Massive gene amplification on a recently formed Drosophila Y chromosome

Shivani Mahajan & Doris Bachtrog

Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA

email: dbachtrog@berkeley.edu

Widespread loss of genes on the Y is considered a hallmark of sex chromosome differentiation. Here we show that the initial stages of Y evolution are driven by massive amplification of distinct classes of genes. The neo-Y chromosome of *Drosophila miranda* initially contained about 2700 protein-coding genes, but has gained over 3200 genes since its formation about 1.5 MY ago, primarily by amplification of proteins ancestrally present on this chromosome. We show that distinct evolutionary processes account for this drastic increase in gene number on the Y. Testis-specific and dosage sensitive genes are amplified on the Y to increase male fitness. A distinct class of meiosis-related multi-copy Y genes is co-amplified on the X, and their expansion is likely driven by conflicts over segregation. Co-amplified X/Y genes are highly expressed in testis, enriched for meiosis and RNAi functions, and are frequently targeted by short RNAs in testis. This suggests that their amplification is driven by X vs. Y antagonism for increased transmission, where sex chromosome drive suppression is likely mediated by sequence homology between the suppressor and distorter, through RNAi mechanism. Thus, newly emerged sex chromosomes are a battleground for sexual and meiotic conflict.

Introduction

Sex chromosomes have originated multiple times from ordinary autosomes (Bachtrog *et al.* 2011). After suppression of recombination, sex chromosomes evolve independently and differentiate (Charlesworth 1978). The complete lack of recombination on Y chromosomes renders natural selection inefficient, and Y evolution is characterized by a loss of the majority of its ancestral genes (Bachtrog 2013). Indeed, old Y chromosomes of various species contain only few functional genes, and Y chromosomes instead have accumulated massive amounts of repetitive DNA (Bellott *et al.* 2017; Mahajan and Bachtrog 2017).

Studies of Y chromosomes are greatly hindered by a lack of high-quality reference sequences, and the Y chromosomes of only a handful of mammalian species have been fully sequenced (Hughes *et al.* 2005; Soh *et al.* 2014). To date, no high-quality sequences of young Y chromosomes have been examined.

Drosophila miranda has a pair of recently formed neo-sex chromosomes that originated ~1.5MY ago, and has served as a model to study the initiation of sex chromosome differentiation (Bachtrog and Charlesworth 2002). The neo-sex chromosomes of *D. miranda* are still homologous over much of their length, and a previous genomic analyses using Illumina short reads confirmed the notion that genes on the Y are rapidly lost (Zhou and Bachtrog 2012). About ½ of the roughly 2700 genes ancestrally present on the neo-Y were found to be pseudogenized, and 100s of genes were entirely missing (Zhou and Bachtrog 2012). However, the high level of sequence similarity between the X and Y, yet drastic accumulation of repeats, prevented assembling the Y chromosome.

We recently generated a high-quality sequence assembly of the neo-Y chromosome of *D. miranda* using single-molecule sequencing and chromatin conformation capture (Mahajan *et al.* 2018). Instead of simply shrinking, our assembly revealed that the young neo-Y chromosome dramatically increased in size relative to the neo-X by roughly 3-fold: we assembled 110.5 Mb of the fused ancestral Y and the neo-Y chromosome (Y/neo-Y sequence), and 25.3 Mb of the neo-X (Mahajan *et al.* 2018). Most of this size increase is driven by massive accumulation of repetitive sequences, and over 50% of the neo-Y derived sequence is composed of transposable elements (Mahajan *et al.* 2018). Here, we carefully annotate the neo-sex chromosomes, using transcriptomes from multiple tissues, and short RNA profiles, to study the evolution of gene content on this recently formed neo-Y chromosome.

A catalogue of genes and transcription units on the neo-Y

With a comprehensive high-quality reference sequence of the neo-Y chromosome of *D. miranda*, we systematically catalogue its genes. Automatic annotation of the highly repetitive Y/neo-Y chromosome is challenging, and we used extensive manual curation to validate and correct our gene models (see Methods). Comparison of the neo-sex gene content with that of *D. pseudoobscura*, a close relative where this chromosome pair is autosomal, allows us to infer the ancestral gene complement and reconstruct the evolutionary history of gene gain and loss along the neo-sex chromosomes (**Figure 1A**). We bioinformatically identified and manually examined all Y/neo-Y genes that were not simple 1:1 orthologs between species and chromosomes. In total, we identified 6,448 genes on the neo-Y, and 3,253 genes on the neo-X, compared to 3,087 genes on the ancestral autosome that gave rise to the neo-sex chromosome. Thus, contrary to the paradigm that Y chromosomes undergo genome-wide degeneration, our analysis reveals a dramatic increase in the number of annotated genes on the neo-Y, compared to its ancestral gene complement.

Overall, we find 1,736 ancestral single-copy orthologs between the neo-sex chromosomes, i.e. roughly 56% of genes are conserved. Furthermore, we find 143 genes that are present in *D. pseudoobscura* and the neo-X of *D. miranda*, but are missing from our neo-Y annotation, and we fail to detect a homolog on the neo-Y by BLAST, i.e. 5% of genes ancestrally present are completely absent on the neo-Y. On the other hand, only 17 genes (0.5% of genes ancestrally present) are absent from the neo-X, a rate of gene loss comparable to autosomes and the ancestral X (**Table S1**). Thus, the neo-Y is indeed losing its ancestral genes at a high rate, consistent with theoretical expectation (Charlesworth and Charlesworth 2000; Bachtrog 2013) and empirical observations of gene poor ancestral Y chromosomes (Hughes *et al.* 2005; Soh *et al.* 2014; Bellott *et al.* 2017; Mahajan and Bachtrog 2017).

Intriguingly, however, for 457 unique *D. pseudoobscura* genes, we find multiple copies in our neo-Y chromosome annotation; 1,697 genes from 363 distinct proteins are single-copy (or missing) on the neo-X and amplified on the neo-Y (two genes of which were gained from Muller E; **Table S2**), and 2,036 Y genes from 94 distinct proteins that co-amplify on the X/neo-X (and harbor a total of 647 copies on the X/neo-X; **Table S3, Figure S1**). Thus, genes on the Y/neo-Y fall into distinct categories (**Figure 1B**), and we refer to them as single-copy ancestral Y genes, multi-copy Y genes (which are single-copy on the X/neo-X), and co-amplified X/Y genes. Genes whose ancestral location could not be determined, or genes with more complex evolutionary histories were not further analyzed (see Methods).

Many amplified gene copies are fragmented, and some have premature stop codons or frame shift mutations (**Table S2, S3**). We find full-length copies for 786 amplified Y genes (46%), 776 co-amplified Y-genes (38%), and 300 of co-amplified X genes (46%). Thus, even if ignoring partial gene copies (which may nevertheless have function as non-coding transcripts, see below), we still find considerably more genes on the neo-Y compared to the neo-X or the ancestral autosome.

Transcriptome analysis shows that most individual gene copies of amplified gene families on the Y/neo-Y are expressed, both for partial genes and full-length transcripts. We detect expression of 71% of individual copies among multi-copy Y genes, and 94% for co-amplified X/Y genes, which suggests that many gene copies may indeed be functional (**Table S4**).

Different evolutionary processes drive amplification of Y genes

What may drive massive gene amplification on the neo-Y chromosome? Y chromosomes are subject to unique evolutionary forces, and fundamentally different processes appear to cause gene family expansion of multi-copy Y genes versus co-amplified X/Y genes (**Figure 1C**). The high repeat content

on the Y makes this chromosome particularly vulnerable to accumulate multi-copy genes for various reasons. Repetitive sequences provide a substrate for non-allelic homologous recombination and thereby promote gene family expansion (Konkel and Batzer 2010). Indeed, we find several cases where repeats appear to have contributed to gene duplications, on both the X and Y chromosome (**Figure S2**). Spreading of heterochromatin from repeats globally dampens expression of neo-Y genes (Zhou *et al.* 2013), and multi-copy gene families may simply be more tolerated on the neo-Y (although many individual gene copies are transcribed, **Table S4**) .

Gene family expansions on the Y can also be beneficial for males. Global transcription is lower from the neo-Y chromosome, and drives the evolution of dosage compensation of homologous neo-X genes (Ellison and Bachtrog 2013). Gene amplification may help compensate for reduced gene dose of neo-Y genes, especially if their neo-X homologs are not yet dosage compensated. Additionally, Y chromosomes are transmitted from father to son, and are thus an ideal location for genes that specifically enhance male fitness (Rice 1984). Y chromosomes of several species, including humans, have been shown to contain multi-copy gene families that are expressed in testis and contribute to male fertility (Skaletsky *et al.* 2003; Cortez *et al.* 2014; Bellott *et al.* 2014).

We find a handful of multi-copy Y gene families that have dozens of gene copies on the Y (six gene families have more than 30 copies, and 14 gene families have >15 copies; **Figure 2B, Figure S3**), while the vast majority of multi-copy Y gene families only have a few copies (90% of multi-copy Y gene families have fewer than 4 copies). Gene expression and chromatin analysis suggest that different evolutionary forces drive the accumulation of low versus high-copy number multi-copy Y gene families.

Multi-copy Y gene families with a high copy number (i.e. >15 copies) are expressed almost exclusively in testis (**Figure 2A, Figure S3**), mimicking patterns of gene family amplification of male fertility genes found in other species (Skaletsky *et al.* 2003; Bellott *et al.* 2010; Cortez *et al.* 2014). Their neo-X homologs, in contrast, are expressed predominantly in ovaries (**Figure S3**), and sex-linkage may have enabled neo-Y and neo-X gametologs to specialize in their male- and female specific function, respectively (Rice 1984). Most multi-copy Y gene families, in contrast, only have few copies and are ubiquitously expressed (**Figure 2A**). Consistent with gene dosage contributing to increased copy number on the Y, we find that the neo-X homologs of multi-copy Y genes are less likely to be dosage compensated compared to single-copy Y genes (**Figure 2C**, p-value Fisher's exact test = 0.007). In particular, male Drosophila achieve dosage compensation by recruiting the MSL-complex to their hemizygous X chromosome (Lucchesi and Kuroda 2015), and we find that neo-X homologs of multi-copy neo-Y genes are less likely to be targeted by the MSL complex than the neo-X homologs of

single-copy neo-Y genes (**Figure 2C**). This suggests that multi-copy Y genes are dosage sensitive, and additional gene copies on the Y contribute to dosage compensation.

X/Y co-amplified genes reveal ongoing conflict over sex chromosome transmission

Genes that co-amplify on the X and Y chromosome, on the other hand, show testis-biased expression, independent of copy number (**Figure 2D**). Functional enrichment patterns and short RNA expression profiles suggest that fundamentally different forces drive their co-amplification on the X and Y chromosome (**Figure 3**). In particular, Y chromosomes compete with the X over transmission to the next generation (Frank 1991; Jaenike 2003), and sex chromosomes may try to cheat fair meiosis to bias their representation in functional sperm (i.e. meiotic drive). Meiotic drive on sex chromosomes, however, reduces fertility and distorts population sex ratios (Jaenike 2003), and creates strong selective pressure to evolve suppressors to silence selfish drivers. The RNAi pathway has been implicated to mediate suppression of sex chromosome drive in Drosophila (Tao *et al.* 2007a; b; Lin *et al.* 2018).

Overall, we identify 2683 co-amplified genes on the neo-sex chromosomes of *D. miranda* from 94 distinct proteins (2036 genes amplified on the Y/neo-Y, and 647 on the X/neo-X). Co-amplified X and Y-linked gene copies are typically both highly expressed in testis (**Figure 2D, Figure 3D**; **Figure S3, S4**). Testis expression of co-amplified X-linked genes is unusual, as testis-genes in *Drosophila* normally avoid the X chromosome (Sturgill *et al.* 2007), but can be understood under intragenomic conflict models (Tao *et al.* 2007a; b; Mueller *et al.* 2008; Meiklejohn and Tao 2010; Mueller *et al.* 2013). In particular, an X-linked gene involved in chromosome segregation may evolve a duplicate that acquires the ability to incapacitate Y-bearing sperm (**Figure 1C**). Invasion of this sex-ratio distorter skews the population sex ratio and creates a selective advantage to evolve a Y-linked suppressor that is resistant to the distorter. Suppression may be achieved at the molecular level by increased copy number of the wildtype function or by inactivation of X-linked drivers using RNAi (Tao *et al.* 2007a; b; Lin *et al.* 2018). If both driver and suppressor are dosage sensitive, they would undergo iterated cycles of expansion, resulting in rapid co-amplification of both driver and suppressor on the X and Y chromosome (Jaenike 2003).

Consistent with ongoing conflicts over chromosome segregation driving co-amplification of X/Y genes, we find that many of the most highly co-amplified genes have well-characterized functions in meiosis (**Figure 2E**, **Table S3**), and are ancestrally expressed in gonads (**Figure 3C**, **Figure S5**). Gene ontology (GO) analysis reveals that co-amplified X/Y genes are significantly overrepresented in biological processes associated with meiosis and chromosome segregation (**Figure 3A, 3B**). In particular, multi-copy Y genes are enriched for GO categories including “nuclear division”, “spindle

assembly", "meiotic spindle midzone assembly", "DNA packaging", "chromosome segregation", or "male gamete generation" (see **Table S3**). Among the most highly co-amplified X/Y genes are *wurstfest* (145 copies on the Y and 5 on the X), a gene involved in spindle assembly in male meiosis I; *mars* (48 Y-linked copies and 6 X-linked copies), a gene involved in kinetochore assembly and chromosome segregation, *orientation disruptor* (18 Y-linked copies and 5 X-linked copies), a chromosome-localized protein required for meiotic sister chromatid cohesion, or *Subito* (8 Y-linked copies and 11 X-linked copies), a gene required for spindle organization and chromosome segregation in meiosis (**Figure 2E**, **Figure 4**, see **Table S3** for additional genes). These important meiosis genes are typically single-copy and highly conserved across insects, but highly co-amplified on the recently evolved *D. miranda* X and Y chromosome.

Additionally, our GO analysis reveals an overrepresentation of co-amplified X/Y genes associated with piRNA metabolism and the generation of small RNA's (**Figure 3A, 3B**). Again, this is expected under recurring sex chromosome drive where silencing of distorters is achieved by RNAi, since compromising the short RNA pathway would release previously silenced drive systems (Lin *et al.* 2018). Noteworthy genes in the RNAi pathway that are typically single-copy but co-amplified on the X and Y include *Dicer-2* (26 Y- and 6 X-linked copies), a double-stranded RNA-specific endonuclease that cuts long double-stranded RNA into siRNAs, *cutoff* (7 Y- and 9 X-linked copies), a gene involved in transcription of piRNA clusters, or *shutdown* (50 Y- and 22 X-linked copies), a co-chaperone necessary for piRNA biogenesis (**Figure 2E**, **Figure 4**, **Table S3**). Thus, functional enrichment strongly supports meiotic conflict driving co-amplification of X/Y genes.

We gathered stranded RNA-seq and small RNA profiles from wildtype *D. miranda* testis, to obtain insights into the molecular mechanism of sex chromosome drive. Consistent with meiotic drive and suppression through RNAi mechanisms causing co-amplification of X/Y genes, we detect both sense and antisense transcripts and short RNA's derived from the vast majority of co-amplified X/Y genes (**Figure 3D**, **Figure 4**). Globally, we find that co-amplified Y genes show significantly higher levels of endo-siRNA production than single-copy Y genes, or multi-copy Y genes (**Figure 4A;** p-value < 10^{-16}, **Table S5**). Likewise, siRNA levels are higher for co-amplified X linked genes, compared to single-copy X genes, or X homologs of multi-copy Y genes (**Figure 4A;** p-value < 10^{-16}, **Table S5**). Targeting of co-amplified X/Y genes by short RNA's in testis demonstrates that siRNA production is not simply a consequence of the repeat-rich environment of the neo-Y but instead a property of co-amplified X/Y genes. This is consistent with cryptic sex chromosome drive having repeatedly led to characteristic patterns of gene amplification of homologous genes on both the X and the Y chromosomes that are targeted by short RNAs.

Conclusions

Contrary to the paradigm that Y chromosomes undergo global degeneration, we document a high rate of gene gain on the recently formed neo-Y chromosome of *D. miranda*, mainly through amplification of protein-coding genes that were ancestrally present on the autosome that created the neo-Y. Our comparative genomic analysis reveals different types of amplified Y genes: multi-copy genes exclusive to the Y, and genes that are co-amplified on the X and Y, and we show that their acquisition likely is driven by different selective pressures. Multi-copy Y genes presumably increase male fitness, and come in two flavors; they are either selected and amplifying on the Y because of their testis-specific function, and these male-beneficial genes often have dozens of copies on the Y. Their neo-X homologs are often expressed in ovaries, and sex linkage may have allowed these former homologs to specialize in their sex-specific roles (Rice 1984). Ubiquitously expressed housekeeping genes also duplicate on the Y, presumably to mitigate gene dose deficiencies of partially silenced neo-Y genes; these genes are present at a much lower copy number, and are targeted less often by the dosage compensation complex on the X.

Co-amplified X/Y genes are highly expressed in testis and often have functions in chromosome segregation and RNAi, and we speculate that their parallel amplification on the X and Y is a result of ongoing X-Y interchromosomal conflicts over segregation. Sequence homology between putative drivers and their suppressors on the sex chromosomes, and their widespread targeting by endo-siRNAs suggests that RNAi mechanisms are involved in silencing rampant sex chromosome drive. While we cannot determine the sequence of evolutionary events with certainty, the X chromosome is *a priori* more likely to acquire segregation distorters, creating strong selection to evolve suppressors on the Y. On one hand, natural selection is impaired on the non-recombining Y (Bachtrog 2013), making drivers more likely to originate on the X. Additionally, the heterochromatic nature of a Y chromosome may render it especially vulnerable to be exploited by selfish elements during meiosis (Helleu *et al.* 2016). If amplified Y genes are involved in a battle with the X over fair transmission, changes in gene copy number may tip the balance over inclusion into functional sperm, and drive repeated co-amplification of distorters and suppressors on the sex chromosomes.

Rampant sex chromosome drive can have important evolutionary consequences. Strong selective pressure to amplify Y-linked suppressors of meiotic drive may indirectly account for the ongoing degeneration of most ancestral genes on the Y chromosome, and ultimately its complete decay. Specifically, since the Y chromosome lacks recombination, strong positive selection for meiotic drive suppressors can propel linked deleterious mutations to fixation(Charlesworth and Charlesworth 2000). Thus, the decay of ancestral Y genes may be a by-product of silencing recurrent meiotic drivers arising on the X. Patterns of molecular variation are suggestive of episodes of recurrent positive selection shaping neo-Y evolution of *D. miranda* (Bachtrog 2004), and natural lines of *D.*

miranda show a wide range of sex-ratio bias (with typically female-biased sex ratios (Dobzhansky 1935)). These observations are consistent with recurrent and ongoing conflicts over segregation affecting the genomic architecture of sex chromosomes in this species.

Genetic conflict between X-Y ampliconic genes may also contribute to hybrid sterility and consequent reproductive isolation (Frank 1991; Hurst and Pomiankowski 1991). Segregation distortion can result in male hybrid sterility in Drosophila (Phadnis and Orr 2009), and further functional characterization of co-amplified, lineage-specific X-Y gene families will be needed to test the proposed link between X-Y genetic conflict and hybrid sterility.

X-Y interchromosomal conflict, and its consequent impact on gene amplification on sex chromosomes, may be widespread. In both human and mouse—two species with high-quality reference sequences for both sex chromosomes—the X and Y have co-acquired and amplified genes, and in both cases, meiotic drive has been invoked to explain this co-amplification (Lahn and Page 2000; Cocquet *et al.* 2009; 2012; Soh *et al.* 2014). Co-amplified genes have also been found in *D. melanogaster* (Balakireva *et al.* 1992), and RNAi mechanisms have been shown to mediate suppression of sex ratio drive in flies (Tao *et al.* 2007a; b; Lin *et al.* 2018). Highly amplified gene families have been detected in other mammals (Murphy *et al.* 2006) and across fruit flies (Ellison *et al.*), suggesting that sex chromosome drive may be prevalent in evolution; to determine the true phylogenetic range of lineage-specific acquisition and amplification of X-Y genes, high-quality sex chromosome assemblies across more taxa are needed.

Materials and methods

De novo transcriptome assembly

To mask repeats in the genome, we used RepeatMasker (Smith *et al.*) with custom de novo repeat libraries, generated using RepeatModeler (Smith and Hubley) and Repdenovo (Chu *et al.* 2016), along with the Drosophila repeat library from Repbase (Bao *et al.* 2015). Paired end RNA-seq reads from several male and female tissues (heads, carcass, whole body, testis, ovary, accessory gland, spermatheca, 3rd instar larvae) were then aligned to the repeat-masked genome using HiSat2 (Kim *et al.* 2015) with the --dta parameter on default settings. The resulting alignment file was used to assemble the transcriptome using the software StringTie (Pertea *et al.* 2015) with default parameters. Fasta sequences of the transcripts were extracted from the gtf output produced by StringTie using the gffread utility.

Gene annotation with Maker

We ran Maker (Campbell *et al.* 2014) three times to iteratively build and improve the gene annotation of the neo-sex chromosomes. For the first Maker run, we used annotated protein sequences for *D. melanogaster* and *D. pseudoobscura* from flybase.org, our de novo assembled *D. miranda* transcripts from the previous section, and the gene predictors Augustus (Stanke and Waack 2003) and SNAP (Korf 2004) to get the initial set of predictions. The parameters est2genome and protein2genome were set to 1 to allow Maker to directly build gene models from the transcript and protein alignments, and we used the Augustus fly gene model and the SNAP *D. melanogaster* hmm file for this first run.

The predictions from the first round were then used to train Augustus using BUSCO (Simão *et al.* 2015) and also to train SNAP. The new Augustus gene model and SNAP hmm file were then used during the second Maker run, with the parameters est2genome and protein2genome set to 0. The maximum intron size was increased to 20000bp (default 10000bp). The results from the second round were then used to train Augustus and SNAP again, before the final round of Maker. This process resulted in a total of 21,524 annotated genes in *D. miranda*.

Orthology detection

Transcript sequences for *D. pseudoobscura* were downloaded from flybase.org and only the largest transcript per gene was retained for downstream analyses. De novo annotated *D. miranda* transcripts were then aligned to this filtered *D. pseudoobscura* transcript set using BLAST (Camacho *et al.* 2009). Alignments with percentage identity <60% were discarded and the best alignment was calculated based on the e-value, score, % identity and alignment lengths. Each *D. miranda* transcript was thus assigned the ortholog that was its best BLAST hit. We identified paralogous genes in the *D. pseudoobscura* genome as those for which at least 80% of the sequence of one aligned to the other and vice versa. Paralogous genes in the *D. miranda – D. pseudoobscura* orthology calls were replaced by a single gene name from the duplicated gene family.

Identifying multicopy genes

The gene annotation produced by Maker had roughly 2,500 more genes annotated on the Y/neo-Y compared to the neo-X, and hundreds of genes had multiple annotated copies on the Y/neo-Y chromosome (and also the X chromosome and autosomes in some cases). Based on the orthology calls from BLAST, 822 Maker annotated genes had more than 2 copies on the Y/neo-Y, and 209 of those genes had more than two copies on both the X/neo-X and Y/neo-Y. In our initial Maker annotation, 366 genes were missing on the neo-Y and 155 genes were missing on the neo-X. However, closer inspection revealed that the annotation was often fragmented, especially on the Y/neo-Y chromosome, which led to an overestimation of the number of distinct genes that had duplicated, but subsequent BLAST searches also revealed that Maker often failed to annotate individual copies of gene families. On the other hand, several genes in the annotation were

"chimeras", where two genes were collapsed into one by Maker and thus one of the genes appeared to be missing from the gene annotation, if it got assigned to the other *D. pseudoobscura* gene during orthology assignment. Thus, the actual number of missing genes is much smaller than our initial Maker annotation suggested. We thus manually verified, and if necessary fixed, each gene model that was annotated by Maker, and inferred to be either duplicated on the Y/neo-Y or missing from the neo-X and/or Y/neo-Y annotations. We used nucmer (Kurtz *et al.* 2004) from the mummer package to individually align (one gene at a time) the sequences of their corresponding *D. pseudoobscura* orthologs to the *D. miranda* genome with the parameters --maxmatch and --nosimplify. Alignment coordinates were manually stitched together to get full gene coordinates. Only fragments that were at least 25% the length of the corresponding *D. pseudoobscura* ortholog were counted as duplicates/paralogs in the *D. miranda* genome. We also performed BLAST searches to identify the genes that had been lost from the neo-sex chromosomes.

In total, we annotate 6,448 genes on the neo-Y. Of these, 1,736 are ancestral single-copy Y genes (i.e. they were present on the ancestral autosome that formed the neo-Y). 1,105 of these genes were readily identified on both the neo-X and neo-Y by our Maker annotation, and are used as the single-copy orthology gene set in our analysis. 631 ancestral single-copy Y genes were initially missed or mis-qualified by our Maker annotation (i.e. 347 neo-Y genes were wrongly annotated as multi-copy by Maker, but our manual inspection revealed that they were present only as single-copy genes, and 114 neo-X genes and 170 neo-Y genes were missing from the Maker annotation, but found to be present on the neo-X and neo-Y, respectively, after manual checking using nucmer (Kurtz *et al.* 2004)).

In addition, we identify 457 genes (with 3,733 gene copies) that have become amplified on the neo-Y: 1,697 multi-copy Y genes (with a single copy on the neo-X), and 2,036 co-amplified neo-Y genes (which also amplified on the X/neo-X). In addition, we detect 959 genes in our Maker annotation that were not further considered. These "other" genes are comprised of 159 neo-Y genes that lack a homolog in *D. pseudoobcura*, 287 neo-Y genes that are present on an unknown location in *D. pseudoobscura*, 189 single-copy neo-Y genes that are present at multiple other locations in the genome (based on the Maker annotation), and 324 genes (from 49 unique proteins) with complicated mapping which could not be included in any categories of our analysis (i.e. genes for which the number of copies were ambiguous based on alignments such as for nested/overlapping genes; genes for which many alignments of variable identity were observed; genes which were amplifying on autosomes and the Y chromosome; chimeric genes).

Thus, after manual verification, we identified 94 genes that have co-amplified on the X/neo-X and the Y/neo-Y, with 647 copies on the X/neo-X and 2036 copies on the Y/neo-Y (and 58 copies on the

autosomes). We also identified 363 genes that have only amplified on the Y/neo-Y chromosome, with a total of 1697 copies on the Y/neo-Y. Thus, the Y/neo-Y chromosome has gained at least 3200 gene copies.

We identified 17 genes that are present on chr3 in *D. pseudoobscura* but are missing on the neo-X in *D. miranda*; 6 of those genes are found on other chromosomes in the *D. miranda* genome and 6 are still present on the Y chromosome in *D. miranda*. We identified 143 genes that are present on chr3 in *D. pseudoobscura* but absent on the neo-Y in *D. miranda* and 138 of those are still present on the neo-X in *D. miranda*. However 5 genes have been lost from both the neo-X and neo-Y chromosomes and BLAST searches failed to identify other chromosomal locations that those genes could have moved to.

Gene expression analysis

Kallisto (Bray *et al.* 2016) was used to quantify gene expression and calculate TPM values for each gene in our annotation using several male and female tissues (whole body, carcass, 3rd instar larvae, gonads, spermatheca and accessory gland). The R function heatmap.2 from the gplots package (https://cran.r-project.org/web/packages/gplots/index.html) was used to plot heatmaps to visualize tissue-specific differences in gene expression. Each row in the heatmap is a different gene and the different columns represent different tissues. The heatmap was row-normalized to indicate the tissue with the highest expression for each gene.

GO analysis

GO analysis was done using GOrilla (Eden *et al.* 2009) and the *D. melanogaster* orthologs of the genes co-amplifying on the X and Y was used as the target set. The GO terms that were enriched and had a p-value less than 10^{-3} and fdr less than 0.05 were visualized using the software Revigo (Supek *et al.* 2011).

Analysis of smRNA and totalRNA data

Stranded total RNA paired end reads were mapped to the *D. miranda* genome using HiSat2 with default parameters and the --rna-strandness parameter set to RF. The sam file was then filtered to obtain sense and antisense transcription estimates. Single end small RNA reads were aligned to the genome using bowtie2. BamCoverage from the deeptools package (Ramírez *et al.* 2014) was used to convert bam alignment files to bigwig format in both cases to be visualized using IGV. The number of small RNA reads mapping to each annotated gene was calculated using bedtools (Quinlan and Hall 2010) and these counts were divided by the gene/fragment length and boxplots were plotted in R for single copy Y genes, genes that have only amplified on the Y and genes that have co-amplified on the X and Y.

References:

Bachtrog D., 2004 Evidence that positive selection drives Y-chromosome degeneration in Drosophila miranda. Nat. Genet. **36**: 518–522.

Bachtrog D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. **14**: 113–124.

Bachtrog D., Charlesworth B., 2002 Reduced adaptation of a non-recombining neo-Y chromosome. Nature **416**: 323–326.

Bachtrog D., Kirkpatrick M., Mank J. E., McDaniel S. F., Pires J. C., Rice W., Valenzuela N., 2011 Are all sex chromosomes created equal? Trends Genet. **27**: 350–357.

Balakireva M. D., Shevelyov YuYa, Nurminsky D. I., Livak K. J., Gvozdev V. A., 1992 Structural organization and diversification of Y-linked sequences comprising Su(Ste) genes in Drosophila melanogaster. Nucleic Acids Res. **20**: 3731–3736.

Bao W., Kojima K. K., Kohany O., 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA **6**: 11.

Bellott D. W., Hughes J. F., Skaletsky H., Brown L. G., Pyntikova T., Cho T.-J., Koutseva N., Zaghlul S., Graves T., Rock S., Kremitzki C., Fulton R. S., Dugan S., Ding Y., Morton D., Khan Z., Lewis L., Buhay C., Wang Q., Watt J., Holder M., Lee S., Nazareth L., Rozen S., Muzny D. M., Warren W. C., Gibbs R. A., Wilson R. K., Page D. C., 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature **508**: 494–499.

Bellott D. W., Skaletsky H., Cho T.-J., Brown L., Locke D., Chen N., Galkina S., Pyntikova T., Koutseva N., Graves T., Kremitzki C., Warren W. C., Clark A. G., Gaginskaya E., Wilson R. K., Page D. C., 2017 Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. Nat. Genet. **49**: 387–394.

Bellott D. W., Skaletsky H., Pyntikova T., Mardis E. R., Graves T., Kremitzki C., Brown L. G., Rozen S., Warren W. C., Wilson R. K., Page D. C., 2010 Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature **466**: 612–616.

Bray N. L., Pimentel H., Melsted P., Pachter L., 2016 Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. **34**: 525–527.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L., 2009 BLAST+: architecture and applications. BMC Bioinformatics **10**: 421.

Campbell M. S., Holt C., Moore B., Yandell M., 2014 Genome Annotation and Curation Using MAKER and MAKER-P. Curr Protoc Bioinformatics **48**: 4.11.1–39.

Charlesworth B., 1978 Model for evolution of Y chromosomes and dosage compensation. Proc. Natl. Acad. Sci. U.S.A. **75**: 5618–5622.

Charlesworth B., Charlesworth D., 2000 The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond., B, Biol. Sci. **355**: 1563–1572.

Chu C., Nielsen R., Wu Y., 2016 REPdenovo: Inferring De Novo Repeat Motifs from Short Sequence Reads. (C Antoniewski, Ed.). PLoS ONE **11**: e0150719.

Cocquet J., Ellis P. J. I., Mahadevaiah S. K., Affara N. A., Vaiman D., Burgoyne P. S., 2012 A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. (MW Nachman, Ed.). PLoS Genet. **8**: e1002900.

Cocquet J., Ellis P. J. I., Yamauchi Y., Mahadevaiah S. K., Affara N. A., Ward M. A., Burgoyne P. S., 2009 The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. (N Hastie, Ed.). PLoS Biol. **7**: e1000244.

Cortez D., Marin R., Toledo-Flores D., Froidevaux L., Liechti A., Waters P. D., Grützner F., Kaessmann H., 2014 Origins and functional evolution of Y chromosomes across mammals. Nature **508**: 488–493.

Dobzhansky T., 1935 Drosophila Miranda, a New Species. Genetics **20**: 377–391.

Eden E., Navon R., Steinfeld I., Lipson D., Yakhini Z., 2009 GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics **10**: 48.

Ellison C. E., Bachtrog D., 2013 Dosage compensation via transposable element mediated rewiring of a regulatory network. Science **342**: 846–850.

Ellison C. E., Leonard C., Landeen E. L., Giblisco L., Phadnis N., Bachtrog D., Cryptic sex chromosome drive in Drosopila.

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

Helleu Q., Gérard P. R., Dubruille R., Ogereau D., Prud'homme B., Loppin B., Montchamp-Moreau C., 2016 Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. Proc. Natl. Acad. Sci. U.S.A. **113**: 4110–4115.

Hughes J. F., Skaletsky H., Pyntikova T., Minx P. J., Graves T., Rozen S., Wilson R. K., Page D. C., 2005 Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature **437**: 100–103.

Hurst L. D., Pomiankowski A., 1991 Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. Genetics **128**: 841–858.

Jaenike J., 2003 Sex Chromosome Meiotic Drive. http://dx.doi.org/10.1146/annurev.ecolsys.32.081501.113958 **32**: 25–49.

Kim D., Langmead B., Salzberg S. L., 2015 HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**: 357–360.

Konkel M. K., Batzer M. A., 2010 A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin. Cancer Biol. **20**: 211–221.

Korf I., 2004 Gene finding in novel genomes. BMC Bioinformatics **5**: 59.

Kurtz S., Phillippy A., Delcher A. L., Smoot M., Shumway M., Antonescu C., Salzberg S. L., 2004 Versatile and open software for comparing large genomes. Genome Biol. **5**: R12.

Lahn B. T., Page D. C., 2000 A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. Hum. Mol. Genet. **9**: 311–319.

Lin C.-J., Hu F., Dubruille R., Vedanayagam J., Wen J., Smibert P., Loppin B., Lai E. C., 2018 The hpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the Drosophila Male Germline. Dev. Cell **46**: 316–326.e5.

Lucchesi J. C., Kuroda M. I., 2015 Dosage compensation in Drosophila. Cold Spring Harb Perspect Biol **7**: a019398.

Mahajan S., Bachtrog D., 2017 Convergent evolution of Y chromosome gene content in flies. Nat Commun **8**: 785.

Mahajan S., Wei K. H.-C., Nalley M. J., Gibilisco L., Bachtrog D., 2018 De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. (C Tyler-Smith, Ed.). PLoS Biol. **16**: e2006348.

Meiklejohn C. D., Tao Y., 2010 Genetic conflict and sex chromosome evolution. Trends Ecol. Evol. (Amst.) **25**: 215–223.

Mueller J. L., Mahadevaiah S. K., Park P. J., Warburton P. E., Page D. C., Turner J. M. A., 2008 The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat. Genet. **40**: 794–799.

Mueller J. L., Skaletsky H., Brown L. G., Zaghlul S., Rock S., Graves T., Auger K., Warren W. C., Wilson R. K., Page D. C., 2013 Independent specialization of the human and mouse X chromosomes for the male germ line. Nat. Genet. **45**: 1083–1087.

Murphy W. J., Pearks Wilkerson A. J., Raudsepp T., Agarwala R., Schäffer A. A., Stanyon R., Chowdhary B. P., 2006 Novel gene acquisition on carnivore Y chromosomes. PLoS Genet. **2**: e43.

Pertea M., Pertea G. M., Antonescu C. M., Chang T.-C., Mendell J. T., Salzberg S. L., 2015 StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**: 290–295.

Phadnis N., Orr H. A., 2009 A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science **323**: 376–379.

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

Ramírez F., Dündar F., Diehl S., Grüning B. A., Manke T., 2014 deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. **42**: W187–91.

Rice W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution **38**: 735–742.

Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V., Zdobnov E. M., 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210–3212.

Skaletsky H., Kuroda-Kawaguchi T., Minx P. J., Cordum H. S., Hillier L., Brown L. G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S.-F., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R. H., Wilson R. K., Rozen S., Page D. C., 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature **423**: 825–837.

Smith A., Hubley R., *RepeatModeler Open-1.0*. httpwww.repeatmasker.org.

Smith A., Hubley R., Green P., *RepeatMasker Open-4.0*. httpwww.repeatmasker.org.

Soh Y. Q. S., Alföldi J., Pyntikova T., Brown L. G., Graves T., Minx P. J., Fulton R. S., Kremitzki C., Koutseva N., Mueller J. L., Rozen S., Hughes J. F., Owens E., Womack J. E., Murphy W. J., Cao Q., de Jong P., Warren W. C., Wilson R. K., Skaletsky H., Page D. C., 2014 Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell **159**: 800–813.

Stanke M., Waack S., 2003 Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19 Suppl 2**: ii215–25.

Sturgill D., Zhang Y., Parisi M., Oliver B., 2007 Demasculinization of X chromosomes in the Drosophila genus. Nature **450**: 238–241.

Supek F., Bošnjak M., Škunca N., Šmuc T., 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. (C Gibas, Ed.). PLoS ONE **6**: e21800.

Tao Y., Araripe L., Kingan S. B., Ke Y., Xiao H., Hartl D. L., 2007a A sex-ratio Meiotic Drive System in Drosophila simulans. II: An X-linked Distorter (D Barbash, Ed.). PLoS Biol. **5**: e293.

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

Zhou Q., Bachtrog D., 2012 Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science **337**: 341–345.

Zhou Q., Ellison C. E., Kaiser V. B., Alekseyenko A. A., Gorchakov A. A., Bachtrog D., 2013 The epigenome of evolving Drosophila neo-sex chromosomes: dosage compensation and heterochromatin formation. (PB Becker, Ed.). PLoS Biol. **11**: e1001711.

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

Author contributions. SM generated the genome assembly and annotation and performed the bioinformatics analysis. D.B. oversaw the project and wrote the manuscript with input from all authors.

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

Figure legends

Figure 1. Gene content evolution of newly formed sex chromosomes. A. Karyotype and gene content evolution on the neo-sex chromosomes of *D. miranda*. Shown are the karyotype of *D. miranda*, and its close relative *D. pseudoobscura*, from which it diverged about 2MY ago. In *D. miranda*, the fusion of an autosome (chromosome 3) with the Y chromosome created the neo-sex chromosomes, about 1.5 MY ago. Shown along the trees are numbers of gene amplifications (in green) and gene losses (in red) of genes ancestrally present on chr3, on the neo-X and neo-Y chromosomes, assuming parsimony. **B**. Gene content of the *D. miranda* neo-Y/Y chromosome. The *D. miranda* neo-Y contains different types of genes: about 1582 ancestral single-copy genes; 1697 multi-copy Y genes from 363 distinct proteins that were ancestrally present on the neo-Y, and 2036 genes from 94 distinct proteins that co-amplified on both the X/neo-X and Y/neo-Y. "Others" refers to genes not present or mapping to an unknown location in *D. pseudoobscura* (446 genes), or genes with

complex mapping (513 genes; see Methods). **C.** Distinct evolutionary processes may drive the accumulation of multi-copy Y genes, or co-amplified X and Y genes.

Figure 2. Characterization of ampliconic Y genes. A. Expression of multi-copy Y genes in different male samples. Multi-copy Y genes with high copy number are primarily expressed in testis, while gene families with low copy are expressed in multiple tissues. B. List of top amplified multi-copy Y genes. C. Dosage compensation status of neo-X homologs of multi-copy Y genes. Genes with multiple copies on the Y are less likely to be dosage compensated on the X. D. Expression pattern of co-amplified Y genes. E. List of top co-amplified X/ Y genes, their copy number, and inferred function.

Figure 3. Characterization of X/Y co-amplified genes. A, B. Gene ontology analysis enrichment. Genes that are co-amplified on both the X and the Y are enriched for meiosis-related and RNAi functions. C. Tissue expression patterns of co-amplified genes in *D. pseudoobscura*. D. Tissue expression patterns, short RNA targeting and copy numbers for co-amplified X and Y genes.

Figure 4 Co-amplified gene families are targeted by short RNAs in testis. A. Enrichment of short RNAs mapping to co-amplified X and Y genes. Shown are testis short RNA counts (normalized by total gene length for all copies of a gene family) for different categories of genes on the X/neo-X and Y/neo-Y chromosome (i.e. ancestral single-copy X and Y genes; genes only amplifying on the Y and their neo-X gametologs; genes co-amplifying on the Y/neo-Y and X/neo-X). **B**.-**D**. Examples of co-amplified X and Y genes (*fest*, *mars*, *zip*). Shown are the genomic architecture of co-amplified gene families on the neo-X and neo-Y (repetitive regions are displayed in gray, and genes in green), and expression profiles from testis (stranded RNA-seq data in red, short RNA profiles in blue). For each gene, only a representative subset of copies is shown.

Table S1. Gene loss on the different Muller element's of *D. miranda*

Table S2. Overview of multi-copy Y genes

Table S3. Overview of co-amplified X/Y genes

Table S4. Expression of individual copies of multi-copy Y genes and co-amplified X/Y genes

Table S5. Small RNA mapping from testis, for single copy X and Y genes, multi-copy Y genes and their X homologs, and co-amplified X and Y genes

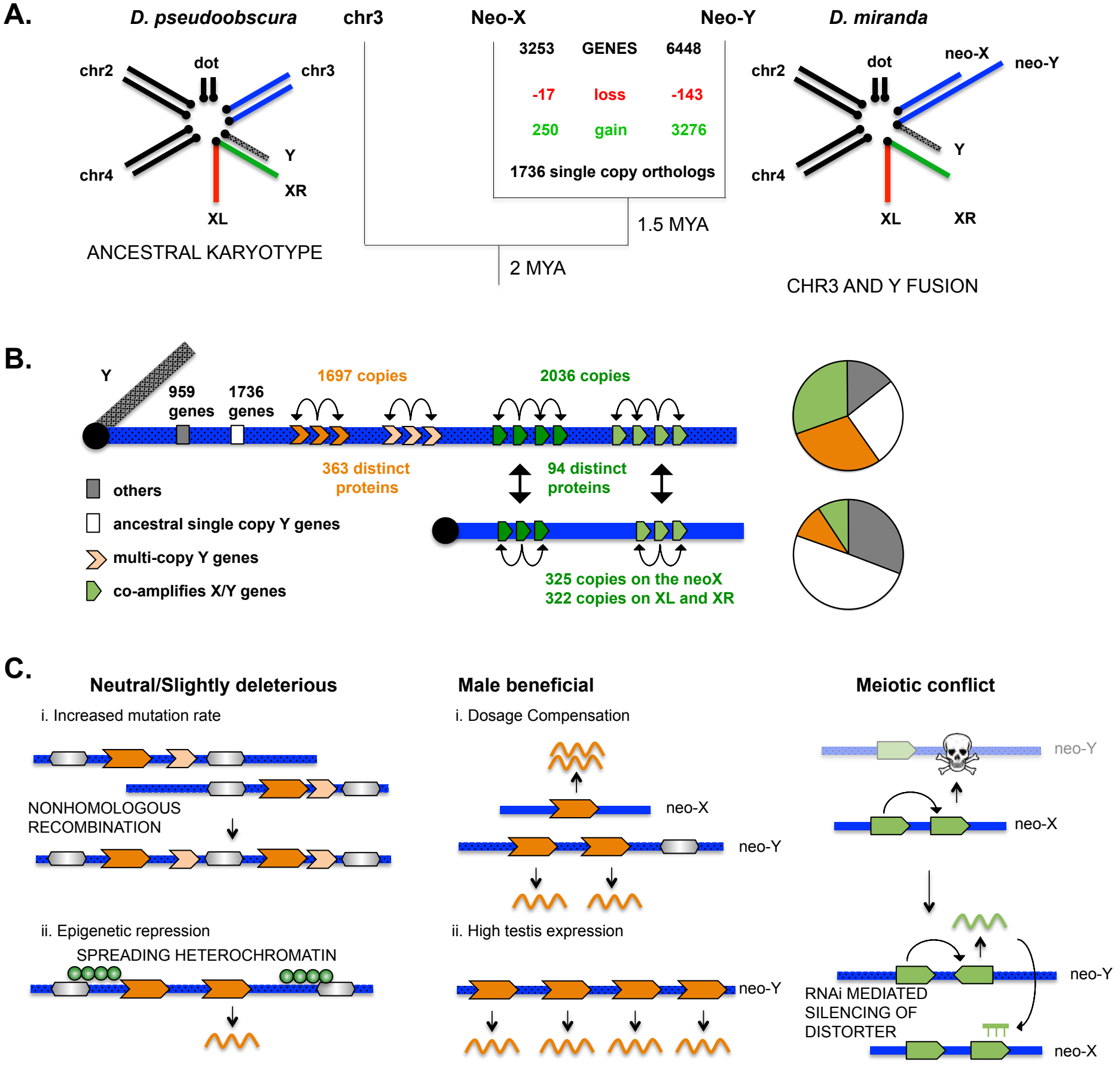
Figure S1. Location of multi-copy Y genes and co-amplified X/Y genes

Figure S2. Repeats contribute to accumulation of multi-copy genes

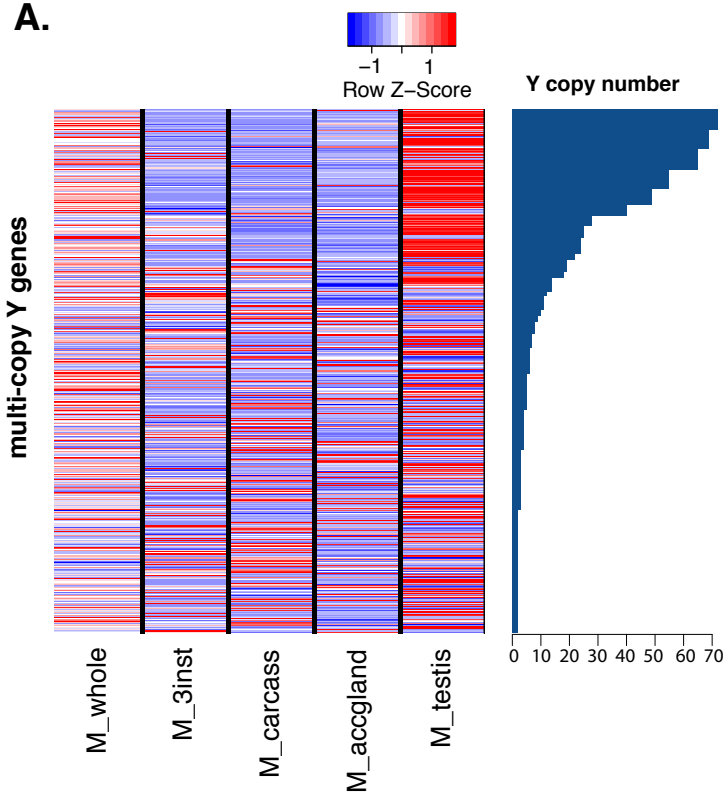
Figure S3. Expression patterns of X- and Y-linked genes for A. single-copy X- and Y-linked genes; B. multi-copy Y-linked genes and their X homologs; C. co-amplified X and Y genes. Expression for individual gene copies is shown.

Figure S4. Expression patterns of X- and Y-linked genes for A. single-copy X- and Y-linked genes; B. multi-copy Y-linked genes and their X homologs; C. co-amplified X and Y genes. Total summed expression for all gene copies of a gene family is shown.

Figure S5. Tissues-specific expression patterns of orthologs of co-amplified X/Y genes in *D. melanogaster*.


A.
D. pseudoobscura
chr3
Neo-X
Neo-Y
D. miranda
chr2
dot
chr3
chr4
Y
XR
XL
ANCESTRAL KARYOTYPE
3253
GENES
6448
-17
loss
-143
250
gain
3276
1736 single copy orthologs
1.5 MYA
2 MYA
chr2
dot
neo-X
neo-Y
chr4
Y
XL
XR
CHR3 AND Y FUSION
B.
Y
959 genes
1736 genes
1697 copies
2036 copies
363 distinct proteins
94 distinct proteins
others
ancestral single copy Y genes
multi-copy Y genes
co-amplifies X/Y genes
325 copies on the neoX
322 copies on XL and XR
C.
Neutral/Slightly deleterious
i. Increased mutation rate
NONHOMOLOGOUS RECOMBINATION
ii. Epigenetic repression
SPREADING HETEROCHROMATIN
Male beneficial
i. Dosage Compensation
neo-X
neo-Y
ii. High testis expression
neo-Y
Meiotic conflict
neo-Y
neo-X
neo-Y
RNAi MEDIATED SILENCING OF DISTORTER
neo-X

A.

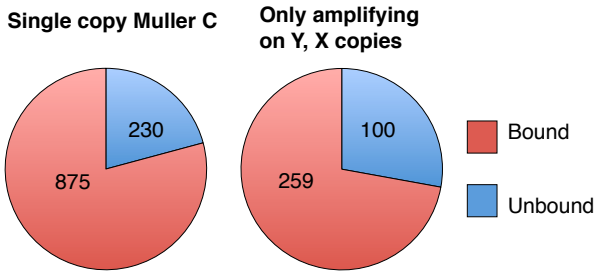


B.

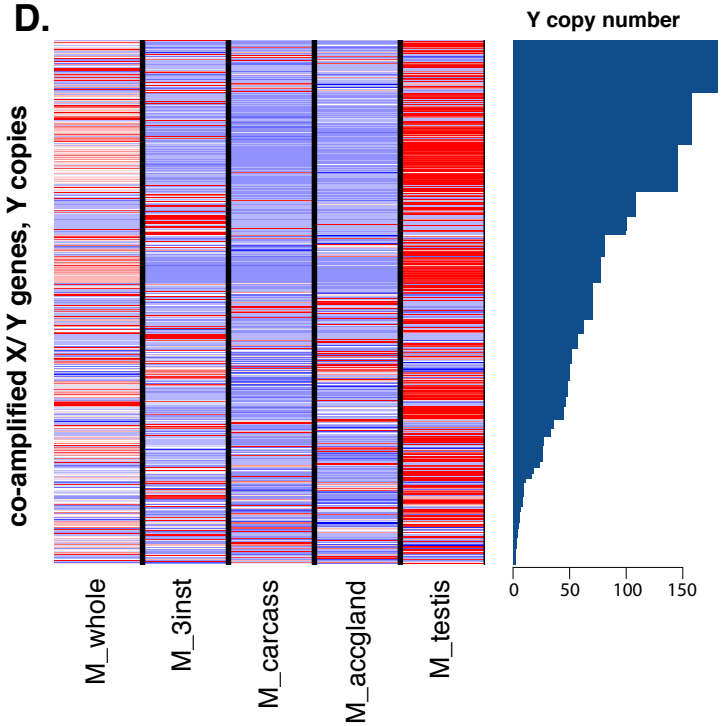
Multi-copy genes with highest Y copy number

D. pse ortholog	*D. mel* ortholog	Copy #	Highest expression in *D. pse*
FBgn0073907	CG15657	72	testis
FBgn0081758	Hmg-2	69	ovary
FBgn0079429	fand	65	ovary
FBgn0246260	RhoGAP54D	55	ovary
FBgn0077065	Drak	49	ovary
FBgn0074178	CG16787	40	testis
FBgn0064414	thr	28	ovary
FBgn0246134	Cap-G	25	ovary
FBgn0073063	CG14490	25	NA
FBgn0249238	CG3955	24	adult male head
FBgn0072254	CG13326	24	testis
FBgn0272696	CG14154	22	testis
FBgn0081263	sand	19	testis & ovary
FBgn0074882	Dg	19	ovary
FBgn0245806	Cg30460	18	adult male

C.



D.

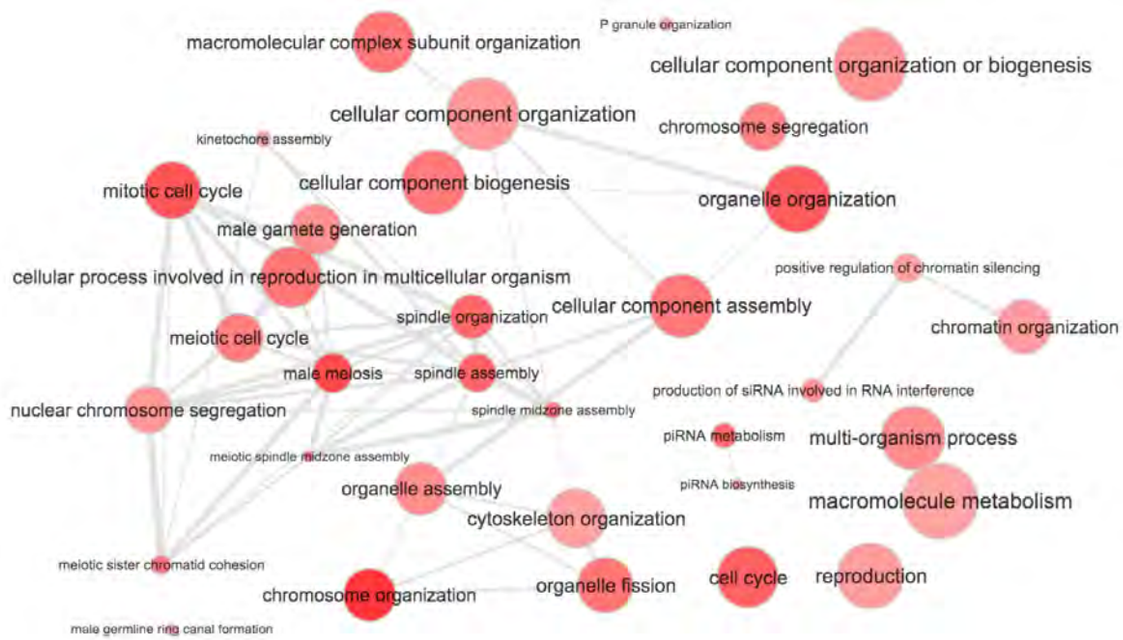


E.

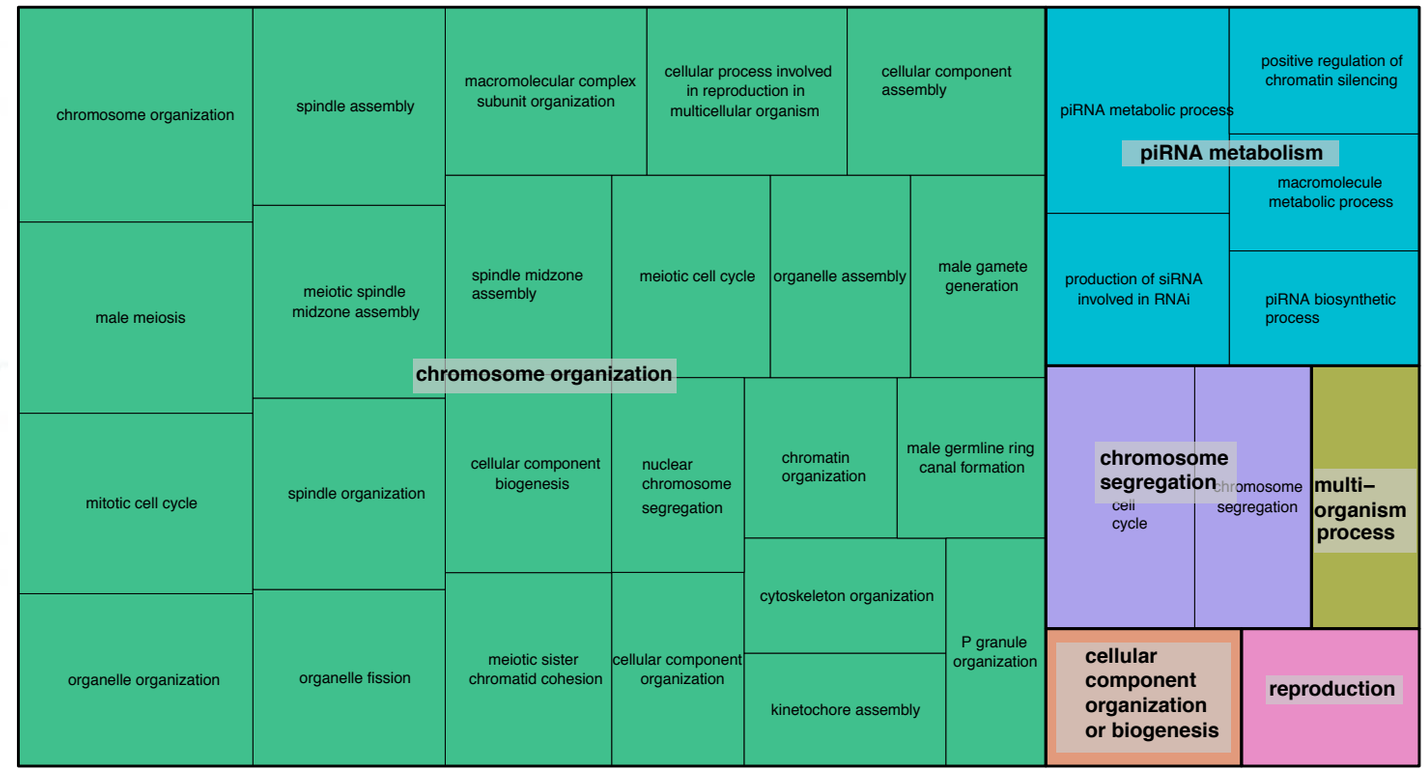
Co-amplified genes with highest Y copy number

D. pse ortholog	*D. mel* ortholog	X #	Y #	Function
FBgn0071596	polo	37	11	meiotic spindle assembly
FBgn0078395	shu	22	50	piRNA biogenesis
FBgn0245975	CG10822	22	45	microtubule-based movement
FBgn0081996	panx	22	2	piRNA-guided silencing
FBgn0248821	AGO2	18	2	siRNA interaction/ RISC
FBgn0243940	Incenp	17	8	microtubule binding, meiotic chromosome condensation
FBgn0246244	FANCI	12	9	DNA polymerase binding
FBgn0245626	subito	11	8	spindle organization, chromosome segregation
FBgn0073991	zip	9	5	ATP binding, ATPase activity
FBgn0245938	Best2	9	70	chloride channel activity
FBgn0244163	cuff	9	7	production of siRNA, piRNA metabolic process
FBgn0081586	Klp61F	8	27	microtubule binding, microtubule motor activity
FBgn0070220	SMC2	7	3	chromosome condensation
FBgn0079631	Dicer-2	6	26	RNAi pathway
FBgn0074335	mars	6	48	spindle organization, kinetochore assembly, chromosome segregation
FBgn0076207	ord	5	18	meiotic sister chromatid cohesion, chromosome segregation
FBgn0082139	fest	5	145	spindle assembly involved in male meiosis
FBgn0075258	scra	2	9	microtubule binding

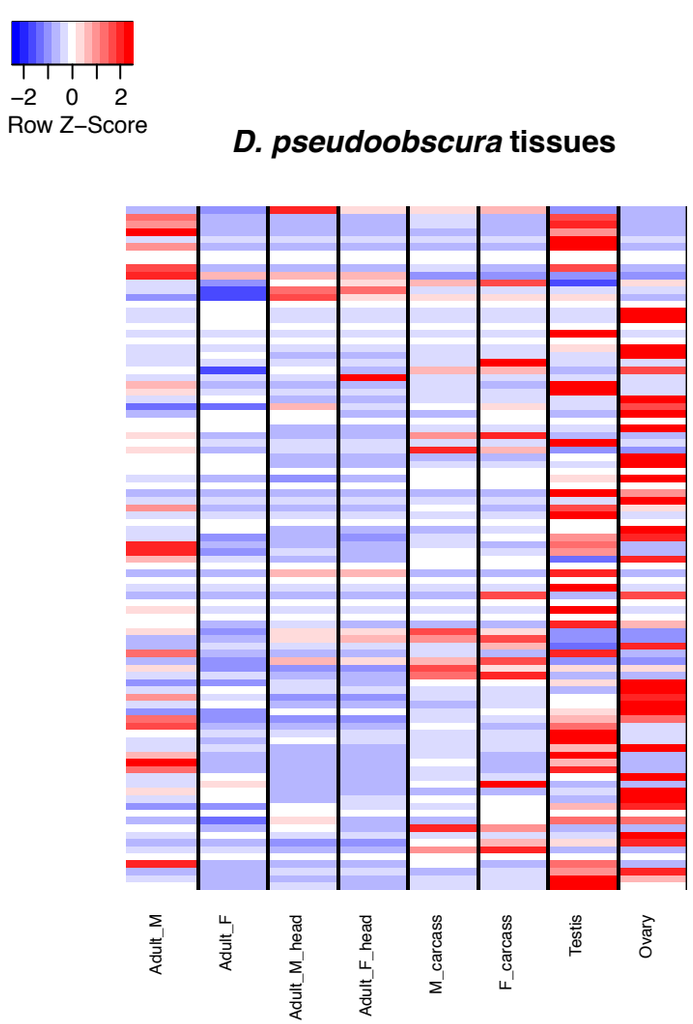
A.


macromolecular complex subunit organization
P granule organization
cellular component organization or biogenesis
cellular component organization
chromosome segregation
kinetochore assembly
mitotic cell cycle
cellular component biogenesis
organelle organization
male gamete generation
cellular process involved in reproduction in multicellular organism
positive regulation of chromatin silencing
spindle organization
cellular component assembly
chromatin organization
meiotic cell cycle
male meiosis
spindle assembly
production of siRNA involved in RNA interference
nuclear chromosome segregation
spindle midzone assembly
piRNA metabolism
multi-organism process
meiotic spindle midzone assembly
piRNA biosynthesis
organelle assembly
macromolecule metabolism
cytoskeleton organization
meiotic sister chromatid cohesion
chromosome organization
organelle fission
cell cycle
reproduction
male germline ring canal formation

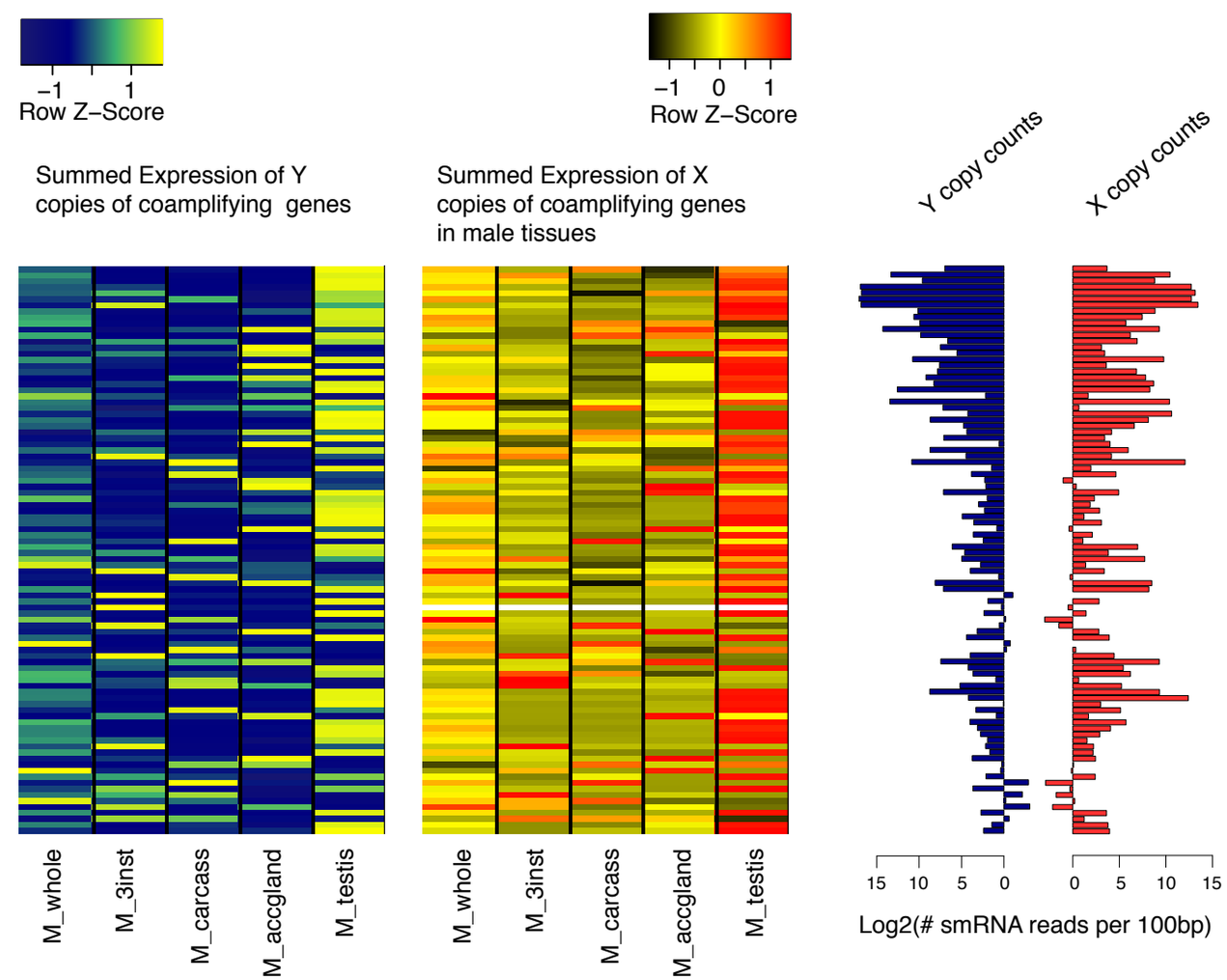
B.


chromosome organization
spindle assembly
macromolecular complex subunit organization
cellular process involved in reproduction in multicellular organism
cellular component assembly
piRNA metabolic process
positive regulation of chromatin silencing
piRNA metabolism
macromolecule metabolic process
male meiosis
meiotic spindle midzone assembly
spindle midzone assembly
meiotic cell cycle
organelle assembly
male gamete generation
production of siRNA involved in RNAi
piRNA biosynthetic process
chromosome organization
mitotic cell cycle
spindle organization
cellular component biogenesis
nuclear chromosome segregation
chromatin organization
male germline ring canal formation
chromosome segregation
cell cycle
chromosome segregation
multi–organism process
cytoskeleton organization
organelle organization
organelle fission
meiotic sister chromatid cohesion
cellular component organization
P granule organization
kinetochore assembly
cellular component organization or biogenesis
reproduction

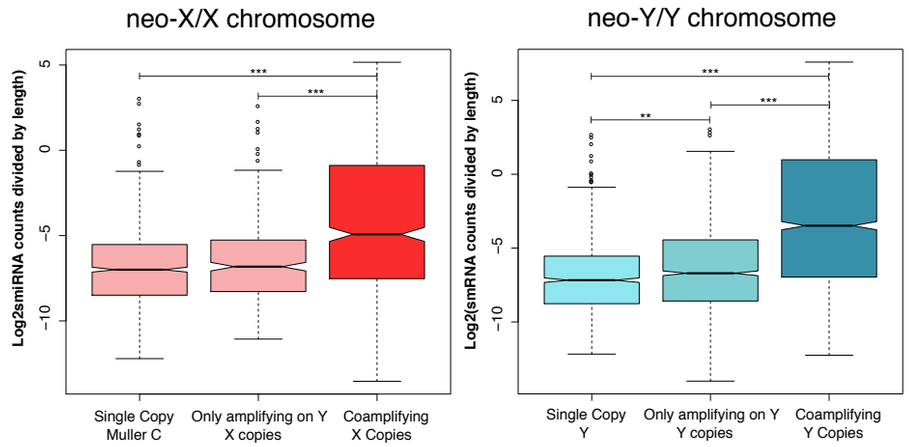
C.


−2 0 2
Row Z−Score
D. pseudoobscura tissues
Adult_M
Adult_F
Adult_M_head
Adult_F_head
M_carcass
F_carcass
Testis
Ovary

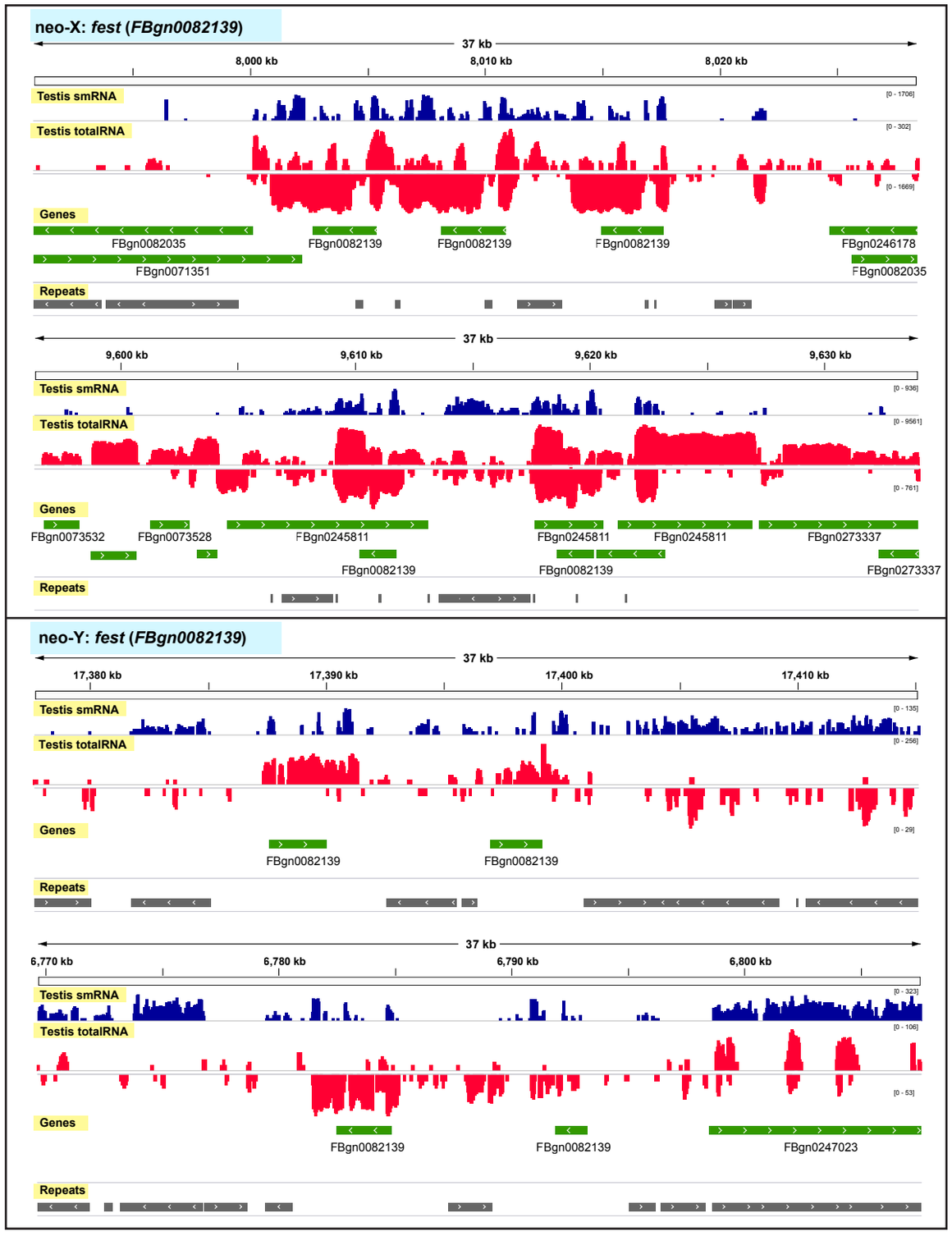
D.


−1 1
Row Z−Score
−1 0 1
Row Z−Score
Y copy number
X copy number
150 100 50 0 10 20 30 40
# of copies
Summed Expression of Y copies of coamplifying genes
Summed Expression of X copies of coamplifying genes in male tissues
M_whole
M_3inst
M_carcass
M_accgland
M_testis
Y copy counts
X copy counts
15 10 5 0 0 5 10 15
Log2(# smRNA reads per 100bp)

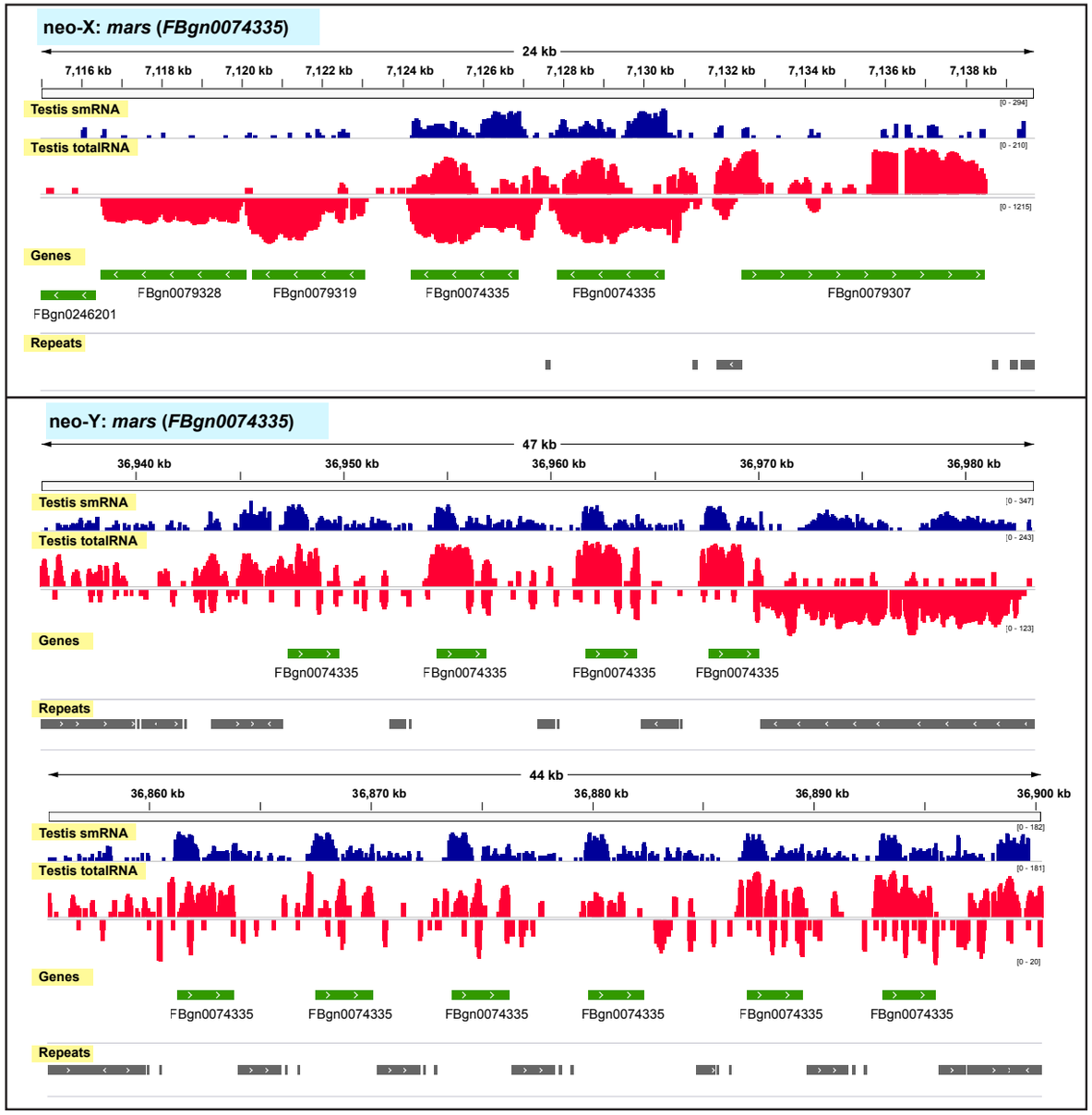
A.


neo-X/X chromosome
neo-Y/Y chromosome
Log2smiRNA counts divided by length)
Log2(smRNA counts divided by length)
Single Copy Muller C
Only amplifying on Y X copies
Coamplifying X Copies
Single Copy Y
Only amplifying on Y Y copies
Coamplifying Y Copies

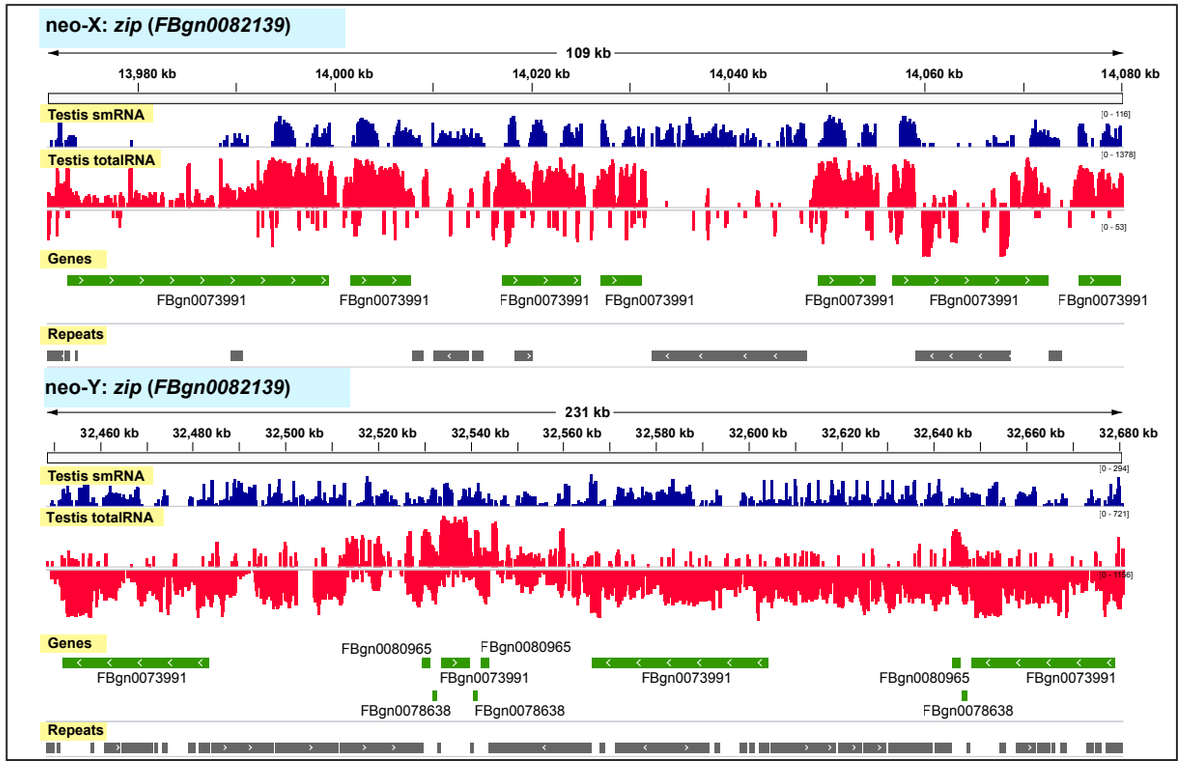
B.


neo-X: fest (FBgn0082139)
neo-Y: fest (FBgn0082139)
Testis smRNA
Testis totalRNA
Genes
Repeats

C.


neo-X: mars (FBgn0074335)
neo-Y: mars (FBgn0074335)
Testis smRNA
Testis totalRNA
Genes
Repeats

D.


neo-X: zip (FBgn0082139)
neo-Y: zip (FBgn0082139)
Testis smRNA
Testis totalRNA
Genes
Repeats