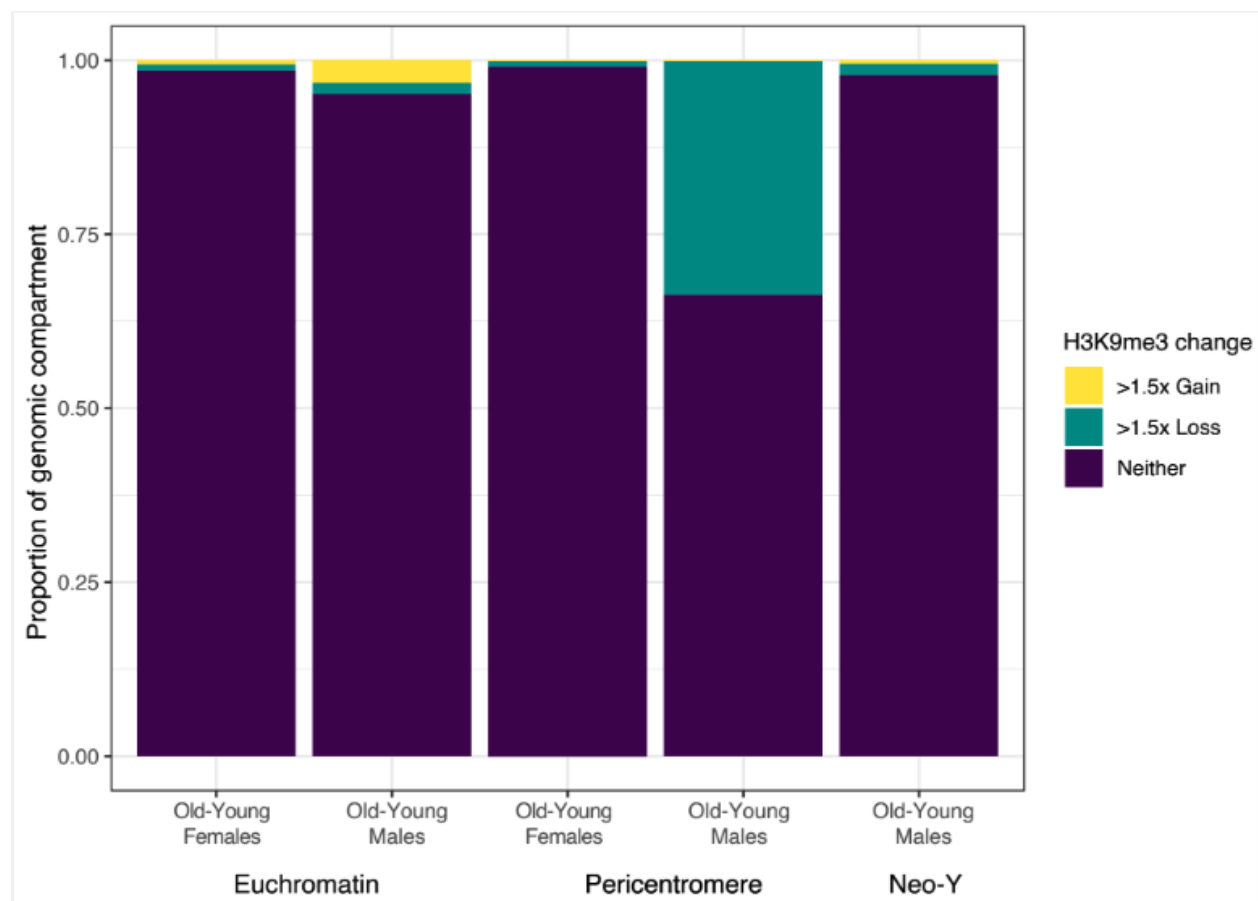


Figure S9.

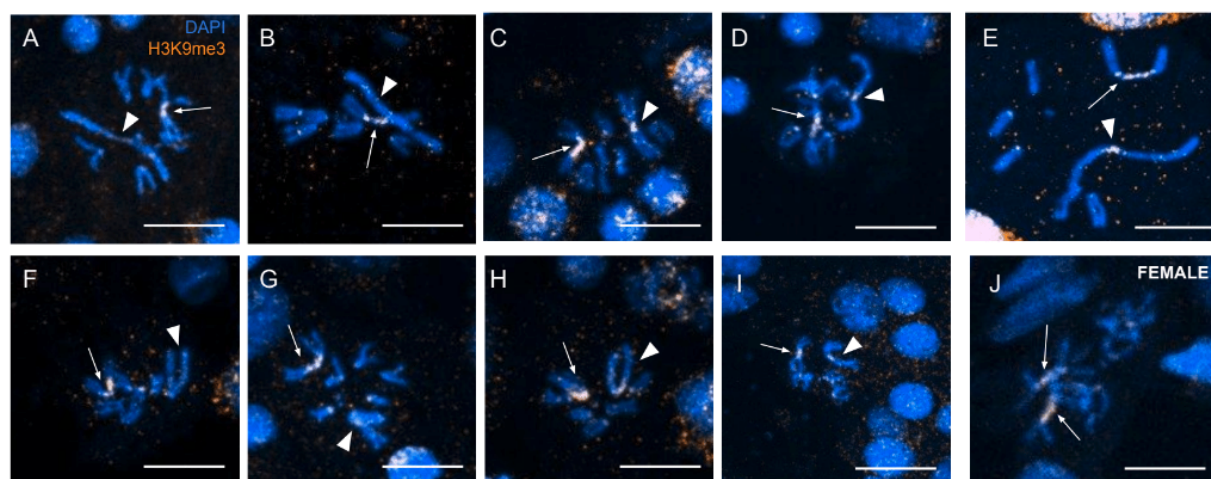


Figure S10.

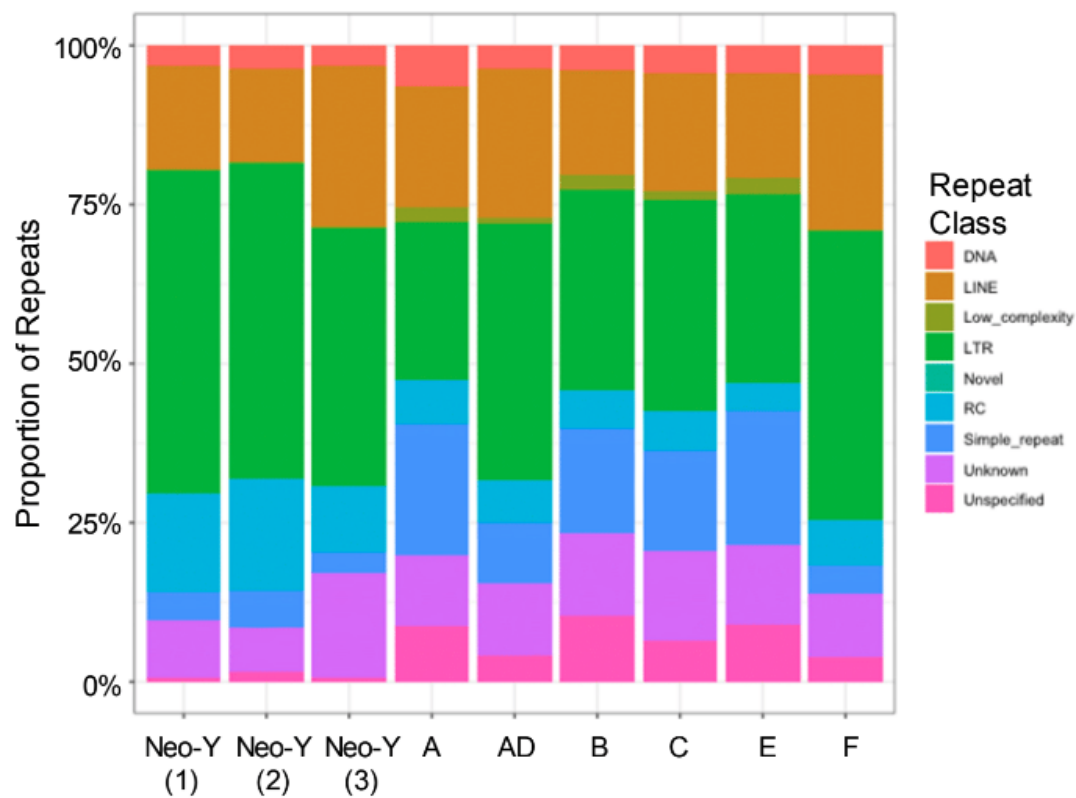


Figure S11.

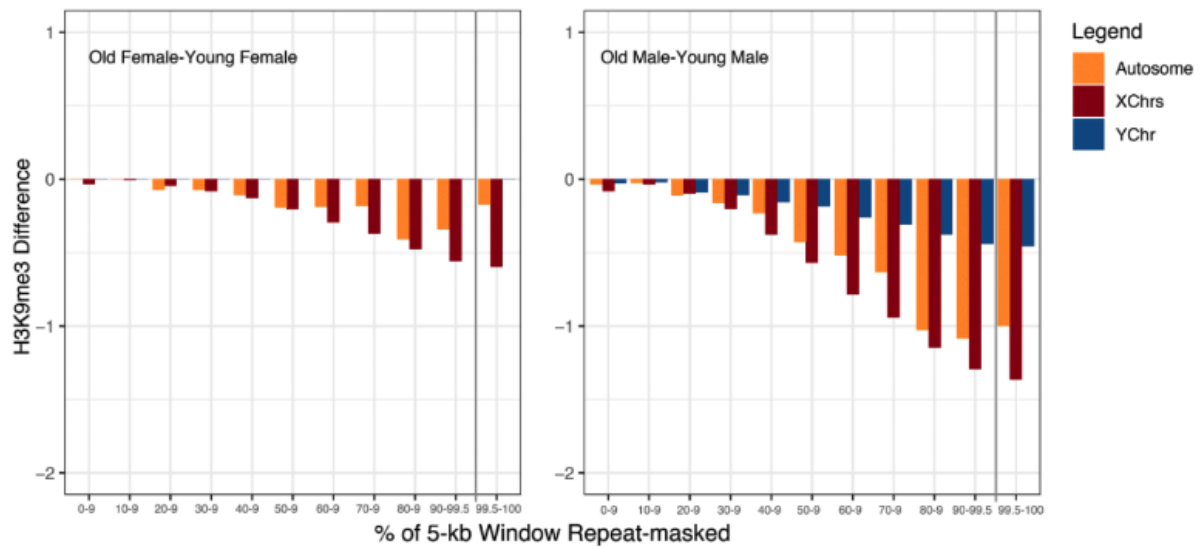


Figure S12.

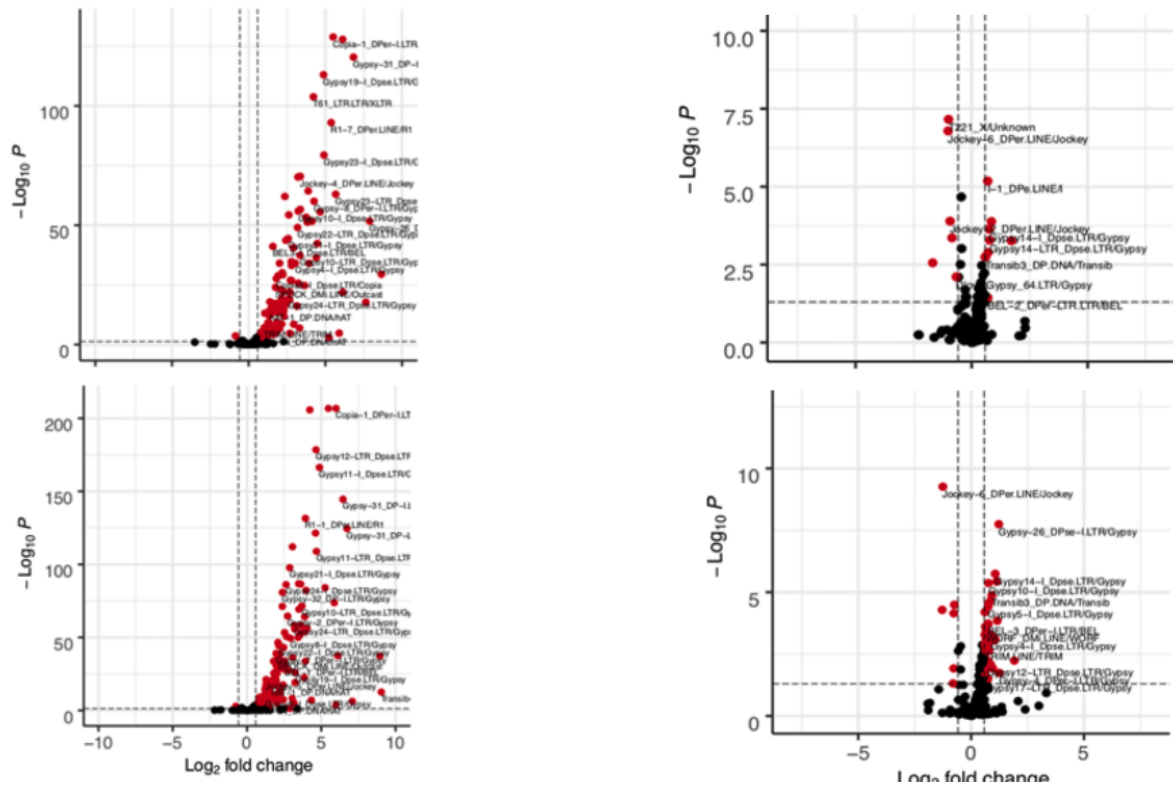


Figure S13.

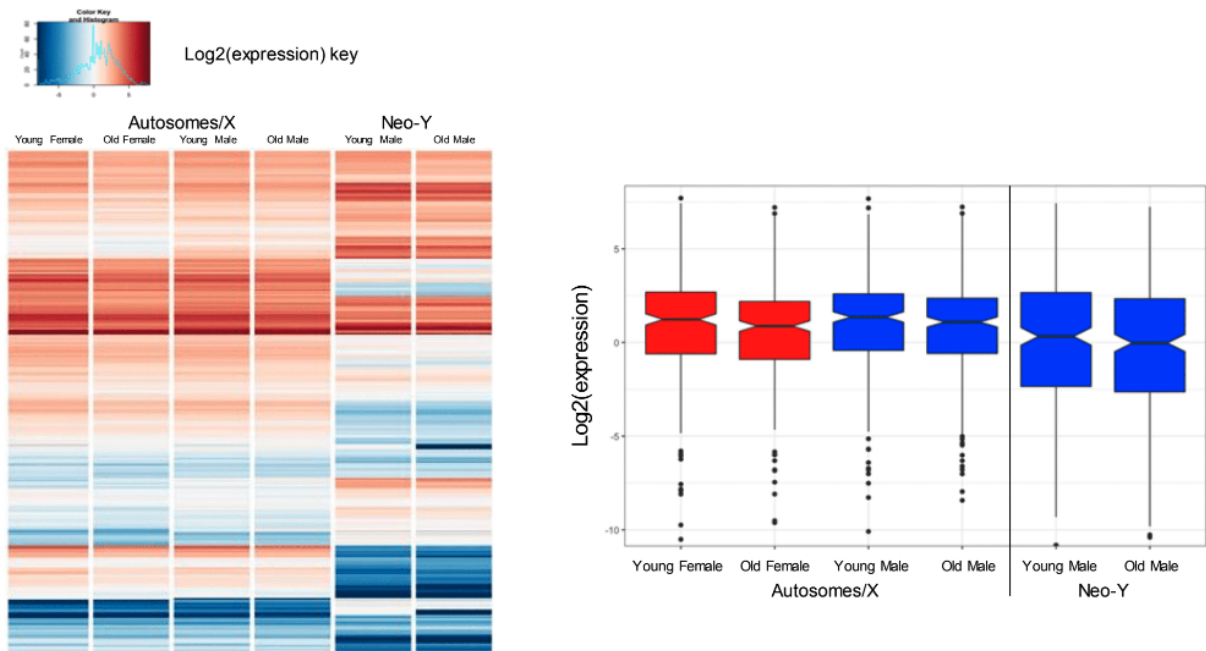


Figure S14.

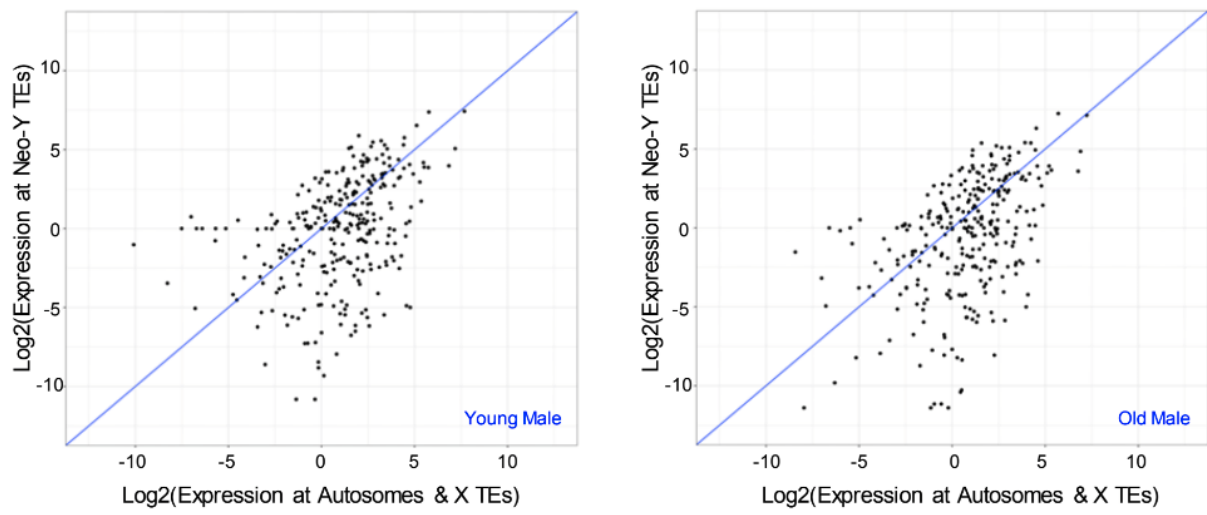


Figure S15.

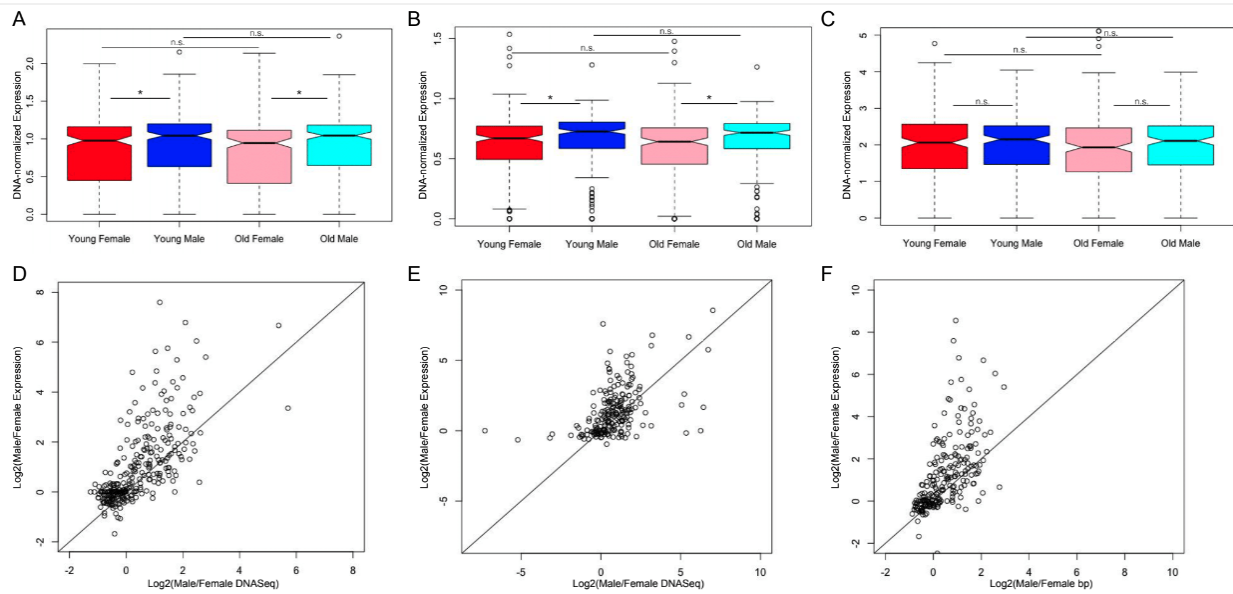


Figure S16.

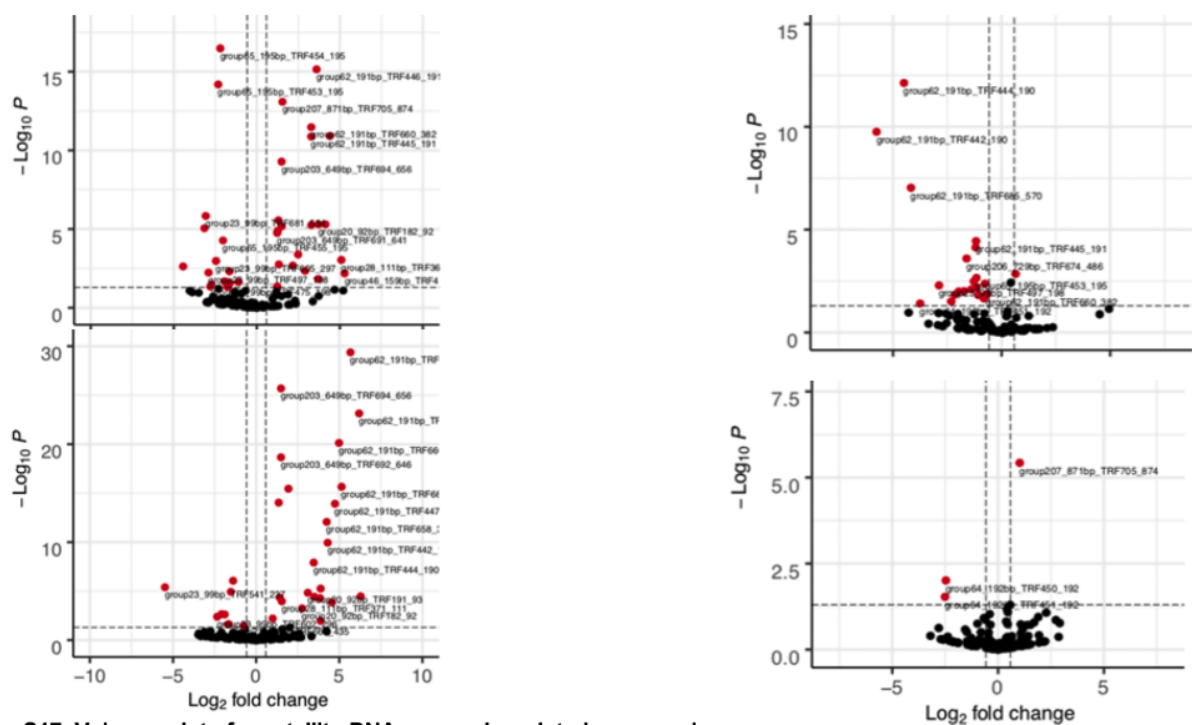


Figure S17.

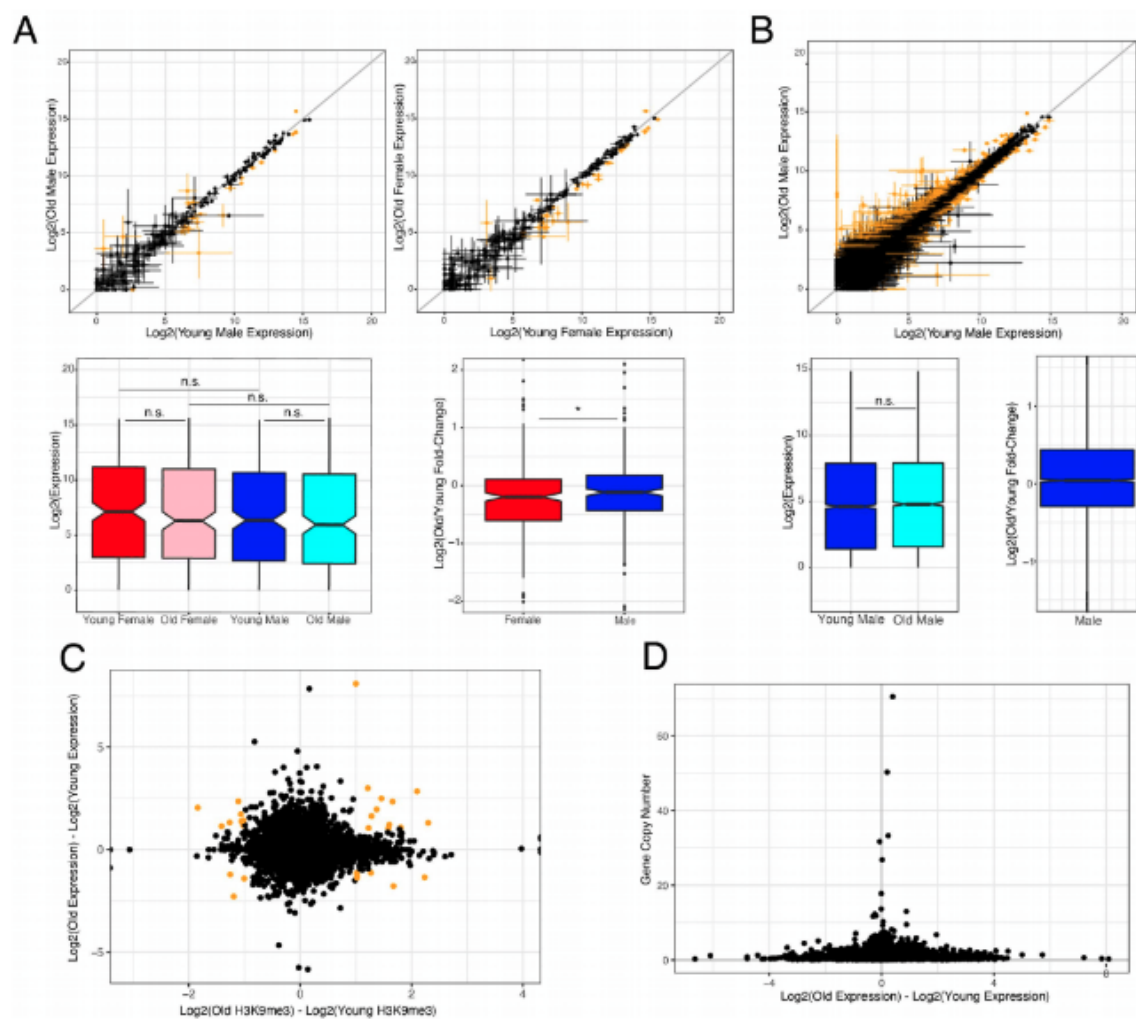


Figure S18.

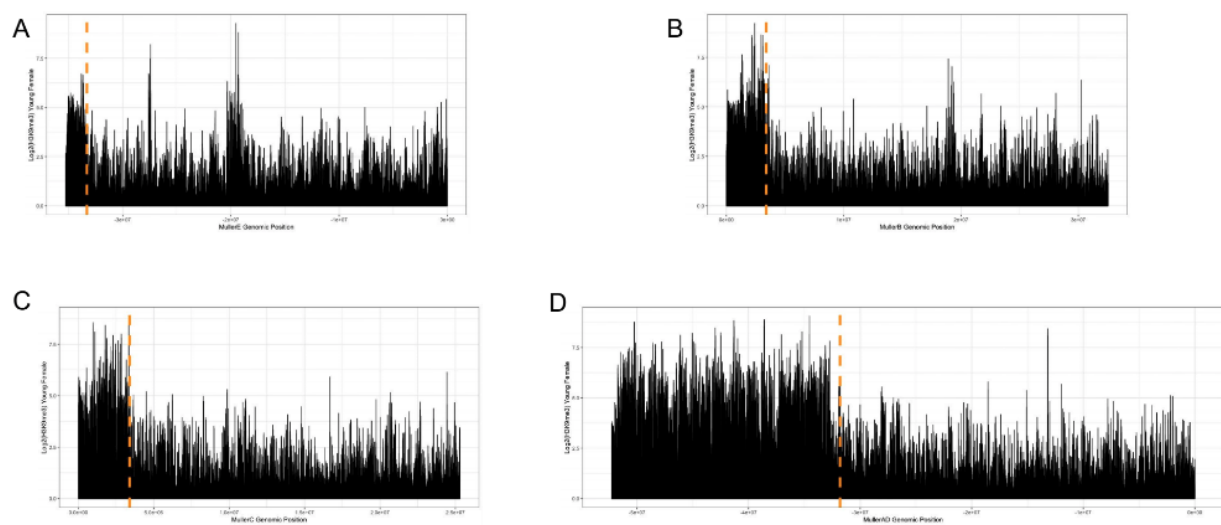


Figure S19.

S1 Table. Overview of ChIP data generated. Description of samples/tissues used for 4 rounds/batches of ChIP-seq experiments.

Sex	Age	Batch	Data type	Antibody	reads	Spike-in	Tissue	Figure	SRA Accession
Female	5 day	1	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192013
Female	5 day	2	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192012
Female	9 day	3	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192045
Female	9 day	4	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192034
Female	98 day	1	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192015
Female	98 day	2	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192014
Female	80 day	3	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192011
Female	80 day	4	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192010
Female	5 day	1	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192023
Female	5 day	2	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192018
Female	9 day	3	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192017
Female	9 day	4	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192016
Female	98 day	1	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192053
Female	98 day	2	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192052
Female	80 day	3	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192051
Female	80 day	4	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192050
Male	5 day	1	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192049
Male	5 day	2	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192048
Male	9 day	3	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192047
Male	9 day	4	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192046
Male	98 day	1	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192040
Male	98 day	2	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192039
Male	80 day	3	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192038
Male	82 day	4	ChIP-Seq	H3K9me3	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192037
Male	5 day	1	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192044
Male	5 day	2	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192043
Male	9 day	3	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192042
Male	9 day	4	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192041
Male	98 day	1	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192036
Male	98 day	2	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192035
Male	80 day	3	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192033
Male	82 day	4	ChIP-Seq	Input	100 bp PE	<i>D. melanogaster</i>	brain	Fig. 2-3, 5	SRR12192032

S2 Table. Mapping statistics of ChIP data. Description of mapping statistics for ChIP-experiment data for each batch. Reported numbers of reads mapped to the sample and spike-in genomes.

Sample	Total reads	Reads map to	Reads map to	Total INPUT	Input reads	Input reads map
			<i>D.</i>	reads	map	Input reads map
	<i>post-trimming</i>	<i>D. miranda</i>	<i>melanogaster</i>	<i>post-trimming</i>	<i>D. miranda</i>	<i>D. melanogaster</i>
Young Female 1	67,704,582	56,928,414	10,662,194	44,807,108	35,615,280	9,159,754
Young Female 2	61,703,392	51,974,874	10,001,605	45,237,472	34,809,823	10,576,212
Young Female 3	48,851,678	39,751,715	8,736,930	42,824,410	28,970,764	12,752,833
Young Female 4	51,493,160	44,725,176	6,770,327	44,628,860	35,458,976	9,229,236
Old Female 1	64,473,240	44,239,808	20,178,734	38,856,540	26,691,510	12,071,580
Old Female 2	56,463,368	33,046,791	23,559,347	46,816,274	22,722,284	24,068,004
Old Female 3	47,613,398	39,574,264	7,780,996	39,235,422	30,364,995	8,806,333
Old Female 4	58,142,168	50,838,062	7,422,156	49,946,734	42,060,051	7,913,083
Young Male 1	57,794,342	50,934,715	6,381,307	47,569,564	39,831,886	7,611,881
Young Male 2	75,280,770	67,181,375	8,044,335	51,955,562	43,878,867	8,176,091
Young Male 3	61,947,110	54,787,100	6,446,831	54,470,572	47,203,948	6,690,023
Young Male 4	54,218,228	48,982,633	4,725,311	47,507,242	40,741,207	6,741,949
Old Male 1	60,294,580	39,653,383	20,288,116	50,722,396	28,781,010	21,744,490
Old Male 2	53,493,186	41,908,320	11,603,888	45,531,982	34,772,933	10,843,562
Old Male 3	63,904,722	54,639,077	8,361,739	39,046,054	30,543,313	8,409,519
Old Male 4	83,095,444	73,518,270	8,853,517	101,375,398	83,614,989	17,641,675

S3 Table. Pearson coefficients between replicates of the same biological sample. We used H3K9me3 enrichment in 5kb windows from the autosomes and X chromosomes to conduct Pearson correlations between each pair of replicates. Pearson correlations were calculated using the base R function `cor.test(replicateA, replicateB, method="pearson")`.

		Replicate 1	Replicate 2	Replicate 3	Replicate 4
Young female	Replicate 1		0.9806	0.9709	0.9716
	Replicate 2	0.9806		0.9523	0.9554
	Replicate 3	0.9709	0.9523		0.9897
	Replicate 4	0.9716	0.9554	0.9897	
Old female	Replicate 1		0.9303	0.9654	0.9799
	Replicate 2	0.9303		0.9209	0.9313
	Replicate 3	0.9654	0.9209		0.9831
	Replicate 4	0.9799	0.9313	0.9831	
Young male	Replicate 1		0.9838	0.9614	0.9735
	Replicate 2	0.9838		0.9601	0.9711
	Replicate 3	0.9614	0.9601		0.9748
	Replicate 4	0.9735	0.9711	0.9748	
Old male	Replicate 1		0.8927	0.8890	0.8927
	Replicate 2	0.8927		0.9280	0.9347
	Replicate 3	0.8890	0.9280		0.9830
	Replicate 4	0.8927	0.9347	0.9830	

S4 Table. Proportion of 5kb-windows in top 20 percentile H3K9me3 enrichment shared between normalization methods. For each normalization method, we identified the top 20% most enriched 5kb windows in euchromatin (N = 4646). We then identified how many 5kb regions were shared between each normalization method and reported those results as a percentage (no. windows shared between two methods / no. windows in top 20% enrichment).

Normalization method	No Spike-In	Brown (2020)	Bonhoure (2014)	Wei (2020)
No Spike-In		99.98%	99.53%	99.14%
Brown (2020)	99.98%		99.53%	99.16%
Bonhoure (2014)	99.53%	99.53%		99.61%
Wei (2020)	99.14%	99.16%	99.61%	

S5 Table. Mapping statistics of ChIP data for MAPQ > 3 alignments.

Sample	Total reads	Reads map to		Total INPUT reads	Input reads map	
	<i>post-trimming</i>	<i>D. miranda</i>	<i>D. melanogaster</i>	<i>post-trimming</i>	<i>D. miranda</i>	<i>D. melanogaster</i>
Young Female 1	67,704,582	40,487,182	6,632,869	44,807,108	29,438,452	7,071,855
Young Female 2	61,703,392	36,914,280	6,312,139	45,237,472	28,552,858	8,250,055
Young Female 3	48,851,678	28,773,644	5,897,662	42,824,410	24,090,157	9,866,222
Young Female 4	51,493,160	31,849,850	4,396,533	44,628,860	29,431,898	6,973,192
Old Female 1	64,473,240	31,467,019	13,417,446	38,856,540	22,071,447	9,532,316
Old Female 2	56,463,368	24,217,457	15,790,171	46,816,274	22,071,447	19,356,315
Old Female 3	47,613,398	28,996,177	5,153,463	39,235,422	25,370,450	6,730,047
Old Female 4	58,142,168	36,984,324	4,719,688	49,946,734	35,064,663	5,898,510
Young Male 1	57,794,342	32,145,574	3,976,855	47,569,564	28,468,786	5,923,154
Young Male 2	75,280,770	42,501,933	5,085,807	51,955,562	31,322,018	6,391,210
Young Male 3	61,947,110	35,698,519	4,251,804	54,470,572	35,083,391	5,080,061
Young Male 4	54,218,228	31,554,397	3,042,336	47,507,242	29,534,573	5,160,270
Old Male 1	60,294,580	27,642,556	13,548,267	50,722,396	20,775,505	17,466,283
Old Male 2	53,493,186	27,186,663	7,733,141	45,531,982	24,972,759	8,654,831
Old Male 3	63,904,722	36,296,614	5,494,746	39,046,054	22,270,233	6,465,585
Old Male 4	83,095,444	48,664,557	5,924,032	101,375,398	60,903,326	13,563,320

S6 Table. Heterochromatin-related genes and their expression change in old males and females.

FlybaseID	Gene Name	GO Biological Process	Regulation	Log2(fold-change)	Adj. P-val	Regulation	Log2(fold-change)	Adj. P-val
FBgn0246261	Atf-2	nuclear heterochromatin	up	0.540	0.014	up	0.098	0.821
FBgn0074791	Trf2	nuclear heterochromatin	down	0.250	0.154	up	0.007	0.989
FBgn0074758	MED26	nuclear heterochromatin	up	0.102	0.904	up	0.614	0.721
FBgn0080711	MED17	nuclear heterochromatin	up	0.195	0.375	up	0.080	0.840
FBgn0249854	HP4	nuclear heterochromatin	down	0.084	0.859	up	0.130	0.818
FBgn0247702	moon	nuclear heterochromatin	up	0.602	0.383	up	0.897	NA
FBgn0244273	Hmt4-20	nuclear heterochromatin	down	0.235	0.101	down	0.118	0.775
FBgn0261712	SuUR	nuclear heterochromatin	up	0.119	0.609	up	0.169	0.749
FBgn0078743	Sirt1	nuclear heterochromatin	down	0.169	0.348	up	0.022	0.946
FBgn0078701	TfIIA-S	nuclear heterochromatin	up	0.344	0.113	up	0.514	0.114
FBgn0081044	HP1A	nuclear heterochromatin	up	0.199	0.499	up	0.091	0.842
FBgn0071193	Arp6	heterochromatin organization	down	0.584	0.003	down	0.436	0.087
FBgn0078014	vig	heterochromatin organization	down	0.375	0.107	down	0.121	0.848
FBgn0071285	vig2	heterochromatin organization	down	0.018	0.960	up	0.175	0.699
FBgn0265021	egg	heterochromatin organization	up	0.124	0.573	down	0.036	0.939
FBgn0074378	Su(var)3-3	heterochromatin organization	up	0.102	0.652	down	0.072	0.847
FBgn0074067	hop	heterochromatin assembly	up	0.109	0.641	up	0.037	0.931
FBgn0079367	piwi	heterochromatin assembly	down	0.658	0.463	up	0.320	0.739
FBgn0247875	Stat92E	heterochromatin assembly	up	0.213	0.098	up	0.356	0.113
FBgn0080534	Mekk1	heterochromatin assembly	up	0.160	0.380	up	0.026	0.943
FBgn0078103	Bap60	heterochromatin assembly	up	0.103	0.723	up	0.028	0.943
FBgn0246408	wde	heterochromatin assembly	up	0.247	0.187	down	0.085	0.802
FBgn0075184	Acf	heterochromatin assembly	up	0.008	0.978	down	0.114	0.682
FBgn0248173	htt	heterochromatin assembly	down	0.058	0.818	down	0.166	0.516
FBgn0078708	Fmr1	heterochromatin assembly	up	0.237	0.085	up	0.186	0.521
FBgn0080049	HP1b	heterochromatin assembly	down	0.139	0.563	down	0.117	0.804
FBgn0245751	HPS4	heterochromatin assembly	up	0.402	0.088	up	0.293	0.331
FBgn0080124	CG7137	heterochromatin assembly	up	0.138	0.705	down	0.097	0.822
FBgn0081790	FBXO11	heterochromatin assembly	down	0.050	0.852	down	0.053	0.904

FBgn0079435	CG6204	heterochromatin assembly down	0.162	0.440	down	0.109	0.758
FBgn0080459	Naa40	heterochromatin assembly down	0.048	0.893	up	0.167	0.683
FBgn0077229	Charc-14	heterochromatin assembly up	0.240	0.418	up	0.010	0.985
FBgn0244102	CoRest	heterochromatin assembly up	0.140	0.495	up	0.034	0.917
FBgn0070268	gpp	heterochromatin assembly down	0.114	0.518	up	0.171	0.756

S7 Table. Overview of RNA data generated. Description of samples/tissues used for 3 rounds/batches of RNA-seq experiments.

Sex	Age	Batch	Data type	Tissue	reads	Figure	reads generated	SRA Accession
Female	9 day	1	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	45,436,492	SRR12192031
Female	9 day	2	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	149,817,434	SRR12192030
Female	9 day	3	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	179,126,850	SRR12192029
Female	80 day	1	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	60,918,286	SRR12192028
Female	80 day	2	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	150,266,062	SRR12192027
Female	80 day	3	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	168,272,836	SRR12192026
Male	9 day	1	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	54,347,756	SRR12192025
Male	9 day	2	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	136,139,696	SRR12192024
Male	9 day	3	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	144,374,744	SRR12192022
Male	80 day	1	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	45,924,434	SRR12192021
Male	80 day	2	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	152,547,756	SRR12192020
Male	80 day	3	RNA-Seq	pooled brains	100 bp PE	Fig. 5-6	159,981,252	SRR12192019

S8 Table. Mapping and counts statistics of RNA data. Description of mapping statistics for RNA-seq experiment data for each batch. Reported gene and repeat counts by Subread featurecounts and Tetrascripts where applicable.

Sample	Total reads	Reads map to <i>post-rRNA</i> <i>removal</i> <i>D. miranda</i>	Gene count totals by		Transposable element count totals by				Satellite DNA count totals by	
			<i>Subread</i> featurecounts	<i>Tetrascript</i> ts	<i>Subread</i> featurecounts	%	<i>Tetrascripts</i>	%	<i>Subread</i> featurecounts	%
YF1	33,396,824	30,496,555	14,091,845	7,225,056	473,202	1.6%	400,725	1.3%	28,386	0.09%
YF2	127,413,478	119,301,094	56,069,961	28,333,230	2,523,245	2.1%	1,786,219	1.5%	137,691	0.12%
YF3	161,923,358	151,229,688	70,388,358	35,588,522	2,689,933	1.8%	2,018,726	1.3%	172,596	0.11%
OF1	21,683,064	18,353,375	8,635,458	4,454,130	236,941	1.3%	201,467	1.1%	15,466	0.08%
OF2	132,515,030	124,508,964	59,588,781	29,818,464	2,103,645	1.7%	1,531,115	1.2%	143,510	0.12%
OF3	121,338,760	99,588,981	47,480,007	23,937,653	1,223,690	1.2%	967,402	1.0%	98,327	0.10%
YM1	33,184,830	30,019,808	13,223,662	6,768,926	1,070,610	3.6%	597,786	2.0%	37,394	0.12%
YM2	125,532,740	115,020,312	51,002,631	24,791,212	5,271,178	4.6%	2,611,345	2.3%	198,342	0.17%
YM3	132,522,320	120,583,108	55,586,184	27,626,956	4,356,357	3.6%	2,234,565	1.9%	186,127	0.15%
OM1	25,670,500	22,481,303	10,001,558	6,768,926	871,614	3.9%	468,626	2.1%	28,016	0.12%
OM2	132,662,164	123,016,943	55,405,264	27,935,265	5,066,965	4.1%	2,488,609	2.0%	177,217	0.14%
OM3	141,307,748	131,079,565	61,138,923	30,732,062	4,270,558	3.3%	2,149,597	1.6%	178,617	0.14%

S9 Table. Mapping and counts statistics of RNA data filtered for MAPQ > 40

Sample	Reads map to <i>D. miranda</i>	Gene count totals by		Transposable element count totals by				Satellite DNA count totals by
		<i>Subread</i> <i>featurecounts</i>	<i>TEtranscripts</i>	<i>Subread</i> <i>featurecounts</i>	%	<i>TEtranscripts</i>	%	<i>Subread featurecounts</i>
YF1	24,470,598	14,091,845	7,069,630	473,368	1.9%	297,510	1.2%	28,386
YF2	95,902,373	56,069,961	27,745,752	2,523,245	2.6%	1,303,256	1.4%	137,691
YF3	123,024,250	70,388,358	34,892,990	2,689,933	2.2%	1,542,943	1.3%	172,596
OF1	14,148,032	8,635,458	4,361,564	237,119	1.7%	147,776	1.0%	15,466
OF2	97,593,658	59,588,781	29,207,255	2,103,645	2.2%	1,146,703	1.2%	143,510
OF3	76,595,230	47,480,007	23,467,539	1,223,690	1.6%	741,431	1.0%	98,327
YM1	23,401,050	13,223,662	6,583,993	1,070,996	4.6%	403,013	1.7%	37,394
YM2	92,043,043	51,002,631	24,071,978	5,271,178	5.7%	1,790,176	1.9%	198,342
YM3	96,286,901	55,586,184	26,862,946	4,356,357	4.5%	1,548,567	1.6%	186,127
OM1	16,957,126	10,001,558	4,980,605	872,043	5.1%	317,498	1.9%	28,016
OM2	93,999,119	55,405,264	27,124,819	5,066,965	5.4%	1,670,406	1.8%	177,217
OM3	99,173,506	61,138,923	29,883,477	4,270,558	4.3%	1,468,000	1.5%	178,617

S10 Table. DNAPipeTE estimates of repeats in RNASeq data

	% TE	Average	% All repeats	Average
Young Female 2	28.1%	21.5%	46.0%	38.3%
Young Female 3	19.9%		36.1%	
Young Female 1	16.6%		32.80%	
Old Female 2	20.8%	22.4%	37.60%	36.5%
Old Female 3	29.9%		41.50%	
Old Female 1	16.5%		30.40%	
Young Male 2	33.4%	23.7%	49.70%	40.3%
Young Male 3	18.9%		35.60%	
Young Male 1	18.7%		35.50%	
Old Male 2	19.3%	23.2%	38.10%	39.2%
Old Male 3	30.7%		44.60%	
Old Male 1	19.5%		34.80%	

Chapter 1: Supplementary Material 1. *D. melanogaster* as a spike-in to normalize *D. miranda* ChIP's

The use of a spike-in allows normalization of individual ChIP experiments, and enables us to compare levels of H3K9me3 enrichment across sexes and time. We chose *D. melanogaster* to serve as a spike-in for *D. miranda*, and below, we demonstrate that this species pair is sufficiently diverged from each other that there is very little ambiguity in the assignment of reads to the correct species.

Table A1 shows how many of the reads that were originally mapped to *D. miranda* also mapped to *D. melanogaster*. Specifically, we took reads that initially mapped to *D. miranda* and mapped them to the *D. miranda* and the *D. melanogaster* genomes simultaneously to estimate the cross-mapping rate (same Bowtie2 parameters as initial mappings). We find that, on average, only about 0.6% of the reads that initially aligned to *D. miranda* will cross-map to *D. melanogaster*.

We similarly mapped genomic DNA from *D. miranda* to only the *D. melanogaster* genome using the same parameters as described previously. We found that of 24,375,599 male reads mapped to *D. miranda*, 275,936 (0.1%) then mapped to *D. melanogaster* and of the 15,563,664 female reads mapped to *D. miranda*, 151,165 (1%) mapped to *D. melanogaster*. These cross-mapping estimates are similar to those in the ChIP-data, suggesting that any reads mapped to *D. miranda* are likely derived from *D. miranda*.

Table A2 shows cross-mapping of reads derived from *D. miranda* heterochromatin. We took reads originally mapped to *D. miranda* heterochromatin and mapped them simultaneously to both *D. miranda* and *D. melanogaster* genomes to infer cross-mapping between the species' repeats (same Bowtie2 parameters). We find that the rate is almost always less than 1% (especially when omitting Old Male Replicate 1) and this suggests low cross-mapping between species-specific repeats. Lower cross-mapping from heterochromatic reads between *D. miranda* and *D. melanogaster* presumably reflects the rapid turn-over of repetitive DNA between species (that is, repeats are typically less conserved between species than unique DNA).

Table A1. Cross-mapping statistics of *D. miranda* ChIP-Seq reads to *D. melanogaster*

Sample	Reads map to <i>D. miranda</i>	Cross-mapping to <i>D. melanogaster</i>	Effective Rate	Input reads map <i>D. miranda</i>	Cross-mapping to <i>D. melanogaster</i>	Effective Rate
YF1	56,928,414	412,013	0.7%	35,615,280	276,386	0.8%
YF2	51,974,874	391,147	0.8%	34,809,823	280,416	0.8%
YF3	39,751,715	284,123	0.7%	28,970,764	249,567	0.9%
YF4	44,725,176	309,167	0.7%	35,458,976	271,407	0.8%
OF1	44,239,808	539,401	1.2%	26,691,510	264,336	1.0%
OF2	33,046,791	533,892	1.6%	22,722,284	355,757	1.6%
OF3	39,574,264	301,770	0.8%	30,364,995	233,662	0.8%
OF4	50,838,062	381,440	0.8%	42,060,051	292,726	0.7%
YM1	50,934,715	128,219	0.3%	39,831,886	114,441	0.3%
YM2	67,181,375	164,696	0.2%	43,878,867	123,689	0.3%
YM3	54,787,100	116,963	0.2%	47,203,948	99,499	0.2%
YM4	48,982,633	101,850	0.2%	40,741,207	99,723	0.2%
OM1	39,653,383	294,845	0.7%	28,781,010	230,548	0.8%
OM2	41,908,320	187,926	0.4%	34,772,933	138,877	0.4%
OM3	54,639,077	135,974	0.2%	30,543,313	100,235	0.3%
OM4	73,518,270	160,215	0.2%	83,614,989	228,989	0.3%

Table A2. Cross-mapping statistics of *D. miranda* ChIP-Seq reads from heterochromatin

Sample	Sample reads	Reads <u>re-map</u> to <i>D. miranda</i>	Cross-mapping to <i>D.</i> <i>melanogaster</i>	Effective Rate	Input reads	Input reads <u>re-map</u> to <i>D. miranda</i>	Cross-mapping to <i>D.</i> <i>melanogaster</i>	Effective Rate
YF1	20,311,258	20,301,290	8,617	0.0%	6,318,194	6,313,904	3,935	0.1%
YF2	17,182,960	17,173,800	8,223	0.0%	5,871,858	5,866,876	4,527	0.1%
YF3	14,143,700	14,138,560	2,903	0.0%	5,164,536	5,159,248	4,771	0.1%
YF4	16,697,004	16,692,800	2,515	0.0%	6,433,086	6,429,049	3,617	0.1%
OF1	15,233,926	15,216,770	16,238	0.1%	4,893,508	4,888,142	5,115	0.1%
OF2	9,833,772	9,814,370	18,684	0.2%	3,881,778	3,871,189	10,218	0.3%
OF3	12,951,250	12,946,530	2,667	0.0%	5,381,206	5,377,031	3,686	0.1%
OF4	16,713,484	16,708,360	3,492	0.0%	7,391,458	7,387,815	3,168	0.0%
YM1	28,196,674	28,084,860	106,150	0.4%	17,060,460	16,971,160	87,009	0.5%
YM2	36,895,602	36,754,400	135,483	0.4%	18,646,480	18,550,720	93,402	0.5%
YM3	29,207,016	29,096,950	99,585	0.3%	18,285,618	18,201,040	78,574	0.4%
YM4	26,575,570	26,479,620	87,559	0.3%	17,185,284	17,102,750	79,363	0.5%
OM1	18,448,058	18,219,970	224,821	1.2%	11,966,652	11,811,810	153,172	1.3%
OM2	21,710,342	21,561,380	145,969	0.7%	14,545,924	14,445,320	98,874	0.7%
OM3	27,623,332	27,498,560	113,585	0.4%	12,708,820	12,629,980	76,299	0.6%
OM4	36,853,832	36,710,190	133,561	0.4%	34,771,568	34,590,330	176,644	0.5%

Chapter 1: Supplementary Material 2. Cross-mapping between the *D. miranda* neo-X and neo-Y chromosome

The neo-Y is recently diverged from the neo-X and thus may potentially produce mis-mapping issues. Several lines of evidence suggest that mis-mapping between the neo-sex chromosomes is of limited concern. **Table B1** shows how many reads that originally mapped to the neo-Y re-map to either the neo-X or the neo-Y by mapping to both contigs simultaneously with Bowtie2. We find that roughly 96% of reads originally mapped to the neo-Y will map back to the neo-Y while only 1% will map to the neo-X. These data suggest that cross-mapping between the two chromosomes is very low since the vast majority of reads will align back to the original chromosome.

We also simulated 100-bp paired-end reads from the neo-X and the neo-Y using ART and mapped both of these simulated read sets to the neo-X and the neo-Y (Huang et al. 2012). The following parameters were used: -ss HS20 -l 100 -f 10 -p -m 375 -s 25. **Table B2** shows that overall, the vast majority of reads maps back to their respective chromosome and even more so when we apply an alignment quality cutoff. In particular, over 99% of simulated neo-Y reads map back to the neo-Y (and 99.9% of reads if using a MAPQ>3).

We further assessed mis-mapping from the neo-X to the neo-Y by mapping female genomic DNA reads from *D. miranda*. We mapped female gDNA reads [parameters] and mapped them to the female genome; we then took all reads mapping to the neo-X and re-mapped them to both the neo-X and the neo-Y. Of the original 2,042,127 neo-X alignments, 2,020,692 (98.9%) were aligned properly to the neo-X and the rest (1.1%) aligned to the neo-Y. Thus, similar to our results of mapping neo-Y reads to the neo-X, the vast majority of reads will be reassigned to their initial chromosome.

Table B1. Cross-mapping statistics of neo-Y vs neo-X reads

Sam ple	ChIP reads mapped <i>to neo-Y</i>	Mapped back <i>to neo-Y</i>	Cross- mapped <i>to neo-X</i>	Cross- mappin g rate	Input reads mapped <i>neo-Y</i>	Mapped back <i>to neo-Y</i>	Cross- mappe d <i>to neo-X</i>	Cross- mappin g rate
YM1	19,693,00 4	19,005,75 5	263,436	1.4%	14,449,50 0	14,101,12 2	164,63 4	1.2%
YM2	25,598,61 4	24,865,38 4	358,057	1.4%	15,687,48 6	15,304,48 9	188,37 9	1.2%
YM3	21,532,46 7	20,417,13 1	275,963	1.3%	15,756,92 9	15,008,15 3	173,94 0	1.1%
YM4	19,203,51 8	18,298,98 9	252,029	1.4%	14,585,24 6	14,177,02 6	161,83 8	1.1%
OM1	15,055,68 5	14,625,25 6	167,524	1.1%	10,219,49 6	9,964,189	112,76 5	1.1%
OM2	16,286,66 8	15,868,11 6	214,098	1.3%	12,230,16 7	11,940,83 6	147,79 8	1.2%
OM3	21,781,73 6	20,563,36 5	266,817	1.3%	10,803,14 2	10,481,63 5	120,29 3	1.1%
OM4	29,375,20 0	27,726,41 9	362,116	1.3%	29,685,14 8	28,798,99 5	331,41 2	1.1%

Table B2. Cross-mapping statistics between neo-X and neo-Y using simulated reads.

	All mappings				Mappings above MAPQ > 3			
	Total reads	Map to neo-X	Map to neo-Y	Cross-mapping %	Total reads	Map to neo-X	Map to neo-Y	Cross-mapping %
Neo-X	2,530,344	2,443,493	86,851	3.432%	2,247,137	2,235,727	11,410	0.51%
Neo-Y	11,041,076	89,021	10,952,055	0.806%	7,535,923	10,685	7,525,238	0.14%

Chapter 2: Transposable element accumulation drives size differences among polymorphic Y chromosomes in *Drosophila*

Originally published with some modifications as Nguyen, Alison H., Weixiang Wang, Emily Chong, Kamalakar Chatla, and Doris Bachtrog. “Transposable Element Accumulation Drives Size Differences among Polymorphic Y Chromosomes in *Drosophila*.” *bioRxiv*, June 3, 2021, 2021.06.02.446622. <https://doi.org/10.1101/2021.06.02.446622>.

Abstract

Y chromosomes of many species are gene poor and show low levels of nucleotide variation, yet often display high amounts of structural diversity. Dobzhansky cataloged several morphologically distinct Y chromosomes in *Drosophila pseudoobscura* that differ in size and shape, but the molecular causes of their dramatic size differences are unclear. Here we use cytogenetics and long-read sequencing to study the sequence content of polymorphic Y chromosomes in *D. pseudoobscura*. We show that Y chromosomes differ by almost 2-fold in size, ranging from 30 to 60 Mb. Most of this size difference is caused by a handful of active transposable elements (TEs) that have recently expanded on the largest Y chromosome, with different elements being responsible for Y expansion on differently sized *D. pseudoobscura* Y's. We show that Y chromosomes differ in their heterochromatin enrichment, expression of Y-enriched TEs, and also influence expression of dozens of autosomal and X-linked genes. Intriguingly, the same helitron element that showed the most drastic amplification on the largest Y in *D. pseudoobscura* independently amplified on a polymorphic large Y chromosome in *D. affinis*, suggesting that some TEs are inherently more prone to become deregulated on Y chromosomes.

Introduction

Y chromosomes are a fascinating part of the genome in species with heteromorphic sex chromosomes (Bachtrog 2020). Not only do Y chromosomes determine maleness in many taxa (Bull 1983; Bachtrog et al. 2014), but their clonal and male-limited transmission is responsible for several unique evolutionary processes that shape their genomic composition (B. Charlesworth and Charlesworth 2000; Bachtrog 2013). Y chromosomes are derived from ordinary autosomes, but over time they often have diverged considerably from their former homolog, the X chromosome (Bull 1983; B. Charlesworth 1978).

Y chromosomes are transmitted through males only, which makes them an ideal location for male-beneficial genes (Rice 1984). However, Y chromosomes also have a reduced effective population size and lack recombination; this dramatically decreases the efficacy of natural selection and results in the erosion of ancestral genes on the Y (Bull 1983; B. Charlesworth 1978). Y chromosomes often accumulate repetitive DNA and evolve a heterochromatic appearance (Brian Charlesworth, Sniegowski, and Stephan 1994). Indeed, old Y chromosomes of many species are typically regarded as repeat-rich genetic wastelands.

A lack of recombination and the small effective population size implies that variation should be low on Y chromosomes, and levels of nucleotide diversity are indeed highly reduced on Y chromosomes of various species (Zurovcova and Eanes 1999; Bachtrog 2004). Yet, polymorphic Y chromosomes in *Drosophila* influence a variety of traits, including male

fertility (Chippindale and Rice 2001), temperature tolerance (Rohmer et al. 2004), lifespan (Griffin et al. 2015) and expression of 100s of genes across the genome (Lemos, Araripe, and Hartl 2008; Lemos, Branco, and Hartl 2010). It has been speculated that structural variation involving repetitive regions influences these traits, by globally modulating the genome-wide balance of heterochromatin within the genome (Francisco and Lemos 2014; Brown, Nguyen, and Bachtrog 2020b). Yet, the molecular basis of structural Y variation has not been established.

Morphologically distinct Y chromosomes that occur in natural *D. pseudoobscura* males have first been described almost a century ago (Lancefield 1929), and Dobzhansky categorized seven morphologically distinct Y chromosomes that differ in size and shape (Th. Dobzhansky 1937). The molecular basis of size variations of *D. pseudoobscura* Y chromosomes and their origins is unknown, but Dobzhansky speculated that polymorphic Y chromosomes derived by losses of sections from the largest Y (Th. Dobzhansky 1937). He also concluded that morphological and physiological characteristics of a male are not affected by the type of its Y chromosome.

Here, we take advantage of recent advances in sequencing technology, to revisit the dramatic difference in Y size among *D. pseudoobscura* strains. We combine cytogenetic techniques with whole-genome sequencing and chromatin and transcriptome profiling to study the sequence composition of different Y chromosomes, and their phenotypic consequences. We show that mobilization of a handful of TEs and expansion of a satellite sequence (the intergenic spacer region of the rDNA repeats) contributed to size differences among polymorphic Y chromosomes, and we detect subtle differences in heterochromatin enrichment, expression of Y-enriched TEs, and expression of dozens of autosomal and X-linked genes among males containing different Y chromosomes.

Results & Discussion

Cytogenetic approaches to identify and characterize different Y chromosomes

We generated Y replacement lines, in order to avoid confounding genomic background effects when contrasting different Y chromosomes (Clark 1990). We crossed males from different localities with virgins from the sequenced reference strain of *D. pseudoobscura* (MV25), and repeatedly backcrossed males resulting from this cross with virgin females of the reference strain for 9 generations (**Figure S1, S2**). This crossing scheme should ensure that resulting fly lines contain different Y chromosomes in an otherwise isogenic background, that is, all differences in sequence composition among lines should be attributed to the Y chromosome. A total of 26 Y-replacement lines were generated, from diverse geographic populations (**Table S1, Figure S1-S3**). To identify Y chromosomes that differ in their morphology and their overall sequence content, we resorted to classical cytogenetic approaches. We first used microscopy to broadly classify Y's that differ in shape and size from natural strains of *D. pseudoobscura*, (**Figure 1A, Table S2, Figure S3**). In particular, we measured total Y size (relative to the X), and Y shape (long vs. short arm length) to identify different Y types (**Table S2, Figure 1B**). Consistent with Dobzhansky's studies (Th. Dobzhansky 1937), we find dramatic variation in Y size and shape: we found 22 acrocentric, 3 submetacentric, and 1 metacentric Y chromosome (**Figure 1B, Table S3**).

To further characterize the size of these Y's, we used flow cytometry to estimate the relative size of the Y against a *D. virilis* standard (**Figure 1C, Table S4**). We found that inferred diploid genome sizes vary drastically among different Y replacement lines, ranging from 318 to 348 Mb. Assuming that the rest of the genome is roughly 291 Mb (Bracewell et al. 2019), this implies that the absolute size of the Y chromosomes vary from 27 to 57 Mb.

Most of this size difference is presumably due to differences in repetitive heterochromatin among Y chromosomes. We used two cytometric approaches to test if males with larger Y chromosomes harbor more heterochromatin. First, we estimated nuclear DNA of cells known to under-replicate heterochromatin. *Drosophila* thoracic cells exhibit endoreplication, the process in which DNA is synthesized in S phase but mitosis does not immediately follow, and the DNA in these cells under-replicate heterochromatin (Johnston, Schoener, and McMahon 2013). We therefore compared the amount of DNA in G1-phase cells to endoreplicated cells with propidium iodide to obtain proxy estimates for heterochromatin content across Y-replacement males. As a control, we compared the amount of DNA in G1-phase and G2-phase cells from brain nuclei, which replicate all DNA. We found that the largest Y-replacement strain male had a higher proportion of DNA under-replicated in thoracic cells compared to the sequenced strain male and female (**Figure 1D, Table S5**).

Heterochromatin is generally AT-rich in contrast to GC-rich euchromatin (Saksouk, Simboeck, and Déjardin 2015). Therefore, we also used DAPI, a stain that preferentially binds to AT sequences, to obtain independent estimates on heterochromatin content (Bosco et al. 2007). We found that both proxy measurements for heterochromatin agreed well with each other; that is, more under-replicated DNA in thoracic cells corresponds to higher %AT content (**Figure 1D, Table S5**). Thus, our cytogenetic work suggests that naturally isolated Y chromosomes in *D. pseudoobscura* differ up to almost 2-fold in size, presumably driven mostly by different repeat composition (i.e. heterochromatin).

Assembly of a small, medium and large Y chromosome

We chose three Y chromosomes that spanned the entire spectrum of sizes for further molecular characterization. In particular, we selected a small, medium and large Y, and we refer to them as Y_S , Y_M and Y_L . We obtained high-coverage Illumina sequencing reads (15-19x coverage, obtained by standard PCR+ libraries done on gDNA from head tissue) for each strain, as well as long-read Nanopore sequencing. Nanopore reads were used for *de novo* assembly of the different Y chromosomes. Assemblies of repetitive regions, including the Y, are challenging and these regions are often collapsed, highly fragmented, or missing in whole-genome assemblies (Hoskins et al. 2007; Chang and Larracunte 2019). We thus followed a previously described procedure to obtain heterochromatin-enriched assemblies, which were shown to perform superior to recover Y-linked fragments in *D. melanogaster* (Chang and Larracunte 2019). We used two different assembly algorithms (Koren et al. 2017; Chin et al. 2016), and merged the resulting assemblies (M. Chakraborty et al. 2016), in order to produce a more contiguous and complete assembly (**Table S6**). Y-linkage of contigs was inferred using male vs. female genomic coverage data (see methods).

Overall, we identified between 139-246 contigs as putative Y-specific across the three different Y's (using a cutoff of $\text{Log}_2(\text{Male/Female}) > 1$ and $>5\times$ depth in males; **Figure 2**). In the Y_M assembly, several Y contigs have increased male coverage, indicating that they contain collapsed sequences. When we account for the inferred copy number of Y-specific contigs based on male coverage, the assembled sizes are $Y_S=38.6\text{Mb}$, $Y_M=43.9\text{Mb}$ and $Y_L=59.5\text{Mb}$. Importantly, the assembled sizes of the three Y chromosomes are in good agreement with our estimated sizes based on flow cytometry (**Table 1**), suggesting that we were able to recover a large fraction of the sequence of each Y chromosome in our assembly.

We used BioNano to validate and further scaffold our assembly for Y_L (**Table S7**). Optical mapping produces restriction maps for long strands of DNA which can then be used to validate and place small contigs onto larger scaffolds. The BioNano hybrid assembly pipeline yielded 32 hybrid scaffolds, 6 of which contained the autosomes and X chromosome. The remaining 26 contained sequences from our Y_L Nanopore assembly. In total, 56 Nanopore contigs were fused into 17 hybrid scaffolds and the remaining 9 scaffolds contained sequences of single contigs. Overall, 63 of the 139 Y_L contigs are supported by optical reads, accounting for 41.7Mb of the 55Mb Y_L assembly, and only 15 contigs were cut. As expected, Nanopore contigs that were validated by optical mapping are significantly longer (mean length 662.5kb) vs. those not covered by optical reads (mean length 175.0kb). Thus, optical mappings validated 75.8% of our Y_L assembly.

Molecular characterization of repeat content of Y chromosomes

The ancestral Y chromosome in *Drosophila* consists mostly of satellite and TE-derived DNA (Chang and Larracuente 2019; Pimpinelli et al. 1995). The Y chromosome of *D. pseudoobscura* is not homologous to the ancestral Y of *D. melanogaster* (Carvalho and Clark 2005) but instead originated from an ordinary autosome about 15MY ago (Bracewell and Bachtrog 2020). Below, we use both assembly-free and assembly based methods to characterize the sequence content of the three different Y chromosomes at the molecular level.

As mentioned, cytogenetic methods suggest that Y_M and Y_L are 11Mb and 26Mb larger than Y_S , with an absolute size of $Y_S=33\text{Mb}$, $Y_M=44\text{Mb}$ and $Y_L=59\text{Mb}$ (**Table 1**). Most of this difference is likely driven by repeat accumulation (**Figure 1**). We used different metrics to quantify the repeat content of the different Y chromosomes, and estimate if differences in repeat content can account for the inferred size difference among Y chromosomes. In particular, we estimate the proportion of short reads assembling to TEs using dnapipeTE (Goubert et al. 2015), and the proportion of short reads mapping to a curated TE library of *D. pseudoobscura* (T. Hill and Betancourt 2018). Since the three different Y chromosomes are isogenic at the rest of the genome, these metrics can thus be used to estimate differences in the repeat content of the different Y's.

As expected, we find that the Y replacement line containing Y_S has a larger fraction of single-copy reads (71.5%) followed by Y_M (69.9%) and Y_L (66.7%) (**Figure S4**). Using the difference in single-copy reads to estimate size differences between the Y's suggests that Y_M and Y_L have an additional 5.6Mb and 16.8Mb of repeats in their genomes, respectively, in comparison to Y_S (**Table 1**). We also mapped raw Illumina reads to a TE reference library from *D. pseudoobscura* (T. Hill and Betancourt 2018), to infer size differences in repeat content

among Y chromosome, by normalizing coverage to the repeat libraries by autosomal coverage. Overall, we found that Y_S males contained about 85.3Mb of repetitive DNA, Y_M males contained 94.0Mb, and Y_L males 109.3Mb (**Table S8, Figure S5**). This suggests that Y_M and Y_L have an additional 8.6Mb and 24.0Mb repeats in their genomes, respectively, in comparison to Y_S (**Table 1**).

Lastly, heterochromatin-enriched genome assemblies recover overall similar differences in repeat content. We repeat-masked our three different Y chromosome assemblies using RepeatMasker version 4.1.0 with the *D. pseudoobscura* family repeat library (T. Hill and Betancourt 2018). Indeed, the vast majority of Y-linked sequence is repetitive in all three assemblies, with total repeat-masked bases on Y_S =34.4Mb (89%), Y_M =39.1Mb (89%) and Y_L =52.4Mb (88%). Retroelements are by far the most abundant sequence class on each of the three different Y chromosomes (**Table 2**). We found that the TE families repeat-masked in our de novo assemblies corroborated the families we found in our Illumina mapping (**Table 1, 2; Table S9, Figure S6**). TE abundance estimated from both Illumina mappings and de novo assemblies had a significant linear correlation for all Y chromosome variants (**Figure S6, p-value < 2e-16**).

Thus, these orthogonal approaches, using both cytogenetics and high-throughput sequencing, as well as assembly-based and assembly-free methods are largely in agreement about the different Y chromosome sizes. These results illustrate that our de novo Y assemblies are of high quality, and that the differences in Y size are largely driven by differences in repeat accumulation.

Recent repeat accumulation drives size differences among Y chromosomes.

The difference in sizes among Y chromosomes could be due to some repeats being mobilized on the Y, or due to structural mutations (such as duplications and deletions) resulting in size variation. Indeed, Dobzhansky hypothesized that smaller Y chromosomes originated via multiple deletions from the largest Y (Th. Dobzhansky 1937). To identify which repeats contribute to the observed size differences among Y chromosomes, we used both assembly-free and assembly-based methods. We mapped short reads to a repeat library, to infer different abundances of repeats on the different Y chromosomes (**Figure 3A, Table S8**), and also identified repeats on our assembled Y chromosomes (**Figure 3B, Table S9**). Intriguingly, we found that only a handful of repeats were responsible for the majority of additional DNA of Y_L and Y_M (**Table 3**). Interestingly, most of the TE families responsible for additional DNA of Y_L are not causing the size increase of Y_M .

A few distinct repeats account for most of the size increase in the large Y chromosome (**Table 3**). In particular, we find that the Helitron-1_DPe transposon has notably higher copy numbers on Y_L compared to Y_M and Y_S . Read coverage suggests that this TE (which is about 7.2kb in size) has roughly 70 copies in the Y_S and Y_M strain, but 1090 copies in Y_L , resulting in an extra 7.4Mb of sequence on Y_L (**Figure 3A, Table 3; Figure S7**). A similar excess of this transposon is found in the assembled Y chromosomes: about 200kb and 150kb are repeat-masked on Y_S and Y_M for Helitron-1_DPe, but 4.7Mb on Y_L (**Table S9**). As expected for an active TE, we find copies of Helitron-1_DPe distributed across many different Y-linked scaffolds (**Figure S7, S8**). Helitron TEs replicate through a rolling-circle mechanism, which can result in the

generation of tandem copies of the element (Pritham and Feschotte 2007). Indeed, genomic distribution patterns suggest that this element can occur in clusters of tandem arrays across the large Y chromosome (**Figure S8, S9**). BLAST searches using queries consisting of the last 50bp of the Helitron-1_DPe element fused to the first 50-bp yielded 50 hits on our Y_L assembly, and mate-pair violations using paired-end sequencing data further support the presence of partial copies of tandemly repeated elements (**Figure S9, Table S10**). Because many of the tandem copies of *Helitron-1_DPe* on Y_L are partial (**Figure S9**), our BLAST search yields a conservative estimate for the number of copies in tandem. We find 8 nearly full-length copies (using an 70% alignment cutoff) of Helitron-1_DPe on the assembled autosomes and X (and 3 copies on unmapped scaffolds). This is in contrast to the 289 near-full length copies we find in the Y_L assembly and the 10 and 14 copies in the Y_M and Y_S assemblies, respectively.

The repeat that is responsible for most of the size increase of Y_M , and ranked second on Y_L , is a 267-bp long satellite sequence that is homologous to the intergenic spacer region (IGS) of the rDNA cluster. In many *Drosophila* species, including *D. melanogaster*, the rDNA cluster is found on both the X and Y chromosome, and functions as a pairing site for the sex chromosomes during meiosis (McKee, Habera, and Vrana 1992). A previous study has shown that the Y of *D. pseudoobscura* – which is not homologous to the Y of *D. melanogaster* – lacks functional rDNA clusters on the Y (Larracunte, Noor, and Clark 2010). Instead, *in situ* hybridization revealed four clusters bearing only the IGS of the rDNA repeats on the Y chromosome of the *D. pseudoobscura* strain investigated (Larracunte, Noor, and Clark 2010). We infer roughly 13,300 copies of the IGS satellite on Y_S , 19,800 copies on Y_M , and 26,100 copies on Y_L , which adds about 1.7Mb of DNA to Y_M , and 3.4 Mb of DNA on Y_L (**Table S8, Table S9, Figure 3A, 3B**). The IGS repeats are highly clustered in our assemblies, as expected for satellite sequences, and consistent with the *in situ* results (Larracunte, Noor, and Clark 2010).

The next most common repeat on Y_L is the Polinton-1_DPe repeat. Polintons are large DNA transposons (Polinton-1_DPe is 15.8kb) that encode several proteins necessary to replicate themselves (Krupovic and Koonin 2015). Read coverage analysis suggests that there are about 67 copies of this element on Y_S and Y_M , but 236 copies on Y_L , adding an additional 2.7Mb of sequence (**Table S8**). This element is distributed widely across different Y-linked scaffolds (**Fig S8**), and even read-coverage across the element suggests that many copies are full-length on the different Y chromosomes (**Fig S7**). The next most common element is the T122 repeat, a 1.5kb long uncharacterized element with roughly 70 copies on Y_S and Y_M , but almost 1700 copies on Y_L (**Table S8**), adding an additional 2.5Mb of sequence. T122 repeats are often found in tandem clusters, mostly separated by HelitronN-1_DPe sequences, which may indicate co-amplification of T122_X by Helitron-initiated gene-capture. The next most common repeat is the Copia2-I_Dpse element (4.2kb in size) that has 270 and 330 copies on Y_S and Y_M , respectively, but 670 on Y_L , thereby adding 1.7Mb, and three different gypsy elements that are 2-3x more abundant on Y_L than Y_S and Y_M , and each adding about another 1Mb of sequence to Y_L . Thus, a total of only 8 repeats are responsible for over 20Mb of additional DNA on Y_L (i.e. >85% of the total size gain), with Helitron-1_DPe by far being the largest contributor to the size gain of Y_L .

On the other hand, a much larger number of TEs contribute to the more modest size gain of Y_M , with the top 8 expanded repeats contributing to <50% of size gain on Y_M . As mentioned, the repeat leading to most of the size increase on Y_M is the IGS satellite (adding about 1.7Mb of DNA to Y_M). The next most common one is the 1.1kb long HelitronN-1_DPe element and the related 4.6kb long MINIME_DP element (**Figure S10**), which have roughly about 6400 (HelitronN-1_DPe) and 900 (MINIME_DP) copies on both Y_S and Y_M , but 7100 (HelitronN-1_DPe) and 1000 (MINIME_DP) copies on Y_M (thereby adding 730kb and 360kb of sequence to Y_M). *Gypsy17-I_Dpse* adds another 320kb to Y_M , and all other amplified TE families add less than 200kb sequence each to Y_M (**Table S8**).

We also used our assembled Y chromosome contigs and used Repeatmasker to annotate repeats and infer differences in abundance (**Figure 3B, Figure S11, Table S9**). Because some contigs in our assemblies were collapsed, we calculated these differences by also considering the copy number of assembled contigs and their respective repeats. Similar to our assembly-free method, we found that Y_M and Y_L had 5.4 and 19.2Mb extra transposable elements, respectively, in comparison to Y_S (**Figure 3B**). Again, the same TE families (*Helitron*, 267bp IGS, *Copia*, *Polinton*) as found in the coverage-based approach are responsible for the size differences between Y chromosomes based on repeat-masking our assemblies. Thus, both methods suggest that only a small subset of repeats drive most of the observed size differences on Y_L , and that there is overall little correspondence in the types of TEs that have accumulated on the different sized Y chromosomes.

Individual copies of recently active TEs are more similar to each other than TEs that are older, and the amount of sequence divergence of individual copies from the consensus can be used to infer the age of when a particular TE family proliferated (Makałowski et al. 2019). Consistent with the size increase on polymorphic Y's being driven by the recent expansion of a few TEs, Y_M and especially Y_L contain younger TEs. We find that Y_M and Y_L have 4.9% and 23.3% more of their genome-wide TEs within 0-1% divergence from the consensus compared to Y_S , respectively (**Figure 3C**). Further, the subset of TEs that are the main contributor to the larger sizes are younger than other TEs (**Figure 3D, Figure S12**), and amplified TEs cluster on phylogenetic trees (**Figure S13**). We further leveraged our assemblies to characterize divergence of TEs strictly located on the Y's. As expected, we found many copies of TEs from different families, especially rolling circle elements (RC; to which the *Helitron* element belongs), were young in the large Y chromosome (**Figure 3D**). Thus, our de novo genome assemblies provide orthogonal confirmation that TEs have recently amplified on the large Y. Together, these results demonstrate that the large Y chromosome is the derived state in *D. pseudoobscura*, and DNA accumulation on these larger Y chromosomes is due to recent TE activity.

Phenotypic consequences of repeat accumulation

While repetitive DNA is often viewed as a genetic wasteland, recent studies have found that differences in repeat content can have profound phenotypic consequences. In particular, it was shown that expression of 100s of genes differed in *D. melanogaster* males with polymorphic Y chromosomes (Lemos, Araripe, and Hartl 2008; Lemos, Branco, and Hartl 2010). Additionally, the presence and the number of Y chromosomes strongly influences genome-wide enrichment patterns of repressive chromatin modifications (Brown and Bachtrog 2014; Brown, Nguyen,

and Bachtrog 2020a), with additional Y chromosomes diminishing the heterochromatin enrichment at highly repetitive regions such as pericentromeres. Larger, more repeat- and gene-rich Y chromosomes were also linked to an up-regulation of transposable elements in males (Wei, Gibilisco, and Bachtrog 2020; Nguyen and Bachtrog 2021). These results are generally interpreted as the Y chromosome acting as a heterochromatin sink that re-distributes repressive chromatin marks genome-wide (Francisco and Lemos 2014). The integrity of heterochromatin deteriorates as individuals age (Sun, Yu, and Dang 2018), and the Y chromosome in *D. melanogaster* has been shown to contribute to heterochromatin loss and shorter longevity in males (Brown, Nguyen, and Bachtrog 2020a).

To test for phenotypic consequences of Y chromosome repeat content, we measured longevity of males that contained a small or large Y chromosome (Y_S and Y_L) versus females, and assayed global transcriptome and chromatin profiles from Y_S and Y_L males. We collected H3K9me3 profiles from brains of young Y_S and Y_L males, in order to compare the global heterochromatin landscape in males with different Y chromosomes. The heterochromatin sink effect of the Y would predict that global H3K9me3 levels may be lower for repeats in males with a larger Y chromosome. **Figure 4A** shows that H3K9me3 enrichment is similar for TEs located on the X and autosomes in males containing differently sized Y chromosomes; however, TEs located on the Y chromosome show significantly lower H3K9me3 levels in Y_L males than Y_S . Thus, the higher repeat content of Y_L may indeed dilute heterochromatin components in males with a large Y. Intriguingly, we find that H3K9me3 levels are generally much lower for TE copies located on the Y chromosome compared to autosomes and the X, irrespective of the Y variant (**Figure 4A, B; Figure S14, S15**). This is consistent with previous findings showing lower levels of heterochromatin enrichment for Y-linked TEs (Wei, Gibilisco, and Bachtrog 2020; Nguyen and Bachtrog 2021), and repeat-rich Y chromosomes were found to be associated with the up-regulation of transposable elements in males (Brown, Nguyen, and Bachtrog 2020a; Wei, Gibilisco, and Bachtrog 2020; Nguyen and Bachtrog 2021).

We obtained replicate expression data from heads of young and old Y_S and Y_L males to quantify TE expression between differently sized Y chromosomes, and during aging. While global expression across all TE families is not significantly different between Y-replacement males, irrespective of age (**Figure S16**), we find that a larger number of TE families are significantly up-regulated in Y_L males compared to Y_S (12 vs 0, at least 50% higher expression, Wald-test, p-value < 0.05; **Figure 5A**). Furthermore, half of these TEs are more abundant on Y_L than Y_S , suggesting that the TEs that have accumulated on Y_L are responsible for increased TE expression (**Figure 5B**). Notably, this result is consistent irrespective of age, suggesting that differential regulation of TEs starts early in adult males (**Figure S17**).

Interestingly, old Y_L males had slightly higher TE expression compared to young Y_L males but this was not true in Y_S males (**Figure S15**). Consistent with previous studies, we find that males live shorter than females (p-value < 0.0001). However, we find no significant difference in longevity between males containing the different sized Y chromosomes (**Figure 5C, Figure S19**); if anything, Y_L males live slightly longer than their counterpart Y_S males (median survival time 77 vs. 74 days, respectively). Thus, while males with a larger Y chromosome show slightly increased levels of TE expression, and more so in old males, these differences are relatively minor, and do not result in faster aging. Small differences in TE expression and

lifespan are consistent with previously published results done in *D. melanogaster* Y replacement lines (Griffin et al. 2015).

Previous studies in *D. melanogaster* have shown that different Y chromosomes can influence the expression of hundreds of genes located on the X and autosomes (Lemos, Branco, and Hartl 2010). We find that 930 genes on the autosomes and X exhibit differential expression between Y_S and Y_L Y-replacement lines, 20 of which are significant ($>50\%$ fold change, Wald-test, p -value < 0.05 , **Figure 5D**, **Table S11**). Many differentially expressed genes are involved in the regulation of chromatin and transcription such as *aub*, *Ulp1*, and *mx* (**Table S11**). Taken together, our results underline the impact Y chromosomes can have on genome-wide expression regulation.

Repeated TE mobilization causes Y expansion in other *Drosophila* species

Our data suggest that modest changes in Y chromosome size are caused by copy number changes of a large number of repeats (Y_M), but dramatic size differences are due to mobilization of a few elements (Y_L). We used flow cytometry to study Y chromosome size polymorphism in five different Y-replacement lines in *D. affinis* (kindly provided by Rob Unckless), another member of the *obscura* species group and distant relative of *D. pseudoobscura*. We found three Y chromosomes that dramatically differed in size (**Figure 6**). Compared to the smallest Y chromosome of *D. affinis* ($Y_{S,aff}$; strain YAF41), the medium-sized Y ($Y_{M,aff}$; strain YAF159) has an additional 12.7Mb of DNA, and the largest Y chromosome ($Y_{L,aff}$; strain YAF79) has an extra 19.0Mb of DNA (**Figure 6A**). We used Illumina sequencing of males with the differently sized Y chromosomes, and mapped reads to the same repeat library used for *D. pseudoobscura* to infer which repetitive elements are responsible for the size increase of the Y in *D. affinis*. Intriguingly, overall patterns of repeat accumulation on the large Y chromosomes are similar between *D. affinis* and *D. pseudoobscura*, with only a small number of TEs accounting for most of the size increase of the large Y chromosome (**Figure 6B**). Most surprisingly, we find that the same TE (Helitron-1_DPe) that expanded on Y_L in *D. pseudoobscura* also is the main contributor to increased DNA content of $Y_{L,aff}$ in *D. affinis*; 8 Mb of additional DNA (42%) on the large Y of *D. affinis* are due to additional copies of the Helitron-1_DPe element. The second most common repeat on $Y_{L,aff}$ of *D. affinis* is Daff_Jockey_18 (3.4Mb), followed by Polinton-1_DPe (2.2Mb), the same element which also amplified on Y_L in *D. pseudoobscura*. Congruent with our findings in *D. pseudoobscura*, TEs on $Y_{M,aff}$ and $Y_{L,aff}$ also show low levels of sequence divergence that are commensurate with Y size (**Figure 6C**). Thus, recent amplification of a small number of TEs is also responsible for Y expansion in other species.

TEs can transfer horizontally between species and this has previously been reported in the *D. pseudoobscura* group (T. Hill and Betancourt 2018). To evaluate whether horizontal transfer between species could have contributed to the dramatic repeat accumulation on the large Y's, we examined patterns of sequence evolution of the amplified Helitron and Polinton elements in *D. affinis* and *D. pseudoobscura*. We extracted Helitron-1_DPe and Polinton-1_DPe copies found on the autosomes/X of *D. pseudoobscura* and *D. affinis* and our Y_L assembly from *D. pseudoobscura*, and found that elements cluster by species (**Figure 6D**). Thus, this suggests that these elements evolved independently in the two species, as found in (T. Hill and Betancourt 2018). Since we lack a high-quality assembly of the large Y chromosome in *D.*

affinis we cannot reconstruct the sequence of individual Y-linked TEs in this species, but our coverage analysis implies that a large majority of the Helitron and Polinton reads are derived from Y_L and $Y_{L,aff}$. We therefore constructed consensus sequences for these two TEs for both species and mapped genomic reads from *D. affinis* and *D. pseudoobscura* males with the large Y's to the consensus sequences to estimate the rate of species-specific mapping. As expected if TEs amplified independently in the two species, most Y_L reads map to the *D. pseudoobscura* consensus and most $Y_{L,aff}$ reads map to the *D. affinis* consensus (**Table 4**). Thus, we find no signatures of horizontal transfer at these elements, which suggests that rampant amplification of autonomous TEs occurred independently on the two different species' Y chromosomes.

Conclusions

In a series of papers, Dobzhansky cataloged variation in size and shape of the *D. pseudoobscura* Y chromosome across its geographic range (Th. Dobzhansky 1937). However, the molecular basis or the mechanism of the origin of variation of Y size was unknown, and Dobzhansky speculated that the polymorphic Y's may have been derived by losses of sections from the largest Y chromosome (Th. Dobzhansky 1937). Here, we use a combination of classical cytogenetic methods and long-read sequencing to identify the molecular basis of Y size variation. We show that in contrast to Dobzhansky's hypothesis, most of the size differences between morphologically distinct Y chromosomes are due to the recent expansion of a few TEs on the largest Y chromosome. Thus, rather than being the ancestral form, this indicates that the large Y chromosome is instead derived from a smaller ancestral Y. Most intriguingly, we find that the same TEs families can contribute to Y expansion independently in different species of *Drosophila*. Dobzhansky also tried to find a correlation between the type of Y chromosome present in a given strain and other properties of that strain, including hybrid sterility (Th. Dobzhansky 1937; Th. Dobzhansky and Boche 1933; Th. Dobzhansky 1933), but concluded that the morphological and physiological characteristics of a particular male are not affected by the type of its Y chromosome. By homogenizing the genetic background of the different Y types, we were able to detect subtle differences among Y chromosomes. We show that overall heterochromatin enrichment is reduced on the Y of males containing a larger Y chromosome, and expression of a subset of TEs, and especially those that have accumulated on Y_L , is increased. Additionally, expression of a few dozen of X-linked and autosomal genes is influenced by Y-type. Thus, Y variation can potentially influence male-related fitness traits.

Morphological variation in Y chromosome size and shape has been identified in many organisms, including humans (B. M. Chakraborty and Chakraborty 1984), but has not yet been studied at the DNA sequence level. Our study demonstrates that novel sequencing approaches allow the assembly of even highly repetitive Y chromosomes, and presents novel insights into the repetitive landscapes of variable Y chromosomes.

Materials & Methods

Y-replacement line backcrosses. A total of 26 different *D. pseudoobscura* males were backcrossed to the sequenced *D. pseudoobscura* stock (MVZ 25, courtesy of Stephen Schaeffer). The list of fly strains used for Y-replacement lines is given in **Table S1**, and **Figure S1** shows the geographic origin of different lines. We backcrossed males to virgin females for 9

generations after which the Y-replacement lines were maintained on standard molasses culture. We verified homozygosity on the autosomes and X chromosome by mapping Illumina data to the masked genome, calling sites with Samtools mpileup, and then dividing the number of sites that are homozygous across all strains by the total number of variant sites (**Table S12**).

Y chromosome karyotyping. We dissected brains from third instar larvae in 0.9% NaCl. First, we incubated brains in 0.9% NaCl with 2 drops of 0.1% colchicine for 10 min at room temperature, followed by another 10 min incubation in 0.1M KCl solution. Then, we fixed brains in 3:1 methanol and acetic acid for 1 hour. We moved the fixed tissues to 50 μ L of 60% acetic acid solution and vigorously disassociated tissues via pipetting for 1 minute. We dispensed the solution onto pre-heated 55°C microscope slides and stained them in 7% Giemsa solution (1x PBS pH 6.5) for 40 min. Slides were imaged on a Leica DM5000B microscope with SPOT imaging software. We then measured chromosome lengths with KaryoType (Altinordu et al. 2016).

Flow cytometry. We estimated the sizes of Y chromosomes using flow cytometry as described (Ellis et al. 2014). Briefly, we dissected and snap-froze heads from flies aged 3-7 days old. For an internal standard, we used heads from a *D. virilis* strain provided by Spencer Johnston (1C = 328Mb). To prepare samples, we dounce homogenized one *D. pseudoobscura* head with one *D. virilis* female head in a 2mL Kontes dounce homogenizer with 1mL Galbraith Buffer (45mM MgCl₂, 30mM sodium citrate, 20mM MOPS, 0.1% Triton X-100 (v/v), pH 7.2) and filtered the nuclei-containing solution through a 100 μ m cell strainer. We then added 25 μ L propidium iodide (1mg/mL) to stain nuclei for 30 min at 4°C and ran the stained nuclei on a BD LSR II cytometer at 30-60 events/sec.

We exported raw fluorescence data on FlowJo (v.10.1) and ran a custom R script to calculate the genome sizes. To calculate the genome size of the unknown Y-replacement line, we assumed fluorescent linearity between the *D. virilis* standard and the Y-replacement line males.

To estimate %AT content relative to *D. virilis*, we followed (Ellis et al. 2014) and used 1 μ L of DAPI (1mg/1mL) in place of propidium iodide. We calculated %AT assuming linearity. We estimated the proportion of under-replicated DNA in polytene tissue following the protocol described (Johnston, Schoener, and McMahon 2013). Briefly, we dissected thoracic tissue from flies and followed the protocol described above without the *D. virilis* standard. We ran the same protocol with head tissue as a positive control. We used a custom R script to calculate the 4C cell fluorescence in relation to 2C fluorescence. All flow cytometry work was done at the UC Berkeley Cancer Research Lab Flow Cytometry Facility.

Lifespan assays. We conducted lifespan assays following (Linford et al. 2013) with the following rearing conditions: 18°C, 60% relative humidity, 12h light, and standard Bloomington food. Briefly, we collected synchronized embryos on a molasses plate with yeast paste for 16-20 hours. We washed embryos with 1x PBS pH 7.2 three times and dispensed 10 μ L of embryos per culture vial by pipette. To obtain synchronized adults, we collected emerging adults over a 3 day window and aged the adults for 3 days to mate and copulate. Male and female flies were separated into separate vials, placing 30 flies per vial. We moved flies to new vials every 2-3 days without CO₂ and recorded the number of deaths. Flies that

escaped were censored. In total, 1196 sequenced strain female flies, 722 sequenced strain Y_S male flies, and 987 Y_L male flies were tracked for the lifespan assays recorded here.

Tissue collection. Tissue for gDNA-Seq was collected by freezing male flies and popping off heads via vortexing. To collect tissue for the RNA-Seq experiment, we censored the entire experiment once it reached 50% survivorship for one line. We collected head tissue for RNA-Seq in a similar manner to the gDNA-Seq. For the ChIP-Seq experiment, we anesthetized young (5-7 day old) flies using CO₂, dissected the brains in 1x PBS pH 7.2, and immediately froze 2 brains per sample using dry ice.

DNA, RNA, and ChIP library preparation. We extracted DNA from pooled heads of each Y-replacement male using the DNAeasy Qiagen Extraction kit. We then prepared gDNA libraries using the Illumina TruSeq Nano DNA library kit (#20015965). For the cDNA libraries, we first extracted RNA from 20 pooled heads for Y-replacement line males using TRIzol (Invitrogen #15596026). We then made libraries using the Illumina TruSeq Stranded Total RNA Ribo-Zero Gold kit (Illumina RS-122-2201).

For the ChIP-Seq libraries, we started with brain tissue and prepared ChIP pull-downs as described in (43). Briefly, we modified the native and ultra-low input ChIP protocol (Brind'Amour et al. 2015) for double-brain ChIP-Seq. First, we added 40μL EZ nuclei lysis buffer to frozen tissues and homogenized them with a pestle grinder on ice. We spun cells down at 1000 g for approximately 10 min and decanted 20μL of the supernatant. We froze the cell pellet in the lysis buffer at -80°C. We then followed the 100,000 cell count digestion protocol from (Brind'Amour et al. 2015) with a few modifications. For each *D. pseudoobscura* sample and *D. melanogaster* spike-in sample, we added 4.4μL of 1% DOC and 1% Triton-X100 to resuspended cells and mixed thoroughly. Briefly, we digested each sample at 37°C for 04:58 (mm:ss) with MNase and quenched the reaction with 4.9μL of 100mM EDTA pH 8.0. We then added the fragmented spike-in samples to each *D. pseudoobscura* sample to account for approximately 20% of the final pooled sample volume (i.e. final sample consisted of 80% *D. pseudoobscura*, 20% *D. melanogaster*). We reserved 10% of the pooled fragmented sample for the input (fragment control) and used the remaining 90% to perform the chromatin pull-down (ChIP sample, target antibody: H3K9me3 polyclonal classic Diagenode C15410056) according to the 100,000 cell specifications. We used the Rubicon Genomics ThruPlex kit to prepare ChIP-Seq libraries for sequencing with 10 PCR amplification cycles for the input samples and 12 cycles for the ChIP pull-down samples. We sequenced libraries on the Illumina HiSeq 4000 Platform at the Vincent J. Coates Genomics Sequencing Center (Berkeley, CA).

Repeat abundance estimations by reference library. We used the *D. obscura* group consensus TE library from (T. Hill and Betancourt 2018) and included a Y chromosome-specific 267bp repeat sequence which was later confirmed as the 267bp intergenic spacer (Larracunte, Noor, and Clark 2010). We masked the *D. pseudoobscura* or *D. affinis* genome using this repeat library with RepeatMasker (-no_is -no_low -norna, v4.1.0). Then, we mapped reads with bwa mem (H. Li and Durbin 2009) to the repeat-masked genome and the TE library. In this way, all reads derived from TEs will map to the consensus library to decrease reference bias for the small (reference) Y. We used bedtools coverage to calculate

coverage per TE and normalized coverage by the median autosomal coverage. The autosomes were confirmed to have approximately 2x coverage (i.e. diploid genome).

We summed the copy numbers of all TEs multiplied by their respective lengths to estimate the difference in TE bp contribution between Y-replacement lines. Note that this sum includes TEs derived from autosomes, however, these contributions should be negligible because the strains have isogenic backgrounds.

Repeat content estimation and divergence by dnaPipeTE. We ran dnaPipeTE on the first strands of samples with the following parameters: -coverage 0.5 -genome-size 175000000 -sample-number 2. Note that we used genome size of 175 Mb even though the strains are of different sizes, which will yield a more conservative repeat % result between the strains because of varying genome sizes. We used the % difference in single-copy elements to estimate the Mb difference between Y's. We also used the estimated divergence to identify recently active TEs on a genome-wide basis.

Nanopore sequencing and Y-chromosome de novo assembly. For the reference strain male (Y_S), we used reads from (Bracewell et al. 2019). For Y_M males, we used one Nanopore MinION flow cell (v9.4.1 RevD) with the ligation sequencing kit (SQK-LSK109) and obtained 540,115 total filtered reads with a mean read length of 15,972bp. We sequenced Y_L males using 3 Nanopore MinION flowcells (v9.4) with the ligation sequencing kit (SQK-LSK108 and SQK-LSK109). We combined the total output from these three runs and obtained 482,379 filtered reads with a mean read length of 19,650bp.

We followed a heterochromatin-enriched assembly approach from (Chang and Larracuenta 2019). Briefly, we mapped reads to the *D. pseudoobscura* assembly (Bracewell et al. 2019) using minimap2 (v2.1) with default parameters. Then we extracted reads that did not map to putative euchromatin in the *D. pseudoobscura* assembly (Bracewell et al. 2019) to make heterochromatin-only assemblies using Canu (v1.6) (Koren et al. 2017) and Falcon (v1.4.2) (Chin et al. 2016) independent of each other.

We then removed contigs containing bacterial DNA using BlastN against the NCBI database and then used QuickMerge to merge contigs between Canu- and Falcon-based assemblies. We settled on assemblies that had a high N50 and generally less contigs, all of which resulted from merges between assemblies from a Canu to Falcon direction with assemblies from a Falcon to Canu direction. To identify Y contigs, we mapped Illumina male and female reads to the corresponding heterochromatin-based assemblies with bwa mem and obtained genomic coverages with bedtools coverage. Y chromosomes can contain repeats found elsewhere in the genome so we used differences in female and male coverages to identify male-bias/Y-linked sequences. Contigs whose $\text{Log}_2(\text{Female/Male})$ coverage was below -1 and whose overall male coverage was at least 5x were considered Y contigs. We then appended these Y contigs to the assembled autosomes and X scaffolds to generate full male genomes. Using our coverage analysis, we also split contigs whose coverage was partially male-biased for the medium and small Y assemblies.

To calculate the estimated amount of DNA and repeats on each Y chromosome variant, we normalized the coverage of the Y contigs by the coverage of the autosomes. We used this normalized coverage as the corresponding copy number (estimated abundance) of each Y contig with which we used to adjust the estimated total DNA length and total repeat abundance (RepeatMasker v4.1.0).

BioNano molecules and hybrid assembly. We used optical mapping (BioNano Genomics) to stitch Y contigs from the large Y chromosome together. Briefly, DNA was extracted from 45mg of frozen male larvae using the Bionano Prep Animal Tissue DNA Isolation Kit (Product #80002) with modifications from a protocol by Susan Brown at Kansas State University. Then, DNA was homogenized and embedded in agarose plugs, treated with Proteinase K and RNase A, and isolated from agarose and cleaned via drop dialysis. DNA was then left at room temperature for several days to homogenize prior to quantitation. 750 ng of DNA was labeled using the DLS Labeling Kit (Product #80005) following the Bionano Prep Direct Label and Stain (DLS) Protocol (Document 30206).

Labeled DNA was run through the Bionano Saphyr Platform. These reads were then assembled into full optical maps and then allowed for stitching of Y chromosome sequences. In total, 334,498 molecules were generated with an N50 of 223.875 kb and this was downsampled to proceed with the hybrid assembly pipeline (i.e. 189,039 molecules and N50 of 273.571 kb). The Bionano hybrid assembly contained 32 hybrid scaffolds, 6 belonging to the autosomes and X chromosome and the remaining 26 to Y_L chromosome. We confirmed the 6 scaffolds were from the autosomes and the X chromosome with NUCmer (v3.1) alignments to the *D. pseudoobscura* genome. The BioNano preparation and hybrid assembly were completed at the UC Davis DNA Technologies Core.

Y-linked TE divergence estimates. To obtain Kimura divergence estimates for each TE family and their copies on the Y's, we used parseRM_getLandscape.pl from (Kapusta, Suh, and Feschotte 2017) with our RepeatMasked Y chromosome assemblies.

Tandemly repeated transposable element analysis. We adapted an approach from (McGurk and Barbash 2018) to infer if transposons were tandemly repeated by identifying violations in paired-end alignments. Paired-end reads were first aligned with bwa mem (H. Li and Durbin 2009) in single-end mode to the masked *D. pseudoobscura* genome and the TE library. We then obtained read pairs where both forward and reverse reads aligned to the same TE and then plotted their alignment coordinates.

To infer the number of tandemly repeated *Helitron-1* *DPe* transposons in the Y_L assembly, we made artificial queries of full head-to-tail sequences as in (Pritham and Feschotte 2007). We used BLASTN (v2.6.0+) to blast this sequence to the Y assemblies and identified tandem junctions based on matches with > 80% aligned.

Transposable element phylogenies. We used BLASTN (2.6.0+) to extract transposon sequences from our Y chromosome assemblies. We kept alignments of at least 70% length with the exception of Gypsy22-I_Dpse, Copia-1_DPer-I, Polinton-1_DPe, and T150_X where we kept alignments mapping to regions of elevated coverage. We used clustalo (v1.2.4) using

default parameters to generate multiple sequence alignments and then RAxML (v8.2.12) using the following parameters: -f a -x 1255 -p 555 -# 100 -m GTRGAMMA. Phylogenies were visualized using ggtree (Yu 2020) and FigTree (Rambaut 2007).

Gene expression analysis. Briefly, we mapped RNA-seq reads for each replicate to a repository of ribosomal DNA scaffolds from NCBI and removed all reads that mapped to this scaffold. Differences in rRNA abundance in sequenced samples are likely to be technical artifacts following RNA library preparation. We then mapped the remaining reads to the *D. pseudoobscura* genome and to the repeat library separately using HISAT2 (Kim et al. 2019, 2) with default parameters. We then used Subread FeatureCounts (Liao, Smyth, and Shi 2014) to calculate gene counts and repeat counts and used DESeq2 (Love, Huber, and Anders 2014) to normalize the libraries and perform differential expression analysis between male/female flies of different ages.

ChIP enrichment analysis. We followed the approach from (Nguyen and Bachtrog 2021). We took ChIP mappings to the entire genome and called enrichment at repetitive elements through TECounts (part of TETranscripts) (Jin et al. 2015). To call for repeat enrichment on a per chromosome basis, we split mappings by chromosome (autosome/X chromosome and Y chromosome) and ran TECounts separately for each split mapping. We then normalized counts at repeats in both the ChIP and input by mean autosomal coverage and calculated ChIP enrichment as ChIP/Input.

Transposon cross-mapping analysis. We identified autosome/X Helitron and Polinton elements in both the *D. pseudoobscura* and *D. affinis* genomes using BLASTN, keeping alignments with at least 70% length. We generated consensus sequences using EMBOSS cons with default parameters (https://www.ebi.ac.uk/Tools/msa/emboss_cons, accessed April 22, 2021). For Y_L and $Y_{L,aff}$ samples, we extracted reads mapped to Helitron-1_DPe/Polinton-1_DPe. Then, we used bwa mem (H. Li and Durbin 2009) to map reads to both consensus copies of the corresponding transposon simultaneously.

Figure Legends

Figure 1. *D. pseudoobscura* male karyotype and Y chromosome size variation. **A.** *D. pseudoobscura* male karyotype. **B.** Measurements of Y chromosome arms (long, short) from chromosome spreads for each Y-replacement line. **C.** Diploid genome size estimations of Y-replacement line males from flow cytometry. Rightmost boxplot (red) is the diploid genome size of the sequenced/backcross female. **D.** Estimates of heterochromatin in 3 Y-replacement line males by staining whole nuclei with DAPI (top) and by staining thoracic cells with propidium iodide (bottom).

Figure 2. De novo assembly of three different sized Y chromosomes (YS, YM, YL). **A.** Coverage and repeat tracks of Y-chromosome assemblies. Top to bottom: male coverage track, female coverage track, Log2(Female/Male) coverage, repeat landscape with 6 classes of TEs. One tick mark corresponds to 5Mb and collapsed contigs were adjusted in the plots, indicated by a lighter color. Landscapes of the most abundant repeats are shown separately in Figure S8 A-C.

B. Top to bottom: karyotypes of YS, YM, YL males with the arrowhead denoting the Y chromosome.

Figure 3. Repeat abundance and age suggests recent mobilization contributes to size differences among the Y's. **A.** TE abundance in YM and YL relative to YS from paired-end mappings to the TE library. TEs where the absolute difference is over 200kb are shown. **B.** TE abundance in YM and YL chromosome assemblies relative to the YS assembly. TEs where the absolute difference is over 200kb are shown. **C.** Divergence in TE copies from whole-genome Illumina sequencing across the Y's estimated by dnaPipeTE. **D.** Shown is the Kimura-based divergence in TE copies from the YL assembly based on the top 50 most abundant TEs (relative to YS), the other (remaining) TEs, and the rolling-circle (RC) transposons. Percentages of TEs sum to 100% for each category.

Figure 4. Decreased H3K9me3 enrichment on YL. **A.** Heatmap with corresponding boxplot showing H3K9me3 enrichment at TEs by chromosome (autosomes/X vs. Y chromosomes). Significance values calculated (* < 0.05, ** < 0.01, *** < 1e-5, Wilcoxon-test). **B.** Scatterplot of H3K9me3 enrichment at TEs by chromosome group for YS and YL (p < 0.05, Two Sample t-test) The grey line indicates similar enrichment levels for both chromosome groups.

Figure 5. Differential transposon and gene regulation by Y chromosome. **A.** Differential TE expression between YS and YL males irrespective of age. Data represents the mean of 4 replicates with standard error bars (50% higher or lower expression, p < 0.05, Wald-test). **B.** Differential TE expression between YS and YL plotted against their differential TE abundance from Illumina mappings. **C.** Lifespan curves of YS and YL males and backcross females. **D.** Differential gene expression between YS and YL irrespective of age. Data represents the mean of 4 replicates with standard error bars (50% higher or lower expression, p < 0.05, Wald-test).

Figure 6. Y chromosome expansions in *D. affinis* by the same TEs as in *D. pseudoobscura*. **A.** Diploid genome size estimations from *D. affinis* Y-replacement lines. Rightmost boxplot (red) is the backcross female strain. **B.** TE abundance in YM_{,aff} and YL_{,aff} relative to YS_{,aff} from paired-end mappings to the TE library, made in a similar manner as Figure 3B. TEs where the absolute difference is over 50kb are shown. **C.** Divergence in TE copies from whole-genome Illumina sequencing across the Y's estimated by dnaPipeTE, made in a similar manner as Figure 3C. **D.** Phylogenetic trees of Helitron-1_DPe (left) and Polinton-1_DPe (right) found in *D. pseudoobscura* (green), *D. affinis* (orange), and *D. pseudoobscura* YL (teal). Repbase sequences are highlighted in blue.

Figure S1. Geographic locations of sampled *D. pseudoobscura* Y chromosomes. Specific locations for fly strains are detailed in Table S1.

Figure S2. Backcrossing scheme to generate Y-replacement lines. Left karyotype represents wild-type male (Y-donor) and the right karyotype represents the sequenced female (MV 25).

Figure S3. Representative chromosome squashes of Y-replacement line males with arrowheads denoting the Y chromosome in each squash. Top row (left to right): a telocentric Y, a

submetacentric Y, a metacentric Y. Bottom row (left to right): Y chromosomes from males with diploid genome sizes of 316, 322, 332, and 346Mb.

Figure S4. Reference-free based TE and single-copy (euchromatin) estimates from DNAPipeTE. Grey area shows proportion of sample that is single-copy (euchromatin) while non-grey area (multiple colors) shows proportion of samples that is multi-copy/repetitive (heterochromatin). Legend shows the breakdown of multi-copy sequences by repeat family.

Figure S5. Total TE abundance (Mb) across Y-replacement line males calculated by mapping Illumina reads to the masked genome and the TE library. The data underlying this figure is contained in Table S8. Note these are absolute amounts compared to the relative amounts which are plotted in Figure 3A.

Figure S6. Correlation between the TE abundances from mapping Illumina reads to the genome (X-axis) and the TE abundances from the de novo Y chromosome assemblies (Y-axis). Each data point represents a TE family and its respective abundance calculated from Illumina mappings (X) and the Y chromosome assemblies (Y). Data underlying these figures are contained in Table S8 and Table S9.

Figure S7. The top 10 most abundant transposons from Illumina mappings for (A) Y_L and (B) Y_M and their overall coverage (Y-axis) along their consensus sequences (X-axis) for each of the Y-replacement lines (denoted as separate colors, Y_S = green, Y_M = yellow, Y_L = blue).

Figure S8 (A). TE distribution on Y-linked scaffolds for Y_L . The X-axis represents the Y chromosome assembly split into 5kb windows (1 unit = 5kb) and each black line denotes the split between assembly contigs. Colors denote where a repeat family is located on a particular contig. Note that assemblies are not adjusted for contig coverage. Y-axis is shown for only visualization purposes.

Figure S8 (B). TE distribution on Y-linked scaffolds for Y_M . The X-axis represents the Y chromosome assembly split into 5kb windows (1 unit = 5kb) and each black line denotes the split between assembly contigs. Colors denote where a repeat family is located on a particular contig. Note that assemblies are not adjusted for contig coverage. Y-axis is shown for only visualization purposes.

Figure S8 (C). TE distribution on Y-linked scaffolds for Y_S . The X-axis represents the Y chromosome assembly split into 5kb windows (1 unit = 5kb) and each black line denotes the split between assembly contigs. Colors denote where a repeat family is located on a particular contig. Note that assemblies are not adjusted for contig coverage. Y-axis is shown for only visualization purposes.

Figure S9. Inferring tandemly repeated TEs using violations of forward and reverse mapping coordinates. For each group of plots: the leftmost represents reads from Y_S , the middle from Y_M , and the rightmost from Y_L . The X-axis denotes the mapping coordinate of the forward read and the Y-axis denotes the mapping coordinate of the reverse read. TEs that have head-to-tail

tandems will have clusters of reads in the bottom right corner. For each TE, reads are downsampled to the smallest Y chromosome sample with the exception of T150_X where Y_M has no reads.

Figure S10. Alignment of HelitronN-1_DPe (X-axis) to MINIME_DP (Y-axis) using nucmer.

Figure S11. Repeat composition on a per chromosome basis as calculated from RepeatMasker for each chromosome.

Figure S12. Divergence of abundant/non-abundant TEs on Y_M relative Y_S . Graph is made in a similar fashion as Fig 3D. Top50 category shows divergence for the top 50 TEs that are higher in Y_M compared to Y_S , other category for the remaining TEs, and the RC for the rolling-circle transposons.

Figure S13. Unrooted phylogenies of TE copies from 6 TE families found in Y_S , Y_M , Y_L assemblies showing TE expansions in Y_L . D. miranda copies are identified for Copia2-I_Dpse, T122_X, and Helitron-1_DPe.

Figure S14. Log2(H3K9me3) enrichment for TEs with 50% more copy number in either Y_S or Y_L as represented by heatmap (top) and boxplot (bottom). P-values calculated by Wilcoxon-test (* < 0.05, ** < 0.01, n.s. = not significant).

Figure S15. Difference in H3K9me3 enrichment on the autosomes/X vs on the Y chromosomes by strain. From left to right: Y_L replicate 1, Y_L replicate 2, Y_S replicate 1, Y_S replicate 2.

Figure S16. TE expression by Y-replacement line and age, sorted by TE abundance. Expression data shown as heatmap (top) and boxplot (bottom). (A) Shown are expression levels for all TEs with at least Log2(expression) > 1. (B) Shown are all TEs with at least Log2(expression) > 1 and at least 50% more copy number abundance in either Y_S or Y_L .

Figure S17. TE abundance differences (X-axis) vs TE expression differences in Y-replacement line males (Y-axis) by age. Left shows TE expression from young samples, right shows TE expression from old samples. Each data point represents a TE with its associated difference in abundance/expression between Y_S and Y_L . Data points with labels indicate TEs whose difference in expression and abundance is over 2-fold. Plots were made from the same data that underlies Figure 4C.

Figure S18. Age-associated TE expression differences within Y-replacement line males. X-axis is TE expression from young samples, Y-axis is TE expression from old samples. Left: Y_S males, right: Y_L males. Data highlighted in blue indicates at least 50% significant differential expression (Wald-test, p-value < 0.05).

Figure S19. Y-replacement line aging trials recorded separately and the sample size breakdown by population (red = Y_S females, blue = Y_S males, pink = Y_L females, light blue = Y_L males). All aging trials are shown with females from both Y-replacement line males plotted as individual tracks.

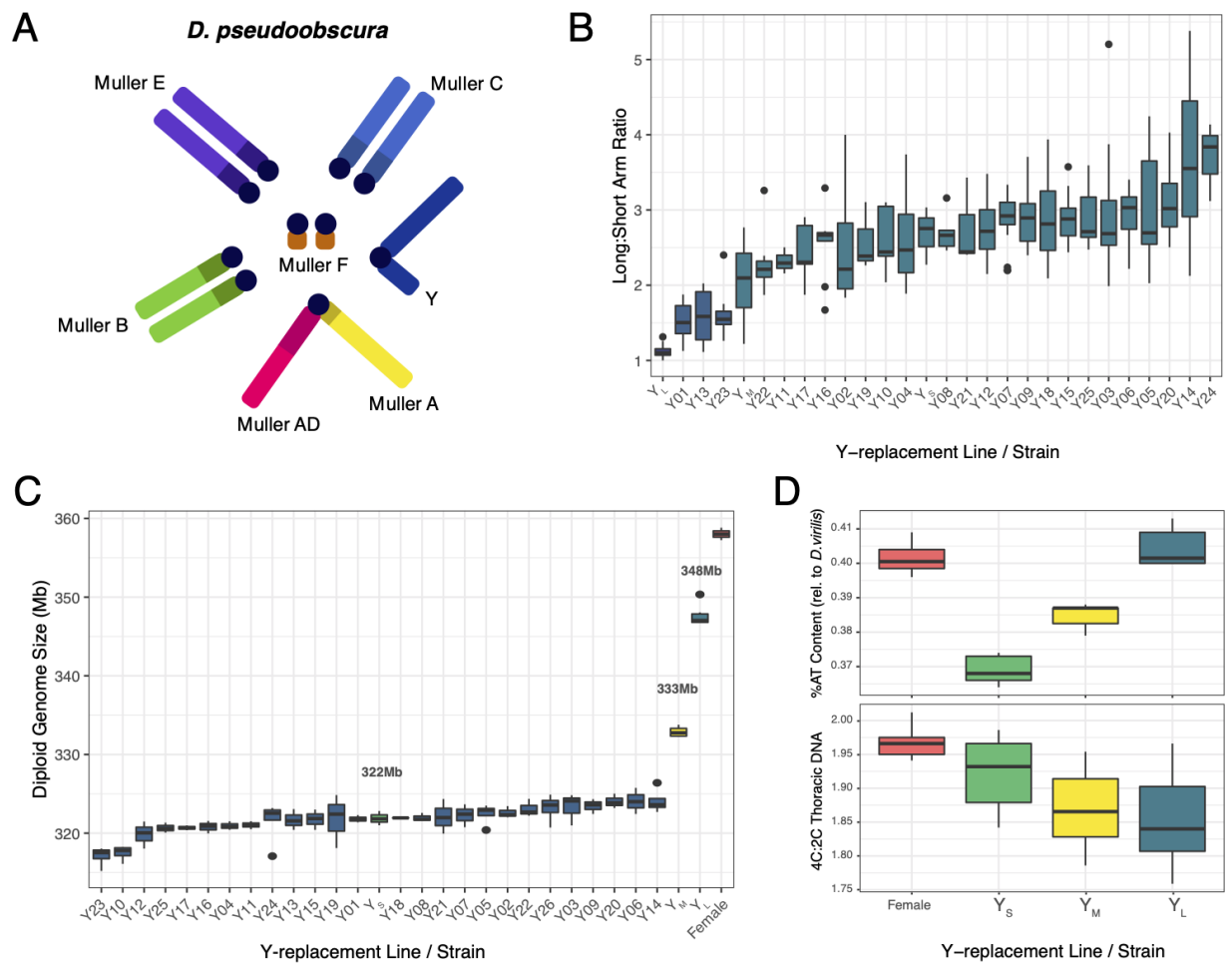


Figure 1.

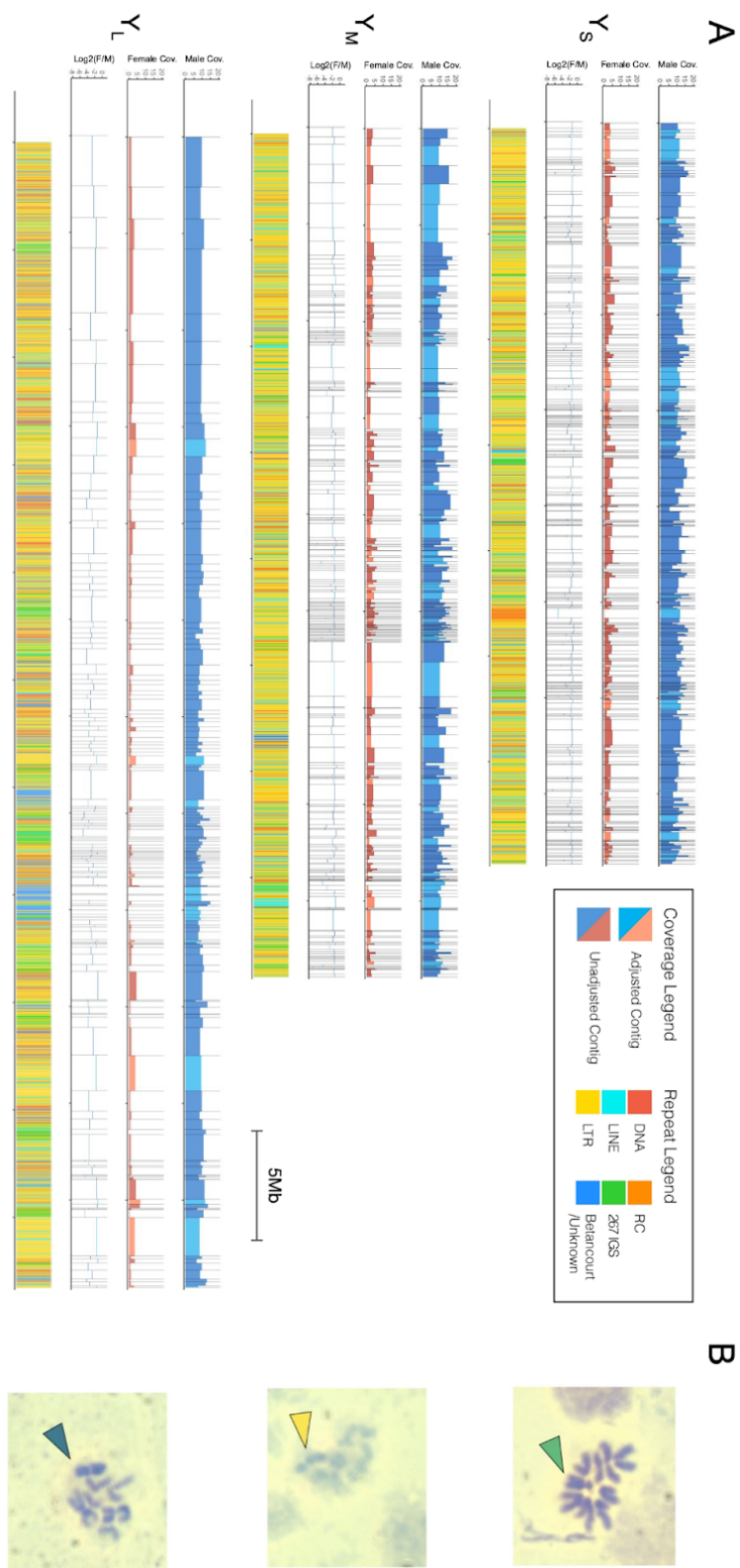


Figure 2.



78

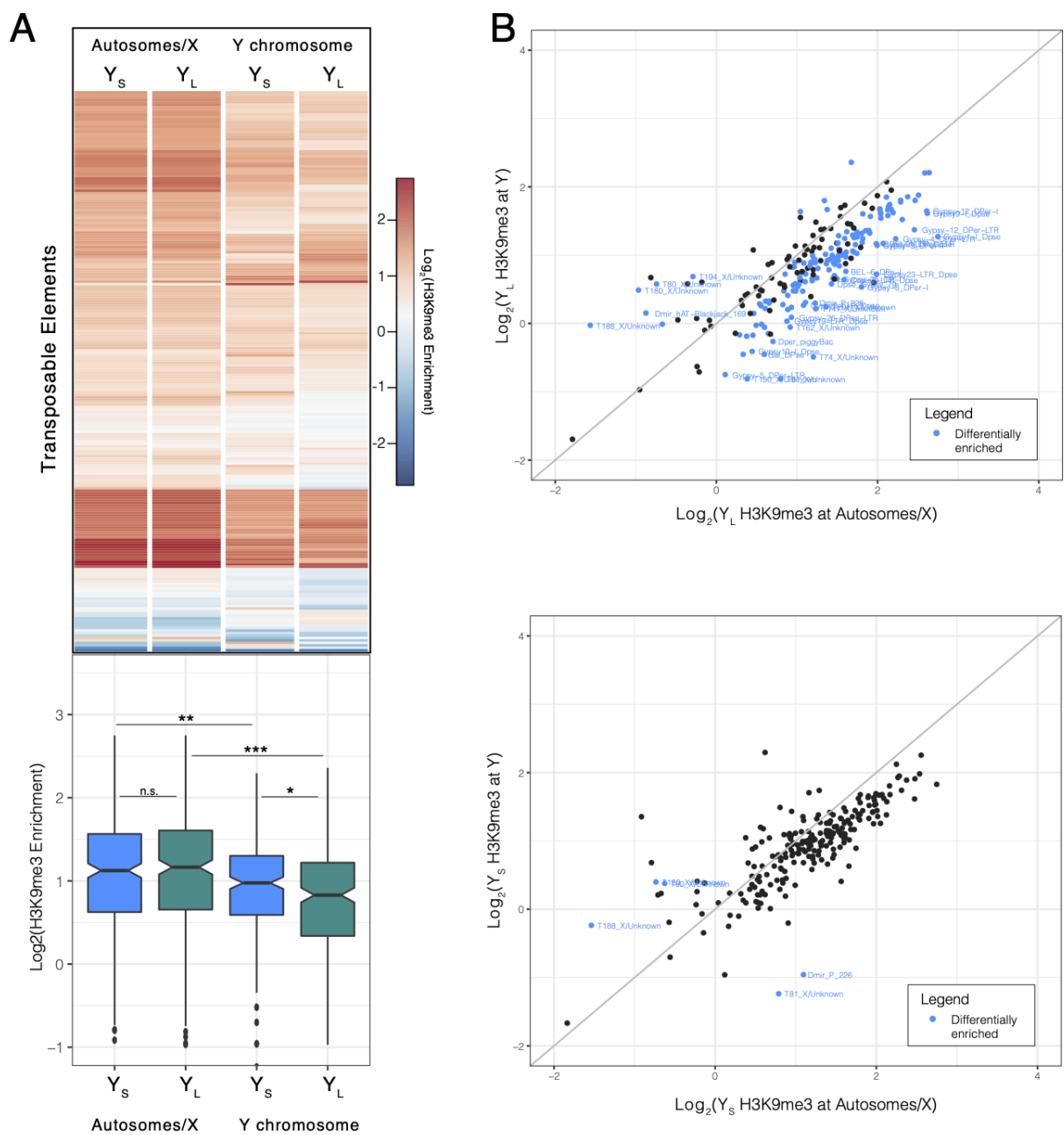
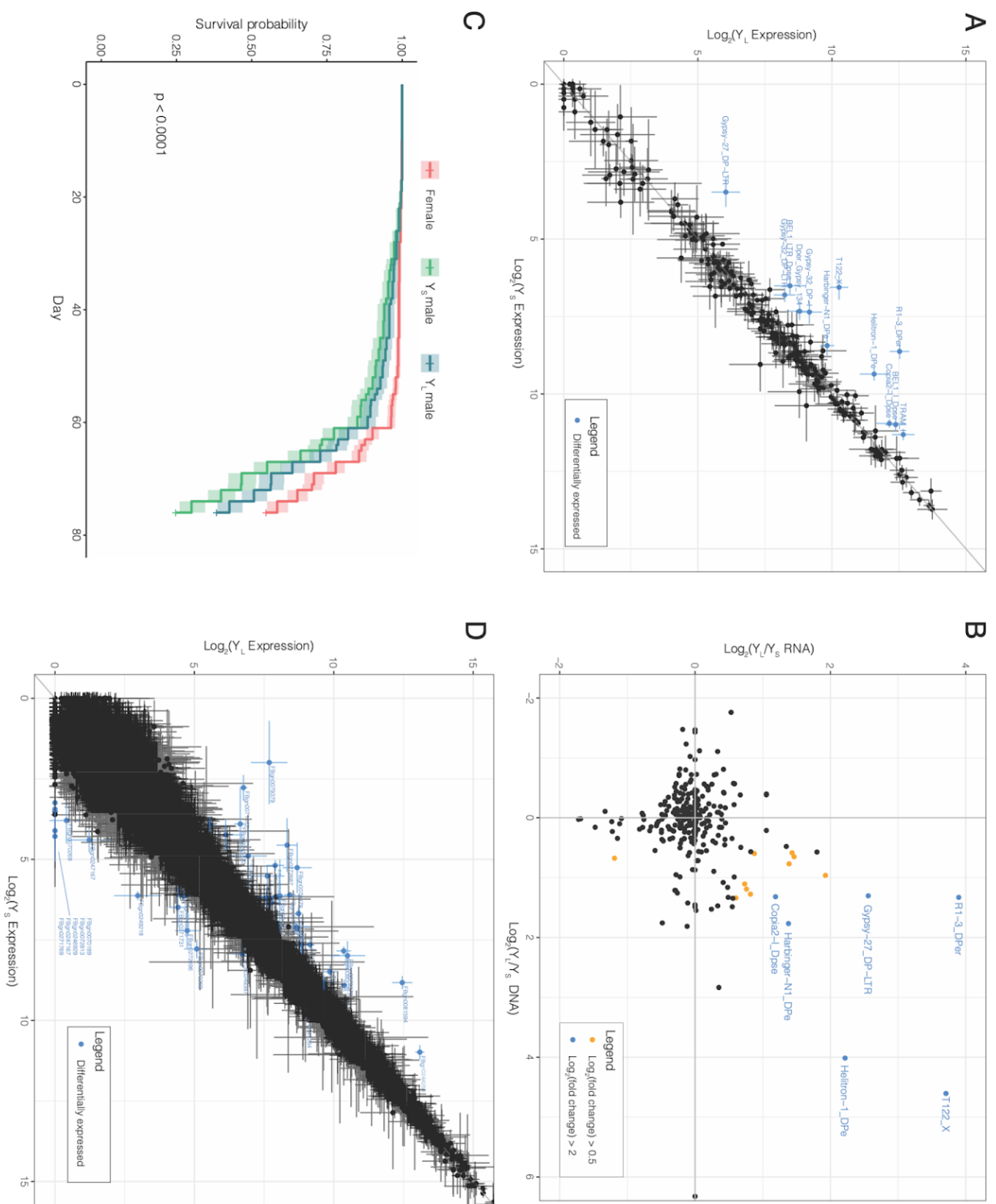


Figure 4.



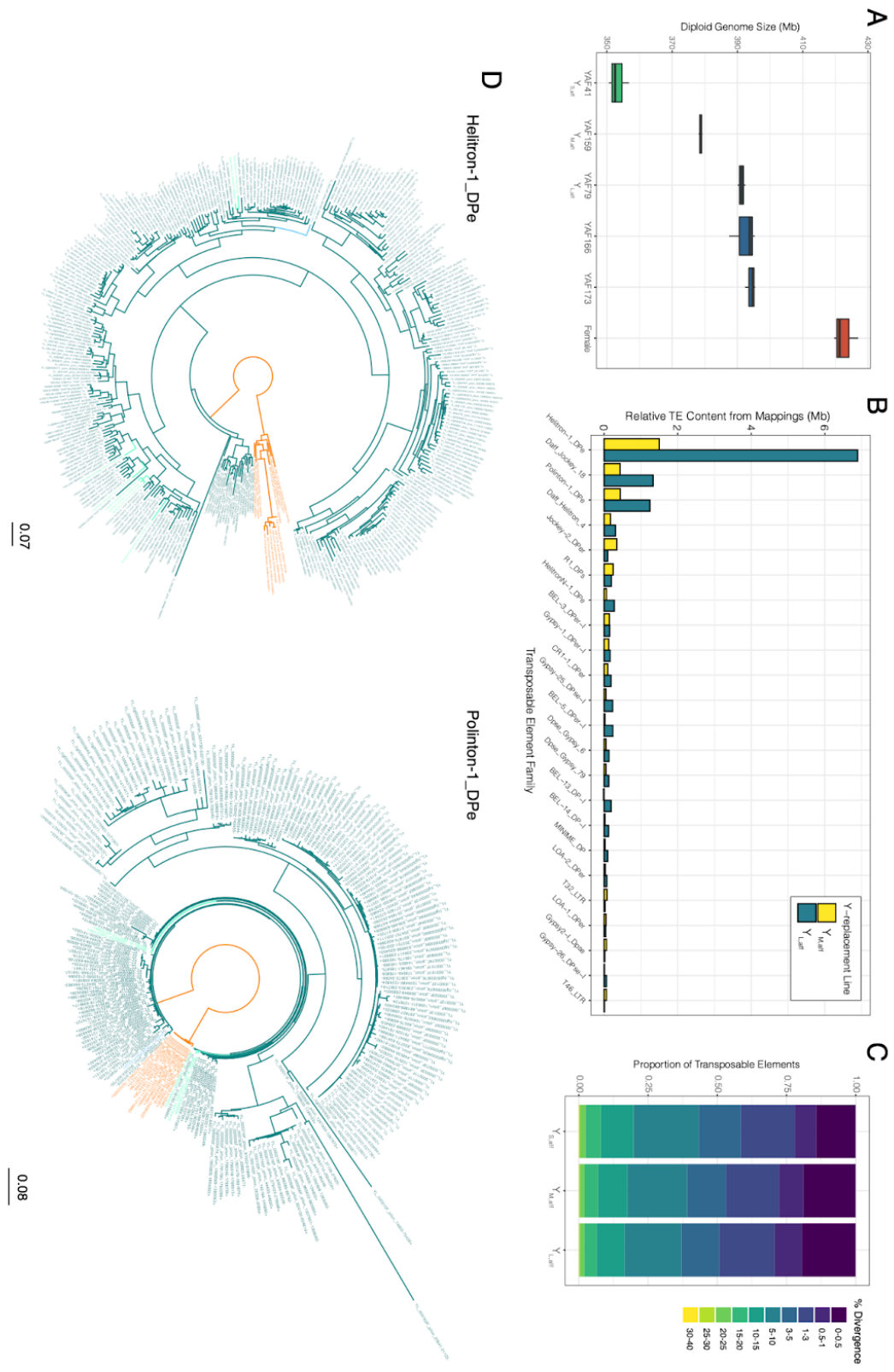


Figure 6.

Table 1. Summary of Y chromosome size estimations (in Mb) in *D. pseudoobscura* using different methods.

Strain	Y _S	Y _M	Y _L
Flow cytometry	33	44	59
Δ YS	--	11	26
Assembly (total)	38.6	43.9	59.5
Δ YS	--	5.3	20.9
dnaPipeTE	--	5.6	16.8
TE library mapping	--	8.6	24
Assembly Repeat Masked	34.4	39.1	52.4
Δ YS	--	4.7	18

Table 2. Summary of sequences repeat-masked in *D. pseudoobscura* Y chromosome assemblies.

	Y _S	Y _M	Y _L
Retroelements	28,695,122	32,680,127	34,049,007
LINEs	6,562,185	7,964,164	7,873,667
LTRs	22,132,937	24,715,963	26,175,340
DNA	618,135	536,136	2,074,574
Rolling-circle	2,203,768	2,434,069	7,312,732
Unclassified	2,580,480	3,087,272	8,479,775
Satellites	-	-	-
Simple Repeats	275,823	321,667	453,173
Low complexity	15,297	14,379	32,770
total bp repeat masked	34,388,625	39,073,650	52,402,031

Table 3. Size contribution (in bp) and inferred copy number of top repeats responsible for size differences in *D. pseudoobscura* Y chromosome.

TE	Y _S copy #	Y _M copy #	Y _L copy #	Y _S bp	Y _M bp	Y _L bp
Helitron-1_DPe	67.5	70.1	1090.7	488,296	506,965	7,884,807
267_IGS	13337.8	19836.5	26140.9	3,561,192	5,296,337	6,979,630
Polinton-1_DPe	67.1	67.5	235.7	1,059,242	1,065,658	3,722,847
T122_X	69.5	75.8	1691.3	105,987	115,548	2,577,517
Copia2-1_Dpse	266.7	328.5	665.2	1,116,006	1,374,972	2,783,998
Gypsy19-I_Dpse	113.0	128.0	313.4	578,010	654,841	1,602,821
Gypsy10-I_Dpse	141.9	175.8	300.2	844,351	1,045,780	1,786,237
Gypsy22-I_Dpse	91.9	107.9	205.7	752,440	883,301	1,684,953
T150_X	16.3	14.7	1303.2	8,937	8,094	715,481
HelitronN-1_DPe	6435.6	7068.8	6392.8	7,433,153	8,164,422	7,383,694
MINIME_DP	919.2	996.9	890.0	4,212,902	4,568,938	4,079,044
Gypsy17-I_Dpse	154.8	194.1	176.8	1,267,758	1,589,474	1,447,682
Gypsy-7_DPer-I	193.8	225.4	165.1	1,587,315	1,845,841	1,351,878
LOA-3_DPer	311.0	355.5	367.7	1,754,905	2,005,649	2,074,626
LOA-2_DPer	314.2	363.4	219.9	1,550,407	1,793,245	1,085,076
R1_DPs	123.1	167.6	92.5	670,278	912,820	503,843

Table 4. Y_L and Y_{L,aff} Helitron/Polinton mappings to the *D. pseudoobscura* and *D. affinis* consensus sequences.

	Helitron-1_DPe derived reads		Polinton-1_DPe derived reads	
Consensus sequence	<i>D. pseudoobscura</i>	<i>D. affinis</i>	<i>D. pseudoobscura</i>	<i>D. affinis</i>
<i>D. pseudoobscura</i>	562,655	2,126	265,181	10,963
<i>D. affinis</i>	5,670	349,819	399	98,907

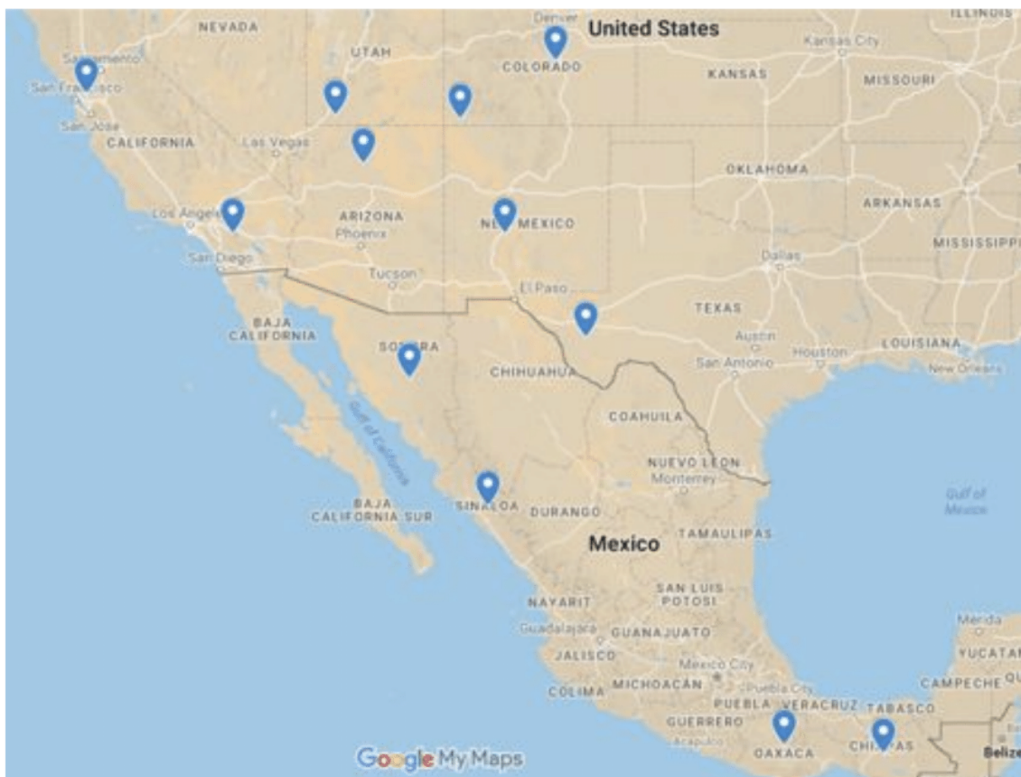


Figure S1.

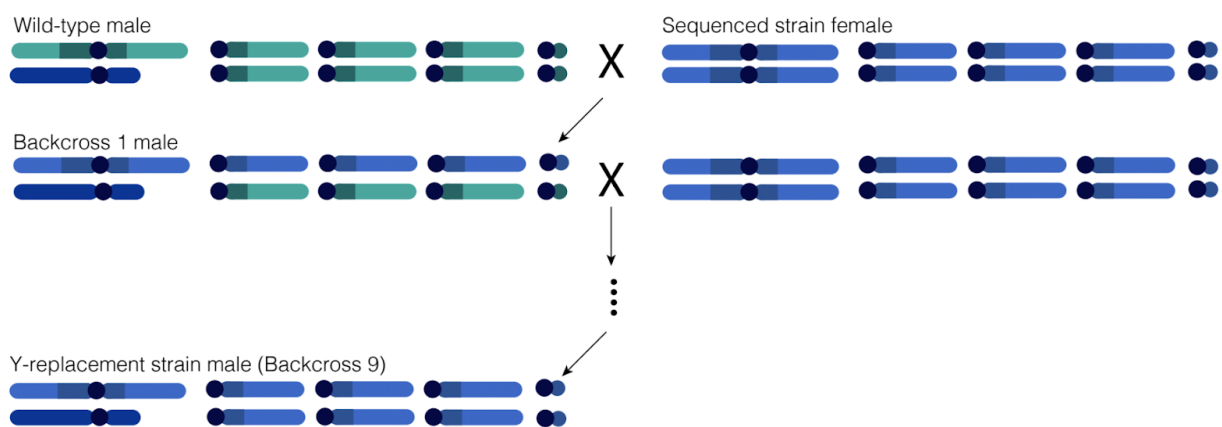


Figure S2.

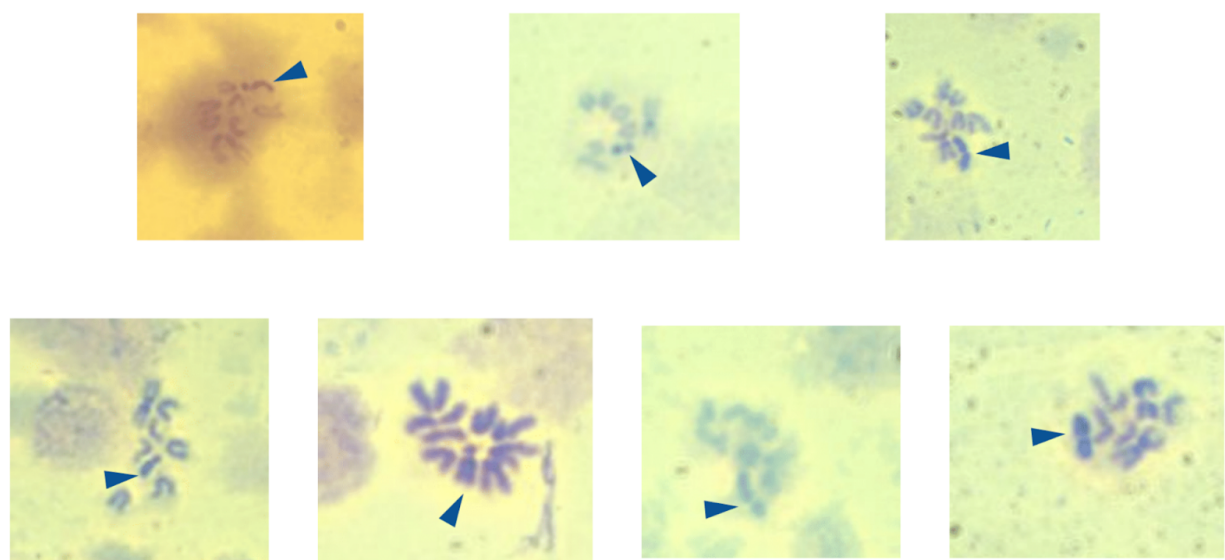


Figure S3.

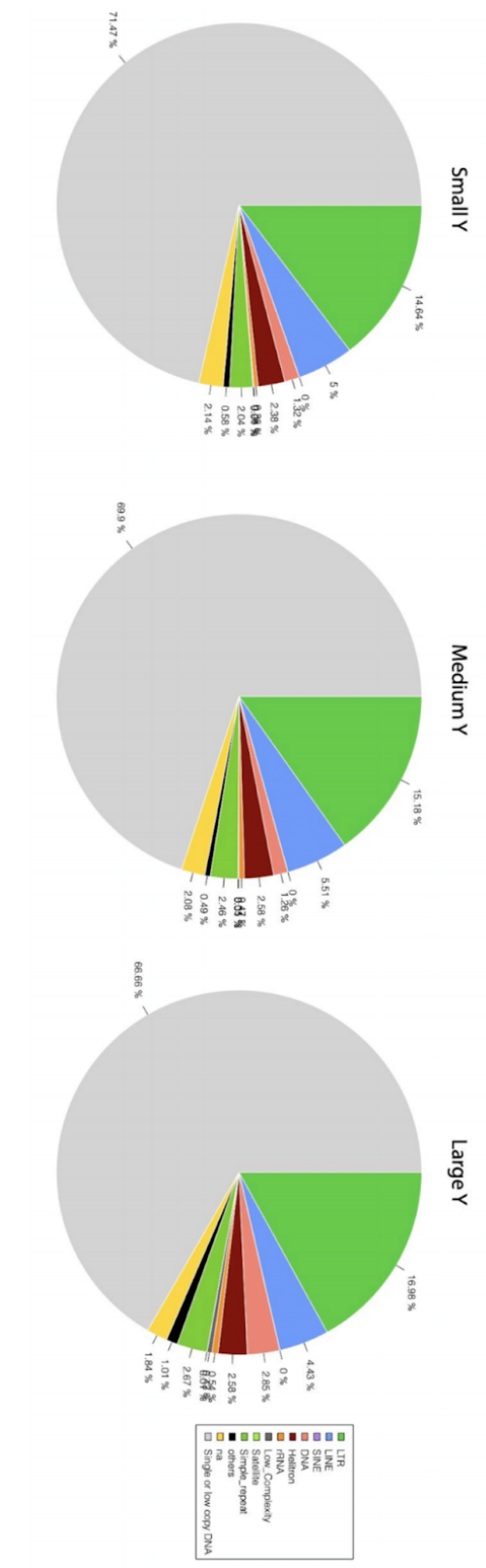


Figure S4.

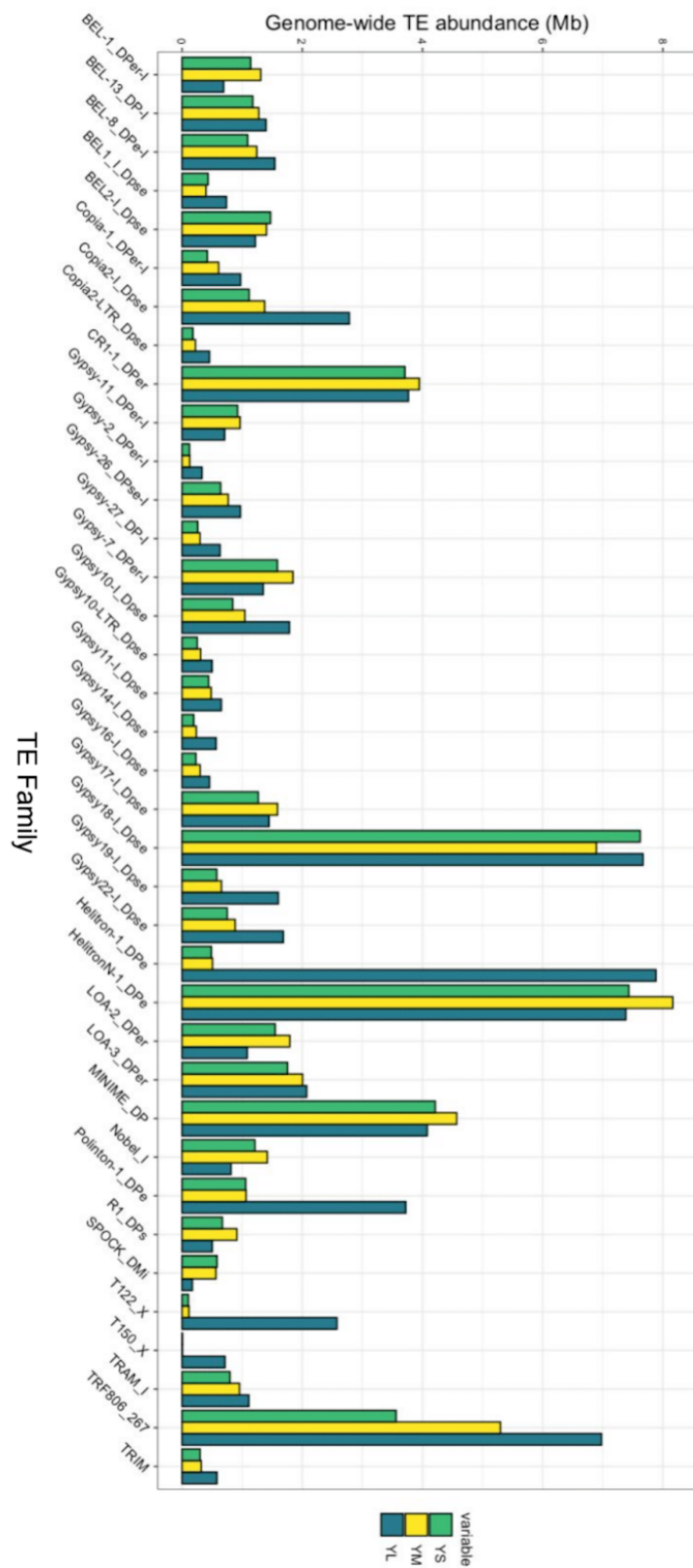


Figure S5.

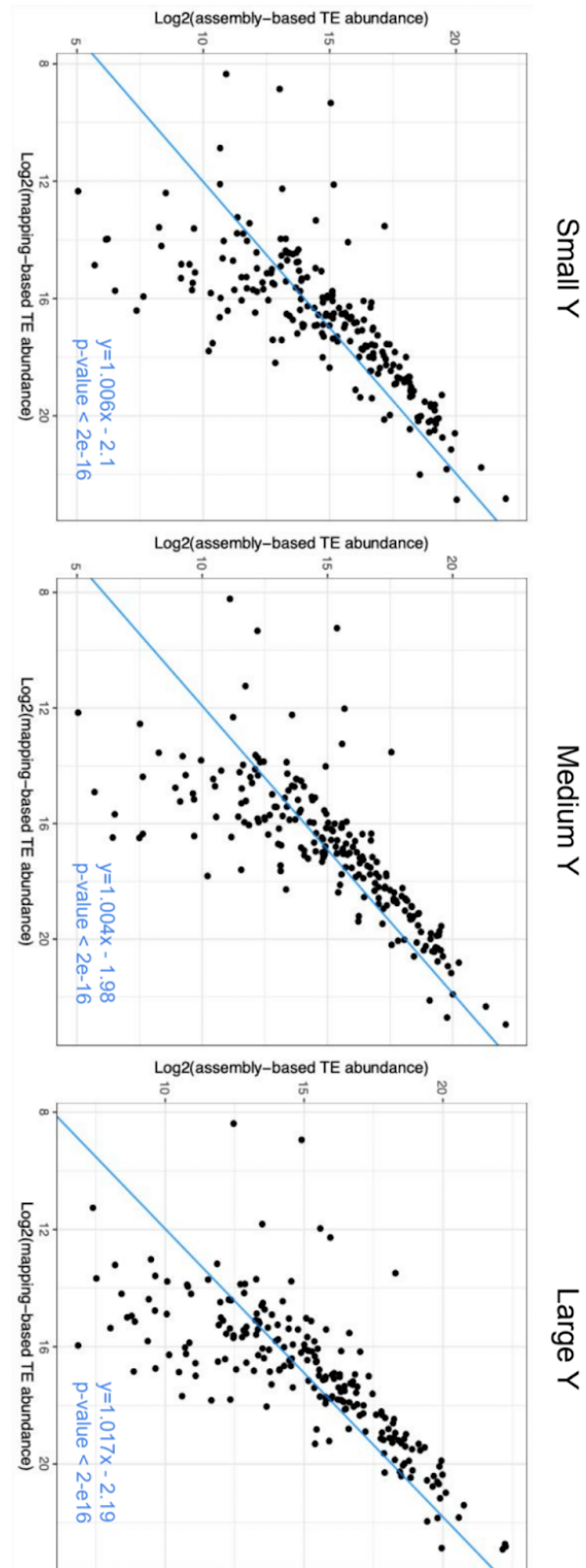


Figure S6.

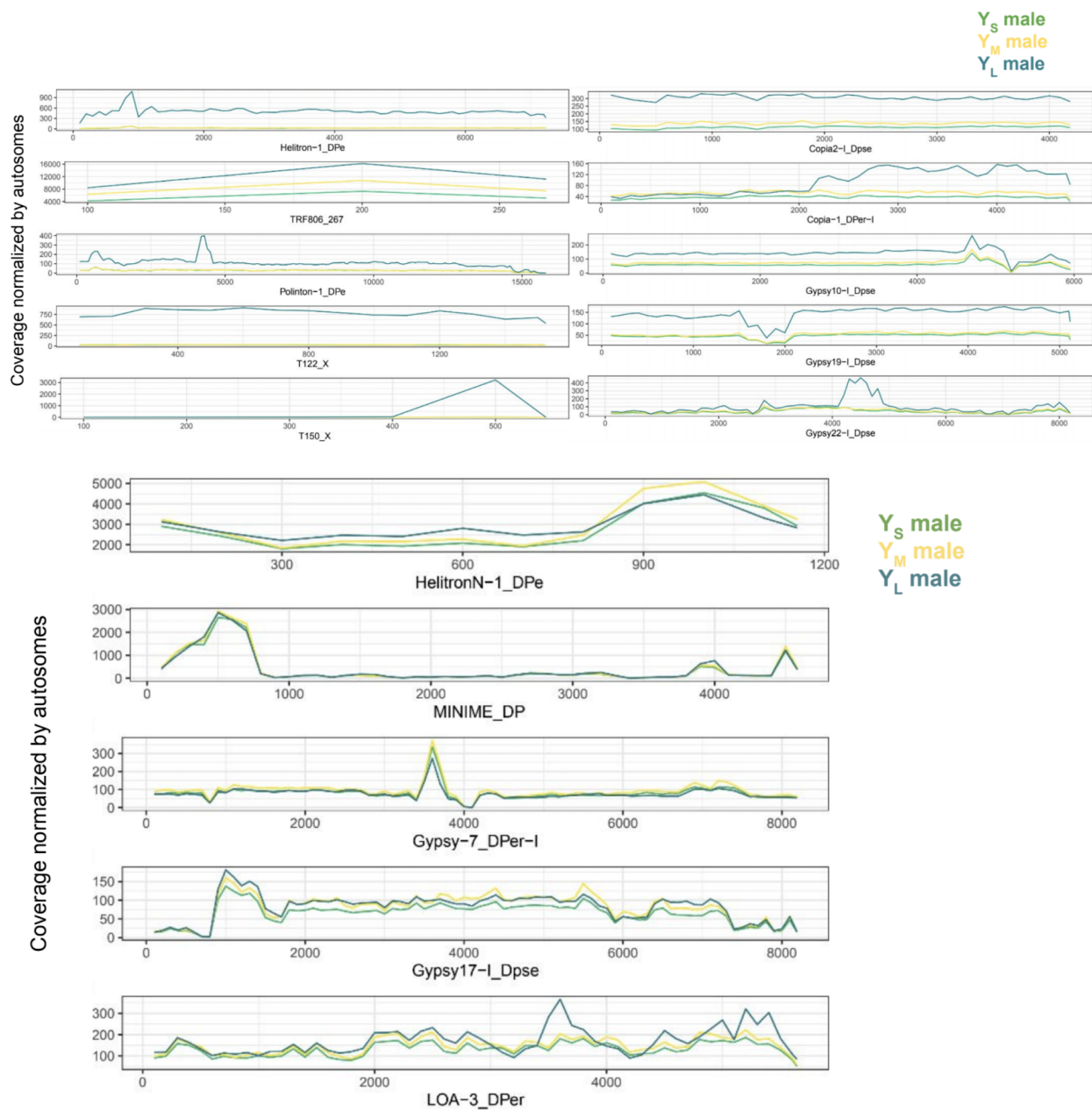


Figure S7.

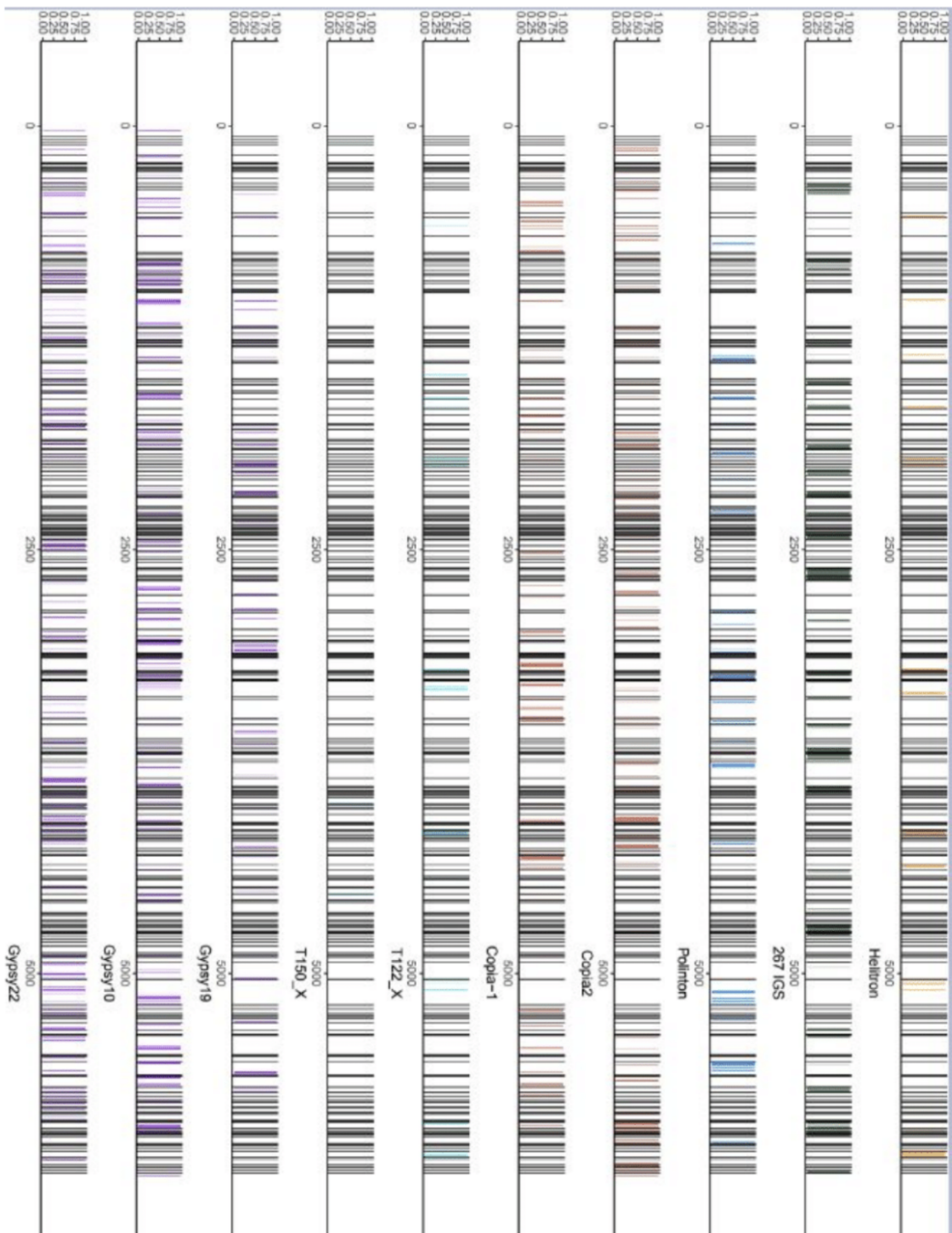


Figure S8A.

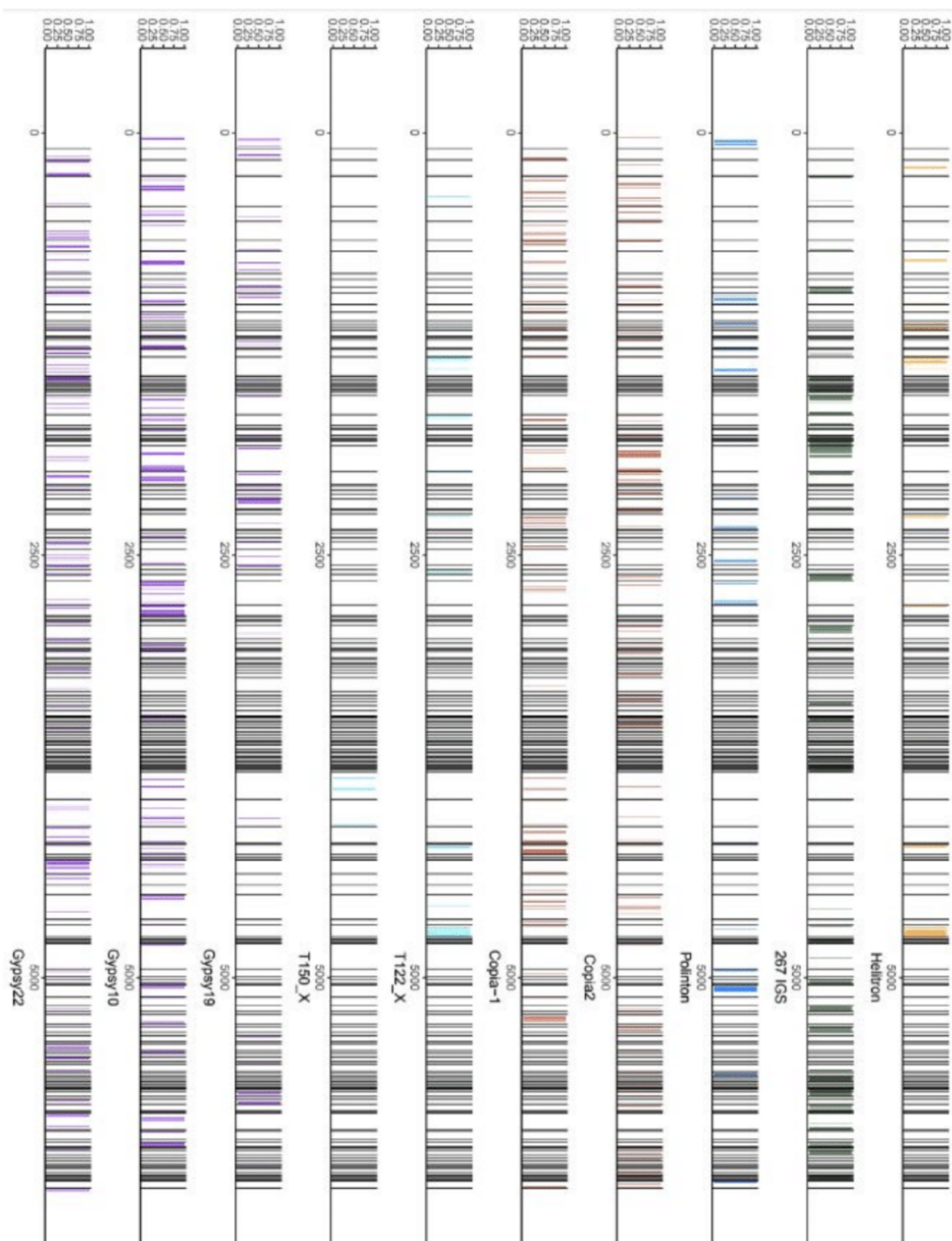


Figure S8B.

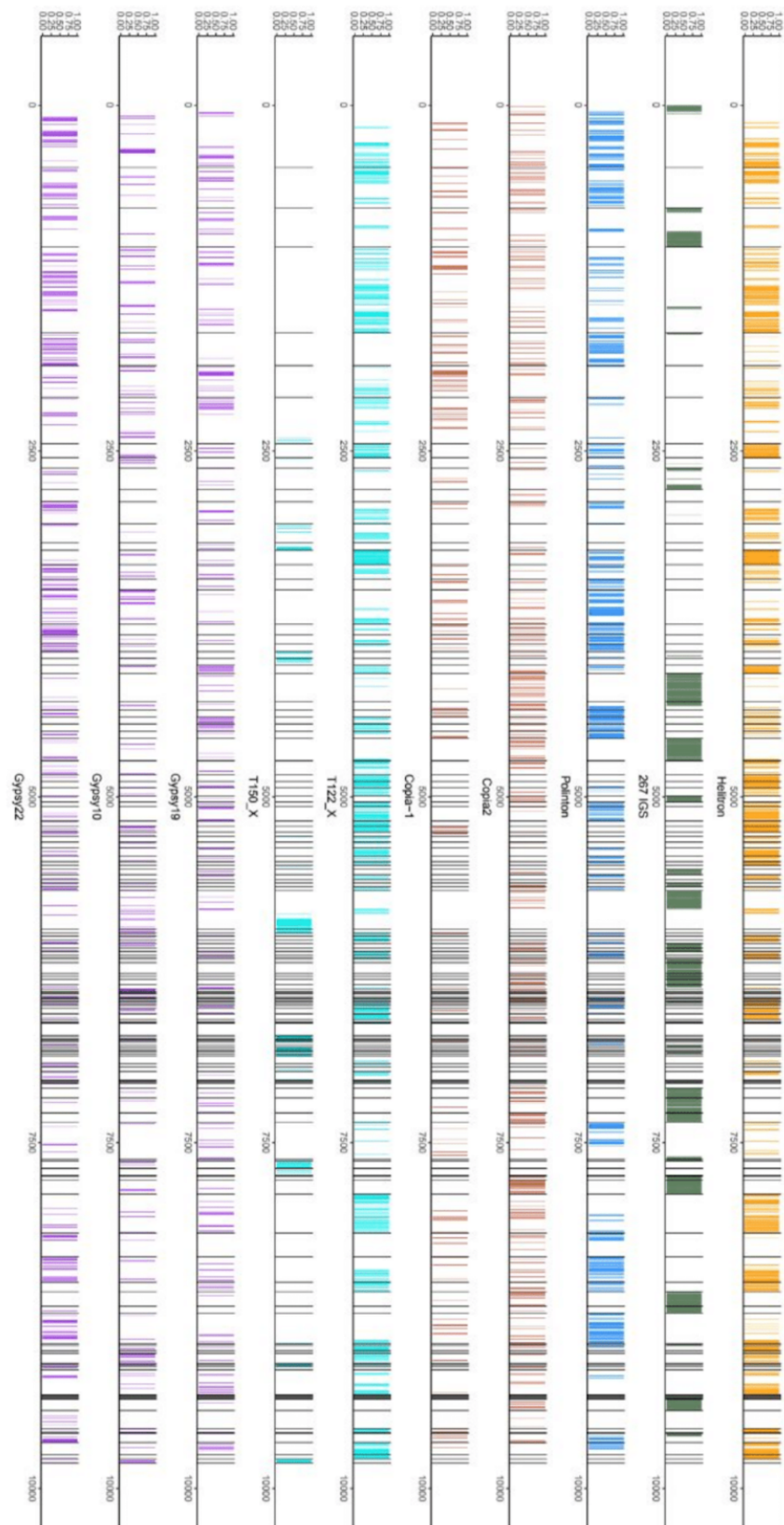


Figure S8C.

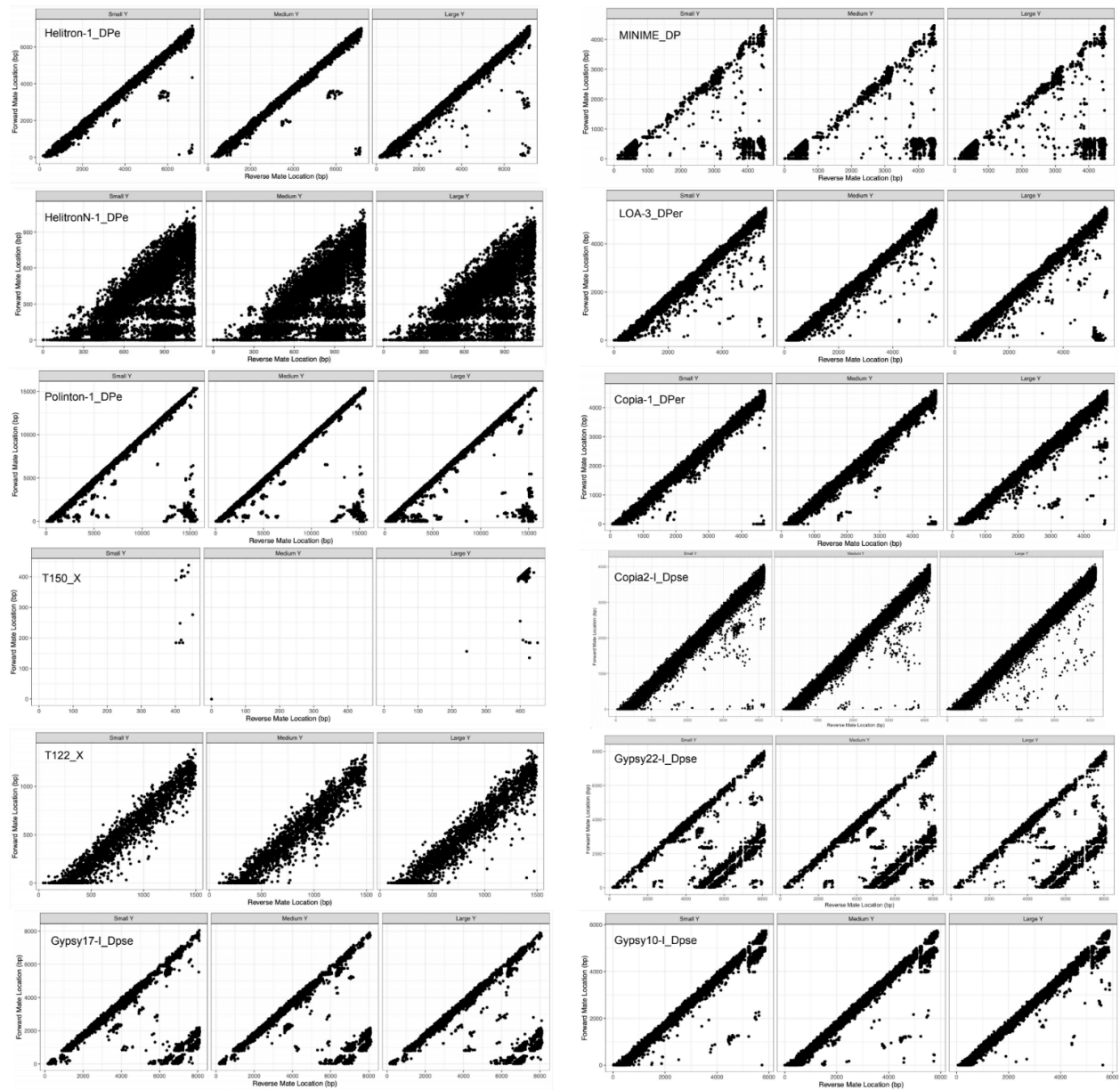


Figure S9.

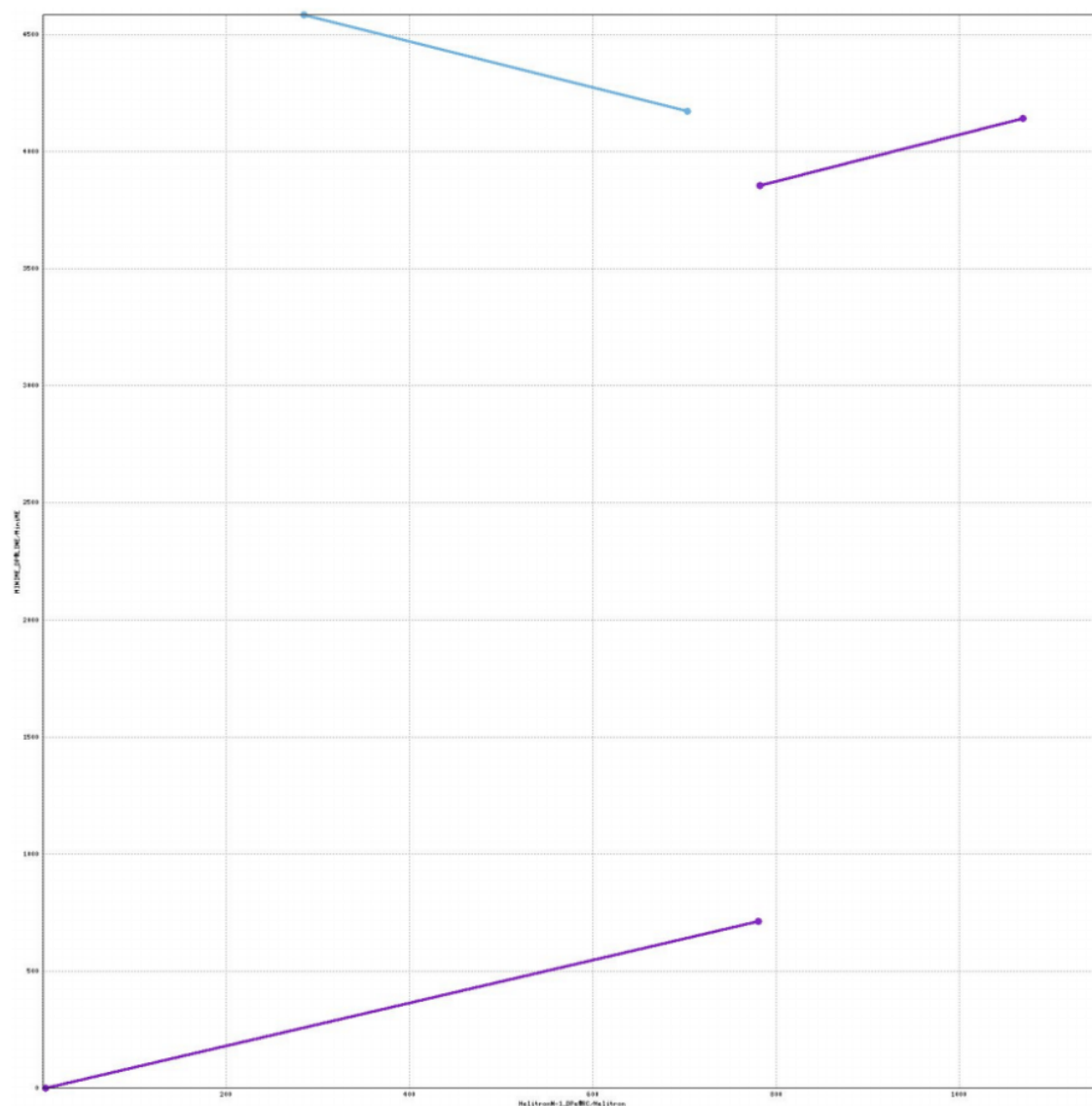


Figure S10.

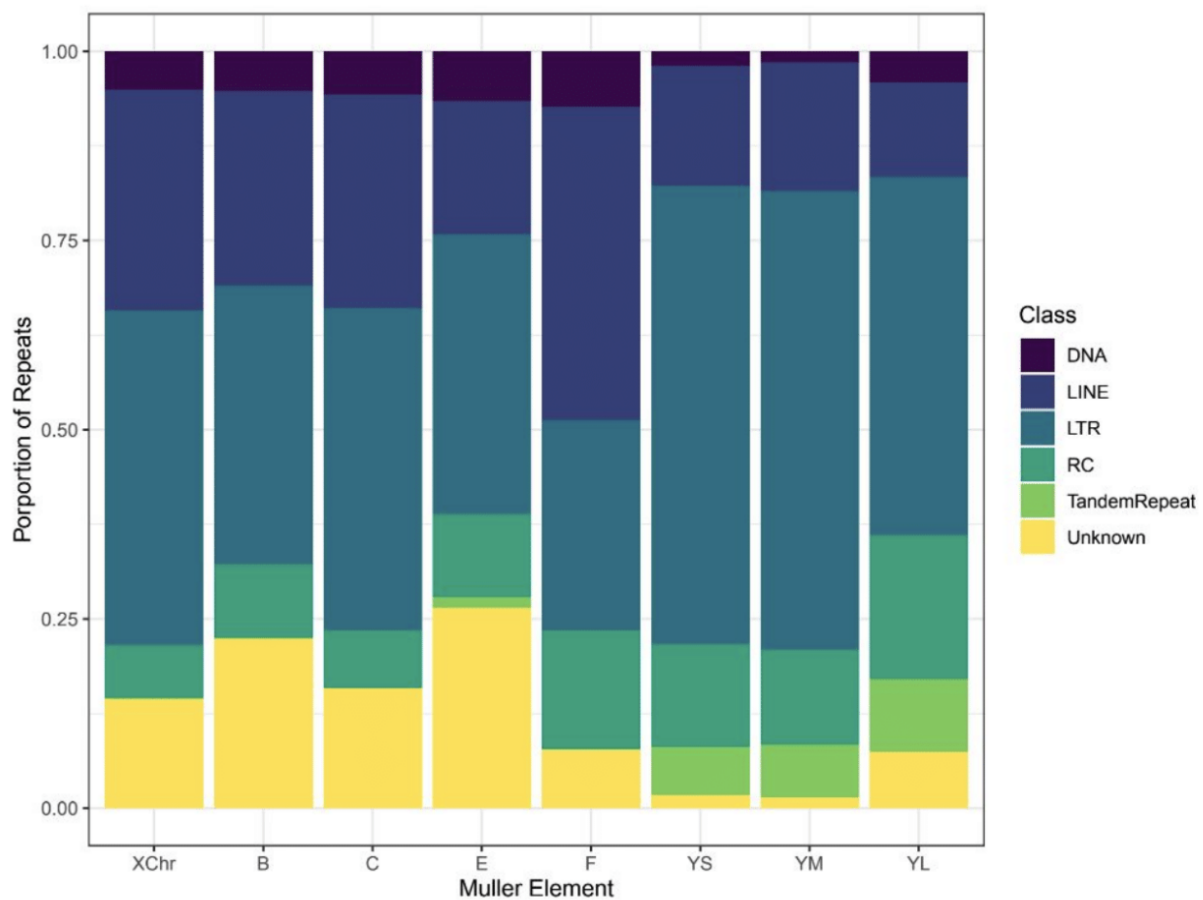


Figure S11.

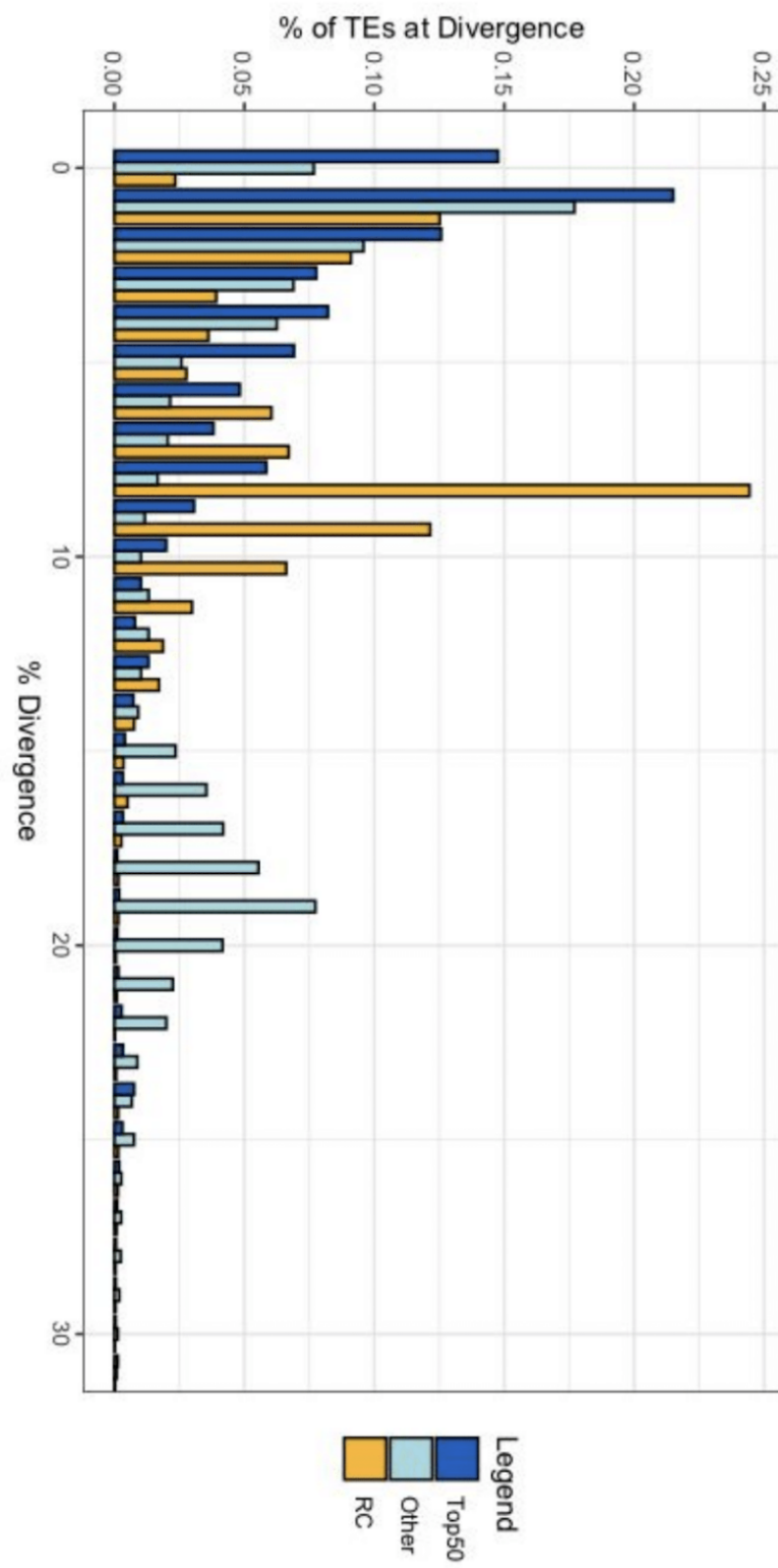


Figure S12.

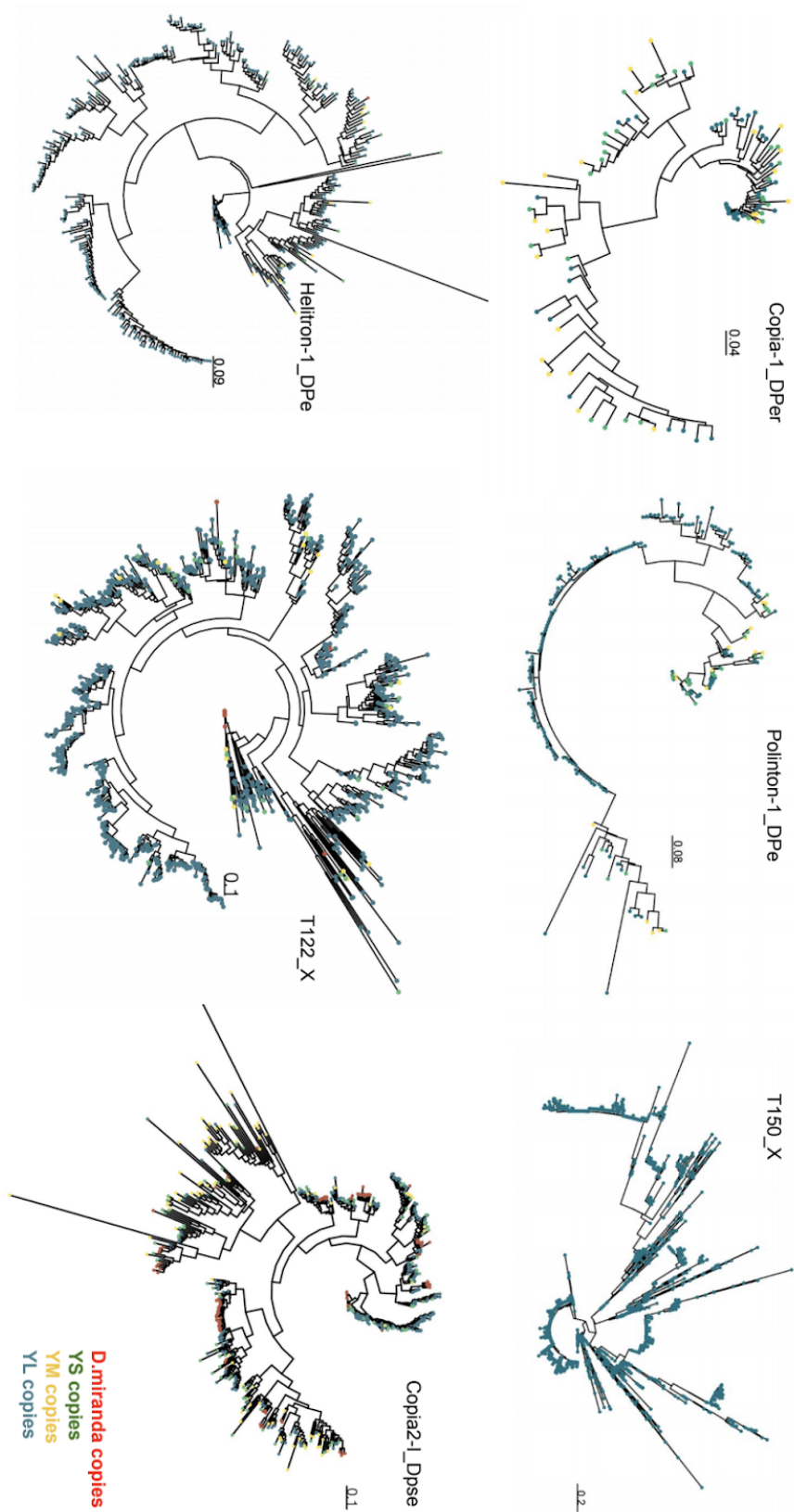


Figure S13.

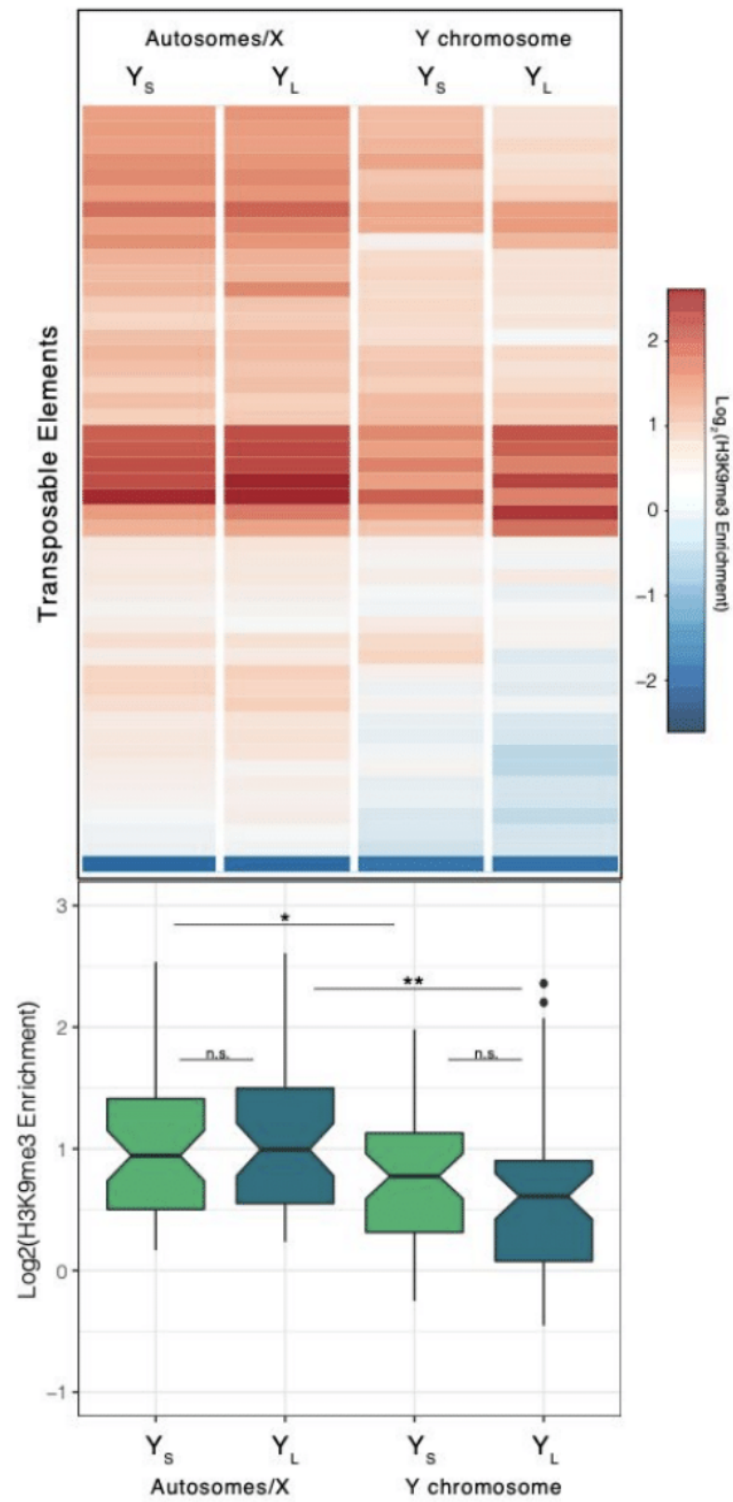


Figure S14.

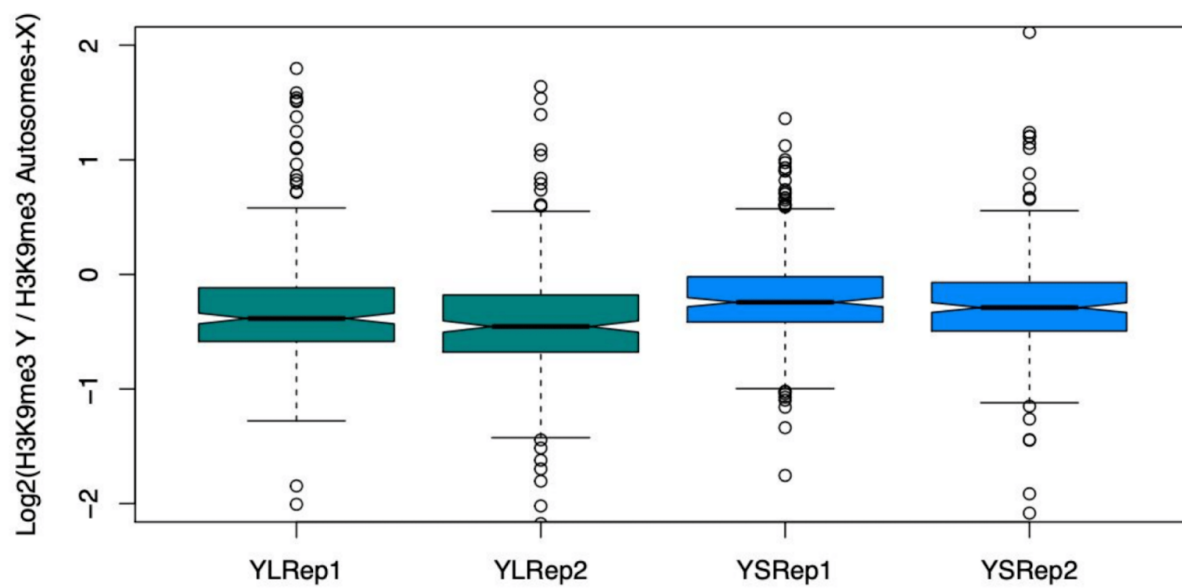


Figure S15.

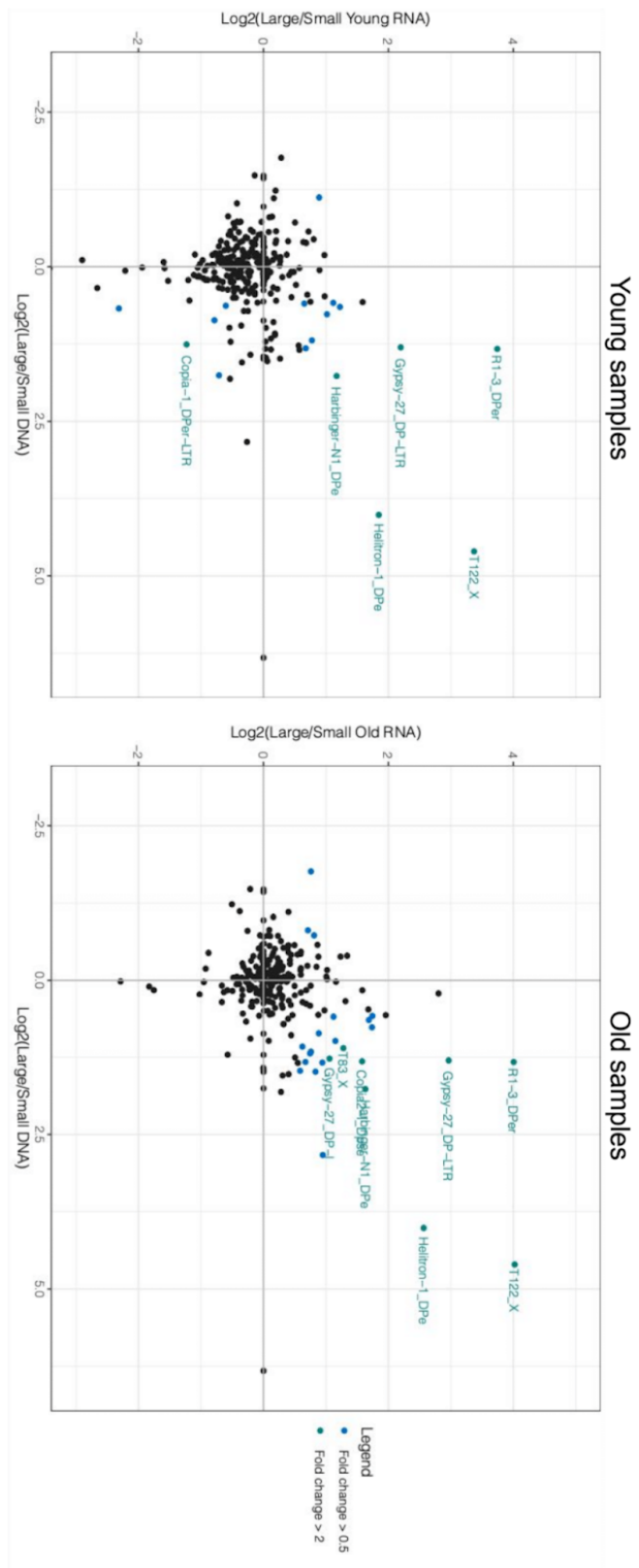
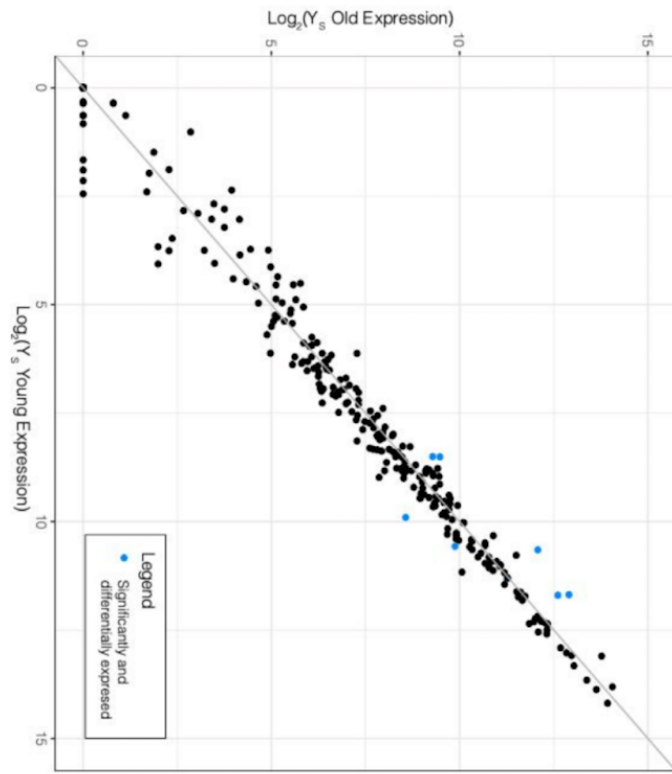


Figure S17.

Small Y-replacement line males



Large Y-replacement line males

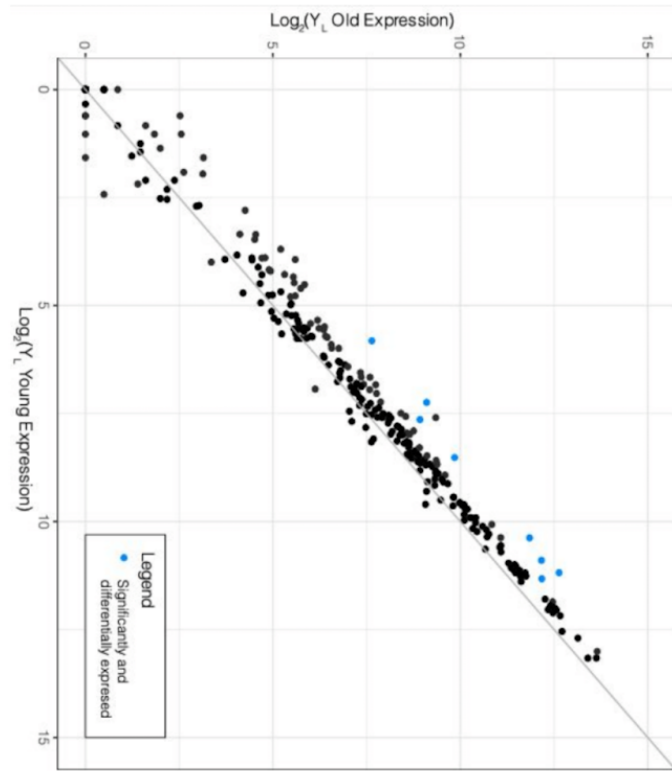


Figure S18.

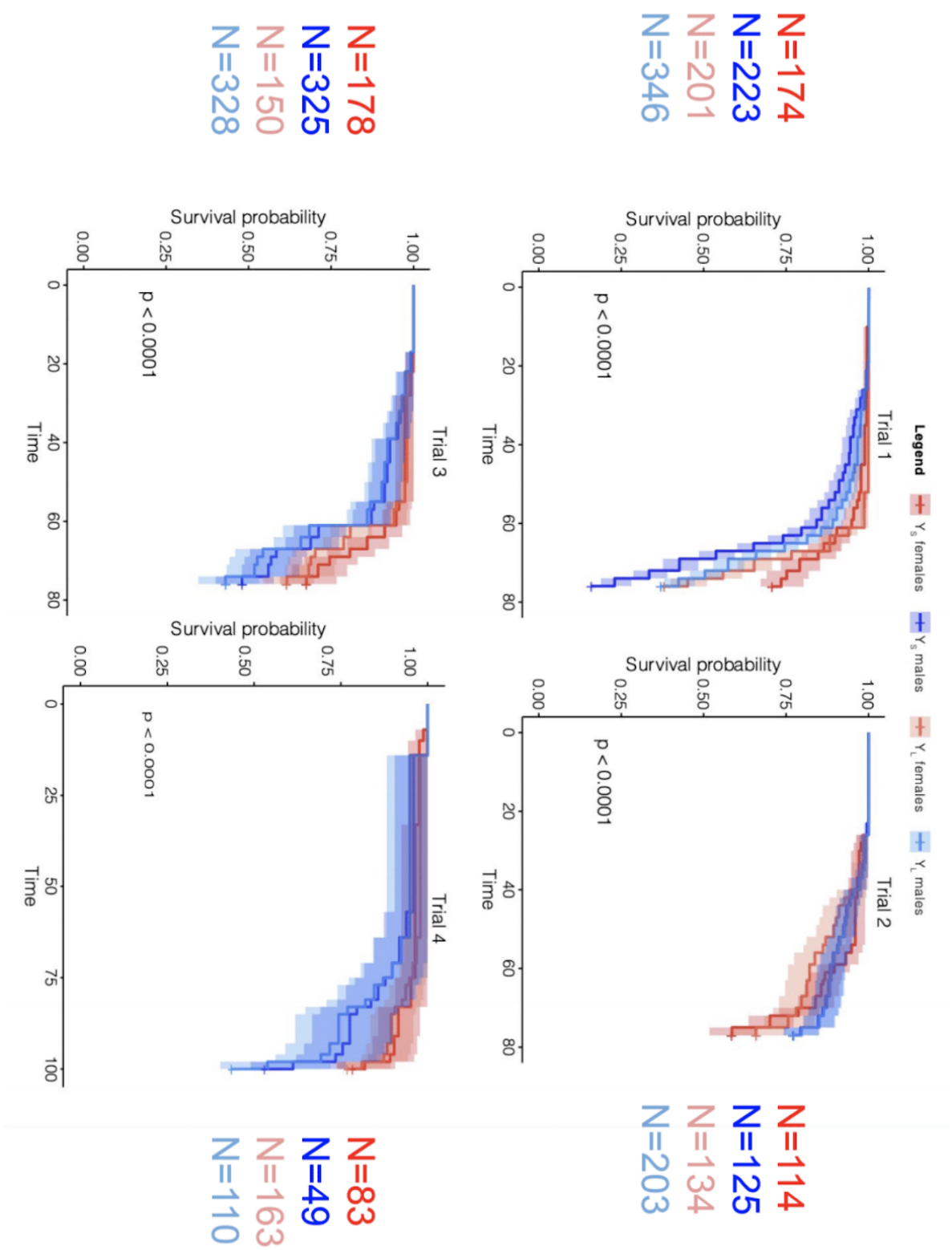


Figure S19.

Table S1. Starting wild-type strain information.

Original Strain	Location
Z17	Zion Park, UT
Z27	Zion Park, UT
Z31	Zion Park, UT
BdA BMC4	Bosque del Apache, NM
BdA BMC5	Bosque del Apache, NM
BdA BMC8	Bosque del Apache, NM
ST JR 84 B	James Reserve, CA
ST JR 138 L	James Reserve, CA
ST JR 158 L	James Reserve, CA
ST JR 209 L	James Reserve, CA
PP DM 1084 B	Daris Mts, TX
PP DM 1054 L	Daris Mts, TX
PP DM 1065 L	Daris Mts, TX
PP DM 1081 L	Daris Mts, TX
KB1 PSU 1097	Kaibab National Forest, AZ
KB3 PSU 1099	Kaibab National Forest, AZ
KB4 PSU 1100	Kaibab National Forest, AZ
KB10 PSU 1105	Kaibab National Forest, AZ
y;gl;or;inc	NA
D .100	Oaxaca, Mexico
DPSE .105	Mesa Verde, CO
DPSE .122	Sonora, Mexico
DPSE .124	Sinaloa, Mexico
DPSE .126	Chiapas, Mexico
DPSE .198	Pike's Peak, CO
RBB 06	Walnut Creek, CA
RBB 03	Walnut Creek, CA
SSR2	Mesa Verde, CO

Table S2. Inferred Y chromosome size by karyotype measurements.

Y donor	Y-replacement Line	Y/X ratio	SD	N
BdA BMC4	Y01	0.460	0.135	7
BdA BMC5	Y02	0.516	0.46	4
BdA BMC8	Y03	0.567	0.95	14
DPSE .100	Y04	0.534	0.079	4
DPSE .105	Y05	0.452	0.056	6
DPSE .122	Y06	0.546	0.085	7
DPSE .124	Large_Y	0.706	0.162	9
DPSE .126	Y07	0.475	0.081	14
DPSE .198	Y08	0.439	0.081	6
KB10 PSU 1105	Medium_Y	0.544	0.107	7
KB1 PSU 1097	Y09	0.523	0.097	8
KB3 PSU 1099	Y10	0.538	0.135	5
KB4 PSU 1100	Y11	0.585	0.034	3
PP DM 1054L	Y12	0.515	0.145	4
PP DM 1065L	Y13	0.496	0.089	4
PP DM 1081L	Y14	0.527	0.103	6
PP DM 1084B	Y15	0.532	0.048	7
RBB 03	Y16	0.52	0.062	8
RBB 06	Y17	0.586	0.103	7
SSR2	Small_Y	0.502	0.111	3
ST JR 138L	Y19	0.500	0.97	3
ST JR 158L	Y20	0.479	0.065	11
ST JR 209L	Y21	0.579	0.0765	3
ST JR 84B	Y18	0.494	0.066	7
y;gl;or;inc	Y22	0.469	0.156	7
Z17	Y24	0.421	0.654	3
Z27	Y27	0.487	0.069	7
Z31	Y23	0.485	0.121	5

Table S3. Inferred Y chromosome morphology by chromosome arm ratios (data underlying Fig. 1B).

Y donor	Y-replacement Line	Arm ratio (L:S)	SD	N	Morphology
BdA BMC4	Y01	1.5	0.3	7	Submetacentric
BdA BMC5	Y02	2.6	1.0	4	Acrocentric
BdA BMC8	Y03	2.9	0.8	16	Acrocentric
DPSE .100	Y04	2.6	0.8	4	Acrocentric
DPSE .105	Y05	3.0	0.8	6	Acrocentric
DPSE .122	Y06	2.9	0.4	8	Acrocentric
DPSE .124	Large_Y	1.1	0.1	12	Metacentric
DPSE .126	Y07	2.9	0.3	19	Acrocentric
DPSE .198	Y08	2.7	0.3	6	Acrocentric
KB10 PSU 1105	Medium_Y	2.1	0.5	11	Acrocentric
KB1 PSU 1097	Y09	2.9	0.4	8	Acrocentric
KB3 PSU 1099	Y10	2.6	0.5	5	Acrocentric
KB4 PSU 1100	Y11	2.3	0.2	3	Acrocentric
PP DM 1054L	Y12	2.8	0.6	4	Acrocentric
PP DM 1065L	Y13	1.6	0.4	7	Submetacentric
PP DM 1081L	Y14	3.7	1.2	6	Acrocentric
PP DM 1084B	Y15	2.9	0.4	8	Acrocentric
RBB 03	Y16	2.5	0.5	9	Acrocentric
RBB 06	Y17	2.5	0.4	7	Acrocentric
SSR2	Small_Y	2.7	0.4	3	Acrocentric
ST JR 138L	Y19	2.6	0.5	3	Acrocentric
ST JR 158L	Y20	3.1	0.5	11	Acrocentric
ST JR 209L	Y21	2.8	0.6	3	Acrocentric
ST JR 84B	Y18	2.9	0.7	7	Acrocentric
y;gl;or;inc	Y22	2.3	0.4	7	Acrocentric
Z17	Y24	3.7	0.5	3	Acrocentric
Z27	Y27	2.9	0.4	7	Acrocentric
Z31	Y23	1.6	0.4	7	Submetacentric

Table S4. Y-replacement male genome size estimates and morphology (data underlies Fig 1C).

Y-replacement line	Y donor	Genome size (Mb)	SD	N	Y Morphology
Female	SSR2 F	179.01	0.4	3	NA
Large_Y	DPSE .124	173.83	0.7	6	Metacentric
Medium_Y	KB10 PSU 1105	166.45	0.35	4	Acrocentric
Small_Y	SSR2 M	160.94	0.46	3	Acrocentric
Y01	BdA BMC4	160.9	0.23	3	Submetacentric
Y02	BdA BMC5	161.3	0.37	3	Acrocentric
Y03	BdA BMC8	161.65	1.02	3	Acrocentric
Y04	DPSE .100	160.46	0.27	3	Acrocentric
Y05	DPSE .105	161.22	0.7	4	Acrocentric
Y06	DPSE .122	162.03	0.83	3	Acrocentric
Y07	DPSE .126	161.14	0.73	3	Acrocentric
Y08	DPSE .198	160.98	0.27	3	Acrocentric
Y09	KB1 PSU 1097	161.73	0.48	3	Acrocentric
Y10	KB3 PSU 1099	159.08	0.19	3	Acrocentric
Y11	KB4 PSU 1100	160.02	1.12	5	Acrocentric
Y12	PP DM 1054 L	159.93	0.86	3	Acrocentric
Y13	PP DM 1065 L	160.84	0.66	3	Submetacentric
Y14	PP DM 1081 L	161.88	0.78	5	Acrocentric
Y15	PP DM 1084 B	160.88	0.64	3	Acrocentric
Y16	RBB03	160.4	0.4	3	Acrocentric
Y17	RBB06	160.34	0.21	2	Acrocentric
Y18	ST JR 84 B	161.55	0.56	3	Acrocentric
Y19	ST JR 138 L	160.89	1.7	3	Acrocentric
Y20	ST JR 158 L	162.01	0.46	3	Acrocentric
Y21	ST JR 209 L	161.02	0.9	4	Acrocentric
Y22	y;gl;or;inc	158.54	0.65	4	Acrocentric
Y23	Z31	161.6	0.89	4	Submetacentric
Y24	Z17	160.74	1.27	5	Acrocentric
Y25	Z27	160.33	0.32	3	Acrocentric

Table S5. Heterochromatin estimates of Y-replacement lines (data underlying Figure 1D).

Line	%AT (rel. to <i>D. virilis</i>)	%AT (rel. to <i>D. virilis</i>) SD	4C:2C DNA Thorax	4C:2C Thorax SD
Large_Y	0.404	0.0061	1.858	0.069
Medium_Y	0.385	0.0038	1.87	0.063
Small_Y	0.369	0.0043	1.923	0.05

Table S6. Summary statistics of Y chromosome assemblies. Note that the assemblies are not adjusted for the copy number of each contig.

Strain	QuickMerge direction	N50	Max Scaffold	Genome Size (without coverage)	No. scaffolds	GC content	Repeat Masked	% Repeat Masked
YS	(Canu to Falcon) to (Falcon to Canu)	215,074	1,097,706	34,545,044	246	44.64	30,461,567	88.18%
YM	(Canu to Falcon) to (Falcon to Canu)	330,156	1,009,159	34,567,851	201	45.01	30,742,053	88.93%
YL	(Canu to Falcon) to (Falcon to Canu)	816,342	3,351,422	55,037,118	139	42.52	48,090,697	87.38%

Table S7. BioNano hybrid scaffolds for Y_L assembly (lengths in Mb). Additional information can be found in Supplementary Material 2.

	Whole-genome	Y chromosome
Contigs	32	26
Min length	0.133	0.133238
Median	1.582	1.3921135
Mean	6.27	1.7411925
N50	30.573	3.272912
Max	35.23	4.459917
Total	200.631	45.271005

Table S8. Transposable element base pair contribution by Illumina data. Numbers represent whole-genome estimates and any differences will be due to the Y. Data underlies Figure 3A.

Transposable element	Class	Length	YS	YM	YL	YS	YM	YL
T150_X	Betancourt_Unknown_TE	549	16.3	14.7	1303.2	8936.9	8094.0	715481.1
T122_X	Betancourt_Unknown_TE	1524	69.5	75.8	1691.3	105986.7	115547.9	2577516.8
BEL-1_DPer-I	BEL	7788	146.5	168.1	89.0	1141203.8	1309470.3	693266.1
BEL-1_DPer-LTR	BEL	367	215.7	234.2	157.5	79150.1	85956.7	57820.1
BEL1_I_Dpse	BEL	5439	79.8	73.4	135.9	433834.6	399024.7	739262.0
BEL1_LTR_Dpse	BEL	300	74.7	66.0	145.5	22406.7	19786.2	43643.5
BEL-13_DP-I	BEL	5493	214.3	232.5	254.5	1176921.4	1277073.1	1397758.9
BEL-13_DP-LTR	BEL	476	429.3	465.8	551.6	204359.1	221715.4	262570.7
BEL-14_DP-I	BEL	5617	101.4	107.7	82.5	569293.8	604987.1	463175.4
BEL-14_DP-LTR	BEL	473	229.7	247.3	197.2	108671.7	116950.9	93265.4
BEL-15_DP-I	BEL	6145	20.0	20.8	18.2	123164.2	128071.7	111855.9
BEL-15_DP-LTR	BEL	288	48.9	55.7	42.7	14076.6	16042.2	12285.8
BEL-2_DPer-I	BEL	5368	126.5	127.7	113.5	679006.3	685587.0	609450.6
BEL-2_DPer-LTR	BEL	564	218.1	221.9	209.6	123016.1	125136.7	118239.4
BEL2-I_Dpse	BEL	7585	194.3	185.3	160.8	1473481.5	1405430.2	1220029.2
BEL2-LTR_Dpse	BEL	207	191.8	184.3	134.9	39694.6	38160.0	27924.9
BEL-3_DPer-I	BEL	5645	412.3	419.5	418.1	2327303.3	2368318.1	2360038.7
BEL-3_DPer-LTR	BEL	351	758.3	780.3	770.3	266152.6	273884.4	270383.4
BEL3-I_Dpse	BEL	5868	36.6	41.5	41.2	214859.3	243475.8	241684.6
BEL3-LTR_Dpse	BEL	319	74.5	92.9	94.1	23777.5	29643.3	30023.8
BEL-4_DPer-I	BEL	6419	107.9	113.3	101.6	692929.7	727257.6	652101.2
BEL-4_DPer-LTR	BEL	660	188.9	196.7	169.9	124656.8	129837.5	112122.0
BEL-4_DPse-I	BEL	6141	15.4	18.5	14.8	94575.7	113400.9	90794.6
BEL-4_DPse-LTR	BEL	300	56.2	68.8	50.6	16856.9	20629.7	15174.0
BEL-5_DPer-I	BEL	5615	211.8	244.4	224.9	1189245.4	1372255.7	1262804.0
BEL-5_DPer-LTR	BEL	371	848.7	1007.2	1078.7	314872.1	373656.0	400209.1
BEL-6_DPe-I	BEL	5707	46.4	60.3	28.0	264576.7	344154.3	159821.0
BEL-6_DPe-LTR	BEL	144	77.0	94.7	57.8	11092.3	13631.5	8322.9

BEL-6_DPer-I	BEL	6160	55.0	63.6	46.2	338772.2	391819.4	284727.2
BEL-6_DPer-LTR	BEL	364	112.5	130.1	96.4	40938.0	47342.6	35099.7
BEL-7_DPe-I	BEL	6378	43.1	50.8	38.3	274865.5	323731.2	244337.2
BEL-7_DPe-LTR	BEL	227	97.8	118.7	81.1	22210.1	26943.0	18403.7
BEL-7_DPer-I	BEL	5639	17.5	23.6	11.9	98595.8	133134.3	66846.3
BEL-7_DPer-LTR	BEL	130	14.4	18.6	10.5	1876.0	2420.8	1361.1
BEL-8_DPe-I	BEL	6104	178.8	203.7	253.2	1091383.5	1243460.3	1545509.5
BEL-8_DPe-LTR	BEL	316	331.6	381.3	422.5	104799.2	120490.5	133516.5
Copia-1_DPer-I	Copia	4739	88.7	128.5	205.6	420482.8	608790.9	974407.3
Copia-1_DPer-LTR	Copia	245	140.3	212.1	335.2	34361.6	51958.6	82116.3
Copia1-I_Dpse	Copia	3027	45.4	43.2	46.2	137361.5	130819.0	139812.2
Copia1-LTR_Dpse	Copia	214	74.8	69.4	70.0	16015.9	14856.6	14973.8
Copia-2_DPer-I	Copia	4008	37.2	36.8	36.5	149213.8	147485.8	146259.9
Copia-2_DPer-LTR	Copia	264	83.8	81.4	82.6	22112.8	21493.0	21800.8
Copia-29_DP-I	Copia	4104	159.0	183.9	144.8	652417.3	754592.4	594184.3
Copia-29_DP-LTR	Copia	207	311.4	379.1	301.0	64458.4	78467.5	62307.4
Copia2-I_Dpse	Copia	4185	266.7	328.5	665.2	1116006.1	1374971.5	2783997.9
Copia2-LTR_Dpse	Copia	494	366.5	449.2	927.3	181026.9	221899.6	458084.0
CR1-1_DPer	CR1	4691	790.4	841.5	802.8	3707869.3	3947276.1	3766048.4
CR1-2_DPer	CR1	3331	68.2	69.3	69.9	227092.7	230942.7	232942.8
CR1-3_DPer	CR1	4396	20.7	21.0	20.0	90974.6	92501.5	87922.5
Daff_Helitron_4	Helitron	2692	6.4	6.2	17.6	17331.6	16642.1	47478.4
Daff_Jockey_18	Jockey	6275	1.9	1.9	1.8	11861.7	11839.0	11520.9
Daff_PIF-Harbinger_199	Harbinger	728	0.1	0.1	0.0	73.7	71.7	27.3
Dbus_piggyBac	piggyBac	1900	0.0	0.0	0.0	9.9	0.0	3.8
Dlow_Gypsy_2	Gypsy	1629	42.7	51.9	45.3	69515.5	84553.8	73787.3
Dlow_Gypsy_64	Gypsy	5272	2.5	2.7	2.5	13205.7	14384.4	13040.4
Dlow_Helitron_184	Helitron	563	53.0	54.7	56.2	29836.0	30820.7	31659.1
Dlow_Helitron_5	Helitron	6444	5.1	5.0	5.5	32831.0	32517.8	35754.4
Dlow_Jockey_185	Jockey	997	4.6	4.9	4.6	4604.9	4868.8	4595.8
Dlow_Loa_58	Loa	6596	1.6	1.5	2.1	10366.8	9708.7	13955.8
Dlow_Mariner_3	Mariner	1569	1.4	1.4	1.6	2180.6	2129.7	2454.4
Dmir_hAT-Blackjac	hAT	1949	96.8	102.1	108.3	188751.9	199013.7	211002.4

k_169								
Dmir_I_149	I	2591	20.7	21.6	22.1	53725.0	55960.2	57202.0
Dmir_P_226	P	1690	7.2	7.1	9.4	12222.6	11997.5	15808.4
Dmir_TcMar-Tc1_1 57	TcMar-Tc1	925	110.7	119.6	129.2	102401.2	110605.9	119540.1
DNA-1_DPe	DNA	732	205.0	207.9	234.7	150056.7	152199.5	171831.2
Dper_Gypsy_134	Gypsy	1097	176.8	211.1	277.6	193931.9	231543.8	304515.6
Dper_Gypsy_138	Gypsy	4063	11.2	10.9	10.4	45355.0	44411.2	42255.7
Dper_hAT-hATm_13 7	hAT	1191	18.4	18.3	21.9	21884.8	21846.6	26122.4
Dper_Jockey_145	Jockey	2369	19.9	23.1	15.3	47111.6	54721.4	36157.4
Dper_Jockey_70	Jockey	7045	22.3	20.8	24.0	156843.9	146775.3	168806.0
Dper_loa_7	LOA	2655	19.6	18.0	49.2	51990.1	47836.4	130597.0
Dper_piggyBac	piggyBac	2829	12.7	13.1	12.5	35971.5	37170.9	35364.1
Dper_TcMar-Tc1_13 6	TcMar-Tc1	1588	55.1	57.2	60.9	87497.1	90836.7	96787.1
Dper_TcMar-Tc1_14 4	TcMar-Tc1	797	36.7	33.6	60.3	29255.6	26808.4	48075.0
Dper_TcMar-Tc1_14 7	TcMar-Tc1	1639	24.1	24.6	31.9	39509.5	40351.1	52230.3
Dpse_CMC-EnSpm _97	CMC-EnSpm	2723	19.7	20.4	29.4	53536.1	55609.6	80137.0
Dpse_Gypsy_6	Gypsy	3663	392.9	433.9	381.8	1439132.2	1589378.1	1398586.5
Dpse_Gypsy_79	Gypsy	2537	124.3	163.3	84.1	315404.2	414410.4	213329.4
Dpse_hAT-hATm_1 25	hAT	2459	14.4	14.9	16.5	35459.4	36741.6	40485.0
Dpse_hAT-Pegasus_ 77	hAT	1807	23.2	21.8	27.5	41942.1	39336.4	49740.5
Dpse_P_121	P	1143	16.6	18.0	18.7	18956.7	20536.7	21356.0
Dpse_TcMar-Tc1_1 20	TcMar-Tc1	1402	23.0	25.3	29.2	32184.9	35460.2	40975.7
Dpse_TcMar-Tc1_9 3	TcMar-Tc1	1634	32.2	32.7	35.2	52662.1	53479.6	57494.8
Gal_DPer	Gal	2797	10.4	11.4	15.2	29194.4	31965.6	42653.2
Gal_DPse	Gal	2439	28.0	30.4	63.9	68276.8	74058.3	155862.6
Gypsy-1_DPer-I	Gypsy	7519	43.8	52.7	28.3	329618.1	395892.6	212606.6

Gypsy-1_DPer-LTR	Gypsy	817	68.2	80.6	51.1	55685.5	65831.4	41724.7
Gypsy-10_DPer-I	Gypsy	5007	46.5	61.7	40.0	232697.5	308818.5	200030.7
Gypsy-10_DPer-LTR	Gypsy	1031	78.5	95.5	70.5	80931.5	98436.4	72734.7
Gypsy10-I_Dpse	Gypsy	5950	141.9	175.8	300.2	844350.5	1045779.7	1786237.1
Gypsy10-LTR_Dpse	Gypsy	1848	136.5	167.6	270.9	252184.5	309725.6	500598.0
Gypsy-11_DPer-I	Gypsy	8190	113.0	117.9	86.6	925334.0	965824.5	709337.1
Gypsy-11_DPer-LTR	Gypsy	2035	235.3	233.9	178.7	478810.4	475984.1	363592.9
Gypsy11-I_Dpse	Gypsy	8190	54.1	59.1	79.8	442753.0	484194.9	653856.3
Gypsy11-LTR_Dpse	Gypsy	817	112.8	122.5	154.5	92192.0	100048.2	126190.5
Gypsy-12_DPer-I	Gypsy	6129	69.1	72.2	41.8	423357.9	442391.5	256189.2
Gypsy-12_DPer-LTR	Gypsy	786	120.3	124.8	91.7	94544.8	98116.0	72066.8
Gypsy12-I_Dpse	Gypsy	8190	51.2	51.0	63.3	419420.0	417525.5	518632.5
Gypsy12-LTR_Dpse	Gypsy	579	175.6	183.3	288.5	101655.9	106133.7	167022.5
Gypsy13-I_Dpse	Gypsy	7268	13.6	19.0	13.0	98847.1	138361.9	94197.6
Gypsy13-LTR_Dpse	Gypsy	374	25.8	34.0	24.7	9642.6	12708.6	9233.2
Gypsy14-I_Dpse	Gypsy	3604	53.9	65.8	157.6	194269.1	237091.6	567823.6
Gypsy14-LTR_Dpse	Gypsy	1179	66.7	81.3	192.1	78613.8	95842.9	226481.3
Gypsy15-I_Dpse	Gypsy	6155	35.8	38.4	35.0	220084.7	236488.7	215692.6
Gypsy15-LTR_Dpse	Gypsy	2568	70.0	72.8	67.6	179790.2	186828.8	173527.1
Gypsy16-I_Dpse	Gypsy	6852	33.8	44.0	66.9	231401.8	301256.7	458721.5
Gypsy16-LTR_Dpse	Gypsy	611	22.5	28.4	52.2	13750.0	17354.5	31912.3
Gypsy17-I_Dpse	Gypsy	8190	154.8	194.1	176.8	1267758.4	1589474.4	1447681.8
Gypsy17-LTR_Dpse	Gypsy	1367	45.8	61.5	46.6	62666.5	84043.9	63669.0
Gypsy18-I_Dpse	Gypsy	7727	986.5	892.1	992.7	7622592.0	6893161.0	7670241.2
Gypsy18-LTR_Dpse	Gypsy	184	29.4	32.6	41.5	5418.1	5998.6	7630.1
Gypsy19-I_Dpse	Gypsy	5115	113.0	128.0	313.4	578010.1	654841.3	1602820.5
Gypsy19-LTR_Dpse	Gypsy	325	213.8	230.5	599.0	69482.1	74898.6	194661.5
Gypsy-1A_DPer-LTR	Gypsy	826	27.0	27.6	11.5	22271.6	22779.4	9488.6
Gypsy1-I_Dpse	Gypsy	4010	11.2	11.0	7.6	44899.0	43938.5	30285.1
Gypsy1-LTR_Dpse	Gypsy	201	25.9	22.8	19.0	5198.5	4581.7	3822.9

Gypsy-2_DPer-I	Gypsy	4308	28.8	29.8	77.4	124218.9	128464.2	333265.1
Gypsy-2_DPer-LTR	Gypsy	110	40.0	46.5	135.1	4399.3	5112.7	14864.1
Gypsy20-I_Dpse	Gypsy	5433	97.6	110.5	102.9	530179.3	600474.8	559263.0
Gypsy20-LTR_Dpse	Gypsy	168	74.7	77.6	83.4	12546.1	13035.1	14006.7
Gypsy21-I_Dpse	Gypsy	3995	151.4	161.8	160.9	605000.5	646411.1	642634.9
Gypsy21-LTR_Dpse	Gypsy	230	268.3	280.1	301.6	61711.3	64423.0	69366.7
Gypsy22-I_Dpse	Gypsy	8190	91.9	107.9	205.7	752440.1	883301.3	1684953.4
Gypsy22-LTR_Dpse	Gypsy	355	45.2	60.2	322.3	16044.1	21357.3	114410.0
Gypsy23-I_Dpse	Gypsy	8190	19.7	20.0	27.9	161256.0	163879.3	228526.8
Gypsy23-LTR_Dpse	Gypsy	2764	15.8	15.9	13.6	43722.3	44080.7	37587.8
Gypsy24-I_Dpse	Gypsy	6428	41.0	44.4	39.1	263757.3	285428.1	251159.4
Gypsy24-LTR_Dpse	Gypsy	1263	28.6	34.6	28.4	36155.2	43756.2	35926.4
Gypsy-25_DP-I	Gypsy	5558	10.0	11.8	4.9	55303.8	65528.4	27197.4
Gypsy-25_DP-LTR	Gypsy	1442	27.6	33.2	15.9	39815.8	47826.1	22879.3
Gypsy-25_DPse-I	Gypsy	4618	28.4	35.8	27.3	131301.1	165113.6	125949.8
Gypsy-25_DPse-LTR	Gypsy	3362	41.3	56.0	41.9	138912.3	188382.7	141031.7
Gypsy-26_DP-I	Gypsy	5621	20.5	24.1	18.3	115432.9	135532.6	103086.3
Gypsy-26_DP-LTR	Gypsy	2159	56.2	63.1	47.8	121327.9	136312.6	103145.3
Gypsy-26_DPse-I	Gypsy	6105	105.3	125.9	159.2	642605.3	768656.7	972041.7
Gypsy-26_DPse-LTR	Gypsy	706	102.2	118.4	186.3	72134.8	83606.5	131506.0
Gypsy-27_DP-I	Gypsy	4725	55.4	63.4	134.1	261781.8	299414.9	633804.8
Gypsy-27_DP-LTR	Gypsy	219	96.6	107.8	238.4	21146.8	23613.9	52201.7
Gypsy2-I_Dpse	Gypsy	4676	51.3	46.6	70.1	239697.7	217960.2	327704.6
Gypsy2-LTR_Dpse	Gypsy	198	116.2	113.7	170.1	23004.5	22511.4	33676.8
Gypsy-3_DPer-I	Gypsy	5851	62.0	79.9	49.4	362692.2	467288.6	289026.2
Gypsy-3_DPer-LTR	Gypsy	2254	115.9	146.1	102.9	261127.8	329348.9	231845.3
Gypsy-31_DP-I	Gypsy	3633	17.6	16.4	18.3	63821.1	59522.6	66313.3
Gypsy-31_DP-LTR	Gypsy	1413	14.6	13.5	15.2	20645.2	19139.0	21546.2
Gypsy-32_DP-I	Gypsy	5705	15.3	15.5	22.7	87268.7	88534.9	129764.8
Gypsy-32_DP-LTR	Gypsy	1556	37.2	37.8	55.7	57805.9	58757.8	86722.5
Gypsy3-I_Dpse	Gypsy	6233	27.2	30.5	18.9	169385.0	189816.5	117704.0
Gypsy3-LTR_Dpse	Gypsy	527	26.9	28.5	16.3	14178.5	15001.5	8566.1

Gypsy-4_DPer-I	Gypsy	3411	9.4	10.3	6.6	32024.2	35261.1	22385.2
Gypsy-4_DPer-LTR	Gypsy	909	17.6	16.5	14.6	16011.8	15042.5	13304.5
Gypsy4-I_Dpse	Gypsy	6028	16.0	17.2	12.9	96624.3	103636.8	77646.3
Gypsy4-LTR_Dpse	Gypsy	1616	31.9	36.1	31.0	51629.2	58304.9	50130.8
Gypsy-5_DPer-I	Gypsy	5234	95.2	101.4	114.8	498120.1	530473.5	601028.7
Gypsy-5_DPer-LTR	Gypsy	199	170.8	174.0	215.7	33992.1	34624.6	42918.9
Gypsy5-I_Dpse	Gypsy	6846	19.1	24.8	21.0	130735.7	169719.2	143499.2
Gypsy5-LTR_Dpse	Gypsy	1223	22.5	30.6	25.1	27486.7	37430.1	30728.9
Gypsy-6_DPer-I	Gypsy	5904	15.9	14.5	13.2	94076.1	85331.0	77894.3
Gypsy-6_DPer-LTR	Gypsy	1868	35.6	33.9	31.9	66450.5	63337.8	59668.5
Gypsy6-I_Dpse	Gypsy	4951	32.6	28.4	22.2	161333.8	140673.8	109666.0
Gypsy6-LTR_Dpse	Gypsy	887	34.5	25.4	21.2	30623.6	22559.5	18786.5
Gypsy-7_DPer-I	Gypsy	8190	193.8	225.4	165.1	1587314.6	1845840.7	1351878.0
Gypsy-7_DPer-LTR	Gypsy	1657	488.6	526.1	458.7	809552.2	871734.7	760092.5
Gypsy7-I_Dpse	Gypsy	7039	16.6	20.7	12.8	117045.9	145594.3	90064.7
Gypsy7-LTR_Dpse	Gypsy	551	30.5	32.3	24.3	16820.0	17817.6	13406.6
Gypsy-8_DPer-I	Gypsy	5638	17.9	20.3	16.4	100980.5	114495.1	92441.1
Gypsy-8_DPer-LTR	Gypsy	2165	43.0	45.0	24.6	93199.0	97373.0	53203.4
Gypsy8-I_Dpse	Gypsy	7193	66.0	79.5	49.1	474414.0	571999.1	353263.8
Gypsy8-LTR_Dpse	Gypsy	1488	133.9	146.7	90.0	199269.5	218256.8	133894.9
Gypsy-9_DPer-I	Gypsy	4896	74.1	85.9	64.9	362785.1	420374.5	317861.9
Gypsy-9_DPer-LTR	Gypsy	1126	126.2	138.9	108.8	142134.4	156352.5	122527.4
Gypsy9-I_Dpse	Gypsy	4981	10.7	11.4	7.9	53381.6	56748.7	39254.0
Gypsy9-LTR_Dpse	Gypsy	762	66.6	65.9	58.2	50757.7	50215.0	44346.8
Harbinger-N1_DPe	Harbinger	1071	52.2	55.7	177.7	55920.2	59610.8	190329.7
hAT-1_DP	hAT	2699	26.7	22.2	25.3	72033.6	59947.0	68389.3
hAT-1N_DP	hAT	557	72.9	69.4	86.6	40611.9	38641.8	48243.5
Helitron-1_DPe	Helitron	7229	67.5	70.1	1090.7	488296.3	506964.6	7884806.5
HelitronN-1_DPe	Helitron	1155	6435.6	7068.8	6392.8	7433153.4	8164421.5	7383694.1
Hopers2	Hopers2	2756	140.8	164.3	208.6	387977.3	452920.9	574768.7
I-1_DPe	I	4166	22.0	22.0	27.8	91592.4	91778.8	115839.0
Jockey-1_DPer	Jockey	7061	27.3	25.6	18.1	192819.5	180899.4	127640.8
Jockey-2_DPer	Jockey	7406	23.6	24.2	18.4	174438.8	179332.0	136084.1

Jockey-3_DPer	Jockey	8190	10.0	8.8	7.6	82003.7	71941.2	62392.2
Jockey-4_DPer	Jockey	4920	14.1	13.4	13.7	69186.6	65907.3	67610.6
Jockey-5_DPer	Jockey	5081	39.3	43.9	36.7	199619.4	222898.5	186386.9
Jockey-6_DPer	Jockey	7822	33.5	51.8	19.1	262122.1	405070.0	149079.9
Jockey-7_DPer	Jockey	3703	26.8	29.0	25.8	99121.0	107317.3	95615.9
Jockey-8_DPer	Jockey	3338	12.9	11.1	13.0	43098.7	37179.8	43417.5
Jockey-9_DPer	Jockey	4770	22.8	15.1	8.2	108743.0	72080.7	39102.6
Kiri-1_DPer	Kiri	4663	78.3	82.2	104.8	365194.4	383421.5	488461.5
Kiri-2_DPer	Kiri	4410	14.4	14.3	15.8	63585.3	63094.2	69841.2
LOA-1_DPer	LOA	5512	187.2	198.6	172.7	1031765.9	1094575.6	951697.2
Loa-1_DPs	LOA	3460	97.3	98.9	78.1	336638.8	342091.5	270216.2
LOA-1B_DPer	LOA	5500	194.9	207.2	188.1	1072152.6	1139394.0	1034671.3
LOA-2_DPer	LOA	4935	314.2	363.4	219.9	1550406.5	1793245.1	1085076.3
LOA-3_DPer	LOA	5642	311.0	355.5	367.7	1754905.0	2005648.8	2074626.2
Mariner-1_DPe	Mariner	1295	42.2	40.5	42.7	54620.8	52493.0	55271.5
Mariner-10_DRh	Mariner	1629	16.5	17.4	20.2	26852.3	28328.7	32956.7
Mariner-3_Dro	Mariner	1648	39.4	38.6	52.1	64854.0	63623.4	85939.2
MINIME_DP	MINIME	4583	919.2	996.9	890.0	4212901.7	4568937.5	4079043.7
Nobel_I	Nobel	6294	192.7	225.5	129.6	1212956.2	1419257.2	815778.7
Nobel_LTR	Nobel	314	305.0	366.5	240.6	95773.0	115077.1	75536.0
PARISa_DPe	PARISa	1742	14.6	14.2	14.3	25392.2	24796.8	24849.6
Polinton-1_DPe	Polinton	15797	67.1	67.5	235.7	1059242.2	1065658.2	3722846.8
R1_DPs	R1	5447	123.1	167.6	92.5	670278.2	912819.9	503842.9
R1-1_DPer	R1	5020	11.0	12.6	11.0	55144.7	63487.7	55164.2
R1-2_DPer	R1	4967	16.5	16.6	14.8	82026.0	82688.3	73335.8
R1-3_DPer	R1	5126	12.5	14.2	31.4	64059.8	73043.6	160956.4
R1-4_DPer	R1	4226	20.3	21.8	16.3	85846.1	92005.4	68877.7
R1-5_DPer	R1	5195	11.7	12.1	8.4	60767.9	62777.0	43620.8
R1-6_DPer	R1	5048	25.6	27.3	64.9	128978.6	137654.5	327591.8
R1-7_DPer	R1	5139	23.7	26.2	22.0	121962.7	134791.2	113106.4
R2_DPs	R2	3557	29.8	42.1	13.8	105996.3	149695.3	49211.0
SPOCK_DMi	SPOCK	4952	117.6	113.9	34.7	582249.4	564049.7	171773.8
T102_X	Betancourt_Unk nown_TE	478	5.2	5.6	4.8	2504.1	2684.0	2305.2

T104_X	Betancourt_Unk nown_TE	1010	40.4	39.5	39.8	40793.8	39899.1	40232.3
T110_X	Betancourt_Unk nown_TE	242	186.7	205.8	175.3	45182.6	49804.8	42423.7
T111_X	Betancourt_Unk nown_TE	1268	112.4	116.3	114.6	142466.1	147457.6	145249.9
T113_X	Betancourt_Unk nown_TE	1131	192.7	205.4	219.9	217938.7	232342.1	248736.1
T117_X	Betancourt_Unk nown_TE	684	27.9	28.8	31.0	19086.8	19702.3	21189.2
T119_X	Betancourt_Unk nown_TE	2699	174.6	164.5	165.7	471244.7	444093.3	447354.3
T123_X	Betancourt_Unk nown_TE	2245	14.4	13.2	12.9	32254.2	29547.2	28924.9
T126_X	Betancourt_Unk nown_TE	425	18.3	19.1	23.5	7788.9	8124.3	10002.9
T129_X	Betancourt_Unk nown_TE	3090	21.4	20.4	20.5	66120.9	63072.3	63324.7
T130_X	Betancourt_Unk nown_TE	1089	43.6	39.9	40.5	47522.9	43494.3	44130.6
T132_X	Betancourt_Unk nown_TE	727	10.6	10.5	12.9	7679.5	7622.4	9400.8
T133_X	Betancourt_Unk nown_TE	747	87.7	81.9	40.4	65481.2	61166.9	30148.1
T139_X	Betancourt_Unk nown_TE	2591	80.6	80.1	74.3	208891.6	207627.9	192441.5
T140_X	Betancourt_Unk nown_TE	3302	21.4	21.5	20.1	70581.4	71009.2	66298.2
T141_X	Betancourt_Unk nown_TE	951	48.8	48.5	62.5	46449.1	46100.8	59431.1
T151_X	Betancourt_Unk nown_TE	424	9.8	9.6	12.7	4150.3	4079.1	5401.0
T153_X	Betancourt_Unk nown_TE	437	10.4	11.0	11.1	4566.4	4807.2	4839.8
T154_X	Betancourt_Unk nown_TE	952	840.1	777.7	807.1	799756.3	740375.8	768373.8
T155_X	Betancourt_Unk nown_TE	553	49.7	48.3	53.0	27495.9	26707.8	29300.8
T158_X	Betancourt_Unk	1198	123.6	123.3	108.5	148117.7	147675.9	129988.8

	noun_TE							
T162_X	Betancourt_Unk noun_TE 981	27.9	29.1	32.6	27335.9	28567.5	32023.2	
T164_X	Betancourt_Unk noun_TE 713	25.9	26.5	27.5	18481.6	18878.5	19599.8	
T166_X	Betancourt_Unk noun_TE 887	69.0	70.0	102.5	61172.1	62126.8	90897.7	
T171_X	Betancourt_Unk noun_TE 3326	22.4	20.4	22.8	74398.7	67783.2	75720.7	
T172_X	Betancourt_Unk noun_TE 2154	151.2	160.4	168.9	325735.8	345418.8	363746.5	
T173_X	Betancourt_Unk noun_TE 688	34.9	32.9	38.8	24000.5	22668.5	26704.1	
T174_X	Betancourt_Unk noun_TE 531	10.7	11.5	11.7	5678.6	6086.7	6213.4	
T177_X	Betancourt_Unk noun_TE 451	27.3	26.8	28.4	12305.6	12065.3	12825.2	
T178_X	Betancourt_Unk noun_TE 643	8.2	8.3	8.1	5304.4	5309.6	5229.1	
T179_X	Betancourt_Unk noun_TE 638	113.5	113.4	112.2	72413.0	72353.7	71605.0	
T180_X	Betancourt_Unk noun_TE 257	201.3	211.4	219.9	51730.1	54323.4	56505.4	
T188_X	Betancourt_Unk noun_TE 493	34.4	40.6	39.5	16976.1	20032.5	19475.2	
T190_X	Betancourt_Unk noun_TE 357	16.2	15.2	19.0	5795.7	5442.5	6794.9	
T192_X	Betancourt_Unk noun_TE 549	8.0	9.4	14.7	4413.7	5147.7	8079.5	
T194_X	Betancourt_Unk noun_TE 934	50.8	52.3	54.7	47429.9	48843.4	51051.6	
T196_X	Betancourt_Unk noun_TE 1206	7.8	6.9	9.9	9384.9	8288.7	11915.0	
T200_X	Betancourt_Unk noun_TE 496	0.0	0.0	0.0	4.4	0.0	12.3	
T201_X	Betancourt_Unk noun_TE 299	6.2	5.8	5.9	1857.5	1735.7	1768.8	
T202_X	Betancourt_Unk noun_TE 8358	3.3	3.6	3.4	27612.4	30357.6	28187.2	

T203_X	Betancourt_Unk nown_TE	402	1.5	1.3	1.8	607.4	512.4	717.4
T206_X	Betancourt_Unk nown_TE	634	0.1	0.1	0.1	68.6	42.8	35.0
T208_X	Betancourt_Unk nown_TE	283	28.9	30.9	28.0	8177.0	8738.1	7918.5
T210_X	Betancourt_Unk nown_TE	663	5.4	8.0	8.6	3591.0	5272.5	5734.4
T211_X	Betancourt_Unk nown_TE	4969	0.2	0.3	0.3	1152.1	1619.5	1314.2
T212_X	Betancourt_Unk nown_TE	752	3.4	3.7	3.3	2557.9	2819.6	2491.9
T213_X	Betancourt_Unk nown_TE	503	10.7	13.8	10.3	5377.6	6930.8	5171.5
T215_X	Betancourt_Unk nown_TE	312	1.8	2.5	2.0	575.3	786.6	621.9
T218_X	Betancourt_Unk nown_TE	1277	21.4	20.4	20.2	27332.0	26074.4	25797.9
T221_X	Betancourt_Unk nown_TE	6168	0.4	0.5	0.6	2718.7	2884.3	3558.7
T222_X	Betancourt_Unk nown_TE	567	1.1	1.1	1.1	621.5	608.2	633.0
T223_X	Betancourt_Unk nown_TE	771	1.4	0.9	1.2	1075.9	701.8	930.6
T224_X	Betancourt_Unk nown_TE	650	4.6	4.6	5.0	2978.8	3008.2	3267.7
T225_X	Betancourt_Unk nown_TE	415	0.2	0.1	0.1	93.3	45.9	33.7
T24_X	Betancourt_Unk nown_TE	5615	1.6	1.6	1.4	8896.9	8947.5	7930.1
T31_LTR	Betancourt_LT R_TE	4278	1.1	1.1	0.9	4905.3	4835.7	4001.7
T32_LTR	Betancourt_LT R_TE	3570	1.0	1.0	1.0	3566.5	3587.7	3610.1
T33_X	Betancourt_Unk nown_TE	1634	0.8	0.8	0.9	1381.6	1344.2	1483.0
T34_X	Betancourt_Unk nown_TE	6395	0.4	0.5	0.4	2743.8	2911.6	2323.0
T35_X	Betancourt_Unk	2636	1.2	1.2	1.2	3158.3	3133.0	3239.8

	noun_TE							
T38_LTR	Betancourt_LT R_TE 2206	0.1	0.1	0.1	325.9	299.4	256.5	
T46_LTR	Betancourt_LT R_TE 5089	0.9	0.8	1.0	4453.1	4167.5	4963.0	
T47_X	Betancourt_Unk noun_TE 1686	0.1	0.1	0.1	191.1	162.5	143.2	
T52_LTR	Betancourt_LT R_TE 2267	0.2	0.3	0.1	464.5	644.0	335.2	
T55_LTR	Betancourt_LT R_TE 4922	0.1	0.1	0.1	645.8	603.2	492.7	
T59_LTR	Betancourt_LT R_TE 2052	13.6	13.2	16.3	27880.5	27083.2	33411.3	
T61_LTR	Betancourt_LT R_TE 11447	10.5	11.5	19.6	120477.5	131609.2	224142.3	
T66_X	Betancourt_Unk noun_TE 2619	4.5	5.4	5.5	11841.7	14028.5	14520.5	
T74_X	Betancourt_Unk noun_TE 5830	2.9	3.0	3.2	17176.2	17259.3	18470.7	
T76_X	Betancourt_Unk noun_TE 615	34.1	35.0	39.5	20966.1	21515.1	24282.3	
T80_X	Betancourt_Unk noun_TE 1642	33.4	33.7	35.8	54823.1	55283.5	58784.8	
T81_X	Betancourt_Unk noun_TE 513	6.3	4.9	6.1	3252.5	2529.3	3152.9	
T82_X	Betancourt_Unk noun_TE 610	38.7	39.8	52.2	23588.4	24296.0	31860.8	
T83_X	Betancourt_Unk noun_TE 707	151.2	156.8	325.3	106885.9	110865.5	229956.9	
T86_X	Betancourt_Unk noun_TE 576	15.6	14.2	17.1	8990.5	8201.4	9877.2	
T88_X	Betancourt_Unk noun_TE 3591	23.7	23.0	21.2	85030.1	82566.8	76010.1	
T89_X	Betancourt_Unk noun_TE 918	18.1	17.7	19.2	16632.2	16210.9	17644.2	
T9_LTR	Betancourt_LT R_TE 5246	5.1	5.0	4.5	26551.2	26489.8	23355.0	
T90_X	Betancourt_Unk noun_TE 789	33.6	37.5	35.1	26486.7	29551.4	27714.4	

T98_X	Betancourt_Unknown_TE	936	49.3	52.5	49.7	46106.9	49144.6	46544.5
T99_X	Betancourt_Unknown_TE	1834	92.5	164.1	56.3	169578.6	300877.4	103304.0
TRAM_I	TRAM	2708	294.5	353.1	410.6	797497.7	956068.6	1111868.7
TRAM_LTR	TRAM	372	470.1	549.8	615.1	174867.6	204508.1	228823.2
Transib1_DP	Transib	2983	9.5	9.3	11.0	28237.0	27786.7	32810.7
Transib2_DP	Transib	2311	19.7	27.8	30.5	45540.2	64329.1	70557.4
Transib3_DP	Transib	2783	24.6	24.4	28.6	68511.0	68041.4	79493.0
Transib4_DP	Transib	1334	2.1	1.9	1.9	2794.8	2589.8	2598.3
TransibN5_DP	Transib	1468	11.1	12.5	12.8	16258.4	18416.5	18756.0
267_IGS	267bp_Y_repeat	267	13337.8	19836.5	26140.9	3561191.7	5296337.3	6979629.8
TRIM	TRIM	3111	96.6	102.5	186.7	300374.9	318866.9	580978.0
WORF_DMi	WORF	4174	274.5	288.6	308.4	1145930.4	1204741.9	1287064.7

Table S9. Transposable element bp contribution from Y chromosome assemblies (i.e. exclusively Y-linked contigs). Estimates are derived from RepeatMasker. Data underlies Fig 3B.

TE	Small Y	Medium Y	Large Y
BEL-1_DPer-I	614579	692160	409704
BEL-1_DPer-LTR	56084	59993	42335
BEL-13_DP-I	527680	654930	833900
BEL-13_DP-LTR	91601	106403	144357
BEL-14_DP-I	66627	78140	44704
BEL-14_DP-LTR	11943	15803	3813
BEL-15_DP-I	23209	23676	17829
BEL-15_DP-LTR	2586	3172	796
BEL-2_DPer-I	75905	77037	61317
BEL-2_DPer-LTR	16210	17221	13918
BEL-3_DPer-I	917926	1006737	977091
BEL-3_DPer-LTR	109337	122233	112715
BEL-4_DPer-I	104407	150340	43049
BEL-4_DPer-LTR	21737	30701	6009
BEL-4_DPse-I	10073	30972	22613
BEL-4_DPse-LTR	1794	4784	1763

BEL-5_DPer-I	372906	469940	357808
BEL-5_DPer-LTR	187766	255866	301282
BEL-6_DPe-I	108242	208369	14528
BEL-6_DPe-LTR	3638	4794	715
BEL-6_DPer-I	129982	135379	92613
BEL-6_DPer-LTR	15253	15659	9742
BEL-7_DPe-I	152706	185318	101168
BEL-7_DPe-LTR	13184	17107	7331
BEL-7_DPer-I	29299	60404	1676
BEL-7_DPer-LTR	1624	3385	0
BEL-8_DPe-I	366522	498148	705430
BEL-8_DPe-LTR	35899	50150	56080
BEL1_I_Dpse	224825	204151	485386
BEL1_LTR_Dpse	10797	10800	28115
BEL2-I_Dpse	583587	596926	490904
BEL2-LTR_Dpse	2922	3435	790
BEL3-I_Dpse	78671	88527	87831
BEL3-LTR_Dpse	8748	11126	10029
Copia-1_DPer-I	284441	354667	449140
Copia-1_DPer-LTR	26159	31774	42752
Copia-2_DPer-I	26623	29300	36538
Copia-2_DPer-LTR	4409	3874	5282
Copia-29_DP-I	317050	402204	237676
Copia-29_DP-LTR	30481	38143	24723
Copia1-I_Dpse	106766	73620	69159
Copia1-LTR_Dpse	8570	5614	6681
Copia2-I_Dpse	651243	756676	1770001
Copia2-LTR_Dpse	40782	48423	99666
CR1-1_DPer	812745	1052286	916802
CR1-2_DPer	1188	1188	3230
CR1-3_DPer	0	180	4560
Daff_Helitron_4	54181	31181	101172
Daff_Jockey_18	148154	192603	320398

Dlow_Gypsy_2	83173	86229	132752
Dlow_Gypsy_64	0	993	183
Dlow_Helitron_184	52	52	442
Dlow_Helitron_5	6712	7112	39039
Dlow_Loa_58	22440	49092	23778
Dlow_Mariner_3	0	0	168
Dmir_hAT-Blackjack_169	1322	3011	1558
Dmir_I_149	752	1504	16165
Dmir_P_226	304	308	1800
Dmir_TcMar-Tc1_157	1605	9102	1440
DNA-1_DPe	13520	19153	24682
Dper_Gypsy_134	169257	167449	225543
Dper_Gypsy_138	767	0	260
Dper_Jockey_145	7192	9485	479
Dper_Jockey_70	122365	113736	123758
Dper_loa_7	0	0	2184
Dper_piggyBac	6584	0	4333
Dper_TcMar-Tc1_136	2001	2293	2171
Dper_TcMar-Tc1_144	701	1460	7735
Dper_TcMar-Tc1_147	3352	3032	6970
Dpse_CMC-EnSpm_97	2399	2777	10405
Dpse_Gypsy_6	296496	359403	370366
Dpse_Gypsy_79	154268	217264	48585
Dpse_hAT-hATm_125	816	816	0
Dpse_P_121	324	648	678
Dpse_TcMar-Tc1_120	4465	10068	7761
Dpse_TcMar-Tc1_93	5318	6633	660
Gal_DPer	560	793	5132
Gal_DPse	6087	8568	103879
Gypsy-1_DPer-I	222517	280075	66078
Gypsy-1_DPer-LTR	35060	40075	13252
Gypsy-10_DPer-I	196343	261573	180756
Gypsy-10_DPer-LTR	100952	109061	82717

Gypsy-11_DPer-I	586147	673521	665128
Gypsy-11_DPer-LTR	293903	335014	309078
Gypsy-12_DPer-I	210335	224681	122950
Gypsy-12_DPer-LTR	45469	42971	36442
Gypsy-1A_DPer-LTR	10150	14129	292
Gypsy-2_DPer-I	32729	48850	131288
Gypsy-2_DPer-LTR	1619	2409	7468
Gypsy-25_DP-I	35502	49247	12243
Gypsy-25_DP-LTR	17651	30054	4035
Gypsy-25_DPse-I	108240	128294	79785
Gypsy-25_DPse-LTR	79927	107857	55438
Gypsy-26_DP-I	68629	56753	23546
Gypsy-26_DP-LTR	50088	50888	19930
Gypsy-26_DPse-I	713136	765682	1009040
Gypsy-26_DPse-LTR	103152	110290	164403
Gypsy-27_DP-I	97213	161201	296299
Gypsy-27_DP-LTR	9805	15854	35421
Gypsy-3_DPer-I	102379	169584	90616
Gypsy-3_DPer-LTR	73799	120791	82503
Gypsy-31_DP-I	33685	13417	19846
Gypsy-31_DP-LTR	13923	2885	4993
Gypsy-32_DP-I	164	820	47464
Gypsy-32_DP-LTR	1258	0	24062
Gypsy-4_DPer-I	22273	23223	19209
Gypsy-4_DPer-LTR	9839	10595	9892
Gypsy-5_DPer-I	194668	279823	345485
Gypsy-5_DPer-LTR	14153	21207	22267
Gypsy-6_DPer-I	29853	6365	0
Gypsy-6_DPer-LTR	14941	4886	1867
Gypsy-7_DPer-I	1018049	1242573	1037021
Gypsy-7_DPer-LTR	589848	726289	595551
Gypsy-8_DPer-I	53253	41267	44756
Gypsy-8_DPer-LTR	36812	36321	5610

Gypsy-9_DPer-I	175476	204647	143656
Gypsy-9_DPer-LTR	70746	88546	69378
Gypsy1-I_Dpse	5197	7603	1062
Gypsy1-LTR_Dpse	33	33	0
Gypsy10-I_Dpse	527555	568228	811337
Gypsy10-LTR_Dpse	144812	159875	228998
Gypsy11-I_Dpse	260605	339123	571501
Gypsy11-LTR_Dpse	50275	70330	116678
Gypsy12-I_Dpse	295243	297301	417381
Gypsy12-LTR_Dpse	59609	54646	89671
Gypsy13-I_Dpse	32719	57074	44048
Gypsy13-LTR_Dpse	2610	4476	3730
Gypsy14-I_Dpse	104078	136052	275624
Gypsy14-LTR_Dpse	17912	28970	56546
Gypsy15-I_Dpse	81565	71540	55972
Gypsy15-LTR_Dpse	57468	54187	42176
Gypsy16-I_Dpse	139689	134435	268989
Gypsy17-I_Dpse	604401	800229	468170
Gypsy17-LTR_Dpse	198	198	116
Gypsy18-I_Dpse	1070862	898170	1016779
Gypsy18-LTR_Dpse	366	183	0
Gypsy19-I_Dpse	301264	326975	909537
Gypsy19-LTR_Dpse	35635	36144	120115
Gypsy2-I_Dpse	108640	126085	197660
Gypsy2-LTR_Dpse	11952	13757	22440
Gypsy20-I_Dpse	306085	319334	309272
Gypsy20-LTR_Dpse	792	597	1075
Gypsy21-I_Dpse	306973	341578	371277
Gypsy21-LTR_Dpse	30872	33440	33359
Gypsy22-I_Dpse	445394	437176	947095
Gypsy22-LTR_Dpse	74	198	17634
Gypsy23-I_Dpse	27840	28256	75973
Gypsy23-LTR_Dpse	9884	8024	10734

Gypsy24-I_Dpse	27082	46435	55887
Gypsy24-LTR_Dpse	5857	10695	10557
Gypsy3-I_Dpse	35861	64169	465
Gypsy3-LTR_Dpse	3065	5078	0
Gypsy4-I_Dpse	24100	22641	1717
Gypsy4-LTR_Dpse	3383	4742	5637
Gypsy5-I_Dpse	48868	80302	52567
Gypsy5-LTR_Dpse	13902	21684	14667
Gypsy6-I_Dpse	40191	17275	800
Gypsy6-LTR_Dpse	8542	1388	1946
Gypsy7-I_Dpse	64182	86114	35489
Gypsy7-LTR_Dpse	3396	4423	2980
Gypsy8-I_Dpse	254316	400981	248037
Gypsy8-LTR_Dpse	113763	166271	79781
Gypsy9-I_Dpse	4012	5742	3889
Gypsy9-LTR_Dpse	12961	15120	7225
Harbinger-N1_DPe	4470	5803	82002
hAT-1_DP	27323	11473	11596
hAT-1N_DP	555	556	4689
Helitron-1_DPe	202682	154200	4655300
HelitronN-1_DPe	4114583	4557948	5033710
Hopers2	123683	135500	222357
I-1_DPe	4251	86	9362
Jockey-1_DPer	156406	98927	65576
Jockey-2_DPer	6891	8948	0
Jockey-3_DPer	42363	47952	32918
Jockey-5_DPer	87377	78127	90095
Jockey-6_DPer	266682	288181	110726
Jockey-7_DPer	10966	8368	7744
Jockey-8_DPer	15089	14500	10335
Jockey-9_DPer	83811	48745	16914
Kiri-1_DPer	99630	90391	151471
LOA-1_DPer	171035	231209	315852

Loa-1_DPs	32739	44174	12836
LOA-1B_DPer	384739	388019	402248
LOA-2_DPer	504151	691530	387499
LOA-3_DPer	723804	917064	1125837
Mariner-1_DPe	91	91	0
Mariner-10_DRh	2318	3057	4100
Mariner-3_Dro	1647	3294	12034
MINIME_DP	389737	554719	708667
Nobel_I	496906	634315	240766
Nobel_LTR	50757	66274	30368
PARISa_DPe	1741	4065	11341
Polinton-1_DPe	402408	274899	1556058
R1_DPs	298806	574462	319042
R1-1_DPer	14627	20077	17133
R1-2_DPer	21574	23927	39841
R1-3_DPer	27021	31556	110137
R1-4_DPer	14557	28110	45878
R1-5_DPer	35560	55550	56688
R1-6_DPer	15730	30437	228048
R1-7_DPer	27424	42456	47925
R2_DPs	57334	80332	9874
SPOCK_DMi	329309	411716	77366
T104_X/Unknown	0	0	939
T110_X/Unknown	8783	10827	0
T111_X/Unknown	603	0	0
T117_X/Unknown	0	0	164
T122_X/Unknown	46259	41416	1464765
T126_X/Unknown	714	715	664
T130_X/Unknown	27	675	3251
T132_X/Unknown	128	128	0
T133_X/Unknown	40203	26750	6641
T139_X/Unknown	152077	91450	19099
T141_X/Unknown	6694	7502	28720

T150_X/Unknown	555	1214	1790794
T151_X/Unknown	1290	1288	7561
T153_X/Unknown	1116	1032	2168
T154_X/Unknown	1313	0	0
T158_X/Unknown	28595	30350	1549
T162_X/Unknown	607	817	1241
T164_X/Unknown	0	0	408
T166_X/Unknown	1727	1508	6733
T172_X/Unknown	11819	11327	13712
T173_X/Unknown	0	93	0
T174_X/Unknown	0	0	566
T177_X/Unknown	236	310	10829
T178_X/Unknown	47	0	0
T180_X/Unknown	4657	3697	3255
T188_X/Unknown	3243	3088	324
T190_X/Unknown	0	54	0
T192_X/Unknown	0	144	160010
T194_X/Unknown	10867	9800	15953
T201_X/Unknown	1744	3598	653
T202_X/Unknown	5376	4129	1968
T206_X/Unknown	0	51	0
T208_X/Unknown	7681	8283	7555
T211_X/Unknown	4519	5349	2912
T212_X/Unknown	309	309	0
T213_X/Unknown	870	1363	3554
T215_X/Unknown	68	0	194
T218_X/Unknown	988	0	0
T221_X/Unknown	14255	15667	21479
T223_X/Unknown	262	396	0
T224_X/Unknown	0	0	94
T24_X/Unknown	0	54	232
T31_LTR	8941	12293	49061
T32_LTR	0	0	11509

T33_X/Unknown	79	158	0
T34_X/Unknown	0	0	1149
T35_X/Unknown	3541	4016	6749
T38_LTR	1918	2205	0
T46_LTR	36746	52542	63109
T47_X/Unknown	29575	34374	33492
T52_LTR	8331	4731	5638
T55_LTR	33825	42651	30799
T59_LTR	13253	13171	34460
T61_LTR	15562	28525	64145
T66_X/Unknown	63	0	1902
T74_X/Unknown	152	469	144
T76_X/Unknown	0	102	0
T80_X/Unknown	3434	2058	3262
T81_X/Unknown	52	104	1984
T82_X/Unknown	1526	1930	18915
T83_X/Unknown	18208	24680	91333
T86_X/Unknown	0	194	1557
T88_X/Unknown	1506	1327	0
T89_X/Unknown	1321	2436	0
T9_LTR	10240	12301	11688
T90_X/Unknown	165	136	1886
T98_X/Unknown	428	812	1404
T99_X/Unknown	125876	164811	106009
TRAM_I	524378	716034	961509
TRAM_LTR	8793	8983	5187
Transib1_DP	0	487	395
Transib2_DP	6861	22347	13643
Transib3_DP	2928	3733	1125
TransibN5_DP	71	1734	343
267_IGS	2099079	2632075	4937536
TRIM	7396	10429	397835
WORF_DMi	146934	193751	244970

Table S10. Head-to-tail chimeric junction locations on Y assemblies.

Assembly	Contig	Start	End	Tandem Junction	
				Start	End
YL	000118F_pilon	218805	218704	1	99
YL	tig00000674_pilon	177146	177051	1	100
YL	tig00000674_pilon	1201832	1201911	1	86
YL	tig00000674_pilon	122097	122185	5	99
YL	tig00000674_pilon	128891	128979	5	99
YL	tig00000201_pilon	20913	21003	1	92
YL	tig00000201_pilon	8743	8839	5	98
YL	000155F_pilon	100043	100136	1	97
YL	000125F_pilon	164071	163978	1	97
YL	000068F_pilon	967992	968074	5	88
YL	000053F_pilon	485299	485378	1	80
YL	000053F_pilon	492566	492645	1	80
YL	000081F_pilon	283628	283546	2	88
YL	000081F_pilon	12130	12051	1	79
YL	000074F_pilon	324838	324925	1	91
YL	000015F_pilon	320695	320776	1	79
YL	000011F_pilon	1970662	1970580	1	80
YL	000011F_pilon	3250158	3250059	5	99
YL	000011F_pilon	2571763	2571859	5	100
YL	000011F_pilon	2062474	2062567	5	94
YL	000011F_pilon	3255915	3255832	5	92
YL	000011F_pilon	2208763	2208854	5	97
YL	tig00000875_pilon	320696	320614	5	88
YL	tig00000875_pilon	324392	324310	5	88
YL	tig00000875_pilon	329194	329112	5	88
YL	tig00000875_pilon	316987	316903	5	88
YL	tig00000889_pilon	22902	22804	5	99
YL	tig00000889_pilon	48894	48796	5	99
YL	tig00000753_pilon	5022	5108	5	92
YL	tig00000194_pilon	35946	35855	5	94

YL	000123F_pilon	121711	121794	5	88
YL	000123F_pilon	118007	118091	5	88
YL	000058F_pilon	378111	378198	5	95
YL	000014F_pilon	380586	380499	5	90
YL	000014F_pilon	791298	791386	5	91
YL	000005F_pilon	892012	891935	6	82
YL	000198F_pilon	893701	893617	5	86
YL	000132F_pilon	184881	184989	1	97
YL	000037F_pilon	395823	395732	1	91
YL	000142F_pilon	115036	115120	5	91
YL	000010F_pilon	608484	608399	5	93
YL	tig00000843_pilon	20619	20700	5	90
YL	tig00000154_pilon	60554	60635	5	89
YL	tig00000154_pilon	5707	5789	5	89
YL	000045F_pilon	120632	120717	5	96
YL	tig00000688_pilon	42070	42164	5	100
YL	tig00000688_pilon	45667	45759	5	100
YL	000042F_pilon	205777	205867	5	100
YL	000126F_pilon	163043	162956	5	100
YL	tig00000809_pilon	22295	22216	5	90
YM	tig00001820_pilon	291409	291495	5	88
YM	tig00006355_pilon	35458	35375	1	86
YM	tig00000116_pilon	460321	460401	5	79
YM	000008F_pilon	128757	128673	5	94
YM	000008F_pilon	106329	106245	5	94
YM	000008F_pilon	205965	205881	5	94
YM	000008F_pilon	361112	361196	5	94
YM	000008F_pilon	167484	167401	5	97
YM	000008F_pilon	216571	216488	5	97
YM	000008F_pilon	116245	116161	5	94
YM	tig00038002_pilon	240525	240620	5	97
YM	000041F_pilon	299842	299735	5	98
YM	tig00002047_pilon	97698	97788	6	93

YM	tig00001632_pilon	41626	41542	5	94
YM	tig00001632_pilon	181998	181914	5	94
YM	tig00001632_pilon	199311	199227	5	94
YM	tig00001632_pilon	176724	176641	5	94
YM	tig00038045_pilon	70523	70607	5	94
YM	tig00002013_pilon	47906	47822	5	94
YM	tig00002013_pilon	58360	58276	5	94
YM	tig00002013_pilon	83079	83164	5	94
YM	tig00001217_pilon	53385	53469	5	94
YM	000129F_pilon	43096	43180	5	94
YM	tig00006468_pilon	35121	35023	5	99
YM	000043F_pilon	39961	40041	5	79
YM	000043F_pilon	3886	3966	5	79
YM	tig00006455_pilon	292781	292698	5	88
YM	tig00001802_pilon	30609	30515	5	99
YM	tig00001802_pilon	204986	205068	5	82
YM	tig00001802_pilon	31869	31786	5	90
YM	tig00006436_pilon	75188	75265	21	96
YM	tig00006431_pilon	253020	252924	5	99
YM	tig00001051_pilon	290485	290575	5	98
YM	tig00006528_pilon	34665	34762	5	100
YS	tig00006602_pilon	436779	436859	5	79
YS	tig00006602_pilon	525035	524947	5	99
YS	tig00000937_pilon	1097652	1097568	5	94
YS	tig00000937_pilon	229375	229267	5	98
YS	tig00006405_pilon	384297	384213	5	94
YS	tig00006405_pilon	361846	361762	5	94
YS	tig00006405_pilon	459487	459403	5	94
YS	tig00006405_pilon	421825	421742	5	97
YS	tig00006405_pilon	371828	371744	5	94
YS	tig00006405_pilon	470057	469975	5	97
YS	tig00006601_pilon	233084	232993	5	97
YS	tig00006553_pilon	54464	54548	5	94

YS	tig00006409_pilon	114677	114593	5	94
YS	tig00006409_pilon	146187	146271	5	94
YS	tig00035849_pilon	45192	45276	5	94
YS	tig00006571_pilon	43050	42960	6	93
YS	tig00006524_pilon	195799	195897	5	99
YS	tig00006474_pilon	102661	102750	5	93
YS	tig00002599_pilon	39938	39854	5	94
YS	tig00000398_pilon	65263	65171	5	91
YS	tig00035856_pilon	72620	72698	5	88
YS	tig00006595_pilon	46577	46657	5	79
YS	tig00006595_pilon	87475	87556	5	79
YS	tig00006548_pilon	9107	9014	5	96
YS	tig00000002_pilon_Y	111892	111972	5	79
YS	tig00000302_pilon	398652	398734	5	82
YS	tig00006488_pilon	9510	9427	5	85
YS	tig00000614_pilon	71185	71271	5	88

Table S11. Differentially expressed genes between Y_S and Y_L strains. Up-regulation means higher expression in Y_L than Y_S. Duplicated genes are due to multiple gene copies on the chromosomes in the *D. pseudoobscura* assembly.

FlybaseID	Gene Name	GO Biological Process	Regulation	Log2(fold change)	adj. p-value	Transcript location
FBgn0245720	Ulp1	SUMO-specific isopeptidase activity	up	2.93	1.71E-11	Muller_A-AD
FBgn0247234	Ulp1	SUMO-specific isopeptidase activity	down	0.62	0.010628069	Muller_A-AD
FBgn0070601	SmD1	mRNA splicing, spliceosomal snRNP assembly	up	1.50	4.55E-14	Muller_A-AD
FBgn0249586	Rnf146	positive regulation of canonical Wnt signaling pathway	up	1.45	4.08E-18	Muller_A-AD
FBgn0250532	Rnf146	positive regulation of canonical Wnt signaling pathway	up	1.42	1.68E-05	Muller_A-AD
FBgn0080026	Nct	Notch signaling pathway	up	2.52	2.56E-37	Muller_E
FBgn0071472	mxs	histone locus body assembly	up	1.56	7.26E-16	Muller_A-AD
FBgn0071472	mxs	histone locus body assembly	up	4.09	6.18E-28	Y_49
FBgn0081594	mh	DNA-binding	up	3.63	1.91E-60	Muller_A-AD
FBgn0079270	isoQC	glutamyl-peptide cyclotransferase activity	up	2.05	5.05E-24	Y_84
FBgn0076039	HDAC11	histone deacetylation	up	2.32	2.23E-27	Muller_E
FBgn0244258	CG43736	N/A	up	2.10	1.07E-31	Muller_A-AD
FBgn0272696	CG42339	predicted polysaccharide binding activity	down	2.53	2.13E-10	Y_45
FBgn0073604	CG42339	predicted polysaccharide binding activity	up	2.10	3.26E-08	Y_37
FBgn0073107	CG14544	tRNA methylation	up	1.45	1.90E-10	Muller_E
FBgn0072813	CG14111	N/A	up	2.16	0.000722917	Y_4
FBgn0072813	CG14111	N/A	down	5.51	0.002022726	Y_4
FBgn0072813	CG14111	N/A	up	1.96	0.000232027	Y_4
FBgn0070269	CG10274	DNA-binding transcription factor activity	down	4.56	0.023627261	Y_107
FBgn0070269	CG10274	DNA-binding transcription factor activity	down	2.79	5.76E-20	Y_74
FBgn0070189	CG10171	N/A	down	6.09	0.0001228	Y_96
FBgn0079379	aub	piRNA binding, chromatin silencing	up	6.11	3.72E-29	Muller_A-AD
FBgn0079379	aub	piRNA binding, chromatin silencing	up	3.41	1.36E-17	Muller_B
FBgn0079379	aub	piRNA binding, chromatin silencing	up	2.50	1.40E-12	Muller_B
FBgn0246885	-		up	0.82	0.039923598	Y_64
FBgn0272687	-		up	3.96	1.24E-23	Muller_A-AD

FBgn0249017 -	up	1.98	0.011701085	Muller_A-AD
FBgn0249387 -	down	1.24	0.000764475	Muller_B
FBgn0271731 -	down	2.11	9.81E-06	Muller_B
FBgn0248340 -	up	2.76	2.25E-24	Muller_E
FBgn0248340 -	up	1.56	2.93E-07	Muller_E
FBgn0249218 -	down	3.14	2.33E-11	Y_138
FBgn0249218 -	down	1.56	0.0001228	Y_138
FBgn0247167 -	down	3.87	0.000661484	Y_23
FBgn0247167 -	down	5.26	0.017481503	Y_3
FBgn0246929 -	down	5.46	0.003806709	Y_3
FBgn0250510 -	up	1.37	4.54E-10	Y_5
FBgn0245408 -	up	1.20	0.001414999	Y_9
FBgn0271769 -	down	6.32	5.70E-05	Y_103
FBgn0249017 -	up	1.76	0.000169323	Y_7

Table S12. Homozygosity on the autosomes/X chromosome for Y-replacement lines.

Chromosome	Homozygous sites	Total sites	% Homozygosity
Muller B	26,290,762	27,182,283	96.72%
Muller C	19,228,238	19,576,218	98.22%
Muller E	29,515,020	30,298,821	97.41%
Muller F	1,147,353	1,171,816	97.91%
Muller A-AD	48,553,691	50,867,144	95.45%

Chapter 3: Dynamic gene content evolution on the *Drosophila pseudoobscura* Y chromosome

Alison H. Nguyen, Kamalakar Chatla, & Doris Bachtrog

Abstract

Y chromosomes in *Drosophila* are typically viewed as gene deserts, and theory and empirical data suggest that nucleotide diversity in protein coding genes is low. Here, we show that the Y chromosome of *Drosophila pseudoobscura* contains hundreds of genes, many of which are multi-copy. Most surprisingly, the gene content of different Y chromosomes is highly variable, and polymorphic Y's contain dozens of unique single-copy and amplified gene families. Thus, counter to the current dogma, Y chromosomes are in fact one of the most variable chromosomal environments within the genome. Many Y-linked genes are expressed in testis, and enriched for spermatogenesis functions, suggesting an important role in males. Y-genes are also enriched for chromatin organization and histone exchange, processes that are known to be targeted by meiotic drivers. Polymorphic Y's could thus allow for male-specific phenotypic variation during spermatogenesis or meiotic conflicts.

Introduction

Y chromosome evolution is characterized by Y degeneration. Sex chromosomes are derived from ordinary autosomes, and over time, Y chromosomes lose their genes and instead accumulate repetitive DNA (B. Charlesworth and Charlesworth 2000; Bachtrog 2013). Indeed, old Y chromosomes in many species mostly contain repetitive DNA and only a handful of genes (Mahajan and Bachtrog 2017; Bellott et al. 2017; Skaletsky et al. 2003).

Until recently, however, direct studies of gene content evolution of Y chromosomes have been limited (Bellott et al. 2017; Skaletsky et al. 2003; Bellott et al. 2014). Repetitive DNA is notoriously difficult to sequence and assemble, and given their low gene count, sequencing of Y chromosomes was often considered not worth the effort. Recent advances in long-read sequencing, however, have allowed the assembly of even highly repeat-rich Y chromosomes (Bachtrog, Mahajan, and Bracewell 2019; Mahajan et al. 2018; Peichel et al. 2020; Xue et al. 2021). Careful analysis of more recently formed Y chromosomes in *Drosophila* have revealed a more dynamic nature of gene content evolution across Y chromosomes (Bachtrog, Mahajan, and Bracewell 2019; C. Ellison and Bachtrog 2019). Neo-sex chromosomes in *Drosophila* were repeatedly formed by fusions of autosomes with the ancestral sex chromosomes, and depending on their age, neo-sex chromosomes show intermediate levels of differentiation (Mahajan and Bachtrog 2017; Mahajan et al. 2018; Wei and Bachtrog 2019). The neo-Y of *D. miranda* originated about 1.5 MY ago and in fact has gained 1000s of genes since its formation, instead of simply degenerating (Bachtrog, Mahajan, and Bracewell 2019). Computational analyses of dozens of *Drosophila* species have shown that amplification of genes on Y chromosomes is common on younger neo-Y chromosomes (C. Ellison and Bachtrog 2019). However, detailed sequence analysis is still lacking for most Y chromosomes. Also, it is unclear how much variation in gene content exists among polymorphic Y chromosomes within a species. Cytogenetic studies have shown dramatic morphological variation of the Y chromosome (Th. Dobzhansky 1937; T. Dobzhansky 1935; B. M. Chakraborty and Chakraborty 1984), but whether this is solely due to differences in repetitive DNA, or also differences in the set or number of genes is unknown.

Here, we study gene content evolution of the *D. pseudoobscura* Y chromosome. Previous work has shown that the Y of *D. pseudoobscura* is not homologous to the Y of *D. melanogaster* (Carvalho and Clark 2005). Instead, it may have been derived from a chromosomal fusion between an autosome (Muller D) and the ancestral sex chromosome about 15MY ago (Larracunte, Noor, and Clark 2010; Bracewell and Bachtrog 2020; Dupim et al. 2018). Cytogenetic investigation has revealed drastic variation in size and shape of the Y chromosome in natural populations of *D. pseudoobscura* (Th. Dobzhansky 1937; T. Dobzhansky 1935). We recently generated high-quality assemblies for three *D. pseudoobscura* Y chromosomes, which dramatically differed in size (ranging from 33 to 59Mb (Nguyen et al. 2021)). We investigated differences in the repeat landscape between these Y's and showed that differences in Y size are largely driven by accumulation of TEs. Here, we study gene content evolution of these polymorphic Y chromosomes. We show that counter to the current dogma, Y chromosomes are in fact one of the most variable chromosomal environments within the genome, with gene content and gene numbers differing drastically among polymorphic Y chromosomes. Functional enrichment and gene expression studies suggest that gene content evolution may reflect spermatogenesis function and intragenomic conflicts over sex chromosome transmission.

Results

Y chromosomes in *D. pseudoobscura* contain hundreds of genes

We employed two different approaches to annotate genes on the Y chromosomes (**Table 1, Table S1, S2**). We used the Comparative Annotation Toolkit (CAT) (Fiddes et al. 2018) to annotate our Y chromosome assemblies and also utilized BLAST searches with *D. pseudoobscura* protein-coding genes to identify genes on the Y chromosomes (see Methods). Previous studies have identified a handful of genes on the *D. pseudoobscura* Y chromosome, and our annotation contains most of them. We verified that our Y assemblies contained three Y-linked genes identified by (Dupim et al. 2018) (CG6661, GA25497, GA27172, **Figure S1**). We also confirmed that the *D. pseudoobscura* Y's contains dozens of *GAPsec* and *S-Lap1* genes, as previously found (C. Ellison and Bachtrog 2019) (Tables S1, S2).

Both approaches identified 100s of genes on the *D. pseudoobscura* Y chromosome (**Table 1, Table S1, S2**). CAT annotates a similar number of genes on Y_S and Y_M ; we detect a total of 168 and 189 genes on Y_S and Y_M . Many Y genes exist in multiple copies, and the identified Y genes belong to 74 unique gene families on both Y_S and Y_M . Of those, 38 and 35 are single copy genes and the other 36 and 39 exist as multiple copies on Y_S and Y_M . The largest Y has about 41% more genes, with a total gene count of 297 that belong to 108 distinct gene families. Of those, 46 are single-copy genes, and 62 have multiple copies on Y_L . BLAST searches reveal similar patterns between Y_S and Y_M , although we generally find a larger number of genes (**Table 1; Table S2**). In particular, we find 914 gene copies on Y_S , 1033 on Y_M and 1440 on Y_L . Note that some of the proteins annotated in *D. pseudoobscura* have homology with transposable elements and the BLAST results probably overestimates the number of genes amplified on the Y.

Thus, contrary to the common dogma that old Y chromosomes only have few genes left, the Y chromosomes of *D. pseudoobscura* contain 100s of genes. Overall, we find that the largest Y

chromosome not only has a higher repeat content, but also contains more genes than the small or medium-sized Y's.

Dynamic gene content evolution of Y chromosomes

The three different Y chromosomes of *D. pseudoobscura* contain a mosaic of single- and multi-copy genes (**Table 1**). Surprisingly, however, many of them are unique to a particular Y chromosome (**Figure 1**). Of the 167 gene families in total that we found on the three different Y chromosomes using CAT, only 27 are shared between all three different Y's (**Figure 1A,B**). Y_S and Y_M have a much larger number of unique genes in common (31 genes) than either shares uniquely with Y_L (4 and 0 genes), and each Y chromosome contains several unique genes and gene families (**Figure 1A**). In particular, Y_S contains 12 genes not found on any other Y chromosome; most of them exist as a single copy, but FBgn0271090 contains 13 copies on Y_S (**Figure S2**). Likewise, Y_M contains 16 genes only found on this chromosome; most are single copy genes, but CG3558 or Sam-S for example have 2 copies on Y_M (**Figure S2**). The largest Y contains 77 unique genes, 36 of which are single-copy on Y_L , and several that have more than 10 copies (i.e., aubergine-like: 10 copies; Caper and CG43322: 11 copies; and CG43736: 12 copies, **Figure S2**). Our BLAST annotation shows similar patterns (**Figure 1A**): Of the 260 genes total, 92 are shared between all three Y chromosomes. Y_S and Y_M uniquely share 68 genes, while either shares many fewer (7 and 5 genes) with Y_L . Again, Y_L contains many more unique genes (60) than does Y_S (9) and Y_M (19). BLAST searches indicate that the most highly amplified gene on both Y_S and Y_M is FBgn0243857 with 83 and 88 copies, respectively. FBgn0243857 is a short protein (87 AA) that is annotated as a Cep78 (centrosomal protein 78) homolog in *D. pseudoobscura* and other *obscura* flies, but absent in *D. melanogaster*. The most abundant gene found by BLAST on Y_L is *mh* (maternal haploid), with 327 copies, followed by *Alg14* (304 copies), which co-amplified with *mh* on the large Y.

Thus, our analysis indicates that a core set of genes is conserved between the different Y chromosomes. Y_S and Y_M have more genes in common, and shared genes have a similar copy number, suggesting that they are more closely related to each other than to Y_L (**Figure 1B,C**). However, we also find that the gene content of the different Y chromosomes is highly variable, and each Y contains many unique genes (**Figure 1A, Figure S2**). This implies that gene amplification and loss occurred independently on the different Y chromosomes. Thus, rather than being uniform gene deserts, the *D. pseudoobscura* Y chromosome is highly dynamic with respect to gene number and content. This finding begs the question of how these genes arose on the Y chromosome, and what evolutionary forces drive their origination and amplification.

Origin of Y genes

Genes on the Y can be derived from multiple sources. Some genes may have been ancestrally present on the autosome that formed the neo-sex chromosomes and so far escaped degeneration. Other genes may have been acquired on the Y chromosome secondarily from an autosomal location. The Y chromosome of *D. pseudoobscura* is not homologous to the Y of *D. melanogaster* (Carvalho and Clark 2005). Instead, the Y chromosome of *D. pseudoobscura* presumably is a degenerate neo-Y chromosome that originated from Muller D, the autosome that fused to the ancestral sex chromosomes (Bracewell and Bachtrog 2020; Mahajan and Bachtrog 2017).

Somewhat surprisingly, the majority of genes located on the three Y chromosomes are not remnants of genes ancestrally located on the neo-Y (i.e. they are not derived from Muller D; **Table 2**). In particular, we find that while there is an excess of Y genes that also have a copy on Muller D (47, $p < 0.05$, two-tailed Test of Proportions), many Y genes also have a homolog on Muller B (41), or Muller A (39, the ancestral X). Thus, most genes that are present now on the Y are not remnants of ancestral neo-Y genes. However, as shown above, genes were gained and lost independently among different Y chromosomes, and only a small fraction of Y genes are shared between them (**Figure 1**). When only considering the core set of shared genes, we find that they have 2.5-fold more homologs on both Muller A and D than other chromosomes ($p > 0.05$, two-tailed Test of Proportions). Thus, shared genes are enriched for genes that have been ancestrally present on the autosome that formed the neo-Y, but genes that are restricted to a particular Y's have been acquired from a variety of genomic sources. Similar conclusions are reached if *D. subobscura*, a related species that lacks the neo-sex chromosome fusion is used to assign chromosomal locations to Y-linked genes in *D. pseudoobscura* (**Table 2**). Genes identified by BLAST show even less evidence for being ancestrally located on the neo-Y (**Table 2**).

Across all three Y chromosomes, we find a total of 167 different protein-coding genes that have one or more copies on the Y, and more than half of those (86 genes) also have a copy on the X chromosome. **Table 3** shows the number of genes from each Y strain that have homologs on Muller A and Muller D (X chromosome). About 50% of the genes located on each Y also have a homolog on the X chromosome (Y_S : 55.4%, Y_M : 44.6%, Y_L : 59.3%).

Mechanisms of gene gain on the Y

Y genes arose from different genomic locations, through diverse molecular mechanisms, and at different time points during the evolution of the Y. We leveraged our assemblies to infer the evolutionary timing of gene copies amongst the three Y's and their homologs elsewhere. Figure 2A shows phylogenetic trees of select gene families that exist in multi-copies on the different Y chromosomes. Again, Y_S and Y_M copies from the same gene family are typically more related to each other, than they are to copies on Y_L (Figure 2A).

Our CAT annotation and BLAST searches showed that many genes exist in multiple copies, and we wanted to identify if they amplified in tandem, or are dispersed across the Y chromosome. In many cases, genes are (co)-amplified in tandem on the different Y chromosomes. **Figure 2B** shows the location of multi-copy genes among some of our largest Y contigs from the different Y assemblies. In most cases, gene copies belonging to the same gene family are found adjacent on the same Y contig (i.e. CG12035 on Y_L or FBgn0245944 on Y_S ; **Figure 2B**), and sometimes different clusters of amplified copies are found on different contigs (i.e. CG43322 on Y_L ; **Figure 2B**). In some cases, two different genes co-amplified on the Y chromosome (i.e. *Larp7* and *amX* on both Y_M or Y_S or *mh* and *Alg14* on Y_L ; **Figure 2B**). In other cases, individual copies of gene families are dispersed across several contigs (i.e. *Caper* or *HDAC11* on Y_L ; **Figure S3**; or *SLap-1* and *GAPsec* on Y_S ; (C. Ellison and Bachtrog 2019)). Thus, the locations of multi-copy genes are highly diverse on the different Y chromosomes, occurring both tandemly and dispersed across contigs, suggesting that the mechanisms generating their amplification are likely diverse as well.

Evolutionary forces driving gene gain on Y

What is driving gene gain on the Y? The Y chromosome of *D. pseudoobscura* is highly repetitive and epigenetically silenced (Wei, Gibilisco, and Bachtrog 2020; Nguyen and Bachtrog 2021), and some genes may accumulate neutrally. Y chromosomes are limited to males, and Y-linked genes could have male-beneficial functions (Rice 1984; B. Charlesworth and Charlesworth 2000). Sex chromosomes are prone to acquire meiotic drive elements, and Y-linked genes may also be involved in a genomic conflict over sex chromosome segregation (Helleu, Gérard, and Montchamp-Moreau 2014; Werren 2011; Taylor and Jaenike 2003).

Gene ontology shows that Y genes are enriched for spermatogenesis functions (GO terms: spermatid development; spermatid differentiation; spermatid nucleus differentiation; **Table 4, 5**). This functional enrichment is consistent with a subset of Y-genes having male-beneficial functions. Furthermore, Y genes are also enriched for chromatin-related functions (sperm chromatin condensation; spermatogenesis, exchange of chromosomal proteins; histone exchange; DNA packaging; chromatin organization; **Tables 4,5**). Meiotic drivers have been shown to target these processes in *Drosophila* (Courret et al. 2019; Lindholm et al. 2016; Montchamp-Moreau 2006), and enrichment of these genes on the Y may indicate past meiotic conflicts during spermatogenesis. The most highly amplified gene on Y_L , *mh*, encodes a conserved protease that is essential for the integrity of paternal chromosomes during the first zygotic mitosis, and is thought to be necessary for the removal of proteins specifically enriched on sperm chromatin (Tang et al. 2017). We also find dozens of copies of *aubergine (aub)* on Y_L ; *aub* acts via the piRNA pathway to silence TEs. In testis, *aub* associates with *Su(Ste)* piRNAs to downregulate *Ste* transcripts, which has been proposed to be a cryptic meiotic drive system in *D. melanogaster* (HURST 1992).

To further investigate the evolutionary forces driving gene gain on the Y chromosome, we utilized gene expression and small RNA data from the reference genome strain (which contains Y_S). **Figure 3** shows that many Y_S genes are currently not expressed; these genes may have accumulated neutrally on the Y or could have been involved in past meiotic conflicts that are now silenced. Genes on the Y are often highly expressed in the testis, suggesting they have male-beneficial functions (Lahn and Page 1997; Zhou and Bachtrog 2012; Carvalho et al. 2001). Indeed, we find another set of Y_S genes that are exclusively expressed in male gonads (**Figure 3, Table 4**), consistent with some Y genes having male-beneficial functions. Finally, we also identify some Y genes that appear to be expressed in both testis and ovaries, consistent with many Y genes having orthologs on the X/autosomes. Intriguingly, many of these genes have functions related to histone modifications and chromatin (**Figure 3A, Table 4**), consistent with the notion that some genes that have amplified on the Y are involved in past or current conflicts over segregation.

Meiotic drive suppression can be achieved by RNAi which involves post-transcriptional silencing via small RNAs (smRNA) (Lin et al. 2018). We utilized testis-derived smRNA libraries to investigate if a subset of Y_S genes in *D. pseudoobscura* shows signatures of previous or current meiotic drive. Overall, we find similar levels of smRNA mappings on and off the Y chromosome (**Figure 3B**, two-tailed Wilcoxon-test). However, Y genes that are highly expressed in testis have fewer smRNA mapped compared to those that were found to have expression in both testis and ovaries (i.e. chromatin-related), though this is not significant (**Figure 3B**). As expected, most smRNAs align to regions flanking Y genes and this likely suggests the repression

of nearby transposable elements instead (**Figure 3C**). Slightly elevated levels of smRNA that target chromatin-related genes on the Y might suggest sex chromosome meiotic drive but further functional work will be required for validation. Thus, our results are consistent with the idea that gene copy number variation on the Y may have resulted from both male-beneficial effects of Y-linked genes, and previous or ongoing intragenomic conflict over sex chromosome transmission.

Conclusions

Polymorphic Y chromosomes in *Drosophila* influence a variety of fitness-related traits, including lifespan (Griffin et al. 2015), male fertility (Chippindale and Rice 2001), temperature tolerance (Rohmer et al. 2004), and expression of 100s of genes across the genome (Lemos, Araripe, and Hartl 2008; Lemos, Branco, and Hartl 2010). Extensive functional variation on the Y chromosome has been somewhat of a paradox since nucleotide diversity of protein-coding genes is low in *Drosophila* (Larracuente and Clark 2013; Zurovcova and Eanes 1999). Y chromosomes have been shown to contain structural polymorphisms (Lyckegaard and Clark 1989), and it has been speculated that structural variation involving repetitive regions influences Y-linked traits, by globally modulating the genome-wide balance of heterochromatin within the genome (Francisco and Lemos 2014; Brown, Nguyen, and Bachtrog 2020b). Here, we uncover a novel source of variation on the Y chromosome: polymorphic gene content. We show that both the identity of genes, and their copy number can differ widely across polymorphic Y chromosomes in *Drosophila*. Individual Y chromosomes contain dozens of different gene families, and hundreds of different gene copies, suggesting that gene identity and copy number can be a key source of cryptic variation amongst polymorphic Y's.

Materials & Methods

Fly strains. We used *D. pseudoobscura* strains MV25, KB10 PSU1105-Y, and DPSE124-Y kept at 18C on standard molasses food culture.

De novo gene identification and gene amplification. We used Comparative Annotation Tool (Fiddes et al. 2018) to annotate our Y chromosome assemblies with RNA-Seq data. Briefly, we made a reference fasta file containing the published *D. pseudoobscura* female genome and the Y_S, Y_M, Y_L contigs. We ran CAT using this reference and male head expression data from Y_S and Y_L. We identified the breakpoint for Muller D in the Muller_A-AD contig by aligning the *D. melanogaster* Chr 3L to our *D. pseudoobscura* genome assembly.

We used the consensus gff3 file to estimate gene family copy number variation between the Y's, accounting for Y contig copy number. Genes identified by CAT with NCBI IDs were converted to FlyBase names. Of all Y genes, 64 were not given an NCBI ID so we assigned names to 26 of them by aligning the corresponding transcripts to the *D. pseudoobscura* CDS database.

Gene copy number estimation by BLAST. For an orthologous method to estimate copy number in the Y assemblies, we used BLASTN (-task blastn) to align Flybase *D. pseudoobscura* exon sequences that were over 100bp in length to the assemblies. We kept matches where 90% of

the exon was aligned to the chromosome to estimate gene copy number. Gene copies were adjusted for differences in contig copy number for each Y assembly.

Phylogenetic trees. We extracted CDS sequences for each gene using a custom Python script (kindly provided by Dat Mai). We then used ClustalO to generate multiple sequence alignments and RAxML to create phylogenetic trees. FigTree (v1.4.4) was used for visualization (Rambaut 2007).

GO analysis. We used PANTHER to identify significant biological pathways in our Y chromosome genes. The enrichment analysis included only Y genes with matches to known *D. melanogaster* orthologs and used the *D. melanogaster* genome as the background list (Fisher's Exact test, Bonferroni corrected). We made Tables 4-5 from the exported .json files.

CAT alternative transcript curation. To generate full annotations for all identified genes, we extracted the alternative annotation coordinates from the consensus gff3 file to obtain a bed file. The resulting bed file was intersected with the intermediate transMap gtf file to obtain the transcript information for alternative copies. For further analysis we used only one random transcript per annotation.

We mapped NCBI gene IDs with FlyBase IDs using the "FBgn <=> DB Accession IDs" database (accessed October 20, 2020). Some genes do not have an assigned Fbgn ID. We therefore ran the OrthoDb pipeline (version 6.2.4) on published *D. melanogaster* and *D. pseudoobscura* protein sets from NCBI and FlyBase to assign the remaining genes (Waterhouse et al. 2013).

Tissue-specific expression. For expression analyses, we used HISAT2 to map subsampled reads from male body, male/female carcass, testis, accessory gland, and ovary (SRP009072) to the male genome (Kim et al. 2019). We used Subread featurecounts (-t exon -g gene_id -Q 40) to get gene counts from the curated transcripts file and RNA-Seq mappings to only the Y chromosome, adding a pseudo-count (+1) for each gene (Liao, Smyth, and Shi 2014). We visualized the data using the R function heatmap.2 from the gplots package. For smRNA mapping, we followed (C. Ellison and Bachtrog 2019) to align testis-derived small RNA reads using Bowtie2 (-L 18) to the entire male genomes of Y_S , Y_M , and Y_L .

JBrowse visualization. Coverage was calculated for each expression dataset using Bedtools coverage with a window size of 100bp. Coverage files were converted to BigWigs.

Figure Legends

Figure 1. Gene content and diversity between Y's. **A.** Venn diagram of CAT (top) and BLAST (bottom) gene families between Y's. **B.** Gene copy number distribution for genes shared between all Y's. Colors represent the different Y chromosomes. **C.** Gene copy number distribution for genes shared between 2 Y's. Colors highlight pairwise-shared genes between the variants. For

panels **B** and **C**, the circle area represents copy numbers. Distribution for Y-unique gene families is shown in **Figure S2**.

Figure 2. Diverse origins and gene gain mechanisms of *D. pseudoobscura* Y genes. A.

Phylogenetic trees of genes that are found on all the Y's in multi-copy. Each branch represents a unique copy of the gene family and the color denotes the genomic location (teal: Y_L , yellow: Y_M , green: Y_S , red: X chromosome). **B.** Gene copies shown for different families on different Y's with each color denoting a particular family. Note: the contigs are not proportional to scale amongst each other so the gene markers can be drawn clearly. Contig lengths are denoted.

Figure 3. Testis-specific genes may drive gene content evolution on the Y. A. Tissue-specific expression profiles of Y chromosome genes shown as (top) heatmap and (bottom) boxplot (p-value * < 0.05, ** < 1e-03, *** < 1e-05, two-tailed Wilcoxon test). Each row in the heatmap corresponds to a gene located on the Y chromosome. **B.** Mappings of small RNA from testis to the Y chromosome genes broken down by the genes expression profile from **Figure 3A** and the autosomal genes as background. All pairwise comparisons are not significant (two-tailed Wilcoxon test). **C.** Coverage tracks for testis expression, accessory gland expression, smRNA expression (forward), and smRNA (reverse) are shown along selected testis-expressed genes: (top) FBgn0071408:Dmel/CG12035, (bottom): FBgn0247139.

Figure S1. Phylogenetic trees of 3 previously identified *D. pseudoobscura* Y genes and their copies on different chromosomes as reported in Dupim et al 2018. The lineages are colored as follows: teal (Y_L), yellow (Y_M), green (Y_S), orange (FlyBase), and red (X chromosome).

Figure S2. Gene copy number distribution of gene families that are unique to each Y chromosome (Y_S , Y_M , Y_L). The circle area corresponds to gene copy number and color corresponds to Y chromosome. Data underlying this figure is contained in Table S1.

Figure S3. (A-D). Additional multi-copy gene locations on the different assemblies. **A.** Gene locations of *Caper* and *HDAC11* on Y_L contigs. **B-D.** All gene locations for the largest contigs on each Y assembly. Note that the contigs are drawn to scale within each panel figure and the length (bp) is denoted per contig.

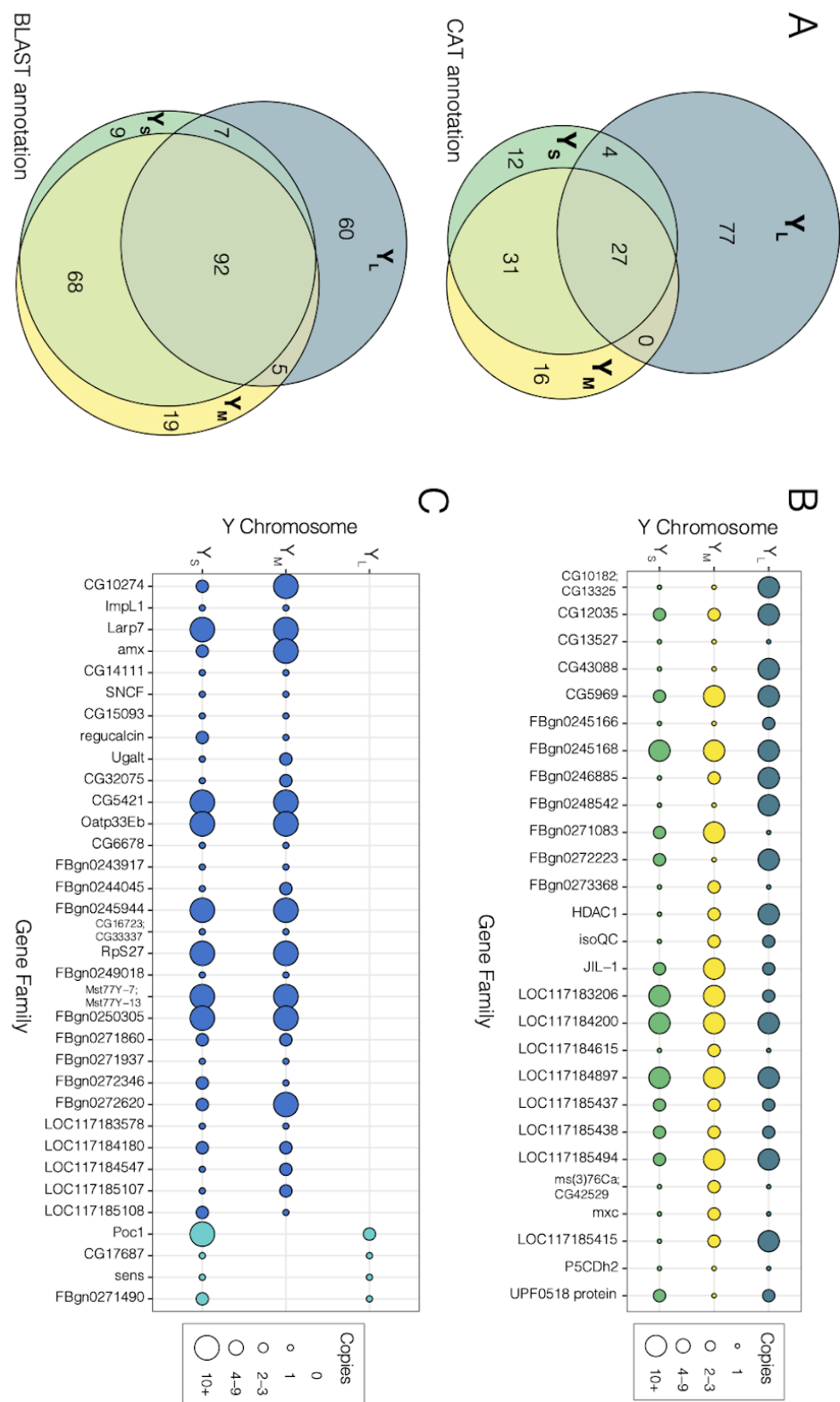


Figure 1.



144

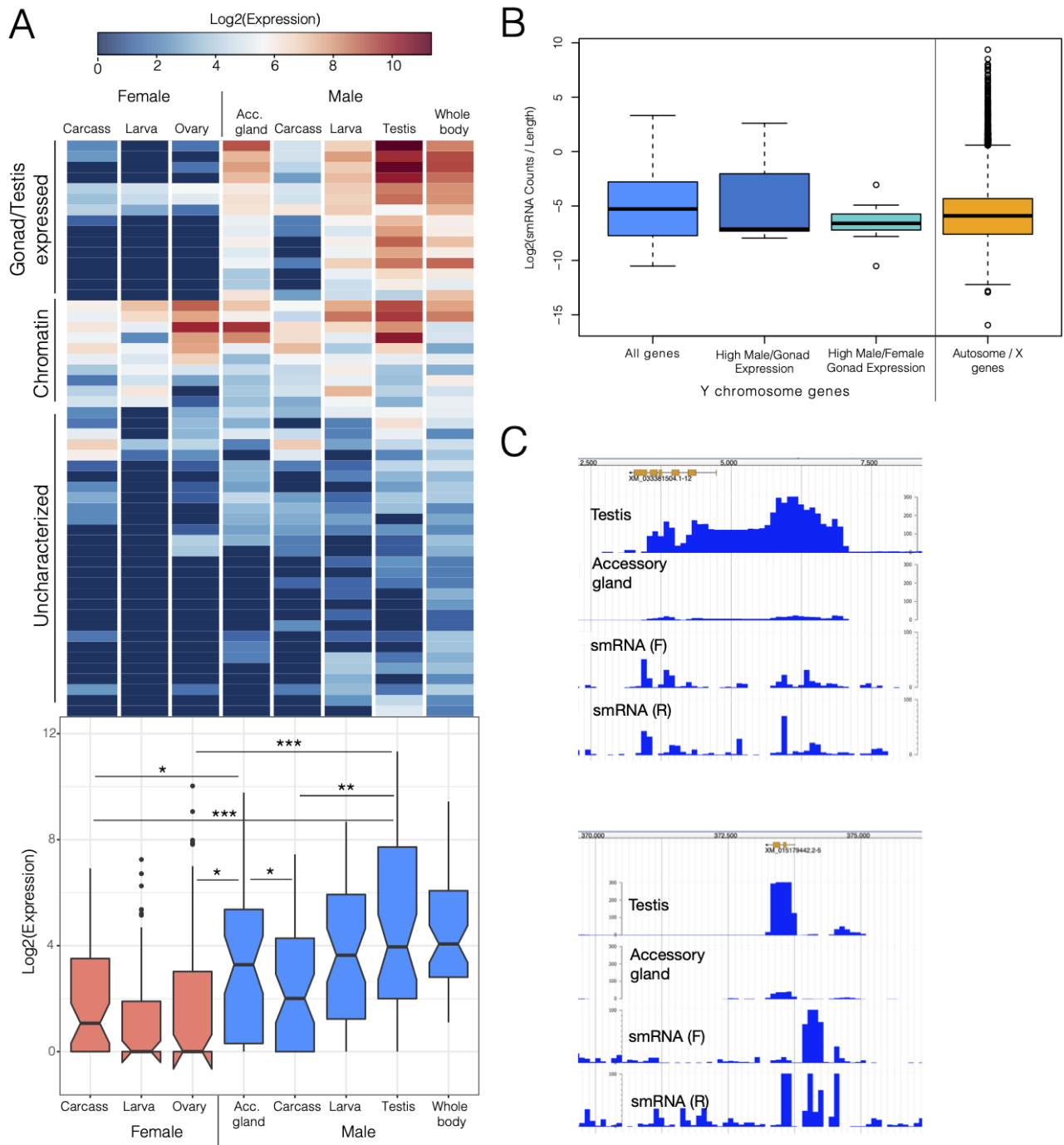


Figure 3.

Table 1. Overview of annotated gene families and copies on the three Y chromosomes

		YS	YM	YL
CAT annotation	Total genes	168	189	297
	Single-copy	38	35	46
	Multi-copy	36	39	62
BLAST annotation	Total genes	914	1033	1440
	Single-copy	41	43	42
	Multi-copy	83	82	76

Table 2. Ancestral location of *D. pseudoobscura* Y genes in *D. pseudoobscura* and *D. subobscura*

		A	D	B	C	E	F
CAT annotation	All Y genes	39	47	41	29	20	3
	Genes shared by all Y's	7	11	3	3	6	2
	All Y genes (<i>D. sub</i>)	33	52	32	24	21	5
	Genes shared by all Y's (<i>D. sub</i>)	4	9	4	5	3	2
BLAST annotation	All Y genes	37	39	37	14	25	4
	Genes shared by all Y's	15	11	3	0	7	2
	All Y genes (<i>D. sub</i>)	40	66	42	19	25	7
	Genes shared by all Y's (<i>D. sub</i>)	12	25	8	2	4	3

Table 3. Distribution of CAT-annotated Y gene families located on Muller A and Muller D by Y chromosome.

	Gene families	Muller A	Muller D	% of Y genes on X
YS	74	17	24	55.4%
YM	74	14	19	44.6%
YL	108	29	35	59.3%
YS, YM, YL	167	39	47	51.5%

Table 4. GO Enrichment of genes identified by PANTHER using CAT matches against D. melanogaster genome (Fisher's Exact, Bonferroni correction).

GO biological process complete	Expected	Over/Under	Fold Enrichment	P-value	Gene_IDs
spermatogenesis, exchange of chromosomal proteins (GO:0035093)	0.036492651	+	109.6111111	6.51E-04	Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
histone exchange (GO:0043486)	0.173340091	+	28.84502924	4.85E-03	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
cellular process (GO:0009987)	33.19918905	+	1.566303319	4.86E-03	PPP1R15,Caper,Rnf146,HDAC11,Mst77Y-4,Rpt6R,Mst77Y-9,RpS27,scramb2,Mst77Y-7,Src64B,mxc,mh,Gart,ed,sens,Parg,PRL-1,regucalcin,isoQC,Poc1,GAPsec,RpS2,Nct,SmD1,Mkk4,HDAC1,Arl1,amx,DnaJ-1,pasi2,wus,eRF3,Mst77F,Larp7,Alg14,RpS19a,Grip128,Atet,P5CDh2,Dys,ko,JIL-1,Oatp33Eb,RpL14,Jafrac2,Ugalt,bdl,Serinc,SLO2,Sam-S,Mst77Y-13
ATP-dependent chromatin remodeling (GO:0043044)	0.250886974	+	19.92929293	2.59E-02	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
sperm chromatin condensation (GO:0035092)	0.077546883	+	64.47712418	1.43E-04	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
spermatid nucleus differentiation (GO:0007289)	0.114039534	+	43.84444444	7.51E-04	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
nucleus organization (GO:0006997)	0.387734415	+	15.4745098	1.19E-02	Parg,Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
DNA packaging (GO:0006323)	0.501773948	+	11.95757576	4.84E-02	HDAC1,Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7

Table 5. GO Enrichment of genes identified by PANTHER using BLAST matches against *D. melanogaster* genome (Fisher's Exact, Bonferroni correction).

GO biological process complete	Expected	Over/Under	Fold Enrichment	P-value	Gene_IDs
spermatogenesis, exchange of chromosomal proteins (GO:0035093)	0.030120918	+	132.7980769	2.98E-04	Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
cellular process (GO:0009987)	27.40250525	+	1.56919959	3.21E-02	SmD1,PPP1R15,Caper,Rnf146,HDAC1,Rpt6,HDAC11,Mst77Y-4,Mst77Y-9,amx,RpS27,scr amb2,Mst77Y-7,DnaJ-1,pasi2,Src64B,Mst77F,Larp7,Alg14,RpS19a,mh,Gart,Atet,sens,Parg,PRL-1,JIL-1,ko,RpL14,Oatp33Eb,Jafrac2,iso QC,bdl,Poc1,aub,ReepA,RpS14b,RpS14a,GA Psec,Sam-S,RpS2,Mst77Y-13,Nct
sperm chromatin condensation (GO:0035092)	0.064006951	+	78.11651584	5.36E-05	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
DNA packaging (GO:0006323)	0.414162624	+	16.90157343	8.71E-04	HDAC1,Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7,aub
DNA conformation change (GO:0071103)	0.523350952	+	13.37534588	3.92E-03	HDAC1,Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7,aub
chromosome organization (GO:0051276)	2.112229382	+	5.207767722	2.43E-02	Parg,JIL-1,HDAC1,Src64B,Mst77F,HDAC11,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7,aub
chromatin organization (GO:0006325)	1.340380856	+	7.460566119	3.00E-03	Parg,JIL-1,HDAC1,Mst77F,HDAC11,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7,aub
spermatid nucleus differentiation (GO:0007289)	0.094127869	+	53.11923077	2.83E-04	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
nucleus organization (GO:0006997)	0.320034755	+	18.7479638	3.79E-03	Parg,Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
histone exchange (GO:0043486)	0.143074361	+	34.94686235	1.84E-03	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
nucleosome organization (GO:0034728)	0.282383607	+	17.70641026	4.15E-02	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7
protein-containing complex subunit organization (GO:0043933)	2.036927087	+	5.400291483	0.01729663	SmD1,RpS14b,RpS14a,Mst77F,Mst77Y-4,RpS19a,Mst77Y-9,RpS27,Mst77Y-13,Mst77Y-7,aub
ATP-dependent chromatin remodeling (GO:0043044)	0.207081312	+	24.1451049	0.009975746	Mst77F,Mst77Y-4,Mst77Y-9,Mst77Y-13,Mst77Y-7

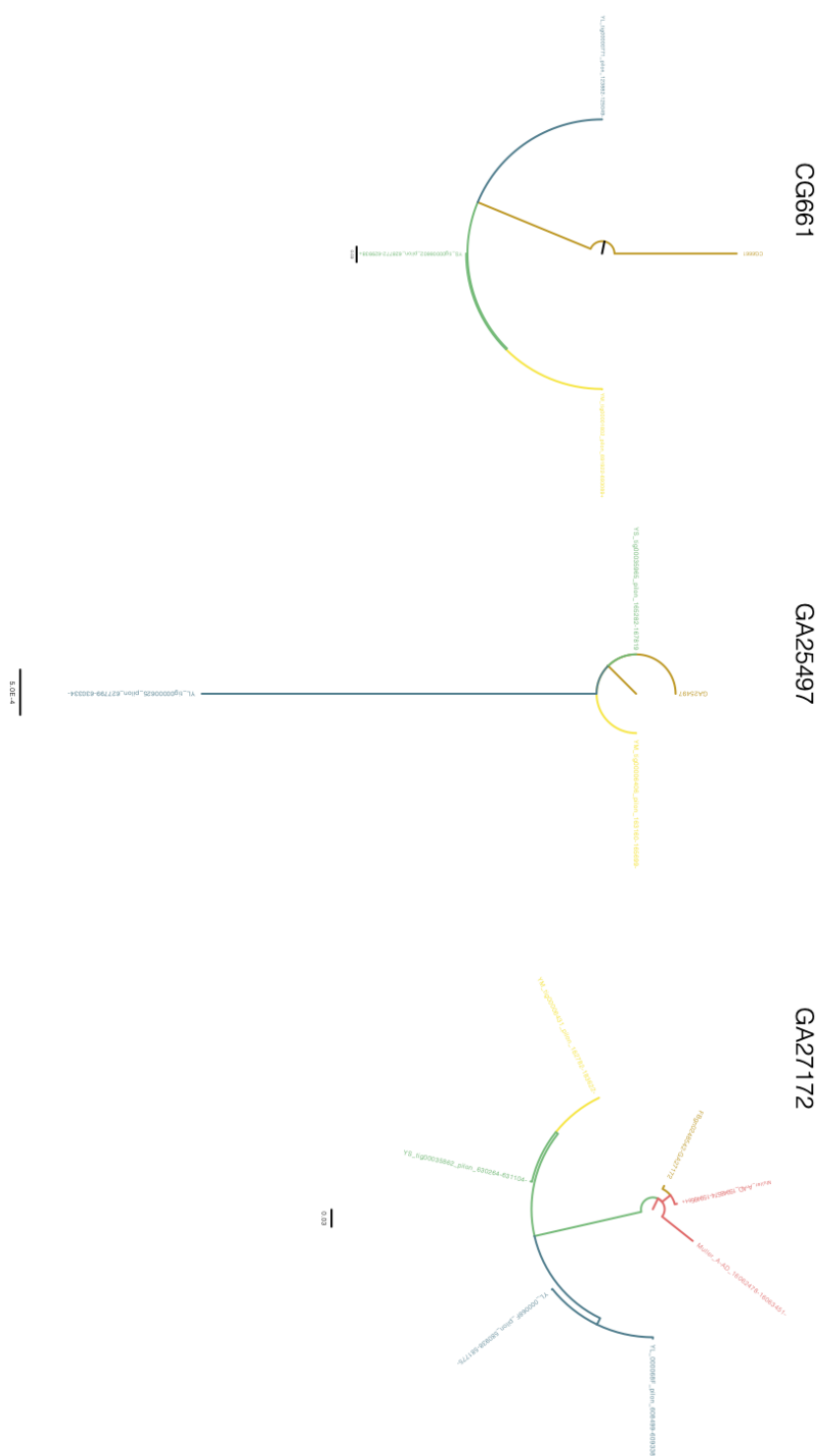


Figure S1.

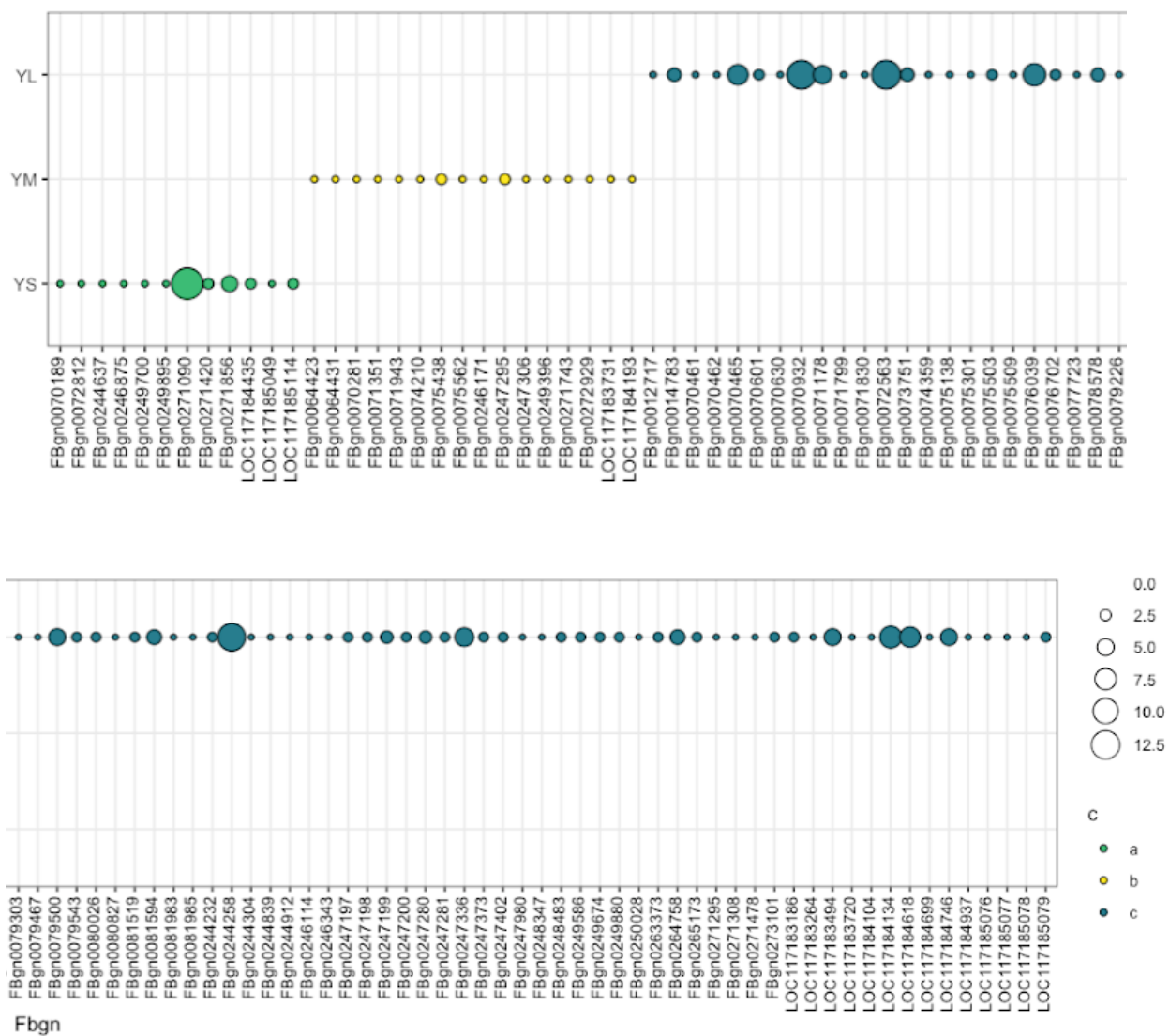


Figure S2.

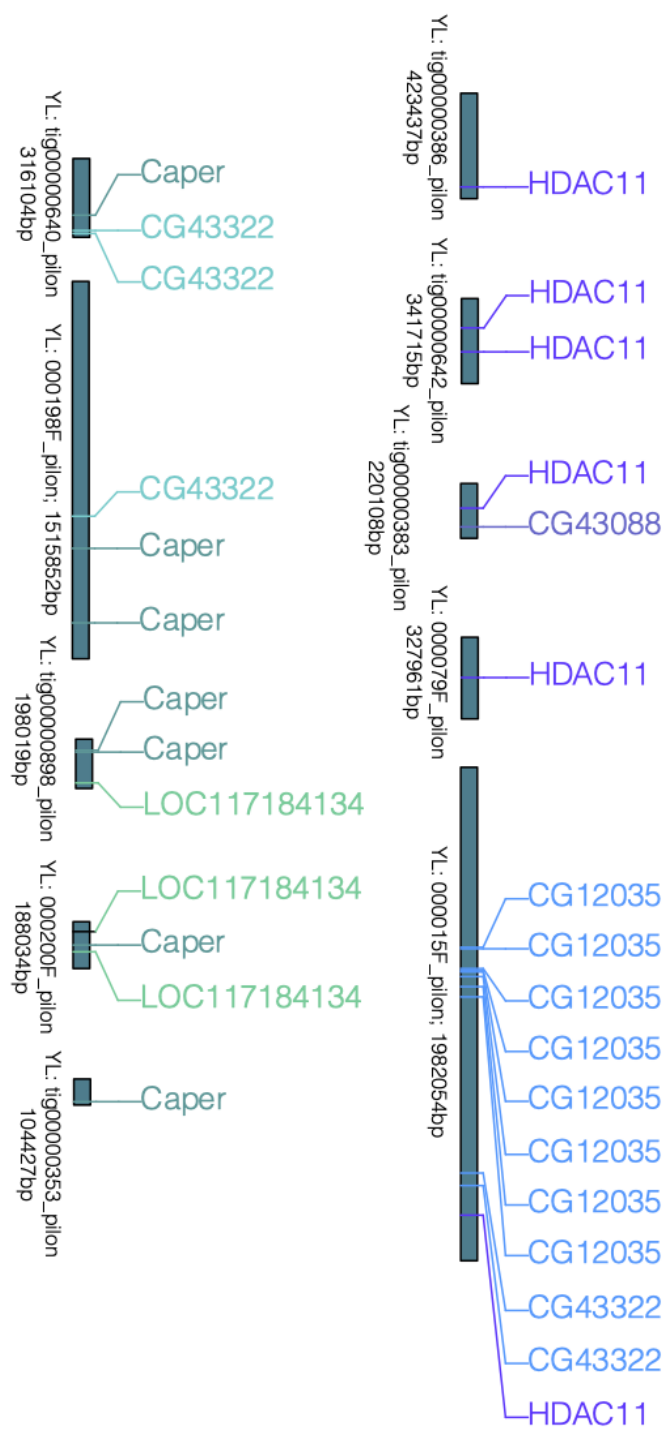


Figure S3A.

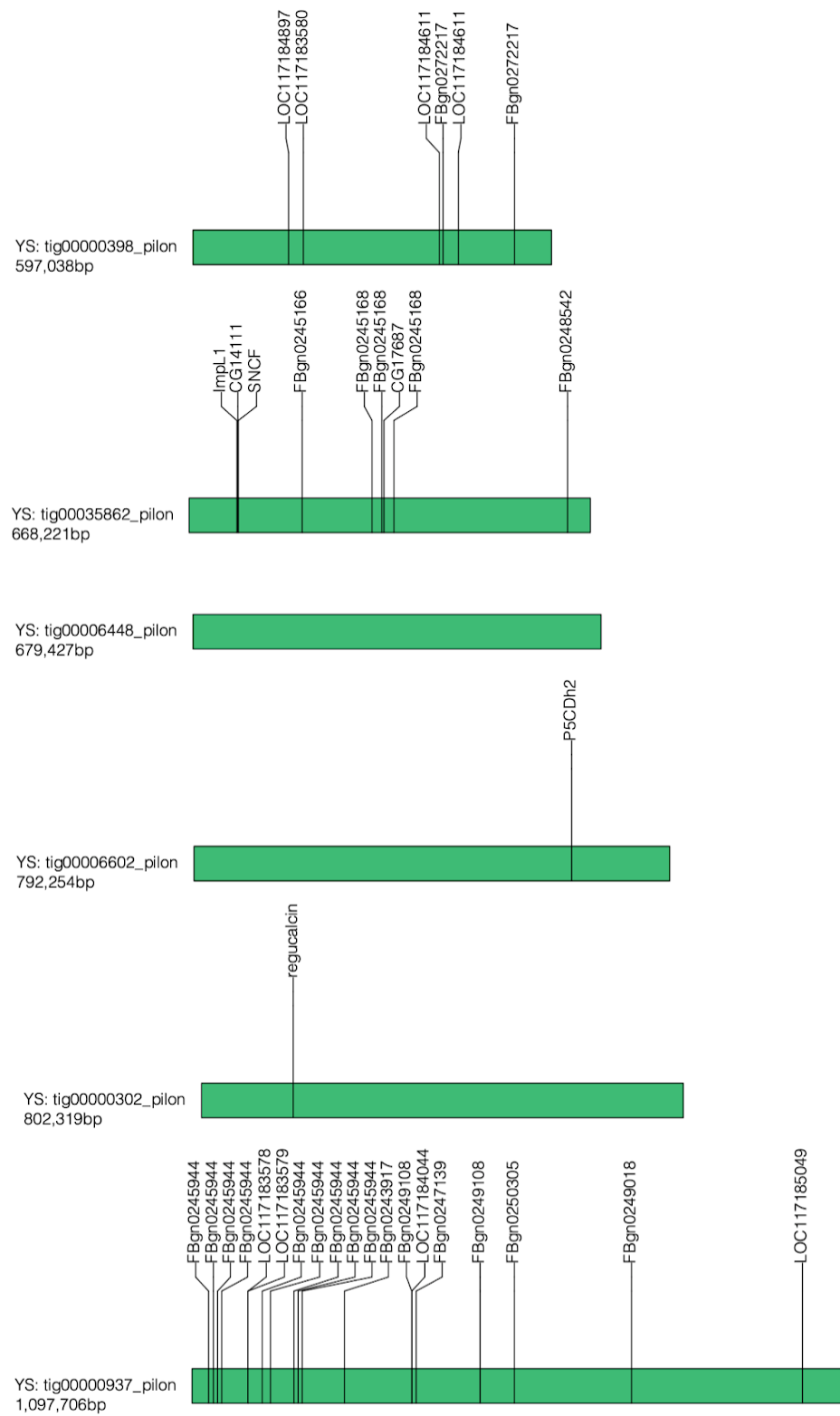


Figure S3B.

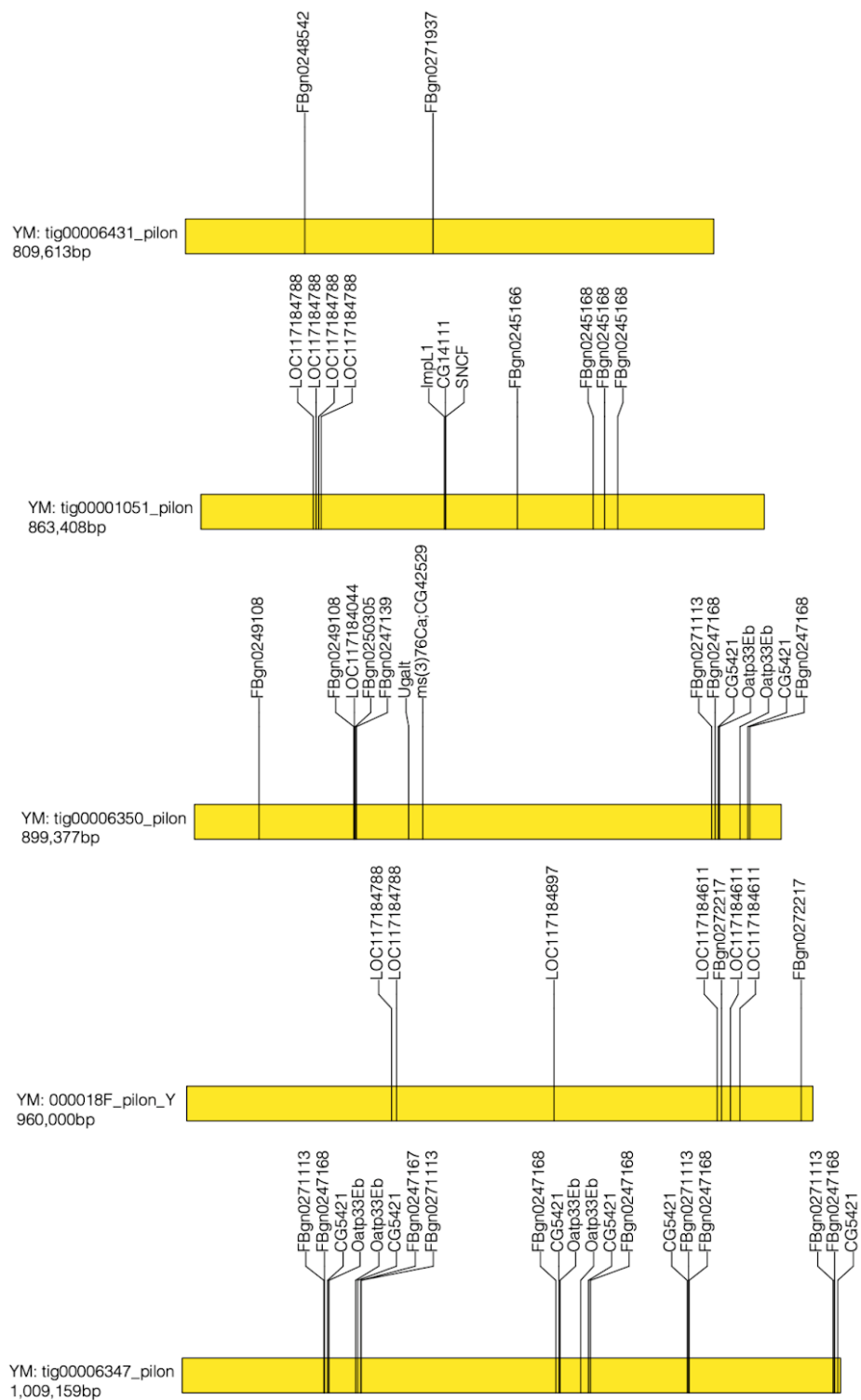


Figure S3C.

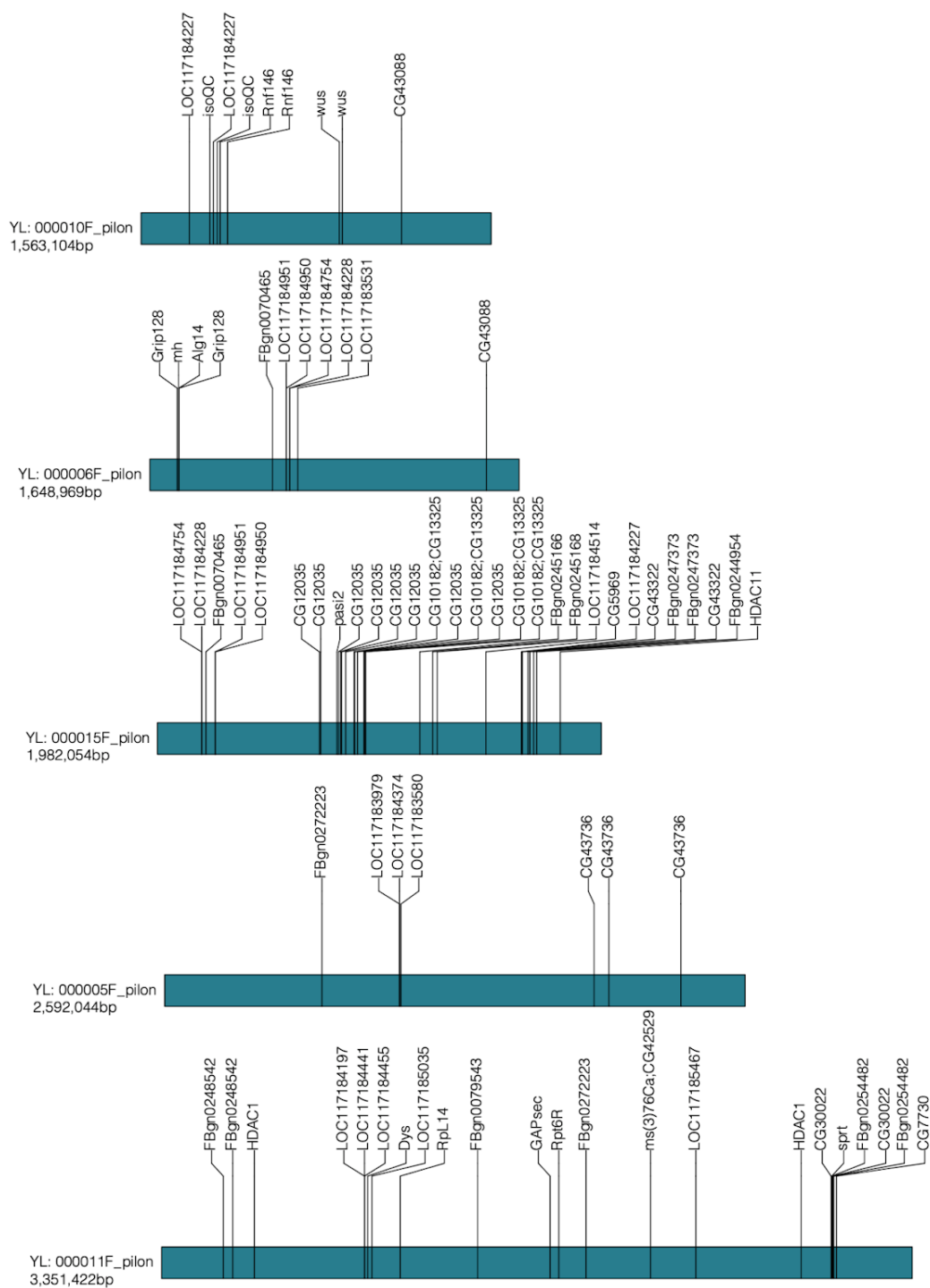


Figure S3D.

Table S1. Y chromosome genes and their copy number as identified by CAT on YS, YM, YL and the rest of the genome. “dpse_pre_b_G” IDs refer to de novo CAT genes.

FBgn/NCBI	A	D	B	C	E	F	YS	YM	YL	D. mel Ortholog	D. subobscura	Feature
FBgn0012717	0	0	1	0	0	0	0	0	1	Gart	Muller_B	protein
FBgn0014783	2	0	0	1	0	0	0	0	3	-	Muller_C	protein
FBgn0064423	0	0	0	1	0	0	0	1	0	spz6	Muller_C	protein
FBgn0064431	0	0	0	1	0	0	0	1	0	CG3880	Muller_C	protein
FBgn0070189	0	1	0	0	0	0	1	0	0	CG10171	Muller_D	protein
FBgn0070203	0	1	0	0	0	0	4	0	2	Poc1	Muller_D	protein
FBgn0070269	0	2	0	0	0	0	3	4	0	CG10274	Muller_D	protein
FBgn0070281	0	1	0	0	0	0	0	1	0	fnt	Muller_D	protein
FBgn0070461	0	1	0	0	0	0	0	0	1	ko	Muller_D	protein
FBgn0070462	0	1	0	0	0	0	0	0	1	I-2	Muller_D	protein
FBgn0070465	0	1	1	0	0	0	0	0	6	DnaJ-1	Muller_D	protein
FBgn0070574	0	1	0	0	0	0	1	1	0	ImpL1	Muller_D	protein
FBgn0070601	0	1	0	0	0	0	0	0	2	SmD1	Muller_D	protein
FBgn0070630	0	1	0	0	0	0	0	0	1	CG10809	Muller_D	protein
FBgn0070932	0	0	1	0	0	0	0	0	11	Caper	Muller_B	protein
FBgn0071178	0	0	0	0	1	0	0	0	5	CG11656	Muller_E	protein
FBgn0071351	0	0	0	1	0	0	0	1	0	CG11961	Muller_C	protein
FBgn0071408	0	1	0	0	0	0	2	2	8	CG12035	Muller_D	protein
FBgn0071420	1	0	0	0	0	0	4	10	0	Larp7	Muller_A	protein
FBgn0071472	1	0	0	0	0	0	1	2	1	mxs	Muller_A	protein
FBgn0071475	1	0	0	0	0	0	3	8	0	amx	Muller_A	protein
FBgn0071799	0	0	2	0	0	0	0	0	1	ed	Muller_B	protein
FBgn0071830	0	1	0	0	0	0	0	0	1	Jafrac2	Muller_D	protein
FBgn0071943	0	0	0	1	0	0	0	1	0	SLO2	Muller_C	protein
FBgn0072563	0	0	1	0	0	0	0	0	11	CG43322	Muller_B	protein
FBgn0072804	0	1	0	0	0	0	1	2	1	ms(3)76Ca;CG42529	Muller_D	protein
FBgn0072812	0	1	0	0	0	0	1	0	0	CG14110	Muller_D	protein
FBgn0072813	0	1	0	0	0	0	1	1	0	CG14111	Muller_D	protein

FBgn0072814	0	1	0	0	0	0	1	1	0	SNCF	Muller_D	protein
FBgn0073520	0	0	0	1	0	0	1	1	0	CG15093	Muller_C	protein
FBgn0073751	0	0	1	0	0	0	0	0	3	CG15422	Muller_B	protein
FBgn0074210	0	0	1	0	0	0	0	1	0	bdI	Muller_B	protein
FBgn0074359	0	0	1	0	0	0	0	0	1	CG17118	Muller_B	protein
FBgn0074640	0	1	0	0	0	0	1	0	1	CG17687	Muller_D	protein
FBgn0074797	1	0	0	0	0	0	2	1	0	regucalcin	Muller_A	protein
FBgn0075138	0	1	0	0	0	0	0	0	1	scramb2	Muller_D	protein
FBgn0075301	0	1	0	0	0	0	0	0	1	CG2162	Muller_D	protein
FBgn0075438	0	0	1	0	0	0	0	2	0	Sam-S	Muller_B	protein
FBgn0075439	1	0	0	0	0	0	1	2	0	Ugalt	Muller_A	protein
FBgn0075503	0	0	0	1	0	0	0	0	2	CG2852	Muller_C	protein
FBgn0075509	1	0	0	0	0	0	0	0	1	Parg	Muller_A	protein
FBgn0075562	0	0	2	0	0	0	0	1	0	Atet	Muller_B	protein
FBgn0076039	0	0	0	0	3	0	0	0	7	HDAC11	Muller_E	protein
FBgn0076674	0	1	0	0	0	0	1	2	0	CG32075	Muller_D	protein
FBgn0076702	0	1	0	0	0	0	0	0	2	CG32109	Muller_D	protein
FBgn0076711	0	1	0	0	0	0	1	0	1	sens	Muller_D	protein
FBgn0077287	0	1	0	0	1	0	1	1	1	P5CDh2	Muller_E	protein
FBgn0077723	0	0	0	1	0	0	0	0	1	PPP1R15	Muller_C	protein
FBgn0078578	0	0	1	0	0	0	0	0	3	PRL-1	Muller_B	protein
FBgn0078868	0	0	1	0	0	0	7	10	0	CG5421	Muller_B	protein
FBgn0079226	1	0	2	0	0	0	0	0	1	RpS2	Muller_B	protein
FBgn0079264	0	1	0	0	0	0	2	4	4	CG5969	Muller_D	protein
FBgn0079270	0	1	0	0	0	0	1	2	2	isoQC	Muller_D	protein
FBgn0079303	0	1	0	0	0	0	0	0	1	Arl1	Muller_D	protein
FBgn0079467	0	1	0	0	0	0	0	0	1	RpL14	Muller_D	protein
FBgn0079500	1	0	0	0	0	0	0	0	5	Alg14	Muller_A	protein
FBgn0079543	0	2	0	0	0	0	0	0	2	S-Lap1	Muller_D	protein
FBgn0079573	0	0	1	0	0	0	4	8	0	Oatp33Eb	Muller_B	protein
FBgn0079769	0	0	0	0	1	0	1	1	0	CG6678	Muller_E	protein

FBgn0080026	0	0	0	0	0	0	0	0	2	Nct	Muller_E	protein
FBgn0080373	0	1	0	0	1	0	1	2	4	HDAC1	Muller_D	protein
FBgn0080827	0	0	0	0	1	0	0	0	1	pasi2	Muller_E	protein
FBgn0081519	1	0	0	0	0	0	0	0	2	wus	Muller_A	protein
FBgn0081594	1	0	0	0	0	0	0	0	4	mh	Muller_A	protein
FBgn0081983	0	0	0	0	1	0	0	0	1	Mkk4	Muller_E	protein
FBgn0081985	0	0	0	0	1	0	0	0	1	Ibf2	Muller_E	protein
FBgn0243917	1	0	1	0	1	1	1	1	0	-	Muller_B	protein
FBgn0244045	1	0	0	0	1	0	1	2	0	-	Muller_A	protein
FBgn0244232	1	0	0	0	0	0	0	0	2	Grip128	Muller_A	protein
FBgn0244258	1	0	0	0	0	0	0	0	12	CG43736	Muller_A	protein
FBgn0244304	1	0	0	0	0	0	0	0	1	CG12112	Muller_A	protein
FBgn0244637	0	0	0	0	0	0	1	0	0	RpS19a	Muller_A	protein
FBgn0244839	0	1	0	0	0	0	0	0	1	iPLA2-VIA	Muller_D	protein
FBgn0244912	0	0	0	0	0	0	0	0	1	-	Muller_D	protein
FBgn0244954	0	1	0	0	0	0	2	4	2	JIL-1	Muller_D	protein
FBgn0245166	0	1	0	0	0	0	1	1	2	-	Muller_E	protein
FBgn0245168	0	1	0	0	0	0	4	4	4	-	Muller_D	protein
FBgn0245944	0	1	0	0	0	0	9	8	0	-	Muller_A	protein
FBgn0246086	0	1	0	1	0	0	1	1	1	CG13527	Muller_C	protein
FBgn0246114	0	0	0	1	0	0	0	0	1	CG5969	Muller_D	protein
FBgn0246171	0	0	0	1	0	0	0	1	0	-	Muller_C	protein
FBgn0246343	0	2	0	1	0	0	0	0	1	CG7730	Muller_D	protein
FBgn0246875	0	0	0	0	0	0	1	0	0	CG6845	Muller_D	protein
FBgn0246885	0	0	0	0	0	0	1	2	9	-	Muller_F	protein
FBgn0247139	0	0	3	0	0	0	6	5	0	-	Muller_A	ncRNA
FBgn0247167	0	0	1	0	0	0	0	1	0	-	Muller_B	pseudogene
FBgn0247168	0	0	2	0	0	0	10	11	0	-	Muller_B	pseudogene
FBgn0247197	0	1	1	0	0	0	0	0	2	-	Muller_D	protein
FBgn0247198	0	0	1	0	0	0	0	0	2	-	Muller_B	protein

FBgn0247199	0	0	1	0	0	0	0	0	3	-	Muller_B	protein
FBgn0247200	0	0	1	0	0	0	0	0	2	-	Muller_B	protein
FBgn0247280	1	0	1	0	0	0	0	0	3	CG2898	Muller_A	protein
FBgn0247281	0	0	2	0	0	0	0	0	2	-	Muller_B	protein
FBgn0247295	0	0	0	0	0	0	0	2	0	CG3558	Muller_B	protein
FBgn0247306	0	0	1	0	0	0	0	1	0	eRF3	Muller_B	protein
FBgn0247336	0	0	1	0	0	0	0	0	6	-	Muller_C	protein
FBgn0247373	0	0	1	0	0	0	0	0	2	-	Muller_D	protein
FBgn0247402	0	0	2	0	0	0	0	0	2	CG17470	Muller_B	protein
FBgn0247980	1	0	0	0	1	0	0	0	1	Rpt6R	Muller_E	protein
FBgn0248189	0	0	0	0	2	0	1	1	0	CG16723;CG33337	Muller_E	protein
FBgn0248191	0	0	0	0	1	0	1	1	4	CG10182;CG13325	Muller_E	protein
FBgn0248347	0	0	0	1	1	0	0	0	1	-	Muller_C	protein
FBgn0248483	1	0	0	0	0	0	0	0	2	-	Muller_A	protein
FBgn0248542	2	0	0	0	0	0	1	1	5	-	Muller_C	protein
FBgn0248937	0	0	0	0	1	0	4	7	0	RpS27	Muller_E	protein
FBgn0249018	0	0	0	0	0	0	1	1	0	-	Muller_A	protein
FBgn0249087	0	0	0	0	0	0	2	2	0	CG45766;Mst77Y-4;Mst77Y-9;CG45765;Mst77F;CG45770;CG45764;Mst77Y-7;Mst77Y-13	Muller_B	protein
FBgn0249108	0	0	2	0	0	0	4	6	0	CG45766;Mst77Y-4;Mst77Y-9;CG45765;Mst77F;CG45770;CG45764;Mst77Y-7;Mst77Y-13	Muller_B	protein
FBgn0249396	0	0	2	1	0	0	0	1	0	-	Muller_B	protein
FBgn0249586	0	1	0	0	0	0	0	0	2	Rnf146	Muller_D	protein
FBgn0249674	0	2	0	0	0	0	0	0	2	-	Muller_D	protein
FBgn0249700	0	1	0	0	0	0	1	0	0	Src64B	Muller_D	protein
FBgn0249880	0	1	0	0	0	0	0	0	2	Serinc	Muller_D	protein
FBgn0249895	0	1	0	0	0	0	1	0	0	CG42540	Muller_D	protein
FBgn0250028	0	2	0	0	0	0	0	0	1	GAPsec	Muller_D	protein
FBgn0250305	0	0	1	0	0	0	9	9	0	-	Muller_C	protein

FBgn0254482	0	0	0	1	0	0	0	0	5	-	Muller_E	rRNA
FBgn0255037	0	1	0	0	0	0	0	0	1	-	Muller_D	snRNA
FBgn0263373	0	0	0	0	1	0	0	0	2	Dys	Muller_E	protein
FBgn0264758	0	0	0	2	0	0	0	0	4	CG30022	Muller_C	protein
FBgn0265173	0	0	0	1	0	0	0	0	2	sppt	Muller_C	protein
FBgn0271075	0	0	1	0	0	0	0	0	2	-	Muller_B	ncRNA
FBgn0271083	0	1	0	0	0	0	2	4	1	-	Muller_C	protein
FBgn0271090	0	0	0	0	0	0	13	0	0	-	Muller_C	protein
FBgn0271113	0	0	2	0	0	0	6	6	0	-	Muller_C	ncRNA
FBgn0271295	0	0	0	0	0	0	0	0	1	-	Muller_A	protein
FBgn0271308	0	0	0	0	0	0	0	0	1	-	Muller_A	protein
FBgn0271420	0	0	0	1	0	0	2	0	0	-	Muller_D	protein
FBgn0271478	0	0	0	0	0	0	0	0	1	-	Muller_A	protein
FBgn0271490	0	0	0	0	1	0	3	0	1	-	Muller_B	protein
FBgn0271743	0	0	0	1	0	0	0	1	0	-	Muller_C	protein
FBgn0271775	1	0	0	1	0	0	1	1	6	CG43088	Muller_F	protein
FBgn0271856	0	0	0	0	0	0	4	0	0	-	Muller_F	protein
FBgn0271860	0	0	0	0	0	0	2	2	0	-	Muller_E	protein
FBgn0271937	0	0	1	0	0	0	1	1	0	-	Muller_B	protein
FBgn0272217	0	0	0	0	0	0	2	4	6	-	Muller_C	ncRNA
FBgn0272223	3	0	1	1	0	0	3	1	10	-	Muller_A	protein
FBgn0272346	0	0	0	0	0	0	2	1	0	-	Muller_F	protein
FBgn0272606	0	1	0	0	0	0	0	0	1	-	Muller_D	ncRNA
FBgn0272620	0	0	0	2	0	0	2	4	0	-	Muller_C	protein
FBgn0272784	0	0	0	4	0	0	0	1	0	-	Muller_C	ncRNA
FBgn0272929	0	1	0	0	0	0	0	1	0	-	Muller_D	protein
FBgn0272986	1	0	0	0	0	0	7	9	1	-	Muller_F	ncRNA
FBgn0273101	0	0	0	0	0	0	0	0	2	CG7824	Muller_D	protein
FBgn0273368	1	0	0	0	1	0	1	3	1	-	Muller_A	protein
LOC117183186	1	0	1	0	0	0	0	0	2	protein aubergine-like	Muller_B	protein
LOC117183199	0	0	1	0	0	0	0	0	2	-	Muller_B	lncRNA

LOC117183206	0	0	0	0	0	3	5	5	3	-	Muller_B	protein
LOC117183263	0	0	0	0	0	0	0	0	1	-	Muller_A	lncRNA
LOC117183264	2	0	1	0	0	0	0	0	1	-	Muller_A	protein
LOC117183494	0	1	1	1	0	0	0	0	5	-	Muller_A	protein
LOC117183525	0	0	0	0	0	0	0	0	1	-	Muller_C	lncRNA
LOC117183531	3	1	1	1	0	0	1	6	6	-	Muller_B	lncRNA
LOC117183532	1	1	1	1	0	0	0	1	0	-	Muller_D	lncRNA
LOC117183578	0	0	0	1	0	0	1	1	0	-	Muller_A	protein
LOC117183579	0	0	0	0	0	0	1	1	0	-	Muller_E	lncRNA
LOC117183580	0	0	0	0	0	0	1	2	4	-	Muller_C	lncRNA
LOC117183647	0	0	0	0	0	0	2	2	0	-	Muller_C	lncRNA
LOC117183720	0	0	0	1	0	0	0	0	1	VMA21 homolog	Muller_D	protein
LOC117183731	0	0	0	1	0	0	0	1	0	101 kDa malaria antigen	Muller_C	protein
LOC117183733	0	0	0	1	0	0	0	1	0	-	Muller_C	lncRNA
LOC117183771	0	0	0	1	0	0	0	1	0	-	Muller_C	lncRNA
LOC117183788	5	0	2	1	1	1	0	3	0	-	Muller_A	lncRNA
LOC117183979	0	0	0	0	0	0	0	0	3	-	Muller_B	lncRNA
LOC117184044	0	0	2	0	0	0	2	3	0	-	Muller_A	lncRNA
LOC117184101	1	0	2	0	0	0	0	0	2	-	Muller_B	lncRNA
LOC117184104	1	0	0	0	0	0	0	0	1	-	Muller_E	protein
LOC117184134	0	0	2	0	0	0	0	0	8	protein aubergine-like	Muller_E	protein
LOC117184180	0	0	1	2	0	0	2	2	0	-	Muller_A	protein
LOC117184193	0	0	1	0	0	0	0	1	0	regulator of chromosome condensation-like	Muller_D	protein
LOC117184197	0	0	0	0	0	0	0	0	1	-	Muller_E	lncRNA
LOC117184200	0	0	0	0	0	3	6	5	4	-	Muller_B	protein
LOC117184227	0	0	0	0	1	1	1	0	7	-	Muller_A	lncRNA
LOC117184228	0	0	1	0	0	0	0	1	3	-	Muller_A	lncRNA
LOC117184344	0	0	0	0	0	1	1	0	0	-	Muller_F	lncRNA
LOC117184351	1	0	0	0	0	0	2	0	1	-	Muller_F	lncRNA

LOC117184374	0	0	0	0	0	0	0	0	3	-	Muller_A	lncRNA
LOC117184435	1	0	0	0	0	0	2	0	0	-	Muller_D	protein
LOC117184436	0	0	0	0	0	0	4	1	0	-	Muller_F	lncRNA
LOC117184441	0	0	0	0	0	0	0	0	1	-	Muller_E	lncRNA
LOC117184455	0	0	0	0	0	0	0	0	1	-	Muller_E	lncRNA
LOC117184514	1	0	1	0	0	0	0	0	3	-	Muller_D	lncRNA
LOC117184547	2	0	0	0	0	0	1	2	0	-	Muller_D	protein
LOC117184611	0	0	0	0	0	0	3	8	0	-	Muller_D	lncRNA
LOC117184615	1	0	1	0	1	0	1	3	1	-	Muller_D	protein
LOC117184618	2	0	0	0	0	0	0	0	7	-	Muller_F	protein
LOC117184637	2	0	0	0	0	0	0	0	1	-	Muller_D	lncRNA
LOC117184670	0	1	0	0	0	0	1	1	2	-	Muller_E	lncRNA
LOC117184699	1	0	0	0	0	0	0	0	1	-	Muller_B	protein
LOC117184700	1	1	0	0	1	0	0	0	3	-	Muller_E	lncRNA
LOC117184724	7	0	3	1	0	0	1	0	0	-	Muller_B	lncRNA
LOC117184746	0	1	1	1	0	0	0	0	5	-	Muller_A	protein
LOC117184754	0	0	1	0	0	0	0	1	3	-	Muller_A	lncRNA
LOC117184788	1	0	0	0	1	0	8	8	0	-	Muller_F	lncRNA
LOC117184798	2	0	0	0	0	0	0	0	1	-	Muller_A	lncRNA
LOC117184826	1	0	0	0	0	0	2	4	0	-	Muller_A	lncRNA
LOC117184829	4	0	1	0	0	0	2	4	0	-	Muller_A	lncRNA
LOC117184897	2	0	0	0	0	0	4	5	4	-	Muller_A	protein
LOC117184932	0	0	0	0	0	0	1	2	1	-	Muller_E	lncRNA
LOC117184935	0	0	0	0	0	0	0	0	1	-	Muller_B	lncRNA
LOC117184937	2	0	1	0	0	0	0	0	1	-	Muller_A	protein
LOC117184942	0	0	0	0	0	0	1	2	1	-	Muller_B	lncRNA
LOC117184950	0	0	0	0	0	0	0	0	5	-	Muller_A	lncRNA
LOC117184951	0	0	0	0	0	0	0	0	5	-	Muller_A	lncRNA
LOC117185022	0	0	0	0	0	0	1	2	0	-	Muller_D	lncRNA
LOC117185035	2	0	0	0	0	1	0	0	1	-	Muller_E	ncRNA
LOC117185049	3	0	2	2	0	0	1	0	0	-	Muller_A	protein

LOC117185076	2	0	0	0	0	0	0	0	1	-	Muller_E	protein	
LOC117185077	1	0	0	0	0	0	0	0	1	-	Muller_E	protein	
LOC117185078	1	0	0	0	0	0	0	0	1	-	Muller_E	protein	
LOC117185079	2	0	0	0	0	0	0	0	2	-	Muller_E	protein	
LOC117185081	2	0	2	0	0	0	0	0	1	40S ribosomal protein S2-like	Muller_B	rRNA	
LOC117185107	0	0	0	0	0	0	1	2	0		Muller_C	protein	
LOC117185108	0	0	0	0	0	0	3	1	0		Muller_A	protein	
LOC117185114	2	0	0	1	0	0	2	0	0		Muller_A	protein	
LOC117185132	0	0	0	0	0	1	7	4	1		Contig_8	lncRNA	
LOC117185134	0	0	0	0	0	0	2	0	0		Muller_F	lncRNA	
LOC117185415	0	0	0	0	0	0	1	2	9	neurofilament heavy polypeptide-like	Muller_C	protein	
LOC117185426	1	0	0	0	0	0	2	0	1	-	Muller_B	lncRNA	
LOC117185431	0	0	0	0	0	0	1	2	1	-	Muller_E	lncRNA	
LOC117185432	0	0	0	0	0	0	3	5	1	-	Muller_E	lncRNA	
LOC117185433	0	0	0	0	0	0	3	5	1	-	Muller_E	lncRNA	
LOC117185437	0	0	0	0	0	0	2	2	2	-	Muller_D	protein	
LOC117185438	0	0	0	0	0	0	2	2	2	-	Muller_C	protein	
LOC117185467		0	0	0	0	0	0	0	0	2	large subunit ribosomal RNA	Contig_2	rRNA
LOC117185494		0	0	0	0	1	0	2	5	7	-	Muller_B	protein
LOC117185652		0	0	0	0	0	0	0	0	1	-	Muller_C	lncRNA
LOC117185664		0	0	0	0	2	0	2	1	0	-	Muller_B	lncRNA
LOC26532143		0	0	1	0	0	0	2	1	3	UPF0518 protein GA25918	Muller_B	protein
dpse_pre_b_G0016335	0	0	0	0	0	0	0	0	0	1	-	Muller_B	unknown
dpse_pre_b_G0016336	0	0	0	0	0	0	0	0	0	1	-	Muller_B	unknown
dpse_pre_b_G0016347	0	0	0	0	0	0	0	0	0	1	-	Muller_A	unknown
dpse_pre_b_G0016349	0	0	0	0	0	0	0	0	0	1	-	Muller_C	unknown

dpse_pre_b_G00163 50	0	0	0	0	0	0	0	0	1	-	Muller_A	unknown
dpse_pre_b_G00163 55	0	0	0	0	0	0	0	0	1	-	Muller_E	unknown
dpse_pre_b_G00163 59	0	0	0	0	0	0	0	0	1	-	Muller_E	unknown
FBgn0072563	0	0	0	0	0	0	0	0	1	CG43322	Muller_B	unknown
dpse_pre_b_G00163 62	0	0	0	0	0	0	0	0	1	-	Muller_B	unknown
dpse_pre_b_G00163 65	0	0	0	0	0	0	0	0	1	-	Muller_D	unknown
dpse_pre_b_G00163 66	0	0	0	0	0	0	0	0	1	-	Muller_A	unknown
dpse_pre_b_G00163 67	0	0	0	0	0	0	0	0	1	-	Muller_E	unknown
dpse_pre_b_G00163 75	0	0	0	0	0	0	0	0	2	-	Muller_A	unknown
dpse_pre_b_G00163 77	0	0	0	0	0	0	0	0	2	-	Muller_A	unknown
dpse_pre_b_G00163 80	0	0	0	0	0	0	0	0	3	-	Muller_A	unknown
dpse_pre_b_G00163 81	0	0	0	0	0	0	0	0	3	-	Muller_C	unknown
dpse_pre_b_G00163 88	0	0	0	0	0	0	0	1	0	-	Muller_A	unknown
dpse_pre_b_G00163 97	0	0	0	0	0	0	0	1	0	-	Muller_D	unknown
FBgn0272684	0	0	0	0	0	0	3	0	0	CG14190	Muller_F	unknown
dpse_pre_b_G00164 22	0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
dpse_pre_b_G00164 23	0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
dpse_pre_b_G00164 24	0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
dpse_pre_b_G00164 26	0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
dpse_pre_b_G00164 27	0	0	0	0	0	0	1	0	0	-	Muller_C	unknown
dpse_pre_b_G00164	0	0	0	0	0	0	1	0	0	-	Muller_A	unknown

28

dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
34											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_D	unknown
35											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
36											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_C	unknown
37											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Contig_8	unknown
38											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_D	unknown
41											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_A	unknown
42											
dpse_pre_b_G00164 0	0	0	0	0	0	1	0	0	-	Muller_C	unknown
43											

Table S2. Y chromosome genes and their copy number as identified by BLAST searches on YS, YM, YL and the rest of the genome. [NKO] stands for no known orthologs and “other” corresponds to unplaced contigs in the *D. pseudoobscura* genome. Note FBgn0249087 was abbreviated for its full ortholog hit of CG45764, CG45765, CG45766, CG45770, Mst77F, Mst77Y-13, Mst77Y-4, Mst77Y-7, Mst77Y-9.

Gene_Name	A	B	C	D	E	F	YS	YM	YL	other	D.mel	Dsub
FBgn0012717	0	1	0	0	0	0	0	0	1	0	Gart	Muller_B
FBgn0064423	0	0	1	0	0	0	0	1	0	0	spz6	Muller_C
											CG42395,CG4	
FBgn0064426	0	0	0	0	1	0	1	1	0	0	686	Muller_E
FBgn0070189	0	0	0	1	0	0	1	0	0	1	CG10171	Muller_D
FBgn0070203	0	0	0	1	1	0	3	0	2	1	Poc1	Muller_D
FBgn0070269	0	0	0	1	0	0	2	2	0	0	CG10274	Muller_D
FBgn0070368	0	0	0	0	1	0	4	6	0	0	RpS27	Muller_E
FBgn0070461	0	0	0	0	0	0	0	0	2	0	ko	Muller_D
FBgn0070462	0	0	0	1	0	0	0	0	1	0	I-2	Muller_D
FBgn0070465	0	0	0	1	0	0	0	0	11	0	DnaJ-1	Muller_D
FBgn0070574	0	0	0	1	0	0	1	1	0	0	ImpL1	Muller_D
FBgn0070601	0	0	0	1	0	0	0	0	1	0	SmD1	Muller_D
FBgn0070630	0	0	0	1	0	0	0	0	1	0	CG10809	Muller_D
FBgn0070772	0	0	0	0	1	0	0	0	1	0	CG1103	Muller_E
FBgn0070932	0	1	0	0	0	0	0	0	8	0	Caper	Muller_B
FBgn0071408	0	0	0	0	0	0	7	5	7	0	CG12035	Muller_D
FBgn0071420	1	0	0	0	0	0	1	0	0	0	Larp7	Muller_A
FBgn0071475	1	0	0	0	0	0	4	4	0	0	amx	Muller_A
FBgn0071830	0	0	0	1	0	0	0	0	1	0	Jafrac2	Muller_D
FBgn0071902	0	0	1	0	0	0	1	1	0	0	CG12848	Muller_C
FBgn0072563	0	1	0	0	0	0	0	0	39	0	CG43322	Muller_B
											CG42529,ms(	
FBgn0072804	0	0	0	1	0	0	1	1	1	0	3)76Ca	Muller_D
FBgn0072812	0	0	0	1	0	0	1	0	0	0	CG14110	Muller_D
FBgn0072813	0	0	0	1	0	0	1	0	0	0	CG14111	Muller_D
FBgn0073364	1	0	0	0	0	0	0	0	1	0	Rpt6	Muller_A

FBgn0073520	0	0	1	0	0	0	1	1	0	0	CG15093	Muller_C
FBgn0073751	0	1	0	0	0	0	0	0	2	0	CG15422	Muller_B
FBgn0073846	0	0	0	0	1	0	1	0	0	1	CG15553	Muller_E
FBgn0074210	0	1	0	0	0	0	0	1	0	0	bdl	Muller_B
FBgn0074640	0	0	0	1	0	0	1	1	2	0	CG17687	Muller_D
FBgn0075138	0	0	0	1	0	0	0	0	1	0	scramb2	Muller_D
FBgn0075438	0	1	0	0	0	0	0	1	0	0	Sam-S	Muller_B
FBgn0075473	0	0	1	0	0	0	1	2	0	0	CG2765	Muller_C
FBgn0075503	0	0	1	0	0	0	0	0	2	0	CG2852	Muller_C
FBgn0075509	1	0	0	0	0	0	0	0	1	0	Parg	Muller_A
FBgn0075562	0	1	0	0	0	0	0	1	0	0	Atet	Muller_B
FBgn0076039	0	0	0	0	3	0	0	0	18	0	HDAC11	Muller_E
FBgn0076674	0	0	0	1	0	0	1	1	0	0	CG32075	Muller_D
FBgn0076702	0	0	0	1	0	0	0	0	2	0	CG32109	Muller_D
FBgn0076711	0	0	0	1	0	0	1	0	1	0	sens	Muller_D
FBgn0077723	0	0	1	0	0	0	0	0	1	0	PPP1R15	Muller_C
FBgn0077766	0	0	1	0	0	0	7	16	0	1	CG3894	Muller_C
FBgn0078578	0	1	0	0	0	0	0	0	19	0	PRL-1	Muller_B
FBgn0078868	0	1	0	0	0	0	3	4	0	0	CG5421	Muller_B
FBgn0079226	0	2	0	0	0	0	0	0	1	0	RpS2	Muller_B
FBgn0079264	1	0	0	1	0	0	2	2	0	0	CG5969	Muller_D
FBgn0079270	0	0	0	1	0	0	1	1	1	0	isoQC	Muller_D
FBgn0079379	0	3	0	0	0	0	0	0	31	0	aub	Muller_B
FBgn0079467	0	0	0	1	0	0	0	0	1	0	RpL14	Muller_D
FBgn0079500	1	0	0	0	0	0	0	0	304	0	Alg14	Muller_A
FBgn0079573	0	1	0	0	0	0	0	3	0	0	Oatp33Eb	Muller_B
FBgn0079769	0	0	0	0	1	0	1	1	0	0	CG6678	Muller_E
FBgn0080026	0	0	0	0	0	0	0	0	2	0	Nct	Muller_E
FBgn0080373	0	0	0	1	0	0	1	1	4	0	HDAC1	Muller_D
FBgn0080827	0	0	0	0	1	0	0	0	1	0	pasi2	Muller_E

FBgn0081594	1	0	0	0	0	0	0	0	327	0	mh	Muller_A
FBgn0081985	0	0	0	0	1	0	0	0	1	0	Ibf2	Muller_E
FBgn0243614	0	0	0	0	0	0	13	7	9	0	[NKO]	Muller_D
FBgn0243857	1	0	0	0	0	0	83	88	0	3	[NKO]	Muller_C
FBgn0243873	0	0	0	0	0	0	3	7	0	0	CG12848	Muller_D
FBgn0243877	0	0	0	0	0	0	1	3	0	0	[NKO]	Muller_D
FBgn0243878	0	0	0	0	0	0	2	3	0	3	[NKO]	Muller_D
FBgn0243879	0	0	0	0	0	0	0	1	0	0	[NKO]	Muller_D
FBgn0243882	0	0	0	0	0	0	0	3	0	1	[NKO]	Muller_C
											RpS14a,RpS14b	
FBgn0244179	1	0	0	0	0	0	1	1	0	0		Muller_A
FBgn0244258	1	0	0	0	0	0	0	0	8	0	CG43736	Muller_A
FBgn0244304	4	0	0	0	0	0	0	0	1	0	CG12112	Muller_A
FBgn0244599	0	0	0	0	0	0	26	28	17	0	[NKO]	Muller_A
FBgn0244602	0	0	0	0	0	0	10	10	0	0	Larp7	Muller_A
FBgn0244603	0	0	0	0	0	0	4	1	3	1	Larp7	Muller_A
FBgn0244637	0	1	0	0	0	0	1	0	0	0	RpS19a	Muller_A
FBgn0244720	0	0	0	0	0	0	2	1	0	0	[NKO]	Muller_D
FBgn0244839	0	0	0	1	0	0	0	0	1	0	iPLA2-VIA	Muller_D
FBgn0244912	0	0	0	1	0	0	3	3	3	0	[NKO]	Muller_D
FBgn0244954	0	0	0	1	0	0	1	3	1	0	JIL-1	Muller_D
FBgn0245168	0	0	0	1	0	0	4	3	3	0	[NKO]	Muller_D
FBgn0245375	0	0	0	0	0	0	32	42	0	1	[NKO]	Muller_F
FBgn0245376	0	0	0	0	0	0	2	2	0	0	[NKO]	Muller_C
FBgn0245377	0	0	0	0	0	0	5	4	0	0	[NKO]	Muller_B
FBgn0245408	0	0	0	0	0	0	4	4	10	0	[NKO]	Muller_B
FBgn0245453	0	0	0	0	0	0	4	5	10	0	[NKO]	Muller_D
FBgn0245925	0	0	0	1	0	0	1	1	0	0	CG10862	Muller_B
FBgn0246171	0	0	1	0	0	0	0	1	0	0	[NKO]	Muller_C
FBgn0246343	0	0	1	0	0	0	0	0	1	0	CG7730	Muller_D
FBgn0246875	0	0	0	0	0	0	1	1	1	0	CG6845	Muller_D

FBgn0246881	1	0	0	0	0	0	6	6	0	0	[NKO]	Muller_B
FBgn0246885	2	0	0	0	0	0	3	3	1	0	[NKO]	Muller_F
FBgn0246929	0	0	0	0	0	0	1	4	0	0	[NKO]	Muller_B
FBgn0247197	0	1	0	0	0	0	0	0	2	0	[NKO]	Muller_D
FBgn0247199	0	1	0	0	0	0	0	0	1	0	[NKO]	Muller_B
FBgn0247200	0	1	0	0	0	0	0	0	2	0	[NKO]	Muller_B
FBgn0247279	0	1	0	0	0	0	0	0	2	0	CG15800	Muller_A
FBgn0247280	0	1	0	0	0	0	0	0	2	0	CG2898	Muller_A
FBgn0247281	0	1	0	0	0	0	0	0	2	0	[NKO]	Muller_B
FBgn0247282	0	2	0	0	0	0	0	0	2	0	CG31752	Muller_B
FBgn0247295	0	1	0	0	0	0	2	2	2	2	CG3558	Muller_B
FBgn0247306	0	1	0	0	0	0	0	1	0	0	eRF3	Muller_B
FBgn0247336	0	1	0	0	0	0	0	0	7	0	[NKO]	Muller_C
FBgn0247402	0	1	0	0	0	0	0	0	1	0	CG17470	Muller_B
FBgn0248189	0	0	0	0	1	0	1	0	0	0	CG16723,CG3337	Muller_E
FBgn0248190	0	0	0	0	1	0	0	0	3	0	[NKO]	Muller_E
FBgn0248191	0	0	0	0	1	0	0	1	1	0	CG10182,CG13325	Muller_E
FBgn0248340	0	0	0	0	3	0	0	0	24	0	[NKO]	Muller_E
FBgn0248483	1	0	0	0	0	0	0	0	2	0	[NKO]	Muller_A
FBgn0248542	1	0	0	0	0	0	1	1	5	0	[NKO]	Muller_C
FBgn0248937	0	0	0	0	0	0	0	1	0	0	RpS27	Muller_E
FBgn0249018	1	0	0	0	0	0	4	3	9	0	[NKO]	Muller_A
FBgn0249087	0	2	0	0	0	0	3	2	0	0	Mst77Y-7,Mst77Y-9	Muller_B
FBgn0249209	0	0	0	0	0	0	1	1	2	0	[NKO]	Muller_A
FBgn0249210	0	0	0	0	0	0	26	26	0	3	[NKO]	Muller_A
FBgn0249212	0	0	0	0	0	0	0	0	1	0	[NKO]	Muller_A
FBgn0249214	0	0	0	0	0	0	5	16	1	0	[NKO]	Muller_B
FBgn0249217	0	0	0	0	0	0	2	0	0	0	[NKO]	Muller_A

FBgn0249218	0	0	0	0	0	0	2	0	0	0	[NKO]	Muller_A
FBgn0249219	0	0	0	0	0	0	5	3	0	0	[NKO]	Muller_A
FBgn0249268	0	0	0	0	1	0	4	4	1	0	[NKO]	Muller_E
FBgn0249348	0	0	0	0	0	0	3	2	0	0	[NKO]	Muller_D
FBgn0249349	0	0	0	0	0	0	0	2	4	0	[NKO]	Muller_D
FBgn0249452	0	1	0	0	0	0	0	1	0	0	[NKO]	Muller_B
FBgn0249496	7	0	0	0	0	0	1	1	12	0	[NKO]	Muller_B
FBgn0249554	0	0	0	1	0	0	3	7	6	0	[NKO]	Muller_D
FBgn0249586	0	0	0	1	0	0	0	0	1	0	Rnf146	Muller_D
FBgn0249674	0	0	0	1	0	0	0	0	1	0	[NKO]	Muller_D
FBgn0249701	0	0	0	2	0	0	2	0	0	0	Src64B	Muller_D
FBgn0249761	0	0	0	2	0	0	0	0	66	0	[NKO]	Muller_D
FBgn0249782	1	0	0	0	0	0	1	2	0	0	[NKO]	Muller_A
FBgn0249880	0	0	0	1	0	0	0	0	1	0	TMS1	Muller_D
FBgn0250028	0	0	0	1	0	0	1	1	1	0	GAPsec	Muller_D
FBgn0250070	2	0	0	0	0	0	0	0	4	0	[NKO]	Muller_B
FBgn0250532	1	0	0	0	0	0	0	0	2	0	Rnf146	Muller_D
FBgn0250538	0	0	0	0	0	0	0	0	7	0	[NKO]	Muller_D
FBgn0250593	0	0	0	1	0	0	27	18	13	1	[NKO]	Muller_C
FBgn0250613	0	0	0	0	0	0	1	1	0	0	SNCF	Muller_D
FBgn0250614	0	0	0	0	0	0	0	1	0	0	CG14111	Muller_D
FBgn0250621	0	0	0	0	0	0	5	10	0	0	[NKO]	Muller_D
FBgn0250622	0	0	0	0	0	0	6	2	0	0	[NKO]	Muller_D
FBgn0250623	0	0	0	0	0	0	7	9	0	0	[NKO]	Muller_B
FBgn0250659	1	2	0	1	0	1	1	0	1	0	[NKO]	Muller_C
FBgn0263811	0	0	1	0	0	0	0	0	21	0	ReepA	Muller_C
FBgn0270978	0	1	3	0	0	0	1	0	0	4	[NKO]	Muller_F
FBgn0270993	0	0	0	0	0	0	1	4	0	0	[NKO]	Muller_B
FBgn0271028	0	1	0	0	0	0	0	1	0	0	[NKO]	Muller_B
FBgn0271056	0	0	0	0	0	0	4	11	0	0	CG10862	Muller_A

FBgn0271083	0	0	0	2	1	0	2	1	0	0	[NKO]	Muller_C
FBgn0271110	5	0	0	0	0	0	1	1	1	0	[NKO]	Muller_F
FBgn0271168	0	2	0	0	0	0	25	30	1	0	[NKO]	Muller_E
FBgn0271195	0	0	0	0	0	0	13	18	11	0	[NKO]	Muller_D
FBgn0271257	0	1	2	0	0	0	1	0	25	0	[NKO]	Muller_A
FBgn0271286	0	0	0	0	1	0	67	75	0	1	[NKO]	Muller_E
FBgn0271308	6	1	0	0	0	0	2	2	2	3	[NKO]	Muller_A
FBgn0271309	0	1	0	0	0	0	0	0	4	0	[NKO]	Muller_B
FBgn0271341	0	1	0	0	0	0	2	4	0	0	[NKO]	Muller_B
FBgn0271369	3	0	0	0	0	0	9	10	16	1	[NKO]	Muller_A
FBgn0271444	0	0	4	0	2	0	0	0	4	0	[NKO]	Muller_E
FBgn0271464	0	0	0	0	0	0	13	22	0	0	[NKO]	Muller_A
FBgn0271484	0	0	0	0	1	0	1	1	0	0	[NKO]	Muller_E
FBgn0271490	0	0	0	0	0	1	1	1	2	0	[NKO]	Muller_B
FBgn0271573	2	0	0	0	0	1	29	22	7	2	[NKO]	Muller_D
FBgn0271637	3	0	0	0	3	0	2	1	9	4	[NKO]	Muller_A
FBgn0271877	1	0	0	0	0	0	12	14	13	1	Larp7	Muller_A
FBgn0271885	0	0	0	0	0	0	31	23	18	2	[NKO]	Muller_D
FBgn0271913	1	0	0	0	0	0	0	0	1	0	CG43088	Muller_C
FBgn0271989	0	0	0	0	0	0	11	10	0	0	[NKO]	Muller_A
FBgn0272067	0	0	0	0	0	0	10	7	2	0	[NKO]	Muller_F
FBgn0272141	0	0	0	0	0	0	0	1	0	0	[NKO]	Muller_E
FBgn0272235	0	0	0	0	0	0	2	2	0	1	[NKO]	Muller_F
FBgn0272289	0	1	0	0	0	0	45	58	0	0	[NKO]	Muller_A
FBgn0272315	0	0	0	0	0	0	8	22	0	0	[NKO]	Muller_E
FBgn0272346	2	1	0	0	0	2	2	1	0	3	[NKO]	Muller_F
FBgn0272390	0	0	0	2	6	0	0	0	92	0	[NKO]	Muller_A
FBgn0272392	2	0	0	0	4	0	18	18	3	5	[NKO]	Muller_A
FBgn0272409	0	0	0	1	0	0	8	8	3	0	CG12035	Muller_D
FBgn0272437	3	0	0	0	0	0	1	1	18	0	CG43088	Muller_B

FBgn0272513	0	0	0	0	0	0	7	12	0	1	[NKO]	Muller_A
FBgn0272517	1	0	0	0	3	0	18	21	13	3	[NKO]	Muller_D
FBgn0272555	1	0	0	0	0	0	28	26	17	2	[NKO]	Muller_D
FBgn0272614	0	0	0	0	0	0	2	2	3	0	[NKO]	Muller_A
FBgn0272618	0	0	0	0	0	0	9	9	0	0	[NKO]	Muller_B
FBgn0272638	0	0	0	0	0	0	6	7	4	0	[NKO]	Muller_D
FBgn0272645	0	0	0	0	0	0	29	29	18	1	[NKO]	Muller_D
FBgn0272696	0	0	0	0	0	0	3	2	10	0	CG14154	Muller_D
FBgn0272721	0	0	0	0	0	0	2	2	1	0	[NKO]	Muller_E
FBgn0272726	0	0	0	0	0	0	6	11	1	0	[NKO]	Muller_D
FBgn0272747	0	0	0	0	1	0	5	2	4	1	[NKO]	Muller_E
FBgn0272754	3	2	0	0	0	0	0	0	3	0	[NKO]	Muller_B
FBgn0272766	1	0	0	0	3	0	15	15	1	6	[NKO]	Muller_A
FBgn0272786	0	0	0	0	0	0	4	5	0	0	Oatp33Eb	Muller_B
FBgn0272906	4	0	0	0	0	0	0	0	1	0	[NKO]	Muller_D
FBgn0272938	0	0	0	0	0	0	3	10	0	0	[NKO]	Muller_E
FBgn0272980	1	0	2	0	0	0	1	0	0	0	[NKO]	Muller_A
FBgn0273013	0	0	0	0	0	0	24	29	13	0	[NKO]	Muller_D
FBgn0273019	0	0	0	0	0	0	2	1	0	0	[NKO]	Muller_E
FBgn0273123	0	0	0	0	0	0	8	5	0	0	[NKO]	Muller_B
FBgn0273155	0	0	0	0	1	0	3	1	2	0	[NKO]	Muller_A
FBgn0273163	0	0	0	0	0	0	12	19	12	0	[NKO]	Muller_B
FBgn0273188	0	0	0	0	0	0	21	25	24	0	S-Lap2	Muller_D
FBgn0273295	0	0	0	0	0	0	1	2	1	0	[NKO]	Muller_B
FBgn0273318	0	0	0	1	0	0	7	17	0	6	[NKO]	Muller_D
FBgn0273369	0	0	0	0	0	0	0	1	0	0	[NKO]	Muller_C

Table S3. Expression profiles of Y chromosome genes in different sexes and tissues.

Expression values are in log2. Note FBgn0249087 was abbreviated for CG45766, Mst77Y-4, Mst77Y-9, CG45765, Mst77F, CG45770, CG45764, Mst77Y-7, Mst77Y-13 and “dpse_pre_b” IDs refer to de novo CAT genes.

Gene ID	Homolog	Female Carcass	Female Larva	Ovary	Accessory Gland	Male Carcass	Male Larva	Testis	Male Body
LOC117183206	uncharacterized	3.71	0.00	0.94	2.67	1.16	4.10	4.00	4.78
LOC117183578	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117183579	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117183580	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117184044	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	4.86	1.72
LOC117184180	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117184227	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117184344	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117184435	uncharacterized	0.00	0.00	0.94	2.67	2.56	1.89	2.32	2.15
LOC117184436	uncharacterized	5.75	0.00	4.28	6.53	5.46	1.89	5.46	3.50
	NADH-ubiquinone oxidoreductase chain								
LOC117184536	5-like	8.41	0.00	3.98	4.40	9.43	1.89	2.00	2.48
LOC117184547	uncharacterized	0.00	0.00	0.00	0.00	0.00	1.23	0.00	1.10
LOC117184611	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.48
LOC117184615	uncharacterized	2.11	0.00	0.00	2.95	2.84	2.68	0.00	2.75
LOC117184670	uncharacterized	0.00	0.00	0.00	0.00	0.00	1.23	0.00	2.98
LOC117184724	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117184788	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.72
LOC117184826	uncharacterized	0.00	0.00	0.00	0.00	1.16	1.23	1.58	1.72
LOC117184829	uncharacterized	0.00	0.00	0.00	1.88	2.56	0.00	1.00	0.00
LOC117184897	uncharacterized	0.00	0.00	3.80	2.95	1.79	0.00	1.00	3.35
LOC117184942	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185022	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185049	uncharacterized	3.83	5.18	2.71	2.33	4.18	4.66	3.58	2.98
LOC117185107	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NADH-ubiquinone oxidoreductase chain								
LOC117185108		1.68	0.00	1.51	1.88	1.79	6.66	0.00	4.71

	2-like								
	NADH-ubiquinone oxidoreductase chain								
LOC117185114	5-like	5.75	0.00	1.51	4.40	7.67	5.17	2.32	4.44
LOC117185132	uncharacterized	0.00	0.00	0.00	6.35	2.56	4.31	3.91	7.34
LOC117185134	uncharacterized	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00
	neurofilament heavy polypeptide-like								
LOC117185415		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185426	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185431	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185433	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185437	uncharacterized	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185438	uncharacterized	1.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LOC117185494	uncharacterized	1.68	0.00	0.00	0.00	2.56	0.00	0.00	0.00
LOC117185664	uncharacterized	2.11	0.00	0.00	3.18	3.06	1.89	3.32	2.15
	UPF0518 protein								
LOC26532143	GA25918	6.60	5.36	7.88	5.17	7.44	3.72	6.87	2.75
LOC26533379	cytochrome b	2.93	0.00	0.00	0.00	4.53	6.51	1.00	4.09
	cytochrome c oxidase subunit 1								
LOC26533497		9.60	0.00	5.99	6.85	11.17	13.93	7.56	13.48
FBgn0249018	NKO	0.00	0.00	0.00	0.00	2.23	0.00	0.00	1.10
FBgn0244637	RpS19a	9.86	6.78	10.18	9.60	9.63	8.94	8.68	7.23
FBgn0271090	NKO	5.42	0.00	3.89	6.42	5.14	1.23	5.73	3.88
FBgn0272346	NKO	2.93	0.00	0.94	4.10	2.56	1.23	3.58	3.64
FBgn0078868	CG5421	0.00	0.00	0.00	4.21	2.23	4.73	6.11	4.71
FBgn0271856	NKO	2.43	0.00	1.51	0.00	1.79	3.18	1.58	3.77
FBgn0271490	NKO	1.07	0.00	0.94	2.95	1.79	1.23	4.00	4.51
FBgn0246875	CG6845	0.00	0.00	0.00	4.80	0.00	5.33	7.82	6.05
FBgn0249700	Src64B	6.27	5.15	10.03	9.78	6.67	6.71	8.26	4.44
FBgn0080373	HDAC1	5.99	7.25	9.06	7.24	5.61	7.71	9.34	7.83
FBgn0072804	ms(3)76Ca;CG42529	0.00	0.00	1.51	7.96	3.98	7.09	11.07	9.44
FBgn0079264	CG5969	1.07	0.00	0.00	5.22	1.79	5.78	7.43	5.95
FBgn0074640	CG17687	1.68	0.00	0.00	1.23	0.00	1.23	1.00	3.50

FBgn0072814	SNCF	0.00	0.00	0.00	1.23	0.00	0.00	0.00	0.00
FBgn0070574	ImpL1	0.00	1.38	0.00	0.00	2.56	4.21	0.00	0.00
FBgn0072812	CG14110	1.68	0.00	0.00	2.67	0.00	3.18	1.58	0.00
FBgn0070189	CG10171	3.59	5.36	4.59	4.21	4.08	6.04	3.46	2.98
FBgn0070203	Poc1	3.29	4.28	4.89	6.47	4.53	6.97	8.83	8.26
FBgn0076711	sens	0.00	0.00	0.00	0.00	0.00	1.23	2.32	1.72
FBgn0076674	CG32075	4.04	4.34	6.29	3.85	4.74	5.63	4.39	4.28
FBgn0071475	amx	3.94	2.89	1.51	6.38	6.31	7.11	5.55	7.49
FBgn0071420	Larp7	1.68	4.45	2.71	3.38	4.28	5.23	3.32	6.05
FBgn0071472	mxs	2.11	0.00	2.90	4.10	3.06	5.06	5.46	5.15
FBgn0075439	Ugalt	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FBgn0074797	regucalcin	3.94	6.25	0.00	3.56	4.45	7.55	3.91	4.44
FBgn0079573	Oatp33Eb	4.45	5.21	0.94	3.38	2.84	5.46	3.00	5.28
FBgn0077287	P5CDh2	0.00	0.00	0.00	6.30	2.84	6.01	9.22	8.68
FBgn0079769	CG6678	5.18	4.69	7.00	3.18	5.54	5.23	5.09	5.05
FBgn0073520	CG15093	7.92	1.38	9.01	7.69	7.92	0.00	7.09	2.98
FBgn0243917	NKO	10.87	8.56	11.64	10.24	9.72	9.15	9.86	6.84
FBgn0249108	Mst77Y-7;Mst77Y-13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FBgn0249087	Mst77Y-7;Mst77Y-13	0.00	0.00	0.00	1.88	0.00	0.00	0.00	3.50
FBgn0250305	NKO	0.00	0.00	0.00	2.95	0.00	1.89	3.17	3.18
FBgn0244045	NKO	5.38	2.07	7.98	8.31	6.65	5.00	10.33	4.58
FBgn0248542	NKO	0.00	0.00	0.00	6.00	0.00	5.98	8.96	7.28
FBgn0248191	CG10182;CG13325	5.94	1.38	2.23	0.00	4.81	0.00	2.00	3.50
FBgn0248189	CG16723;CG33337	0.00	0.00	0.00	0.00	0.00	2.95	0.00	2.15
FBgn0248937	RpS27	3.94	9.71	5.38	3.98	4.68	9.38	3.70	9.60
FBgn0246086	CG13527	0.00	0.00	0.00	5.06	1.79	4.66	8.02	6.08
FBgn0247139	NKO	0.00	0.00	0.00	7.47	2.56	7.41	10.21	8.80
FBgn0271775	CG43088	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FBgn0245944	NKO	0.00	0.00	0.00	0.00	0.00	2.34	4.17	2.15
FBgn0070269	CG10274	3.83	4.45	5.59	6.15	4.60	7.07	8.58	8.16
FBgn0079270	isoQC	5.38	5.91	6.34	6.33	5.42	4.50	4.52	4.44

FBgn0271860	NKO	0.00	0.00	0.00	0.00	0.00	3.56	2.58	3.35
FBgn0272223	NKO	0.00	0.00	3.59	0.00	1.79	0.00	1.00	2.98
FBgn0272620	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FBgn0245166	NKO	1.07	3.70	0.00	0.00	0.00	2.68	0.00	0.00
FBgn0245168	NKO	2.43	0.00	0.00	7.53	4.53	7.96	9.93	9.38
FBgn0072813	CG14111	0.00	2.07	0.94	0.00	0.00	1.89	0.00	0.00
FBgn0071408	CG12035	0.00	0.00	0.00	5.42	1.79	6.94	8.65	8.86
FBgn0249895	CG42540	6.92	3.41	3.98	1.88	7.06	2.68	4.17	4.36
FBgn0273368	NKO	4.99	0.00	3.07	4.80	4.28	1.89	5.32	2.15
FBgn0244954	JIL-1	5.22	6.72	7.82	5.94	4.18	8.67	9.63	8.52
FBgn0272217	NKO	0.00	0.00	0.00	3.56	2.84	0.00	0.00	0.00
FBgn0271083	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FBgn0271113	NKO	2.11	0.00	1.51	9.19	4.53	6.94	11.33	8.46
FBgn0271937	NKO	1.68	0.00	3.22	2.95	0.00	1.89	6.55	4.51
FBgn0272986	NKO	2.11	0.00	0.94	2.67	0.00	0.00	2.58	3.64
FBgn0271420	NKO	0.00	0.00	0.00	1.23	0.00	0.00	0.00	0.00
dpse_pre_b_G00 16413	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00 16422	NKO	0.00	0.00	0.00	3.71	0.00	4.31	6.07	5.05
dpse_pre_b_G00 16423	NKO	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
dpse_pre_b_G00 16424	NKO	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
dpse_pre_b_G00 16425	NKO	0.00	0.00	0.00	0.00	1.16	0.00	1.58	0.00
dpse_pre_b_G00 16426	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00 16427	NKO	0.00	0.00	0.00	1.88	0.00	3.72	1.58	2.98
dpse_pre_b_G00 16428	NKO	1.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00 16434	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00	NKO	0.00	0.00	0.00	2.33	0.00	0.00	1.58	0.00

16435									
dpse_pre_b_G00									
16436	NKO	0.00	0.00	0.00	3.56	0.00	5.23	6.74	6.15
dpse_pre_b_G00									
16437	NKO	0.00	0.00	0.00	1.23	0.00	0.00	3.70	0.00
dpse_pre_b_G00									
16438	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00									
16441	NKO	0.00	0.00	0.00	0.00	0.00	0.00	1.58	0.00
dpse_pre_b_G00									
16442	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dpse_pre_b_G00									
16443	NKO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

References

- Allshire, Robin C., and Hiten D. Madhani. 2018. "Ten Principles of Heterochromatin Formation and Function." *Nature Reviews Molecular Cell Biology* 19 (4): 229–44. <https://doi.org/10.1038/nrm.2017.119>.
- Altinordu, Fahim, Lorenzo Peruzzi, Yan Yu, and Xingjin He. 2016. "A Tool for the Analysis of Chromosomes: KaryoType." *Taxon* 65 (3): 586–92.
- Bachtrog, Doris. 2003. "Accumulation of Spock and Worf, Two Novel Non-LTR Retrotransposons, on the Neo-Y Chromosome of *Drosophila Miranda*." *Molecular Biology and Evolution* 20 (2): 173–81. <https://doi.org/10.1093/molbev/msg035>.
- . 2004. "Evidence That Positive Selection Drives Y-Chromosome Degeneration in *Drosophila Miranda*." *Nature Genetics* 36 (5): 518–22. <https://doi.org/10.1038/ng1347>.
- . 2006. "Expression Profile of a Degenerating Neo-y Chromosome in *Drosophila*." *Current Biology: CB* 16 (17): 1694–99. <https://doi.org/10.1016/j.cub.2006.07.053>.
- . 2013. "Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration." *Nature Reviews Genetics* 14 (2): 113–24. <https://doi.org/10.1038/nrg3366>.
- . 2020. "The Y Chromosome as a Battleground for Intragenomic Conflict." *Trends in Genetics* 36 (7): 510–22. <https://doi.org/10.1016/j.tig.2020.04.008>.
- Bachtrog, Doris, and Brian Charlesworth. 2002. "Reduced Adaptation of a Non-Recombining Neo-Y Chromosome." *Nature* 416 (6878): 323–26. <https://doi.org/10.1038/416323a>.
- Bachtrog, Doris, Emily Hom, Karen M Wong, Xulio Maside, and Pieter de Jong. 2008. "Genomic Degradation of a Young Y Chromosome in *Drosophila Miranda*." *Genome Biology* 9 (2): R30. <https://doi.org/10.1186/gb-2008-9-2-r30>.
- Bachtrog, Doris, Shivani Mahajan, and Ryan Bracewell. 2019. "Massive Gene Amplification on a Recently Formed *Drosophila* Y Chromosome." *Nature Ecology & Evolution* 3 (11): 1587–97. <https://doi.org/10.1038/s41559-019-1009-9>.
- Bachtrog, Doris, Judith E. Mank, Catherine L. Peichel, Mark Kirkpatrick, Sarah P. Otto, Tia-Lynn Ashman, Matthew W. Hahn, et al. 2014. "Sex Determination: Why So Many Ways of Doing It?" *PLOS Biology* 12 (7): e1001899. <https://doi.org/10.1371/journal.pbio.1001899>.
- Bellott, Daniel W., Jennifer F. Hughes, Helen Skaletsky, Laura G. Brown, Tatyana Pyntikova, Ting-Jan Cho, Natalia Koutseva, et al. 2014. "Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators." *Nature* 508 (7497): 494–99. <https://doi.org/10.1038/nature13206>.
- Bellott, Daniel W., Helen Skaletsky, Ting-Jan Cho, Laura Brown, Devin Locke, Nancy Chen, Svetlana Galkina, et al. 2017. "Avian W and Mammalian Y Chromosomes Convergently Retained Dosage-Sensitive Regulators." *Nature Genetics* 49 (3): 387–94. <https://doi.org/10.1038/ng.3778>.
- Bonhoure, Nicolas, Gergana Bounova, David Bernasconi, Viviane Praz, Fabienne Lammers, Donatella Canella, Ian M. Willis, et al. 2014. "Quantifying ChIP-Seq Data: A Spiking Method Providing an Internal Reference for Sample-to-Sample Normalization." *Genome Research* 24 (7): 1157–68. <https://doi.org/10.1101/gr.168260.113>.
- Bosco, Giovanni, Paula Campbell, Joao T. Leiva-Neto, and Therese A. Markow. 2007. "Analysis of *Drosophila* Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species." *Genetics* 177 (3): 1277–90. <https://doi.org/10.1534/genetics.107.075069>.

- Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, et al. 2018. "Ten Things You Should Know about Transposable Elements." *Genome Biology* 19 (1): 199. <https://doi.org/10.1186/s13059-018-1577-z>.
- Bousios, Alexandros, Hans-Wilhelm Nützmann, Dorothy Buck, and Davide Michieletto. 2020. "Integrating Transposable Elements in the 3D Genome." *Mobile DNA* 11 (1): 8. <https://doi.org/10.1186/s13100-020-0202-3>.
- Bracewell, Ryan, and Doris Bachtrog. 2020. "Complex Evolutionary History of the Y Chromosome in Flies of the *Drosophila Obscura* Species Group." *Genome Biology and Evolution* 12 (5): 494–505. <https://doi.org/10.1093/gbe/evaa051>.
- Bracewell, Ryan, Kamalakhar Chatla, Matthew J Nalley, and Doris Bachtrog. 2019. "Dynamic Turnover of Centromeres Drives Karyotype Evolution in *Drosophila*." Edited by Steven Henikoff, Diethard Tautz, and Steven Henikoff. *ELife* 8 (September): e49002. <https://doi.org/10.7554/eLife.49002>.
- Brenman-Suttner, Dova B., Ryley T. Yost, Ariel K. Frame, J. Wesley Robinson, Amanda J. Moehring, and Anne F. Simon. 2020. "Social Behavior and Aging: A Fly Model." *Genes, Brain, and Behavior* 19 (2): e12598. <https://doi.org/10.1111/gbb.12598>.
- Brind'Amour, Julie, Sheng Liu, Matthew Hudson, Carol Chen, Mohammad M. Karimi, and Matthew C. Lorincz. 2015. "An Ultra-Low-Input Native ChIP-Seq Protocol for Genome-Wide Profiling of Rare Cell Populations." *Nature Communications* 6 (1): 6033. <https://doi.org/10.1038/ncomms7033>.
- Brown, Emily J., and Doris Bachtrog. 2014. "The Chromatin Landscape of *Drosophila*: Comparisons between Species, Sexes, and Chromosomes." *Genome Research* 24 (7): 1125–37. <https://doi.org/10.1101/gr.172155.114>.
- Brown, Emily J., Alison H. Nguyen, and Doris Bachtrog. 2020a. "The Y Chromosome May Contribute to Sex-Specific Ageing in *Drosophila*." *Nature Ecology & Evolution* 4 (6): 853–62. <https://doi.org/10.1038/s41559-020-1179-5>.
- Brown, Emily J., Alison H. Nguyen, and Doris Bachtrog. 2020b. "The *Drosophila* Y Chromosome Affects Heterochromatin Integrity Genome-Wide." *Molecular Biology and Evolution* 37 (10): 2808–24. <https://doi.org/10.1093/molbev/msaa082>.
- Bull, J. J. 1983. "Evolution of Sex Determining Mechanisms." *Evolution of Sex Determining Mechanisms*. <https://www.cabdirect.org/cabdirect/abstract/19840180646>.
- Carvalho, Antonio Bernardo, and Andrew G. Clark. 2005. "Y Chromosome of *D. Pseudoobscura* Is Not Homologous to the Ancestral *Drosophila* Y." *Science (New York, N.Y.)* 307 (5706): 108–10. <https://doi.org/10.1126/science.1101675>.
- Carvalho, Antonio Bernardo, Bridget A. Dobo, Maria D. Vibranovski, and Andrew G. Clark. 2001. "Identification of Five New Genes on the Y Chromosome of *Drosophila Melanogaster*." *Proceedings of the National Academy of Sciences* 98 (23): 13225–30. <https://doi.org/10.1073/pnas.231484998>.
- Chakraborty, B. M., and R. Chakraborty. 1984. "Variations in the Length of Human Y Chromosome: A Statistical Study." *Acta Anthropogenetica* 8 (3–4): 269–75.
- Chakraborty, Mahul, James G. Baldwin-Brown, Anthony D. Long, and J. J. Emerson. 2016. "Contiguous and Accurate de Novo Assembly of Metazoan Genomes with Modest Long Read Coverage." *Nucleic Acids Research* 44 (19): e147. <https://doi.org/10.1093/nar/gkw654>.
- Chang, Ching-Ho, and Amanda M Larracuente. 2019. "Heterochromatin-Enriched Assemblies

- Reveal the Sequence and Organization of the *Drosophila Melanogaster* Y Chromosome.” *Genetics* 211 (1): 333–48. <https://doi.org/10.1534/genetics.118.301765>.
- Charlesworth, B. 1978. “Model for Evolution of Y Chromosomes and Dosage Compensation.” *Proceedings of the National Academy of Sciences of the United States of America* 75 (11): 5618–22. <https://doi.org/10.1073/pnas.75.11.5618>.
- Charlesworth, B., and D. Charlesworth. 2000. “The Degeneration of Y Chromosomes.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 355 (1403): 1563–72. <https://doi.org/10.1098/rstb.2000.0717>.
- Charlesworth, B., and C. H. Langley. 1986. “The Evolution of Self-Regulated Transposition of Transposable Elements.” *Genetics* 112 (2): 359–83.
- Charlesworth, Brian, Paul Sniegowski, and Wolfgang Stephan. 1994. “The Evolutionary Dynamics of Repetitive DNA in Eukaryotes.” *Nature* 371 (6494): 215–20. <https://doi.org/10.1038/371215a0>.
- Chen, Haiyang, Xiaobin Zheng, and Yixian Zheng. 2014. “Age-Associated Loss of Lamin-B Leads to Systemic Inflammation and Gut Hyperplasia.” *Cell* 159 (4): 829–43. <https://doi.org/10.1016/j.cell.2014.10.028>.
- Chin, Chen-Shan, Paul Peluso, Fritz J. Sedlazeck, Maria Nattestad, Gregory T. Concepcion, Alicia Clum, Christopher Dunn, et al. 2016. “Phased Diploid Genome Assembly with Single Molecule Real-Time Sequencing.” *Nature Methods* 13 (12): 1050–54. <https://doi.org/10.1038/nmeth.4035>.
- Chippindale, Adam K., and William R. Rice. 2001. “Y Chromosome Polymorphism Is a Strong Determinant of Male Fitness in *Drosophila Melanogaster*.” *Proceedings of the National Academy of Sciences* 98 (10): 5677–82. <https://doi.org/10.1073/pnas.101456898>.
- Clark, A. G. 1990. “Two Tests of Y Chromosomal Variation in Male Fertility of *Drosophila Melanogaster*.” *Genetics* 125 (3): 527–34.
- Cosby, Rachel L., Ni-Chen Chang, and Cédric Feschotte. 2019. “Host–Transposon Interactions: Conflict, Cooperation, and Cooption.” *Genes & Development* 33 (17–18): 1098–1116. <https://doi.org/10.1101/gad.327312.119>.
- Courret, Cécile, Ching-Ho Chang, Kevin H.-C. Wei, Catherine Montchamp-Moreau, and Amanda M. Larracunte. 2019. “Meiotic Drive Mechanisms: Lessons from *Drosophila*.” *Proceedings of the Royal Society B: Biological Sciences* 286 (1913): 20191430. <https://doi.org/10.1098/rspb.2019.1430>.
- De Cecco, Marco, Steven W. Criscione, Edward J. Peckham, Sara Hillenmeyer, Eliza A. Hamm, Jayameenakshi Manivannan, Abigail L. Peterson, Jill A. Kreiling, Nicola Neretti, and John M. Sedivy. 2013. “Genomes of Replicatively Senescent Cells Undergo Global Epigenetic Changes Leading to Gene Silencing and Activation of Transposable Elements.” *Aging Cell* 12 (2): 247–56. <https://doi.org/10.1111/accel.12047>.
- De Cecco, Marco, Takahiro Ito, Anna P. Petrashen, Amy E. Elias, Nicholas J. Skvir, Steven W. Criscione, Alberto Caligiana, et al. 2019. “L1 Drives IFN in Senescent Cells and Promotes Age-Associated Inflammation.” *Nature* 566 (7742): 73–78. <https://doi.org/10.1038/s41586-018-0784-9>.
- Dobzhansky, T. 1935. “The Y Chromosome of *Drosophila Pseudoobscura*.” *Genetics* 20 (4): 366–76.
- Dobzhansky, Th. 1933. “On the Sterility of the Interracial Hybrids in *Drosophila Pseudoobscura*.” *Proceedings of the National Academy of Sciences of the United States of America* 19 (4): 397–403.

- . 1937. “Further Data on the Variation of the Y Chromosome in *Drosophila Pseudoobscura*.” *Genetics* 22 (3): 340–46.
- Dobzhansky, Th., and R. D. Boche. 1933. “Intersterile Races of *Drosophila Pseudoobscura*.” *Biol. Zbl.* 53.
- Dupim, Eduardo G., Gabriel Goldstein, Thyago Vanderlinde, Suzana C. Vaz, Flávia Krsticevic, Aline Bastos, Thadeo Pinhão, et al. 2018. “An Investigation of Y Chromosome Incorporations in 400 Species of *Drosophila* and Related Genera.” *PLOS Genetics* 14 (11): e1007770. <https://doi.org/10.1371/journal.pgen.1007770>.
- Elgin, Sarah C. R., and Gunter Reuter. 2013. “Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in *Drosophila*.” *Cold Spring Harbor Perspectives in Biology* 5 (8): a017780. <https://doi.org/10.1101/cshperspect.a017780>.
- Ellis, Lisa L., Wen Huang, Andrew M. Quinn, Astha Ahuja, Ben Alfrejd, Francisco E. Gomez, Carl E. Hjelman, et al. 2014. “Intrapopulation Genome Size Variation in *D. Melanogaster* Reflects Life History Variation and Plasticity.” *PLOS Genetics* 10 (7): e1004522. <https://doi.org/10.1371/journal.pgen.1004522>.
- Ellison, Christopher, and Doris Bachtrog. 2019. “Recurrent Gene Co-Amplification on *Drosophila* X and Y Chromosomes.” *PLOS Genetics* 15 (7): e1008251. <https://doi.org/10.1371/journal.pgen.1008251>.
- Ellison, Christopher E., and Doris Bachtrog. 2013. “Dosage Compensation via Transposable Element Mediated Rewiring of a Regulatory Network.” *Science* 342 (6160): 846–50. <https://doi.org/10.1126/science.1239552>.
- Fanti, Laura, Barbara Perrini, Lucia Piacentini, Maria Berloco, Enzo Marchetti, Gioacchino Palumbo, and Sergio Pimpinelli. 2008. “The Trithorax Group and Pc Group Proteins Are Differentially Involved in Heterochromatin Formation in *Drosophila*.” *Chromosoma* 117 (1): 25–39. <https://doi.org/10.1007/s00412-007-0123-7>.
- Fiddes, Ian T., Joel Armstrong, Mark Diekhans, Stefanie Nachtweide, Zev N. Kronenberg, Jason G. Underwood, David Gordon, et al. 2018. “Comparative Annotation Toolkit (CAT)—Simultaneous Clade and Personal Genome Annotation.” *Genome Research* 28 (7): 1029–38. <https://doi.org/10.1101/gr.233460.117>.
- Francisco, Flávio O., and Bernardo Lemos. 2014. “How Do Y-Chromosomes Modulate Genome-Wide Epigenetic States: Genome Folding, Chromatin Sinks, and Gene Expression.” *Journal of Genomics* 2: 94–103. <https://doi.org/10.7150/jgen.8043>.
- Frost, Bess, Martin Hemberg, Jada Lewis, and Mel B. Feany. 2014. “Tau Promotes Neurodegeneration through Global Chromatin Relaxation.” *Nature Neuroscience* 17 (3): 357–66. <https://doi.org/10.1038/nn.3639>.
- Gatti, M., and S. Pimpinelli. 1992. “Functional Elements in *Drosophila Melanogaster* Heterochromatin.” *Annual Review of Genetics* 26: 239–75. <https://doi.org/10.1146/annurev.ge.26.120192.001323>.
- Goubert, Clément, Laurent Modolo, Cristina Vieira, Claire ValienteMoro, Patrick Mavingui, and Matthieu Boulesteix. 2015. “De Novo Assembly and Annotation of the Asian Tiger Mosquito (*Aedes Albopictus*) Repeatome with DnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (*Aedes Aegypti*).” *Genome Biology and Evolution* 7 (4): 1192–1205. <https://doi.org/10.1093/gbe/evv050>.
- Griffin, R. M., D. Le Gall, H. Schielzeth, and U. Friberg. 2015. “Within-Population Y-Linked Genetic Variation for Lifespan in *Drosophila Melanogaster*.” *Journal of Evolutionary Biology* 28 (11): 1940–47. <https://doi.org/10.1111/jeb.12708>.

- Haithcock, Erin, Yaron Dayani, Ester Neufeld, Adam J. Zahand, Naomi Feinstein, Anna Mattout, Yosef Gruenbaum, and Jun Liu. 2005. "Age-Related Changes of Nuclear Architecture in *Caenorhabditis Elegans*." *Proceedings of the National Academy of Sciences* 102 (46): 16690–95. <https://doi.org/10.1073/pnas.0506955102>.
- Helleu, Quentin, Pierre R. Gérard, and Catherine Montchamp-Moreau. 2014. "Sex Chromosome Drive." *Cold Spring Harbor Perspectives in Biology* 7 (2): a017616. <https://doi.org/10.1101/cshperspect.a017616>.
- Hill, Tom, and Andrea J. Betancourt. 2018. "Extensive Exchange of Transposable Elements in the *Drosophila Pseudoobscura* Group." *Mobile DNA* 9 (1): 20. <https://doi.org/10.1186/s13100-018-0123-6>.
- Hill, W. G., and Alan Robertson. 1966. "The Effect of Linkage on Limits to Artificial Selection." *Genetics Research* 8 (3): 269–94. <https://doi.org/10.1017/S0016672300010156>.
- Hoskins, Roger A., Joseph W. Carlson, Cameron Kennedy, David Acevedo, Martha Evans-Holm, Erwin Frise, Kenneth H. Wan, et al. 2007. "Sequence Finishing and Mapping of *Drosophila Melanogaster* Heterochromatin." *Science (New York, N.Y.)* 316 (5831): 1625–28. <https://doi.org/10.1126/science.1139816>.
- Hoskins, Roger A., Joseph W. Carlson, Kenneth H. Wan, Soo Park, Ivonne Mendez, Samuel E. Galle, Benjamin W. Booth, et al. 2015. "The Release 6 Reference Sequence of the *Drosophila Melanogaster* Genome." *Genome Research* 25 (3): 445–58. <https://doi.org/10.1101/gr.185579.114>.
- Huang, Weichun, Leping Li, Jason R. Myers, and Gabor T. Marth. 2012. "ART: A next-Generation Sequencing Read Simulator." *Bioinformatics (Oxford, England)* 28 (4): 593–94. <https://doi.org/10.1093/bioinformatics/btr708>.
- Hurst, Gregory D. D., and John H. Werren. 2001. "The Role of Selfish Genetic Elements in Eukaryotic Evolution." *Nature Reviews Genetics* 2 (8): 597–606. <https://doi.org/10.1038/35084545>.
- HURST, L. D. 1992. "Is Stellate a Relict Meiotic Driver?" *Genetics* 130 (1): 229–30.
- Jin, Ying, Oliver H. Tam, Eric Paniagua, and Molly Hammell. 2015. "TEtranscripts: A Package for Including Transposable Elements in Differential Expression Analysis of RNA-Seq Datasets." *Bioinformatics (Oxford, England)* 31 (22): 3593–99. <https://doi.org/10.1093/bioinformatics/btv422>.
- Johnston, J., Molly Schoener, and Dino McMahon. 2013. "DNA Underreplication in the Majority of Nuclei in the *Drosophila Melanogaster* Thorax: Evidence from Suur and Flow Cytometry." *Journal of Molecular Biology Research* 3 (1): p47. <https://doi.org/10.5539/jmbr.v3n1p47>.
- Kaplan, E. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53 (282): 457–81. <https://doi.org/10.2307/2281868>.
- Kapusta, Aurélie, Alexander Suh, and Cédric Feschotte. 2017. "Dynamics of Genome Size Evolution in Birds and Mammals." *Proceedings of the National Academy of Sciences* 114 (8): E1460–69. <https://doi.org/10.1073/pnas.1616702114>.
- Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. "HISAT: A Fast Spliced Aligner with Low Memory Requirements." *Nature Methods* 12 (4): 357–60. <https://doi.org/10.1038/nmeth.3317>.
- Kim, Daehwan, Joseph M. Paggi, Chanhee Park, Christopher Bennett, and Steven L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping with HISAT2 and

- HISAT-Genotype.” *Nature Biotechnology* 37 (8): 907–15.
<https://doi.org/10.1038/s41587-019-0201-4>.
- Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.
<https://doi.org/10.1101/gr.215087.116>.
- Krupovic, Mart, and Eugene V. Koonin. 2015. “Polintons: A Hotbed of Eukaryotic Virus, Transposon and Plasmid Evolution.” *Nature Reviews Microbiology* 13 (2): 105–15.
<https://doi.org/10.1038/nrmicro3389>.
- Kuhn, G. C. S. 2015. “Satellite DNA Transcripts Have Diverse Biological Roles in *Drosophila*.” *Heredity* 115 (1): 1–2. <https://doi.org/10.1038/hdy.2015.12>.
- Lahn, B. T., and D. C. Page. 1997. “Functional Coherence of the Human Y Chromosome.” *Science (New York, N.Y.)* 278 (5338): 675–80.
<https://doi.org/10.1126/science.278.5338.675>.
- Lancefield, D. E. 1929. “A Genetic Study of Crosses of Two Races or Physiological Species Of *Drosophila Obscura*.” *Zeitschrift Für Induktive Abstammungs- Und Vererbungslehre* 52 (1): 287–317. <https://doi.org/10.1007/BF01847272>.
- Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Larracuente, Amanda M, and Andrew G Clark. 2013. “Surprising Differences in the Variability of Y Chromosomes in African and Cosmopolitan Populations of *Drosophila Melanogaster*.” *Genetics* 193 (1): 201–14. <https://doi.org/10.1534/genetics.112.146167>.
- Larracuente, Amanda M., Mohamed A. F. Noor, and Andrew G. Clark. 2010. “Translocation of Y-Linked Genes to the Dot Chromosome in *Drosophila Pseudoobscura*.” *Molecular Biology and Evolution* 27 (7): 1612–20. <https://doi.org/10.1093/molbev/msq045>.
- Larson, Kimberly, Shian-Jang Yan, Amy Tsurumi, Jacqueline Liu, Jun Zhou, Kriti Gaur, Dongdong Guo, Thomas H. Eickbush, and Willis X. Li. 2012. “Heterochromatin Formation Promotes Longevity and Represses Ribosomal RNA Synthesis.” *PLOS Genetics* 8 (1): e1002473. <https://doi.org/10.1371/journal.pgen.1002473>.
- Lemos, Bernardo, Luciana O. Araripe, and Daniel L. Hartl. 2008. “Polymorphic Y Chromosomes Harbor Cryptic Variation with Manifest Functional Consequences.” *Science (New York, N.Y.)* 319 (5859): 91–93. <https://doi.org/10.1126/science.1148861>.
- Lemos, Bernardo, Alan T. Branco, and Daniel L. Hartl. 2010. “Epigenetic Effects of Polymorphic Y Chromosomes Modulate Chromatin Components, Immune Response, and Sexual Conflict.” *Proceedings of the National Academy of Sciences* 107 (36): 15826–31. <https://doi.org/10.1073/pnas.1010383107>.
- Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.
<https://doi.org/10.1093/bioinformatics/btp324>.
- Li, Wanhe, Lisa Prazak, Nabanita Chatterjee, Servan Grüniger, Lisa Krug, Delphine Theodorou, and Josh Dubnau. 2013. “Activation of Transposable Elements during Aging and Neuronal Decline in *Drosophila*.” *Nature Neuroscience* 16 (5): 529–31.
<https://doi.org/10.1038/nn.3368>.
- Li, Xiao-Yong, Melissa M Harrison, Jacqueline E Villalta, Tommy Kaplan, and Michael B Eisen. 2014. “Establishment of Regions of Genomic Activity during the *Drosophila* Maternal to Zygotic Transition.” Edited by Robb Krumlauf. *ELife* 3 (October): e03737.

- <https://doi.org/10.7554/eLife.03737>.
- Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
- Lin, Ching-Jung, Fuqu Hu, Raphaele Dubruille, Jeffrey Vedanayagam, Jiayu Wen, Peter Smibert, Benjamin Loppin, and Eric C. Lai. 2018. “The HpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the *Drosophila* Male Germline.” *Developmental Cell* 46 (3): 316–326.e5. <https://doi.org/10.1016/j.devcel.2018.07.004>.
- Lindholm, Anna K., Kelly A. Dyer, Renée C. Firman, Lila Fishman, Wolfgang Forstmeier, Luke Holman, Hanna Johannesson, et al. 2016. “The Ecology and Evolutionary Dynamics of Meiotic Drive.” *Trends in Ecology & Evolution* 31 (4): 315–26. <https://doi.org/10.1016/j.tree.2016.02.001>.
- Linford, Nancy J., Ceyda Bilgir, Jennifer Ro, and Scott D. Pletcher. 2013. “Measurement of Lifespan in *Drosophila Melanogaster*.” *Journal of Visualized Experiments: JoVE*, no. 71 (January). <https://doi.org/10.3791/50068>.
- Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Lyckegaard, E M, and A G Clark. 1989. “Ribosomal DNA and Stellate Gene Copy Number Variation on the Y Chromosome of *Drosophila Melanogaster*.” *Proceedings of the National Academy of Sciences of the United States of America* 86 (6): 1944–48.
- Mahajan, Shivani, and Doris Bachtrog. 2017. “Convergent Evolution of Y Chromosome Gene Content in Flies.” *Nature Communications* 8 (1): 785. <https://doi.org/10.1038/s41467-017-00653-x>.
- Mahajan, Shivani, Kevin H.-C. Wei, Matthew J. Nalley, Lauren Gibilisco, and Doris Bachtrog. 2018. “De Novo Assembly of a Young *Drosophila* Y Chromosome Using Single-Molecule Sequencing and Chromatin Conformation Capture.” *PLoS Biology* 16 (7). <https://doi.org/10.1371/journal.pbio.2006348>.
- Makałowski, Wojciech, Valer Gotea, Amit Pande, and Izabela Makałowska. 2019. “Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics.” *Methods in Molecular Biology (Clifton, N.J.)* 1910: 177–207. https://doi.org/10.1007/978-1-4939-9074-0_6.
- Maklakov, Alexei A., and Virpi Lummaa. 2013. “Evolution of Sex Differences in Lifespan and Aging: Causes and Constraints.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (8): 717–24. <https://doi.org/10.1002/bies.201300021>.
- Marais, Gabriel A.B., Jean-Michel Gaillard, Cristina Vieira, Ingrid Plotton, Damien Sanlaville, François Gueyffier, and Jean-Francois Lemaitre. 2018. “Sex Gap in Aging and Longevity: Can Sex Chromosomes Play a Role?” *Biology of Sex Differences* 9 (July). <https://doi.org/10.1186/s13293-018-0181-y>.
- McGurk, Michael P., and Daniel A. Barbash. 2018. “Double Insertion of Transposable Elements Provides a Substrate for the Evolution of Satellite DNA.” *Genome Research* 28 (5): 714–25. <https://doi.org/10.1101/gr.231472.117>.
- McKee, B. D., L. Habera, and J. A. Vrana. 1992. “Evidence That Intergenic Spacer Repeats of *Drosophila Melanogaster* RRNA Genes Function as X-Y Pairing Sites in Male Meiosis, and a General Model for Achiasmatic Pairing.” *Genetics* 132 (2): 529–44.
- Mills, Wilbur Kyle, Yuh Chwen G Lee, Antje M Kochendoerfer, Elaine M Dunleavy, and Gary

- H Karpen. 2019. "RNA from a Simple-Tandem Repeat Is Required for Sperm Maturation and Male Fertility in *Drosophila Melanogaster*." Edited by Howard Y Chang, James L Manley, and Yukiko M Yamashita. *ELife* 8 (November): e48940. <https://doi.org/10.7554/eLife.48940>.
- Molaro, Antoine, and Harmit S. Malik. 2016. "Hide and Seek: How Chromatin-Based Pathways Silence Retroelements in the Mammalian Germline." *Current Opinion in Genetics & Development* 37 (April): 51–58. <https://doi.org/10.1016/j.gde.2015.12.001>.
- Montchamp-Moreau, C. 2006. "Sex-Ratio Meiotic Drive in *Drosophila Simulans*: Cellular Mechanism, Candidate Genes and Evolution." *Biochemical Society Transactions* 34 (Pt 4): 562–65. <https://doi.org/10.1042/BST0340562>.
- Nguyen, Alison H., and Doris Bachtrog. 2021. "Toxic Y Chromosome: Increased Repeat Expression and Age-Associated Heterochromatin Loss in Male *Drosophila* with a Young Y Chromosome." *PLOS Genetics* 17 (4): e1009438. <https://doi.org/10.1371/journal.pgen.1009438>.
- Nguyen, Alison H., Weixiang Wang, Emily Chong, Kamalakar Chatla, and Doris Bachtrog. 2021. "Transposable Element Accumulation Drives Size Differences among Polymorphic Y Chromosomes in *Drosophila*." *BioRxiv*, June, 2021.06.02.446622. <https://doi.org/10.1101/2021.06.02.446622>.
- O'Sullivan, Roderick J., and Jan Karlseder. 2012. "The Great Unravelling: Chromatin as a Modulator of the Aging Process." *Trends in Biochemical Sciences* 37 (11): 466–76. <https://doi.org/10.1016/j.tibs.2012.08.001>.
- Peichel, Catherine L., Shaugnessy R. McCann, Joseph A. Ross, Alice F. S. Naftaly, James R. Urton, Jennifer N. Cech, Jane Grimwood, et al. 2020. "Assembly of the Threespine Stickleback Y Chromosome Reveals Convergent Signatures of Sex Chromosome Evolution." *Genome Biology* 21 (1): 177. <https://doi.org/10.1186/s13059-020-02097-x>.
- Pimpinelli, S, M Berloco, L Fanti, P Dimitri, S Bonaccorsi, E Marchetti, R Caizzi, C Caggese, and M Gatti. 1995. "Transposable Elements Are Stable Structural Components of *Drosophila Melanogaster* Heterochromatin." *Proceedings of the National Academy of Sciences of the United States of America* 92 (9): 3804–8.
- Pipoly, Ivett, Veronika Bókony, Mark Kirkpatrick, Paul F. Donald, Tamás Székely, and András Liker. 2015. "The Genetic Sex-Determination System Predicts Adult Sex Ratios in Tetrapods." *Nature* 527 (7576): 91–94. <https://doi.org/10.1038/nature15380>.
- Pritham, Ellen J., and Cédric Feschotte. 2007. "Massive Amplification of Rolling-Circle Transposons in the Lineage of the Bat *Myotis Lucifugus*." *Proceedings of the National Academy of Sciences* 104 (6): 1895–1900. <https://doi.org/10.1073/pnas.0609601104>.
- Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics (Oxford, England)* 26 (6): 841–42. <https://doi.org/10.1093/bioinformatics/btq033>.
- Rambaut, Andrew. 2007. *FigTree*. <https://tree.bio.ed.ac.uk/software/figtree/>.
- Rice, William R. 1984. "Sex Chromosomes and the Evolution of Sexual Dimorphism." *Evolution* 38 (4): 735–42. <https://doi.org/10.2307/2408385>.
- Riddle, Nicole C., Christopher D. Shaffer, and Sarah C. R. Elgin. 2009. "A Lot about a Little Dot – Lessons Learned from *Drosophila Melanogaster* Chromosome Four." *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire* 87 (1): 229–41. <https://doi.org/10.1139/o08-119>.
- Rohmer, Céline, Jean R. David, Brigitte Moreteau, and Dominique Joly. 2004. "Heat Induced

- Male Sterility in *Drosophila Melanogaster*: Adaptive Genetic Variations among Geographic Populations and Role of the Y Chromosome.” *The Journal of Experimental Biology* 207 (Pt 16): 2735–43. <https://doi.org/10.1242/jeb.01087>.
- Saksouk, Nehmé, Elisabeth Simboeck, and Jérôme Déjardin. 2015. “Constitutive Heterochromatin Formation and Transcription in Mammals.” *Epigenetics & Chromatin* 8 (1): 3. <https://doi.org/10.1186/1756-8935-8-3>.
- Skaletsky, Helen, Tomoko Kuroda-Kawaguchi, Patrick J. Minx, Holland S. Cordum, LaDeana Hillier, Laura G. Brown, Sjoerd Repping, et al. 2003. “The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes.” *Nature* 423 (6942): 825–37. <https://doi.org/10.1038/nature01722>.
- Steinemann, M., and S. Steinemann. 1992. “Degenerating Y Chromosome of *Drosophila* Miranda: A Trap for Retrotransposons.” *Proceedings of the National Academy of Sciences of the United States of America* 89 (16): 7591–95. <https://doi.org/10.1073/pnas.89.16.7591>.
- Sun, Luyang, Ruofan Yu, and Weiwei Dang. 2018. “Chromatin Architectural Changes during Cellular Senescence and Aging.” *Genes* 9 (4). <https://doi.org/10.3390/genes9040211>.
- Tang, Xiaona, Jinguo Cao, Liang Zhang, Yingzi Huang, Qianyi Zhang, and Yikang S. Rong. 2017. “Maternal Haploid, a Metalloprotease Enriched at the Largest Satellite Repeat and Essential for Genome Integrity in *Drosophila* Embryos.” *Genetics* 206 (4): 1829–39. <https://doi.org/10.1534/genetics.117.200949>.
- Tarka, Maja, Anja Guenther, Petri T. Niemelä, Shinichi Nakagawa, and Daniel W.A. Noble. 2018. “Sex Differences in Life History, Behavior, and Physiology along a Slow-Fast Continuum: A Meta-Analysis.” *Behavioral Ecology and Sociobiology* 72 (8): 132. <https://doi.org/10.1007/s00265-018-2534-2>.
- Taylor, Jesse E., and John Jaenike. 2003. “Sperm Competition and the Dynamics of X Chromosome Drive in Finite and Structured Populations.” *Annales Zoologici Fennici* 40 (2): 195–206.
- Topp, Christopher N., Cathy X. Zhong, and R. Kelly Dawe. 2004. “Centromere-Encoded RNAs Are Integral Components of the Maize Kinetochore.” *Proceedings of the National Academy of Sciences* 101 (45): 15986–91. <https://doi.org/10.1073/pnas.0407154101>.
- Tower, John, and Michelle Arbeitman. 2009. “The Genetics of Gender and Life Span.” *Journal of Biology* 8 (4): 38. <https://doi.org/10.1186/jbiol141>.
- Usakin, Lev, José Abad, Vasily V. Vagin, Beatriz de Pablos, Alfredo Villasante, and Vladimir A. Gvozdev. 2007. “Transcription of the 1.688 Satellite DNA Family Is Under the Control of RNA Interference Machinery in *Drosophila Melanogaster* Ovaries.” *Genetics* 176 (2): 1343–49. <https://doi.org/10.1534/genetics.107.071720>.
- Van Meter, Michael, Mehr Kashyap, Sarallah Rezazadeh, Anthony J. Geneva, Timothy D. Morello, Andrei Seluanov, and Vera Gorbunova. 2014. “SIRT6 Represses LINE1 Retrotransposons by Ribosylating KAP1 but This Repression Fails with Stress and Age.” *Nature Communications* 5 (September): 5011. <https://doi.org/10.1038/ncomms6011>.
- Waterhouse, Robert M., Fredrik Tegenfeldt, Jia Li, Evgeny M. Zdobnov, and Evgenia V. Kriventseva. 2013. “OrthoDB: A Hierarchical Catalog of Animal, Fungal and Bacterial Orthologs.” *Nucleic Acids Research* 41 (Database issue): D358–65. <https://doi.org/10.1093/nar/gks1116>.
- Wei, Kevin H -C, Sarah E Lower, Ian V Caldas, Trevor J S Sless, Daniel A Barbash, and Andrew G Clark. 2018. “Variable Rates of Simple Satellite Gains across the *Drosophila*

- Phylogeny.” *Molecular Biology and Evolution* 35 (4): 925–41.
<https://doi.org/10.1093/molbev/msy005>.
- Wei, Kevin H.-C., and Doris Bachtrog. 2019. “Ancestral Male Recombination in *Drosophila Albomicans* Produced Geographically Restricted Neo-Y Chromosome Haplotypes Varying in Age and Onset of Decay.” *PLOS Genetics* 15 (11): e1008502.
<https://doi.org/10.1371/journal.pgen.1008502>.
- Wei, Kevin H.-C., Lauren Gibilisco, and Doris Bachtrog. 2020. “Epigenetic Conflict on a Degenerating Y Chromosome Increases Mutational Burden in *Drosophila* Males.” *Nature Communications* 11 (1): 5537. <https://doi.org/10.1038/s41467-020-19134-9>.
- Werren, John H. 2011. “Selfish Genetic Elements, Genetic Conflict, and Evolutionary Innovation.” *Proceedings of the National Academy of Sciences* 108 (Supplement 2): 10863–70. <https://doi.org/10.1073/pnas.1102343108>.
- Wood, Jason G., Sara Hillenmeyer, Charles Lawrence, Chengyi Chang, Suzanne Hosier, Will Lightfoot, Eric Mukherjee, et al. 2010. “Chromatin Remodeling in the Aging Genome of *Drosophila*.” *Aging Cell* 9 (6): 971–78.
<https://doi.org/10.1111/j.1474-9726.2010.00624.x>.
- Wood, Jason G., Brian C. Jones, Nan Jiang, Chengyi Chang, Suzanne Hosier, Priyan Wickremesinghe, Meyrolin Garcia, et al. 2016. “Chromatin-Modifying Genetic Interventions Suppress Age-Associated Transposable Element Activation and Extend Life Span in *Drosophila*.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (40): 11277–82. <https://doi.org/10.1073/pnas.1604621113>.
- Xirocostas, Zoe A., Susan E. Everingham, and Angela T. Moles. 2020. “The Sex with the Reduced Sex Chromosome Dies Earlier: A Comparison across the Tree of Life.” *Biology Letters* 16 (3): 20190867. <https://doi.org/10.1098/rsbl.2019.0867>.
- Xue, Lingzhan, Yu Gao, Meiyang Wu, Tian Tian, Haiping Fan, Yongji Huang, Zhen Huang, Dapeng Li, and Luohao Xu. 2021. “Telomere-to-Telomere Assembly of a Fish Y Chromosome Reveals the Origin of a Young Sex Chromosome Pair.” *Genome Biology* 22 (1): 203. <https://doi.org/10.1186/s13059-021-02430-y>.
- Yasuhara, Jiro C., Christine H. DeCrease, and Barbara T. Wakimoto. 2005. “Evolution of Heterochromatic Genes of *Drosophila*.” *Proceedings of the National Academy of Sciences* 102 (31): 10958–63. <https://doi.org/10.1073/pnas.0503424102>.
- Yasuhara, Jiro C., and Barbara T. Wakimoto. 2008. “Molecular Landscape of Modified Histones in *Drosophila* Heterochromatic Genes and Euchromatin-Heterochromatin Transition Zones.” *PLOS Genetics* 4 (1): e16. <https://doi.org/10.1371/journal.pgen.0040016>.
- Yoon, Jong Sik, Kathleen Pausic Gagen, and Ding L. Zhu. 1990. “Longevity of 68 Species of *Drosophila*,” March. <https://kb.osu.edu/handle/1811/23379>.
- Yu, Guangchuang. 2020. “Using Ggtree to Visualize Data on Tree-Like Structures.” *Current Protocols in Bioinformatics* 69 (1): e96. <https://doi.org/10.1002/cpbi.96>.
- Zhou, Qi, and Doris Bachtrog. 2012. “Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in *Drosophila*.” *Science* 337 (6092): 341–45.
<https://doi.org/10.1126/science.1225385>.
- Zhou, Qi, Christopher E. Ellison, Vera B. Kaiser, Artyom A. Alekseyenko, Andrey A. Gorchakov, and Doris Bachtrog. 2013. “The Epigenome of Evolving *Drosophila* Neo-Sex Chromosomes: Dosage Compensation and Heterochromatin Formation.” *PLoS Biology* 11 (11). <https://doi.org/10.1371/journal.pbio.1001711>.
- Zurovcova, M., and W. F. Eanes. 1999. “Lack of Nucleotide Polymorphism in the Y-Linked

Sperm Flagellar Dynein Gene Dhc-Yh3 of *Drosophila Melanogaster* and *D. Simulans*.”
Genetics 153 (4): 1709–15.