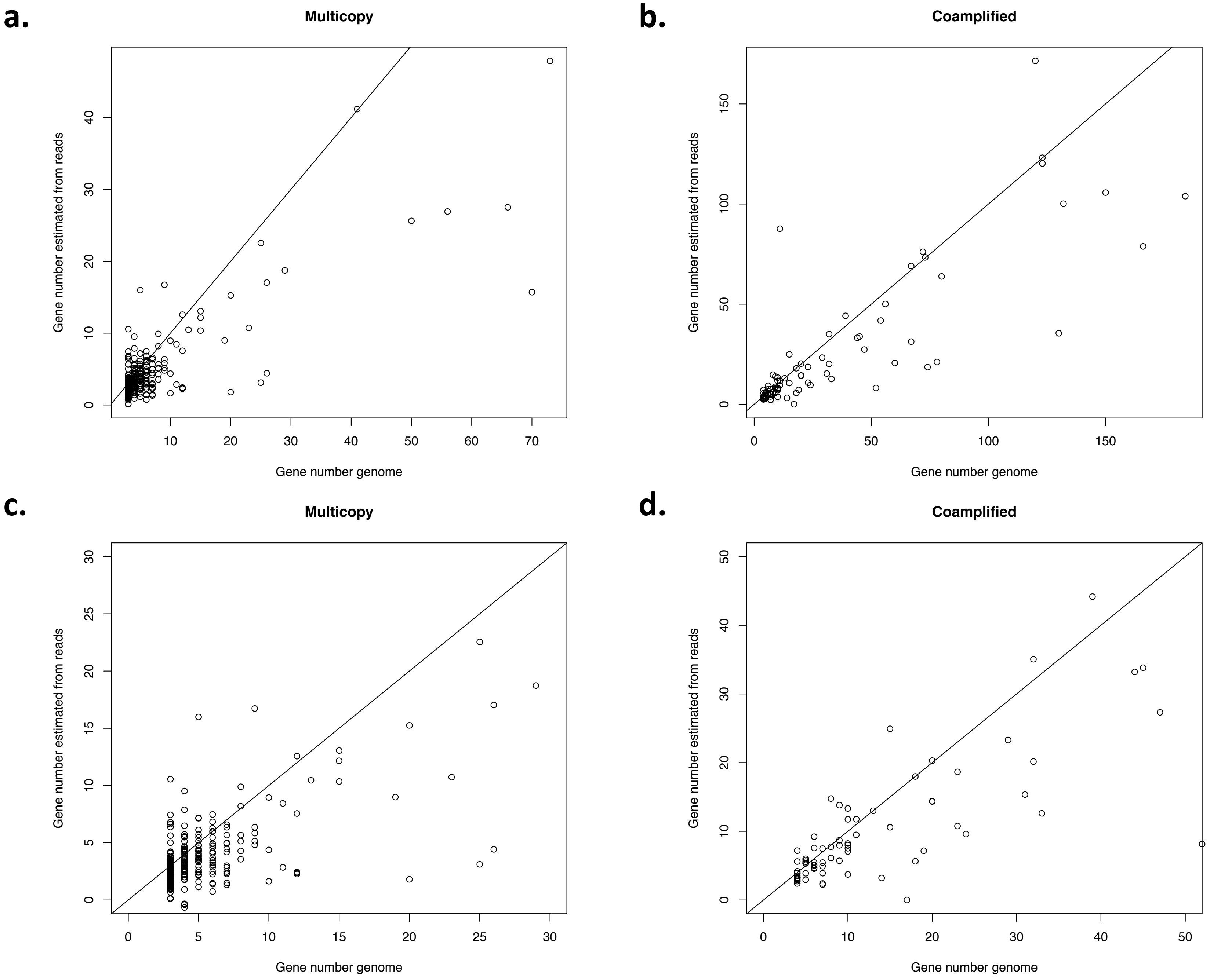
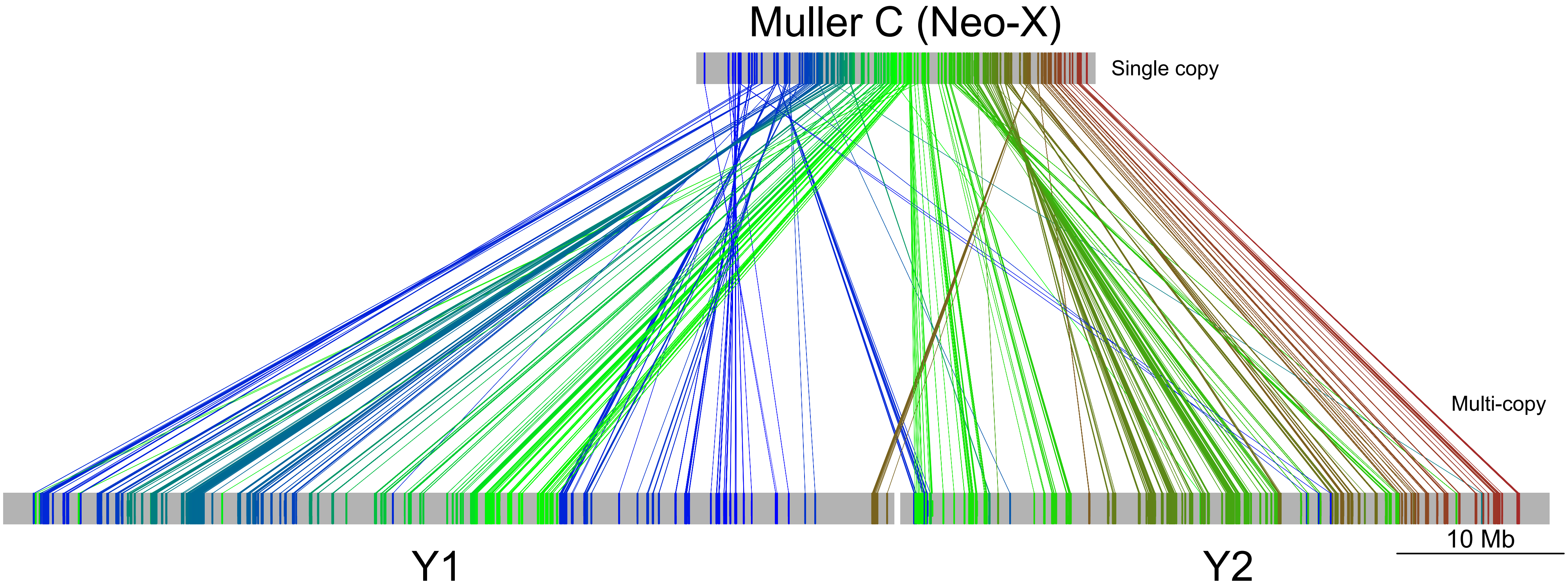
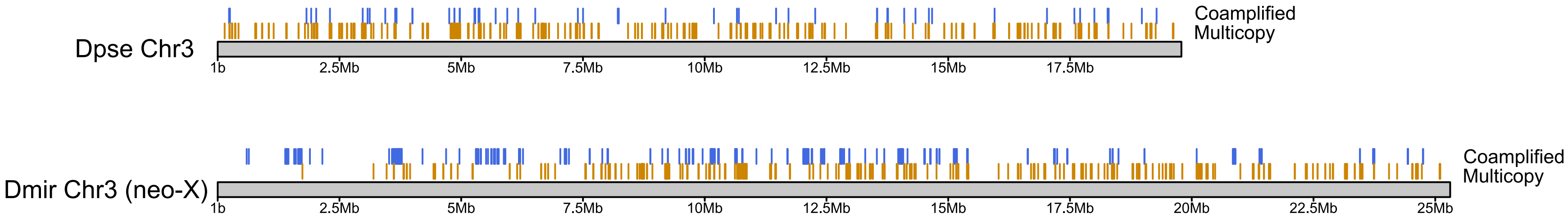


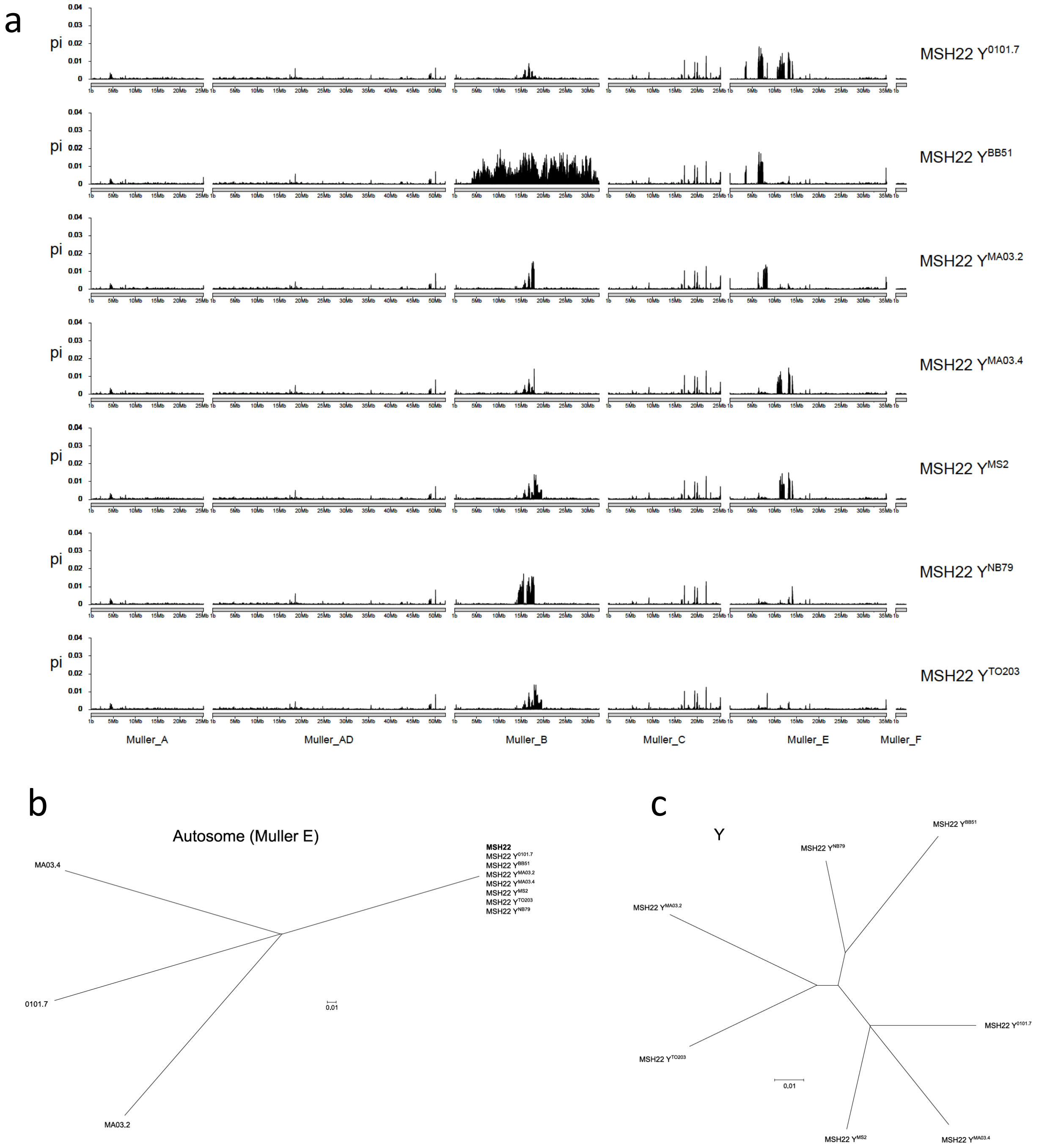
Supplementary Figure 1. Validation of multi-copy and co-amplified Y genes in the *D. miranda* genome assembly using Illumina read coverage analysis from the reference strain (MSH22). Shown is the gene copy number from the genome assembly and the estimated copy number based on read mapping to A) multi-copy genes (Pearson correlation = 0.82) and B) co-amplified genes (Pearson correlation = 0.84) with the 1:1 line shown. Zoomed in regions for multi-copy and co-amplified genes are shown in C) and D). Note that our coverage analysis typically underestimates the number of Y-linked copies found in the assembly, presumably due to many multi-copy genes being fragmented in the assembly.



Supplementary Figure 2. Location of multi-copy Y genes. The genomic coordinates of genes identified as multi-copy on the two largest Y-chromosome scaffolds (Y1 and Y2) and their corresponding single copy location on the neo-X (Muller C). Most multi-copy genes on the Y exist in tandem and therefore there is a substantial overlap of plotting lines. Note that the Y/neo-Y-chromosomes in our current assembly in *D. miranda* consists of three major scaffolds (Y1, Y2, Y3, with Y3 possibly corresponding to the ancestral Y).



Supplementary Figure 3. Location of orthologs of multi-copy Y genes and co-amplified X/Y genes in *D. pseudoobscura* and *D. miranda*. Each line shows the location of the ortholog of a co-amplified X/Y gene (in blue), or a multi-copy Y gene (in orange) along *D. pseudoobscura* chr3, or the neo-X of *D. miranda*. Of the 363 multi-copy genes identified on the Y/neo-Y of *D. miranda*, 349 are located on chr3 of *D. pseudoobscura*, 2 genes are on chr2, 7 genes with unknown location, 1 gene on XL, and 4 genes on XR. Of the 94 co-amplified X/Y genes, 59 are located on chr3 of *D. pseudoobscura*, 4 genes are on chr2, 1 gene on chr4, 16 genes with unknown location, 6 genes on XL, and 8 genes on XR.

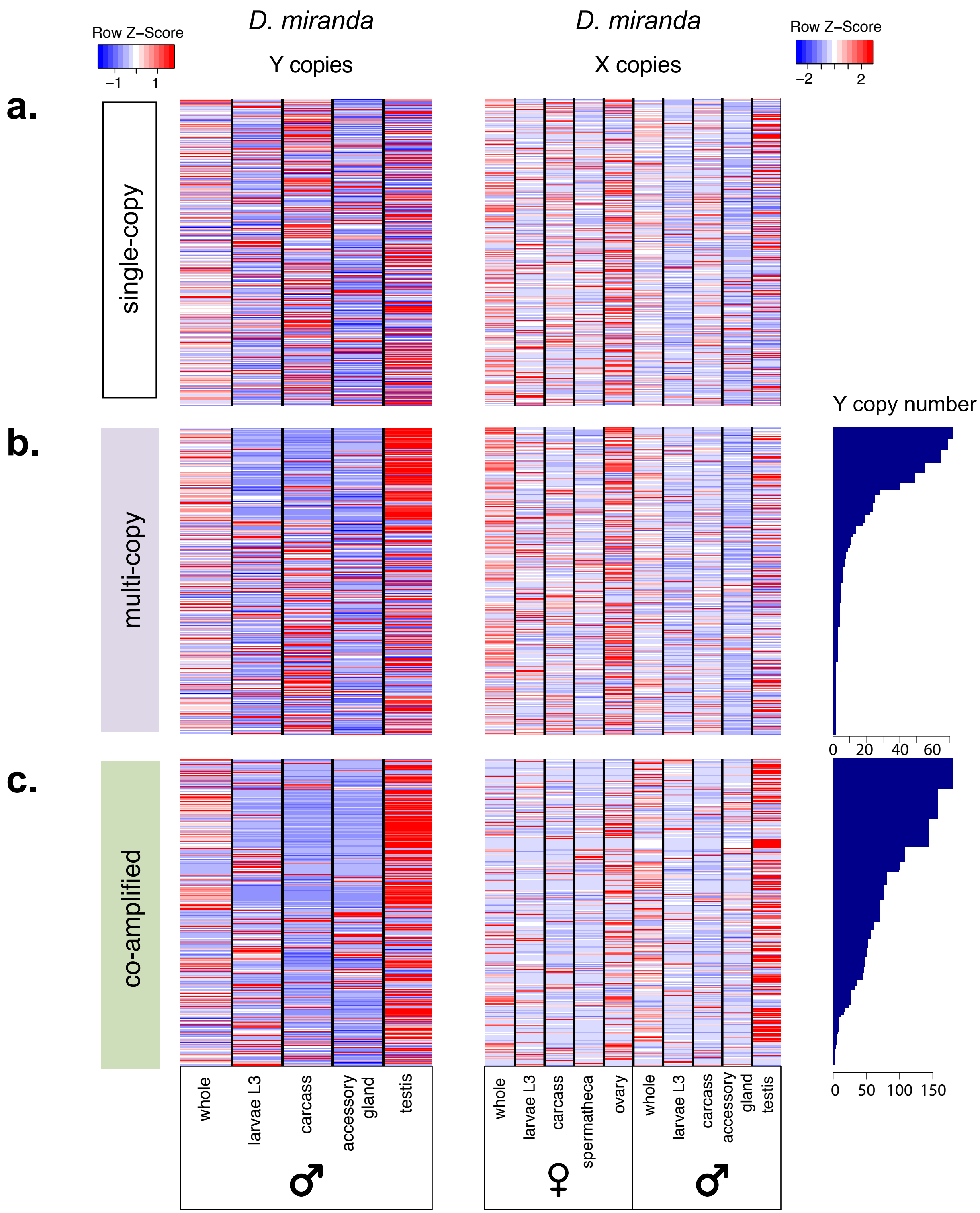


Supplementary Figure 4. Properties of Y-chromosome replacement lines. A) Individual nucleotide diversity (pi) shown for each replacement line. Peaks identify regions of residual heterozygosity in each Y-chromosome replacement line. B) Phylogenetic networks of autosomal SNPs of the seven Y-chromosome replacement lines and three lines from which some Y-chromosomes were derived. As expected, all Y replacement lines cluster with MSH22. C) Phylogenetic network of SNPs showing seven distinct Y types.

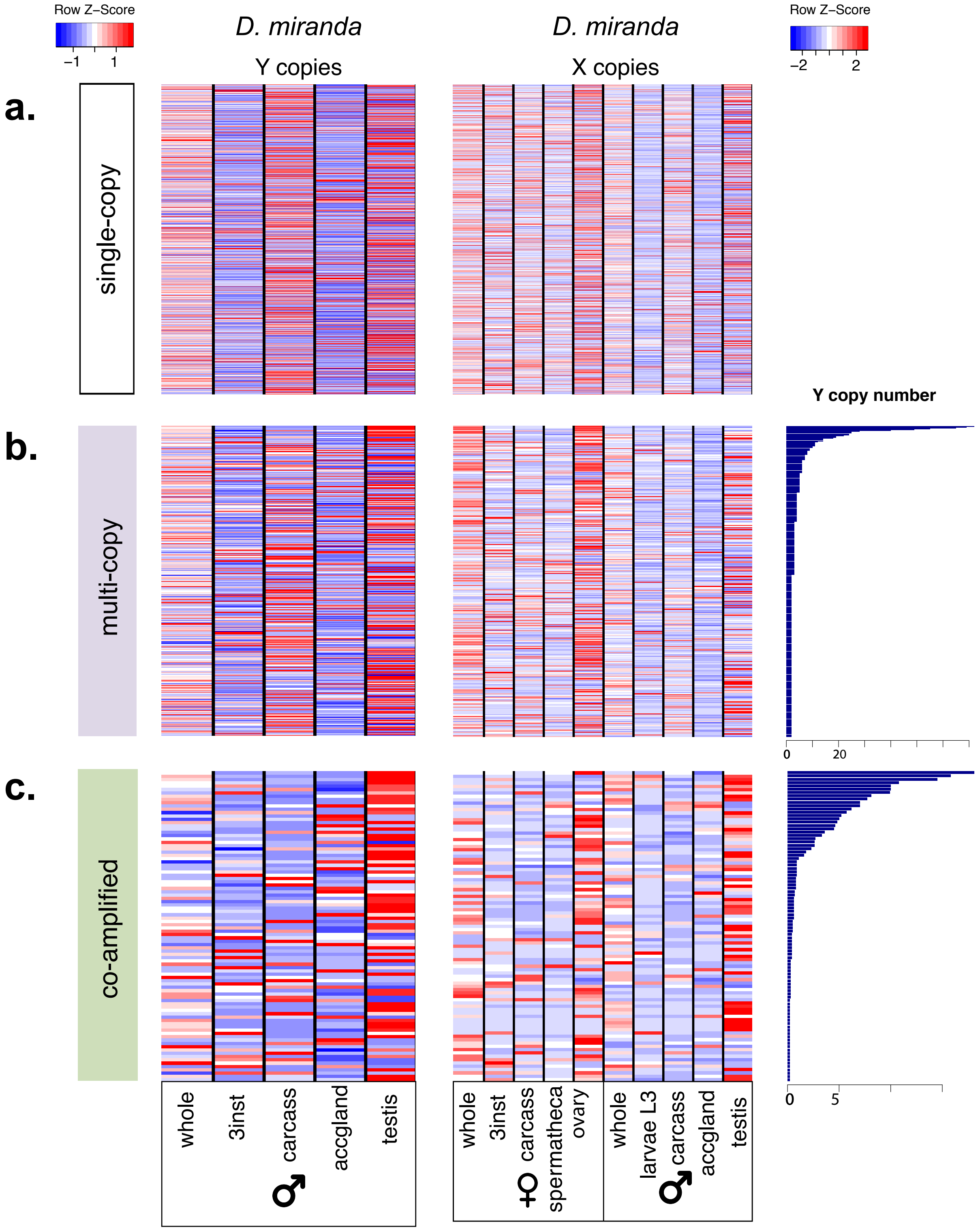
neo-X - FBgn0079631 (Dicer-2)

neoY - FBgn0079631 (Dicer-2)

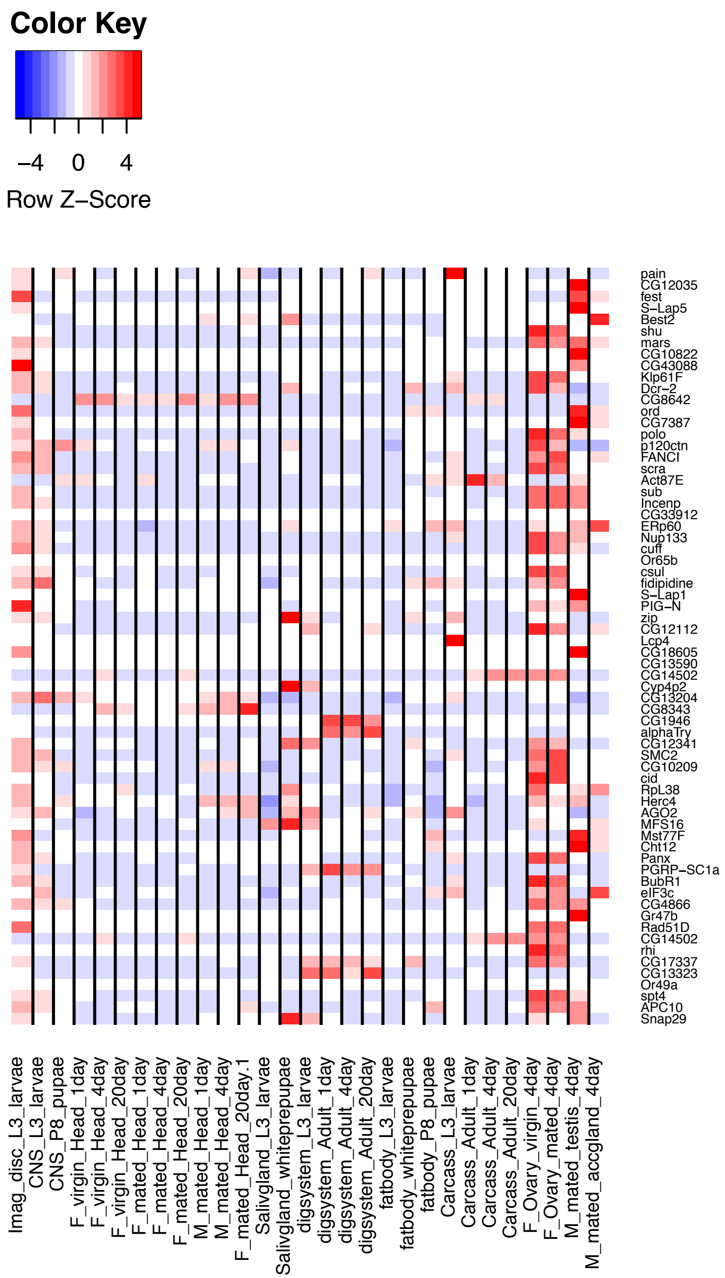
Supplementary Figure 5. Repeats may contribute to accumulation of multi-copy genes. Transposable elements often flank multi-copy Y genes, and may have contributed to amplification of genes on the X and Y. Genes are shown in green, and TEs are shown in gray.



Supplementary Figure 6. 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. Expression values were row-normalized to obtain Z-scores with a mean of 0 and standard deviation of 1, using the built-in scale = 'row' argument in the heatmap.2 function from the package gplots in R. The data shown are presented in **Data Supplement 16**.



Supplementary Figure 7. 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. Expression values were row-normalized to obtain Z-scores with a mean of 0 and standard deviation of 1, using the built-in scale = ’row’ argument in the heatmap.2 function from the package gplots in R. The data shown are presented in **Data Supplement 17**.



Supplementary Figure 8. Tissues-specific expression patterns of orthologs of co-amplified X/Y genes in *D. melanogaster*. FPKM values were downloaded from flybase.org. The data shown are presented in **Data Supplement 18**.

Additional Supplementary Files (as Spreadsheet-based Tables)

Supplementary Table 1. Gene loss on the different Muller elements of *D. miranda*. Shown are genes in *D. pseudoobscura* for different chromosomes that are absent on their homologous chromosome arm in *D. miranda*. The table gives *D. pseudoobscura* gene ID (FBgn), *D. melanogaster* ortholog, the chromosomal location of the gene in *D. pseudoobscura*, the tissue of highest expression in *D. pseudoobscura*, the category of the gene, and whether the gene is present somewhere else in the *D. miranda* genome.

Supplementary Table 2. Overview of multi-copy Y genes. Shown are total copy numbers for multi-copy Y genes, as well as the number of full-length (>90%) and partial Y copies (50-90%; 25-50%; and less than 25% compared to the length of the orthologous gene in *D. pseudoobscura*). The expression spreadsheet shows expression of orthologs of multi-copy Y genes in *D. pseudoobscura*, and the GO analysis shows the orthologous genes in *D. melanogaster*. No significant GO enrichment terms were detected.

Supplementary Table 3. Overview of co-amplified X/Y genes. Shown are total copy number for co-amplified X and Y genes, as well as the numbers of full-length (>90%) and partial X and Y copies (50-90%; 25-50%; and less than 25% compared to the length of the orthologous gene in *D. pseudoobscura*). The expression spreadsheets show expression of orthologs of co-amplified X/Y genes in *D. pseudoobscura* and *D. melanogaster*, and GO analysis shows the orthologous genes in *D. melanogaster* and GO terms that were significantly enriched among co-amplified X/Y genes (using either Gorilla or PantherDB).

Supplementary Table 5. Clustering of multi-copy Y and co-amplified X/Y genes. Genome intervals of clustered (within 100kb of each other) multi-copy genes and co-amplified genes and their *Drosophila pseudoobscura* gene ID (FBgn).

Supplementary Table 4

Expression of individual copies of multi-copy Y genes and co-amplified X/Y genes. Shown is the inferred fraction of individual gene copies expressed, depending on different cut-offs.

Number of reads mapping (with mapq > 20)	Fraction of individual gene copies expressed Multi-copy Y genes	 Co-amplified Y genes
2 unique reads	71%	94%
5 unique reads	60%	88%
10 unique reads	50%	79%

Supplementary Table 6

Y-chromosome replacement lines used in the study.

Shown is the collection location for the different Y chromosome replacement lines generated.

Y-replacement line	Collection location of Y line	Collection year of Y line	Isofemale (Y) line originally described in
MSH22 $Y^{0101.7}$	Port Coquitlam, B.C., Canada	N/A	Yi and Charlesworth 2000
MSH22 Y^{BB51}	Big Bear, California	2017	present study
MSH22 $Y^{MA03.2}$	Mather, California	2003	Bartolomé and Charlesworth 2006
MSH22 $Y^{MA03.4}$	Mather, California	2003	Bartolomé and Charlesworth 2006
MSH22 Y^{MS2}	Mt. Seymour, B.C., Canada	2017	present study
MSH22 Y^{NB79}	Mt. Hood, Oregon	2017	present study
MSH22 Y^{TO203}	Wellington, Washington	2017	present study

Supplementary Table 7

List of top amplified multi-copy Y genes (i.e. high copy number Y genes), their copy number on the Y, and tissue of highest expression on *D. pseudoobscura* (as a proxy for ancestral expression).

***D. melanogster* ortholog**	**Copy #**	**Highest expression in *D. pseudoobscura***
CG15657	72	testis
Hmg-2	69	ovary
fand	65	ovary
RhoGAP54D	55	ovary
Drak	49	ovary
CG16787	40	testis
thr	28	ovary
Cap-G	25	ovary
CG14490	25	NA
CG3955	24	adult male head
CG13326	24	testis
CG14154	22	testis
sand	19	testis & ovary
Dg	19	ovary
Cg30460	18	adult male

Supplementary Table 8

List of top co-amplified X/ Y genes with well-characterized roles in meiosis or siRNA, their copy number, and inferred function in *D. melanogaster*.

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

Supplementary Table 9

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

smRNA Categories	median (counts)	median (counts/length)	Log2(median (counts))	Log2(median (counts/length))
neo-X copies				
Single-copy Y	21	0.00782661	4.392317	-6.997397
Multi-copy Y	23	0.008840616	4.523562	-6.821637
Co-amplified X/Y	43	0.03284672	5.426265	-4.928107
P-values				
(Log2(counts/length))	Single-copy	Multi-copy	Co-amplified X/Y	
Single-copy Y	-	0.3159	< 2.2e-16	
Multi-copy Y	-	-	< 2.2e-16	
Co-amplified X/Y	-	-	-	
neo-Y copies				
Single-copy Y	17	0.006932409	4.087463	-7.172428
Multi-copy Y	17	0.009580838	4.087463	-6.705632
Co-amplified X/Y	134	0.08941301	7.066089	-3.483377
P-values				
(Log2(counts/length))	Single-copy	Multi-copy	Co-amplified X/Y	
Single-copy Y	-	8.46E-09	< 2.2e-16	
Multi-copy Y	-	-	< 2.2e-16	
Co-amplified X/Y	-	-	-	

Supplementary Table 10

Sequence data generated and SRA accession numbers.

Species	strain	Sex	Tissues	Library type	Accession number	Reference
D. miranda	MSH22	mixed	third instar larvae	ChIP-seq, MSL-tap	SRR789659	Alekseyenko et al. 2013
D. miranda	MSH22	male	third instar larvae	ChIP-seq, Input Control	SRR789646	Alekseyenko et al. 2013
D. miranda	MSH22	female	Female third instar larvae	total RNA	SRR899848	Zhou & Bachtrog 2012
D. miranda	MSH22	female	virgin female carcass	total RNA	SRR364804	Zhou & Bachtrog 2012
D. miranda	MSH22	female	whole virgin female	total RNA	SRR364800	Zhou & Bachtrog 2012
D. miranda	MSH22	female	Female spermatheca	total RNA	SRR4416187	Zhou & Bachtrog 2012
D. miranda	MSH22	female	virgin female ovary	total RNA	SRR364801	Zhou & Bachtrog 2012
D. miranda	MSH22	male	Male third instar larvae	total RNA	SRR899847	Zhou & Bachtrog 2012
D. miranda	MSH22	male	virgin male carcass	total RNA	SRR364803	Zhou & Bachtrog 2012
D. miranda	MSH22	male	whole virgin male	total RNA	SRR364798	Zhou & Bachtrog 2012
D. miranda	MSH22	male	virgin male accessory gland	total RNA	SRR364802	Zhou & Bachtrog 2012
D. miranda	MSH22	male	virgin male testis	total RNA	SRR364799	Zhou & Bachtrog 2012
D. miranda	MSH22	male	virgin male testis	total RNA, stranded	SRR9204764	this study
D. miranda	MSH22	male	virgin male testis	small RNA	SRR9204764	this study
D. miranda	MSH22 $Y^{0101.7}$	male	whole adult	gDNA	SRR9211917	this study
D. miranda	MSH22 Y^{BB51}	male	whole adult	gDNA	SRR9211916	this study
D. miranda	MSH22 $Y^{MA03.2}$	male	whole adult	gDNA	SRR9211919	this study
D. miranda	MSH22 $Y^{MA03.4}$	male	whole adult	gDNA	SRR9211918	this study
D. miranda	MSH22 Y^{MS2}	male	whole adult	gDNA	SRR9211921	this study
D. miranda	MSH22 Y^{NB79}	male	whole adult	gDNA	SRR9211920	this study
D. miranda	MSH22 Y^{TO203}	male	whole adult	gDNA	SRR9211922	this study
D. miranda	MA03.4	male	whole adult	gDNA	SRR9648609	this study
D. miranda	101.7	male	whole adult	gDNA	SRR9648608	this study
D. miranda	MA03.2	male	whole adult	gDNA	SRR9648607	this study

Additional Supplementary Files

Data Supplement 1. Repeat library used for masking the *D. miranda* genome (fasta file).

Data Supplement 2 Repeat annotation of the *D. miranda* genome (gff file).

Data Supplement 3 Gene annotation (all genes) of the *D. miranda* genome (gff file).

Data Supplement 4 Gene annotation of multi-copy Y genes and their orthologs in the *D. miranda* genome (gff file).

Data Supplement 5 Gene annotation of co-amplified X and Y genes in the *D. miranda* genome (gff file).

Data Supplement 6. Gene expression values (TPM) for multi-copy Y genes, and co-amplified Y genes, in different *Drosophila miranda* male tissues, and tissue-specificity index tau.

Data Supplement 7. Gene copy numbers for co-amplified X/Y genes.

Data Supplement 8. Gene expression values (TPM) for co-amplified X and Y gene families, in different *Drosophila miranda* male tissues.

Data Supplement 9. Gene expression values (FPKM) for orthologs of co-amplified X/Y genes in different *Drosophila pseudoobscura* male and female tissues.

Data Supplement 10. Sense and anti-sense testis total RNA summed counts for co-amplified genes.

Data Supplement 11. Sense and anti-sense testis small RNA summed counts for co-amplified genes.

Data Supplement 12. Testis total RNA counts for all copies of a gene family for different categories of genes on the X/neo-X and Y/neo-Y-chromosome.

Data Supplement 13. Testis small RNA raw counts for different categories of genes on the X/neo-X and Y/neo-Y-chromosome .

Data Supplement 14. Fasta alignment of co-amplified X/Y genes and their *D. pseudoobscura* ortholog.

Data Supplement 15. MSL ChIP and Input counts (normalized to library size) for neo-X genes whose homolog is classified as single-copy or multi-copy Y/neo-Y.

Data Supplement 16. Gene expression values (TPM) for single-copy Y genes, multi-copy Y genes, and co-amplified Y genes, and their neo-X/X homologs in different *Drosophila miranda* tissues. Expression values for all copies of a gene family are shown individually.

Data Supplement 17. Gene expression values (TPM) for multi-copy Y genes, and co-amplified X and Y genes in different *Drosophila miranda* tissues. Expression values for all copies of a gene family are summed.

Data Supplement 18. Gene expression values (FPKM) for orthologs of co-amplified X/Y genes in different *Drosophila melanogaster* tissues.