UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The Evolution of Expression in Drosophila

Permalink
https://escholarship.org/uc/item/0c1511n1

Author
Gibilisco, Lauren

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org
Powered by the California Digital Library
University of California

The Evolution of Expression in *Drosophila*

by

Lauren B. Gibilisco

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 Rasmus Nielsen
Professor Donald Rio

Fall 2017

Abstract

The Evolution of Expression in *Drosophila*

by

Lauren B. Gibilisco

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Doris Bachtrog, Chair

Genomes have evolved multiple ways to regulate gene expression. Alternative splicing of genes is a strategy by which more than one transcript and/or protein can be encoded by a single gene. For cases in which expression is harmful, genomes have evolved strategies to repress expression. Epigenetic modifications, and in particular repressive heterochromatin, can repress expression. RNAi, which uses sequence homology to target transcripts for degradation, is another strategy that has evolved to control expression.

I characterized alternative splicing (“AS") within and between *Drosophila* species, sexes, tissues, and developmental stages. Alternative pre-mRNA splicing (“AS”) greatly expands proteome diversity, but little is known about the evolutionary landscape of AS in Drosophila and how it differs between embryonic and adult stages or males and females. I studied the transcriptomes from several tissues and developmental stages in males and females from four species across the Drosophila genus. I found that 20–37% of multi-exon genes are alternatively spliced. While males generally express a larger number of genes, AS is more prevalent in females, suggesting that the sexes adopt different expression strategies for their specialized function. While the number of total genes expressed increases during early embryonic development, the proportion of expressed genes that are alternatively spliced is highest in the very early embryo, before the onset of zygotic transcription. This indicates that females deposit a diversity of isoforms into the egg, consistent with abundant AS found in ovary. Cluster analysis by gene expression levels shows mostly stage-specific clustering in embryonic samples, and tissue-specific clustering in adult tissues. Clustering embryonic stages and adult tissues based on AS profiles results in stronger species-specific clustering, suggesting that diversification of splicing contributes to lineage-specific evolution in Drosophila. Most sex-biased AS found in flies is due to AS in gonads, with little sex-specific splicing in somatic tissues.

Large portions of eukaryotic genomes consist of repetitive DNA, including transposable elements (“TEs”), whose mobilization can have deleterious effects on the host genome. The establishment of transcription-repressing heterochromatin at repetitive regions during early development safeguards genome integrity. Heterochromatin formation coincides with genome-wide activation of zygotic expression. Males often contain substantially more heterochromatic DNA than females, due to the presence of a large, repeat-rich Y chromosome. I characterized repeat expression over embryonic development in two *Drosophila* species and found that repeats are expressed at a higher level in males than in females during early embryogenesis. I carried out chromatin immunoprecipitation (ChIP-seq) on male and female early embryos targeting a repressive histone mark associated with transposable elements (H3K9me3), and showed that heterochromatin formation is delayed in early male Drosophila embryos. This coincides with the increase in repeat expression in male versus female embryos, and more *de novo* insertions of repeats in males. Thus, the Y chromosome may indirectly create a mutational burden in males by postponing the establishment of silencing chromatin marks genome-wide, and permitting expression and mobilization of transposable elements during early embryogenesis.

Piwi-interacting RNAs (piRNAs) are short (23-29 nt) RNAs that bind to PIWI proteins and direct post-transcriptional and transcriptional silencing of target transposons. Many piRNAs are sense or antisense to repeats, and are produced by loci called piRNA clusters. These clusters often contain fragments of many different repeats and act as traps for transposable elements, and piRNAs produced from these clusters use sequence homology to target transcripts of transposable elements for degradation, and induce transcriptional silencing of repeats through heterochromatin formation. Repeat content evolves quickly across species, but less is known how piRNA clusters co-evolve. I sequenced piRNAs from *D. melanogaster*, *D. pseudoobscura*, and *D. miranda* ovaries, testes, and 0-1 hour embryos, to identify and characterize piRNAs and piRNA clusters in these three species. Comparison of piRNA clusters reveals a dynamic picture of their evolution. We identify many piRNA clusters that are unique to a species, or that have newly emerged on the repeat-rich neo-Y chromosome of *D. miranda*. We also find several homologous clusters that are syntenic between *D. miranda* and *D. pseudoobscura*, but whose sequence composition and repeat content has dramatically changed between species. Thus, piRNA clusters co-evolve with the rapidly changing repeat content of species, both by *de novo* formation of clusters, and by turn-over of the repeats contained within a cluster.

DEDICATION

To Jud, Mom, and Dad.

Acknowledgements

Funding for this work was provided by Doris Bachtrog, the National Institutes of Health, and UC Berkeley Department of Integrative Biology.

Thanks to my PI, Professor Doris Bachtrog.

Thanks to my fellow members of the Bachtrog Lab.

Thanks to my committee members, Professor Rasmus Nielsen and Professor Donald Rio.

Thanks to my qualifying exam committee, Professor Michael Eisen, Professor Michael Nachman, Professor Rasmus Nielsen, and Professor Donald Rio.

Thanks to Shivani Mahajan, for your support academically and for your friendship.

Thanks to Dat Mai, for your patience and help with all things computational.

Thanks to Zaak Walton and Carolus Chan, for your help in the wet lab.

Thanks to Karen Lundy at the Functional Genomics Laboratory, Shana McDevitt at the Vincent J. Coates Genomic Sequencing Laboratory, and Lydia Smith at the Evolutionary Genetics Lab.

Thanks to John, for your mentorship and friendship. You are missed.

Thanks to Amanda, for your friendship.

Thanks to Mom and Dad, for everything.

Thanks to Ashby, for your love, for your morning snuggles, and for greeting me with excitement every day when I get home.

Thanks to my fiancé, Jud, for moving to the Bay Area with me, for your support, for your love, and for the joy you bring into my life. I can't wait to marry you.

Chapter 1: Alternative Splicing Within and Between *Drosophila* Species, Sexes, Tissues, and Developmental Stages

Originally published as Gibilisco, L., Zhou, Q., Mahajan, S. and Bachtrog, D., 2016. Alternative splicing within and between Drosophila species, sexes, tissues, and developmental stages. *PLoS genetics*, *12*(12), p.e1006464.

Abstract

Alternative pre-mRNA splicing (“AS”) greatly expands proteome diversity, but little is known about the evolutionary landscape of AS in *Drosophila* and how it differs between embryonic and adult stages or males and females. Here we study the transcriptomes from several tissues and developmental stages in males and females from four species across the *Drosophila* genus. We find that 20-37% of multi-exon genes are alternatively spliced. While males generally express a larger number of genes, AS is more prevalent in females, suggesting that the sexes adopt different expression strategies for their specialized functions. While the number of total genes expressed increases during early embryonic development, the proportion of expressed genes that are alternatively spliced is highest in the very early embryo, before the onset of zygotic transcription. This indicates that females deposit a diversity of isoforms into the egg, consistent with abundant AS found in ovary. Cluster analysis by gene expression (“GE”) levels shows mostly stage-specific clustering in embryonic samples and tissue-specific clustering in adult tissues. Clustering embryonic stages and adult tissues based on AS profiles results in stronger species-specific clustering, suggesting that diversification of splicing contributes to lineage-specific evolution in *Drosophila*. Most sex-biased AS found in flies is due to AS in gonads, with little sex-specific splicing in somatic tissues.

Summary

Alternative pre-mRNA splicing (“AS”) greatly expands proteome diversity within and between species by creating different combinations of exons from the same genomic loci. Recent comparisons of transcriptomes of equivalent adult organs in several vertebrate species revealed that AS differs significantly in its complexity across tetrapods and may contribute to lineage-specific adaptation in vertebrates. AS is also prevalent in *Drosophila* and has been most extensively studied in *D. melanogaster*. However, the spatial and temporal evolution of AS between *Drosophila* species remains to be explored. Here we study the transcriptomes from several tissues and developmental stages in males and females from four *Drosophila* species, spanning a major phylogenetic range of *Drosophila*. We show that while males generally express a larger number of genes, AS is more prevalent in females. This suggests that the sexes adopt different expression strategies for their specialized functions – i.e. males tend to employ different genes, while females tend to utilize different isoforms. We also find that the dynamics of AS change over early development; while the number of total genes expressed increases during early embryonic development, the proportion of expressed genes that are alternatively spliced is highest in the very early embryo, before the onset of zygotic transcription. This indicates that females deposit a diversity of isoforms into the egg, consistent with abundant AS found in ovary. Like in mammals, cluster analysis suggests that diversification of splicing also significantly contributes to lineage-specific adaptation within the *Drosophila* genus.

Introduction

Alternative pre-mRNA splicing ("AS") greatly expands proteome diversity within species by creating different combinations of exons from the same genomic loci (Nilsen & Graveley, 2010). The resulting mRNA isoforms are usually expressed in a tissue or developmental stage- specific manner and underlie numerous essential biological processes like sex determination (Cline and & Meyer, 1996; Schutt & Nothiger, 2000), tissue development (Wang et al., 2008a), and stress response (Staiger & Brown, 2013). AS can also greatly increase proteome diversity between species with similar repertoires of protein-coding genes (Barbosa-Morais et al., 2012; Merkin, Russell, Chen, & Burge, 2012). Given the correlation of AS with the evolution of organismal complexity (Nilsen & Graveley, 2010), its dynamics across developmental stages, tissues, and species have attracted great attention (Barbosa-Morais et al., 2012; Merkin et al., 2012). Recent comparisons of transcriptomes of equivalent adult organs in several vertebrate species revealed that AS differs significantly in its complexity across the studied tetrapods, with primates showing the highest abundance of AS events (Barbosa-Morais et al., 2012; Merkin et al., 2012). Interestingly, AS shows a greater level of interspecific divergence and lineage-specific turnover across tissues than absolute gene expression levels, suggesting that the diversification of splicing significantly contributes to lineage-specific adaptation (Barbosa-Morais et al., 2012; Merkin et al., 2012).

AS is also prevalent in *Drosophila*, and has been most extensively studied in *D. melanogaster* (Brooks et al., 2011; Chang, Dunham, Nuzhdin, & Arbeitman, 2011; Cherbas et al., 2011; Daines et al., 2011; Gan et al., 2010; Hartmann et al., 2011; McIntyre et al., 2006; Telonis-Scott, Kopp, Wayne, Nuzhdin, & McIntyre, 2009; Venables, Tazi, & Juge, 2012). Over half of the spliced *D. melanogaster* genes encode two or more transcript isoforms, with about 50 genes capable of encoding over 1000 transcript isoforms each (Brown et al., 2014). The complexity of AS also differs across developmental stages and tissues (Duff et al., 2015; Graveley et al., 2011) and different environmental perturbations (Brown et al., 2014). However, the spatial and temporal evolution of AS between *Drosophila* species remains to be explored. Here we study the transcriptomes from several tissues and developmental stages in males and females from four *Drosophila* species, spanning a major phylogenetic range of *Drosophila*. In particular, we analyze two species pairs that diverged from each other at varying evolutionary distances: *D. pseudoobscura* and its sister species

D. miranda split about 2 million years ("MY") ago (Bachtrog & Charlesworth, 2002), and they diverged from *D. melanogaster* roughly 25 MY ago (R. Robinson, 2012); *D. nasuta* and *D. albomicans* split only about 0.1 MY ago (Bachtrog, 2006), and diverged from *D. melanogaster* over 60 MY ago (Tamura, Subramanian, & Kumar, 2004). The presence of genomic resources for these species (Lott, Villalta, Zhou, Bachtrog, & Eisen, 2014; Q. Zhou et al., 2013; Q. Zhou & Bachtrog, 2012; Q. Zhou et al., 2012) combined with their varying split times allows us to study the evolution of AS on different time scales in different lineages and enables us to address novel aspects of transcriptome diversity, such as the evolutionary landscape of AS in *Drosophila* and how it differs between embryonic and adult stages and males and females.

Results

Transcriptome diversity across species

We used RNA-seq data for different tissues in *D. pseudoobscura*, *D. miranda*, *D. albomicans,* and *D. nasuta* from males and females (one female and one male larval stage and five female and five male adult samples; **Table S1**). In addition, we used published data derived from sexed early embryonic stages (eight female and eight male embryonic stages) in *D. pseudoobscura* and *D. miranda* spanning the onset of zygotic transcription (Lott et al., 2014). We analyzed a total of 144 samples from different tissues or developmental stages, summing up to a total of 6,453,796,999 reads. An overview of the data used for each species is displayed in **Figure 1**.

The large number of stages and tissues allowed us to comprehensively annotate the transcriptomes of the four species, and a summary of the gene annotations and AS events is outlined in **Figure 1**. We considered skipped exons, alternative 5' and 3' splice sites, mutually exclusive exons, and retained introns. We annotated between 15,357 and 19,007 genes and between 66,747 and 84,491 exons for each species (**Figure 1**). We find that between 20 and 37% of all multi-exon genes are alternatively spliced in at least one tissue or stage and between 5,084 and 10,172 exons (8-12% of all exons) are alternatively spliced (**Figure 1**). These values are similar to reports in *D. melanogaster* (where 31% of genes were found to be alternatively spliced; (Daines et al., 2011)), but considerably lower than the degree of AS in mammals, where almost 100% of multi-exon genes are alternatively spliced (Merkin et al., 2012). Less AS in *Drosophila* is consistent with findings that species further from primates have lower proportions of exons undergoing AS (Barbosa-Morais et al., 2012).

As expected, we detected more genes, exons, and AS events for the two species for which we had both more comprehensive sampling (i.e. more samples and RNA-seq reads) and higher quality genome assemblies (i.e., *D. pseudoobscura* and *D. miranda*). Thus, while differences in the numbers of genomic features detected may in part reflect real species differences, less power to annotate genes, exons, and AS events in *D. albomicans* and *D. nasuta* probably largely contributes to these differences. Note that genomic features shown in **Figure 1** represent our within-species annotations (see **Methods**), and differences in these numbers between species should not greatly affect our comparisons of different tissues/samples within single species. For interspecific comparisons, we focused our analysis on a subset of expressed genes and exons for which we could identify orthologs from all four species (see **Methods)**.

Skipped exons are the most abundant AS event in all four species, and mutually exclusive exons are the least abundant, consistent with previous studies in *D. melanogaster* (Daines et al., 2011; McManus, Coolon, Eipper-Mains, Wittkopp, & Graveley, 2014). The relative percentage of each type of AS is generally similar between the studied *Drosophila* species (**Figure 1**), despite their evolutionary distance or different numbers of tissues and developmental stages sampled, indicating a high level of conservation of AS composition.

Temporal and spatial expression dynamics within species

We examined the transcriptome composition across corresponding samples of the four species using both the abundance of expressed genes and abundance of different alternatively spliced exons. Consistent with previous findings (Zhang, Sturgill, Parisi, Kumar, & Oliver, 2007), males generally express more genes than females across almost all tissues (**Figure 2A**, left panel). For example, 63-80% of all genes are expressed in male whole body while only 47-64% of all genes are expressed in female whole body. The largest discrepancy in the number of expressed genes between sexes is found in adult gonads, with testis expressing 1.6-1.7x more genes than ovary (58-76% of all genes are expressed in testis while only 35-47% of all genes are expressed in ovary). That, however, does not necessarily mean ovary or female tissues have lower transcriptome diversity. While female samples usually show fewer expressed genes, the fraction of expressed genes that are alternatively spliced is generally higher in females than in males (**Figure 2A**, right panel). Ovary and spermatheca have the highest percentage of expressed genes annotated as alternatively spliced (24-45% in ovary and 24-40% in spermatheca; **Figure 2A**, right panel) despite showing among the lowest percentage of expressed genes. This indicates that male and female reproductive tissues may adopt different expression strategies for their specialized functions: male tissues increase their transcriptome diversity by expressing different types of genes, while female tissues rely more on AS to increase the number of transcripts.

Across the four species, between 54-71% of all genes are expressed in head, and head generally shows relatively high proportions of all annotated alternative exons expressed (8-13%; **Figure 2A**, middle panel). This is consistent with earlier studies in *Drosophila* and humans, which show high transcriptional diversity in brain (Mohr & Hartmann, 2014; Wang et al., 2008b). Note that while the general trends reported above hold for all four species, exact proportions of genes and exons expressed and alternatively spliced vary between specific tissues and sexes. Differences in genome assembly quality and sampling among the four species, as well as species-specific idiosyncrasies, probably contribute to these differences.

In addition, we also analyzed developmental time course data in *D. miranda* and *D. pseudoobscura*, focusing on early embryonic development. The analyzed data encompass the maternal to zygotic transition (which happens during stage 5, see **Figure 2B**), when maternal transcripts begin to degrade and widespread zygotic transcription is initiated (Lott et al., 2014). This allows us to contrast the landscape of AS between genes contributed maternally with those transcribed in the early embryo. Over development, both the percentage of total genes expressed (**Figure 2B**, left panel) and the percentage of alternatively spliced exons expressed (**Figure 2B**, middle panel) increase as development proceeds in both males and females. However, the proportion of expressed genes that are alternatively spliced is the highest in early embryonic development, before the onset of zygotic transcription, and drops in later stages as single-transcript gene expression increases (**Figure 2B**, right panel). This indicates that the mother deposits a diversity of isoforms into the egg and that early zygotic transcription increases the number of genes expressed, but most of those genes do not encode multiple isoforms. Many genes annotated in *D. melanogaster* as maternally deposited (vs. zygotically expressed or both maternal and zygotic) (Lott et al., 2011) such as *bbc*, a phosphotransferase, and *Dhc64C*, a gene with ATPase activity, have multiple maternally deposited isoforms, as AS is detected before zygotic expression begins in embryos of *D. pseudoobscura* and *D. miranda*.

Of maternal, zygotic, and both maternal and zygotic genes defined in *D. melanogaster* (Lott et al., 2011) for which we recovered orthologs in all four species, zygotic genes show the lowest proportion of AS. However, genes annotated as being in one category in *D. melanogaster* are not necessarily in the same category in other species (e.g. a maternal gene in *D. melanogaster* may be maternal + zygotic in *D. pseudoobscura*). We therefore simply distinguished between genes that are maternally deposited (and that may or may not also show zygotic expression) as genes expressed at developmental stage 2 and zygotic genes (genes not expressed at stage 2 but expressed at later stages of embryonic development) for which we recovered orthologs in *D. pseudoobscura* and *D. miranda*. Maternally deposited genes show a higher proportion of AS than zygotic genes in both species (*D. miranda:* 180 of 2173 (8.3%) maternally deposited genes alternatively spliced vs. 7 of 647 (1.1%) zygotic genes alternatively spliced; *D. pseudoobscura*: 159 of 2149 (7.4%) maternally deposited genes alternatively spliced vs. 2 of 704 (0.3%) zygotic genes alternatively spliced). Thus, maternally deposited mRNA may comprise higher transcriptome diversity than previously appreciated, consistent with the abundance of alternatively spliced transcripts that we detect in ovaries. Among the zygotic genes, we confirmed sex-specific AS events of the *Sxl* gene, the master sex determining gene of *Drosophila*, during early embryogenesis in *D. pseudoobscura* and *D. miranda* (see below).

PCA analysis of splicing over embryonic development shows that different embryonic stages have distinct splicing profiles, which form a clock-like pattern in the PCA plot corresponding to the developmental time course (**Figure 3**; shown is *D. pseudoobscura*, and similar trends are seen for the other species; see **Figures S1**-**S3**). This differs from gene expression, which does not differentiate the different embryonic stages to the same degree as splicing (**Figure 3**). Remarkably, in PCAs based on AS profiles, ovary is the closest of all post-embryonic tissues to embryonic stages while PCAs based on gene expression ("GE") do not show ovary as being particularly close to embryonic stages. This together with the observation that both ovary and prezygotic embryos express few genes but have a great proportion of alternatively spliced transcripts (**Figure 2**) suggests that AS of maternally deposited transcripts plays an important role in early embryonic development.

Temporal and spatial expression dynamics across species

We compared gene expression levels and AS profiles across developmental stages (between *D. pseudoobscura* and *D. miranda*) and tissues (among *D. pseudoobscura*, *D. miranda*, *D. albomicans*, and *D. nasuta*). We recovered 3,005 orthologous genes among all four species (and 6,707 between *D. pseudoobscura* and *D. miranda*) and 472 orthologous exons expressed and annotated as alternatively spliced in at least one sample in all four species (and 1,122 between *D. pseudoobscura* and *D. miranda*). Recovery of fewer genes and exons among all four species compared to just *D. pseudoobscura* and *D. miranda* is expected, due to greater divergence times and less power to identify genomic features in species with lower quality genome assemblies and fewer RNA-seq data (i.e. *D. albomicans* and *D. nasuta*, see **Table S1**). As a measure of gene expression for a gene in the interspecific comparisons, we used TPM values (Transcripts Per Million), while AS was quantified using Ψ ("PSI", Percent Spliced In), the proportion of isoforms containing an alternatively spliced exon (or retained intron).

When clustering adult samples on the basis of how gene expression levels correlate in pairwise comparisons, there is no strong species-specific clustering, and clustering is almost completely tissue-specific (**Figure 4A**). Clustering embryonic stages based on gene expression levels (**Figure 4B**) results in stage-specific clustering across embryogenesis, and within broad groups of stages, species-specific clustering. Gene expression during early embryogenesis is conserved over particular groups of developmental stages, and in adults, tissue-specific expression patterns dominate.

When clustering adult AS profiles using Ψ, species-based clustering tends to be stronger than tissue-based clustering (**Figure 4C**). During embryonic development, samples segregate broadly by developmental stage: prezygotic (2 and 4) and postzygotic (mid stage 5 – stage 12) stages cluster, and within those categories, the samples cluster by species (**Figure 4D**). The more tissue-based clustering in adults observed for gene expression and the stronger species-based clustering in all samples based on splicing (for gene expression, stages 2 & 4 and stages 10 & 12 cluster by developmental stage, while only the prezygotic stages 2 and 4 cluster for splicing) is consistent with the strong species-specific clustering observed for splicing and “tissue-dominated clustering” observed for gene expression among vertebrates (Merkin et al., 2012). A comparison between gene expression and splicing has not been done over embryonic development in vertebrates, and it will be of great interest to see if embryogenesis in vertebrates follows the same patterns we see in *Drosophila*.

Sex-biased and sex-specific alternative splicing

AS mediates sex determination in Drosophila, and our RNA-Seq data confirm sex-biased splicing of genes involved in sex determination. For example, we detect one of the exons included in males and spliced out in females for *Sxl* in the two *Drosophila* species for which we have developmental expression data (**Figure S4**). We used ΔΨ values ($\Delta\Psi = |\Psi_{\text{female}} - \Psi_{\text{male}}|$) to assess sex-biased AS in various male and female tissues and stages (**Figure 5**), and find sex-biased exons from genes previously observed to have sex-biased isoforms (Chang et al., 2011), such as the male-biased isoform of *thin*, a gene involved in protein ubiquitination. Proportionally, gonads show more pronounced sex-biased splicing, and the splicing pattern for gonadectomized whole flies is skewed towards weaker sex-biased splicing, for all species (**Figures 5**, **S5**-**S7)**. For example, in *D. pseudoobscura* carcasses, 8.5% of alternatively spliced exons are strongly sex-biased ($\Delta\Psi > 0.7$) while in *D. pseudoobscura* gonads (ovary and testis), 14.8% of alternatively spliced exons are strongly sex-biased (**Figure 5**). Similar patterns are also seen for the other three species, with alternatively spliced exons being more strongly sex-biased in gonads (11.5-18.3%) than in carcasses (7.6-16.4%; see **Figures S5**-**S7** for ΔΨ analyses for other species). As mentioned, heads show the highest number of AS events, but most of them show only very weak sex-bias. For example, in *D. pseudoobscura* heads, only 56.0% of sex-biased exons have a $\Delta\Psi > 0.10$, while in gonads 73.6% of sex-biased exons have a $\Delta\Psi > 0.10$ (**Figure 5**). Conversely, more *D. pseudoobscura* exons are strongly sex-biased ($\Delta\Psi > 0.7$) in gonads (14.8%) than in heads (6.1%). Similar patterns are apparent in the other three species: within *D. miranda* and *D. nasuta*, more exons have a $\Delta\Psi > 0.10$ in gonads (77.5% and 74.1%, respectively) than in heads (61.9% and 64.3%), while those fractions are similar for *D. albomicans* (70.8% in gonads vs. 71.4% in heads). However, for all three species, gonads have a higher proportion of strongly sex-biased ($\Delta\Psi > 0.7$) exons (11.5-18.3%) than heads (4.78%-6.9%) (**Figures S5**-**S7**). Thus, most of the sex-biased AS found in flies can be attributed to AS in gonads.

Discussion

Consistent with previous studies, we find that AS significantly contributes to increasing the transcriptome diversity in all *Drosophila* species examined, and approximately 40% of genes are alternatively spliced. In every species studied, we find that head tissue harbors the largest number of alternatively spliced exons, and the highest fraction of expressed genes that are alternatively spliced is found in ovary. High transcriptional diversity has also been reported in brain tissues of mammals and vertebrates (Barbosa-Morais et al., 2012; Merkin et al., 2012; Mohr & Hartmann, 2014; Wang et al., 2008b), and it will be interesting to see how ovarian transcriptional diversity compares to head in other species as well.

While males generally show a higher number of genes expressed than females, female-specific tissues (ovary and spermatheca) have the highest percentage of alternatively spliced genes. Thus, male tissues may increase their transcriptome diversity by expressing more genes, while female tissues increase their number of transcripts through AS.

Gene expression levels have been found to cluster by tissue across different mammalian and vertebrate species (Barbosa-Morais et al., 2012; Merkin et al., 2012). Our analysis of *Drosophila* adult tissues reveals similar strong tissue-specific clustering, with only one exception (*D. nasuta* accessory gland clustering with *D. nasuta* spermatheca). This is interesting because the mammalian and vertebrate studies mostly examined somatic tissues (such as liver or kidney) with the exception of testis. In contrast, most of our samples (with the exception of male and female head) contain gonad tissue (ovary and testis) or sex-specific tissues (spermatheca and accessory gland), which have been shown to be rapidly evolving (Begun, Whitley, Todd, Waldrip-Dail, & Clark, 2000; Grassa & Kulathinal, 2011; Prokupek et al., 2008; Swanson & Vacquier, 2002; Tsaur & Wu, 1997). However, while sex-specific tissues may evolve more rapidly than somatic tissues, our analyses show that gene expression profiles of those tissues are conserved among species. Note that the vertebrate studies similarly compare species with highly varying divergence times (between 6 and 350 MY (Barbosa-Morais et al., 2012)) as our *Drosophila* study does (between 0.1 (Bachtrog, 2006) and 60 MY (Tamura et al., 2004)). None of the evolutionary patterns reported differ between distantly and closely related species – i.e. species- vs. tissue-specific clustering for AS vs. gene expression does not differ between closely vs. distantly related species (see **Figure 4**). This, divergence time does not appear to affect our analysis.

Similar to the findings in mammals and vertebrates, we see strong species-specific clustering for AS in *Drosophila*. One caveat is that the one purely somatic tissue, head, clusters by tissue. This could mean that AS in gonad tissue- and sex-specific tissues may be evolving more rapidly, and AS may be more conserved in somatic tissue. Also, the lack of lineage-specific clustering in head may be due to shorter divergence times in flies (though the number of generations is probably higher). Our interspecies AS analyses focused on all exons annotated as alternatively spliced in at least one sample (for adult tissue splicing analyses, n = 472 exons). However, we only recovered 49 exons alternatively spliced in at least one sample in all four species and 330 alternatively spliced exons unique to a species. Note that we used stringent cut-offs to identify orthologous alternatively spliced exons, that is, all exons in our interspecific analyses have an FPKM > 1 in at least one sample per species. While it is possible that we may miss some lowly expressed shared alternatively spliced exons, this suggests that many exons in flies that are alternatively spliced in one species are constitutive in the others. Species-specific splicing differences in *Drosophila* therefore may be based mostly on the binary category of whether an exon is alternatively spliced or constitutive rather than on differences in Ψ values of orthologous exons among species.

While species-specific clustering for AS is consistent with lineage-specific adaptive evolution (Barbosa-Morais et al., 2012; Merkin et al., 2012), it may also support the hypothesis that much splicing is due to erroneous splice site choice, producing non-functional isoforms targeted for degradation/nonsense-mediated decay (Lareau, Brooks, Soergel, Meng, & Brenner, 2007; Pickrell, Pai, Gilad, & Pritchard, 2010). These presumably deleterious splicing events would therefore be unlikely to be evolutionarily conserved among species.

We find that when comparing all tissues, tissue differences in gene expression patterns in adults are more dominant than sex differences; for example, male heads are closer to female heads than other male tissues. Note that head is a composite structure, but this finding is consistent with previous studies that found sex differences in *D. melanogaster* brain gene expression to be low (Catalán, Hutter, & Parsch, 2012). When restricting our analysis to sex-specific tissues (ovary, spermatheca, testis, accessory gland), we see clustering primarily by sex and within sex, clustering by tissue (**Figure S8**). We also find evidence for considerable sex-specific splicing; however, most of the extreme differences in splicing between sexes are due to differences in splicing profiles in gonads.

In the middle of embryonic development (mid stage 5 - late stage 8), stages cluster by species based both on gene expression and AS. After embryogenesis, gene expression becomes more conserved among tissues and species-specific clustering is less dominant. AS profiles cluster samples by species rather than tissue (with the exception of head) into adulthood, suggesting that AS has diverged more than gene expression levels among *Drosophila* species. Note that we only had embryonic data for two species, and only over the early stages of embryogenesis (i.e. roughly the first half of development), and it will be of great interest to see whether other species show similar patterns during early development.

We show that females deposit a diversity of isoforms into the egg, and maternally deposited genes show a higher proportion of AS than zygotic genes in both species investigated. This relates to a well-known property of early development of *D. melanogaster*, that the first zygotic transcripts tend to be short and lacking introns (Heyn et al., 2014). The shortness/low complexity of early transcripts is thought to reflect time constraints on producing longer (or spliced) transcripts in quickly dividing early embryos (Jeffares, Mourier, & Penny, 2006; Mourier & Jeffares, 2003) during development (Foe, Odell, & Edgar, 1993). Thus, longer and more complex zygotically expressed transcripts may not reach high levels of expression during short mitotic cycles due to the time required for full transcription and splicing. In *D. melanogaster*, zygotic genome activation starts from intronless genes (De Renzis, Elemento, Tavazoie, & Wieschaus, 2007) and RNA-seq coverage along transcripts show patterns consistent intron delay and inability to fully transcribe long transcripts (Artieri & Fraser, 2014). On the other hand, maternally deposited mRNAs may be especially long and complex (i.e. many introns and alternatively spliced exons) because zygotes may not be able to produce them early on (Artieri & Fraser, 2014).

Similar to findings in *D. melanogaster* (Artieri & Fraser, 2014; De Renzis et al., 2007), we do see evidence of shorter genes and fewer introns during early zygotic transcription. In particular, comparisons of the two earlier stages that consist of mostly maternally deposited transcripts (stage 2 and stage 4) with later embryonic stages where zygotic transcription is occurring (mid stage 5 – stage 12) show that transcripts present during the earlier stages are longer and have more exons than those expressed during later stages (**Figure S9**).

Methods

Data

The accession numbers for the previously publicly accessible RNA-seq datasets are listed in **Table S2**.

The remaining samples are RNA-seq datasets generated by us from *D. pseudoobscura* (MV25) (spermatheca, male and female 3rd instar larva), *D. miranda* (MSH22) (male and female head, spermatheca), *D. albomicans* (KM55) (male and female gonadectomized carcass, male and female head, ovary, spermatheca, testis, accessory gland, male and female 3rd instar larva), and *D. nasuta* (15112-1781.00) (male and female gonadectomized carcass, male and female head, ovary, spermatheca, testis, accessory gland, male and female 3rd instar larva). We extracted total RNA from whole body or dissected tissues (Qiagen) and prepared RNA-seq libraries following the standard Illumina protocol. Briefly, we used Dynal oligo(dT) beads (Invitrogen) to isolate poly(A) mRNA from the total RNA samples. We then fragmented the mRNA by using the RNA fragmentation kit from Ambion, followed by first- and second-strand cDNA synthesis using random hexamer primers (Invitrogen). We complemented the cDNA synthesis by an end repair reaction using T4 DNA polymerase and Klenow DNA polymerase for 30 min at 20°C. We then added a single A base to the cDNA molecules by using 3'-to-5' exonuclease and ligated the Illumina adapter.

The fragments were subjected to size selection on a 2% gel and purification (Qiagen). We finally amplified the cDNA fragments by PCR reaction and examined the libraries by Bioanalyzer (Aglient). Paired-end cDNA sequencing was performed on the Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

More information on how to access this data can be found in DATA ACCESS.

Within-species analyses

Within-species analyses are analyses where only samples within a species were directly compared (**Figures 1-3**, **5**, **S1**-**S7**, **S9A** and **S9B**, **S9E** and **S9F,** and **Table S4**). We used seqtk (https://aur.archlinux.org/packages/seqtk-git) to randomly choose paired-end RNA-seq reads so that within each species, each sample used had the same number of reads (**Table S1**). The pooled data from all four species had a total of 957,749,296 reads. For samples with read lengths > 50 bp, we used read quality data from FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) to determine cutoffs and trimmed the reads to 50 bp.

We used bowtie2-build and TopHat v2.0.13 (Trapnell, Pachter, & Salzberg, 2009) to map the reads to each genome (*D. pseudoobscura* Release 3.1 downloaded from http://www.flybase.org, *D. miranda* assembly (Q. Zhou et al., 2013), *D. albomicans* assembly (Q. Zhou et al., 2012), *D. nasuta* assembly, unpublished, **Table S3**) using the parameters --b2-sensitive, --coverage-search, and --microexon-search and then ran Cufflinks v2.1.1 (Trapnell et al., 2010) without a reference annotation on each sample. For each species, we used Cuffmerge to combine all of the Cufflinks-generated annotations to create a master annotation. We used Cuffdiff (Trapnell et al., 2013) with this master annotation to obtain expression data for each sample. We use the “commonly adopted arbitrary inclusion threshold” of FPKM = 1 (Younus et al., 2014) that is used in other *Drosophila* gene expression studies (Graveley et al., 2011).

We used picard v1.106 (http://picard.sourceforge.net) to obtain insert size mean and standard deviation for each sample. We used MATS (Park, Tokheim, Shen, & Xing, 2013; Shen et al., 2012) to on one hand detect and annotate putative AS events from our aligned RNA-seq data (bam files) and Cuffmerge annotations. MATS was also used to get Ψ (“PSI”, Percent Spliced In) values for each alternatively spliced exon by running pairwise comparisons between all samples within each species and taking an average Ψ value for each exon calculated using reads on target and junction counts, excluding samples for which the exon was not classified as alternatively spliced. The program was run using default parameters except setting the “-analysis” parameter to “P” for our paired-end sequencing reads. The anchor/overhang length was the TopHat2 default of 8 bp, so at least eight nucleotides had to map to each end of a given junction. The annotation of AS events by MATS included intron retention events that overlapped with splice sites of other annotated AS events. To validate our pipeline, we also tested a few tissues using an alternative pipeline. We used AltEventFinder (A. Zhou et al., 2012) to annotate skipped exon AS events and MISO (Katz, Wang, Airoldi, & Burge, 2010; Wang et al., 2008a) to compute Ψ values. **Table S4** shows correlations of the results of both pipelines for post-embryonic tissues in *D. miranda* computed in R using the *corr* function*.*

heatmap.2 was used in R to generate heatmaps, which uses *hclust* to cluster samples. The R function *prcomp* was used to perform the PCAs. The R function *wilcox.test* was used to perform Wilcoxon rank sum tests (**S9 Fig**).

Note that for analyses within species, we directly compare datasets subsampled down to the same number of reads (**Table S1**) that are mapped to the same genome assembly. Therefore, differences in genome assembly quality among the four genomes and the differences in the numbers of reads used among species are not expected to impact our within-species inferences.

Interspecific comparisons

Interspecific comparisons are analyses in which samples among different species were compared directly (**Figures 4**, **S8**, **S9C** and **D**, and **S10** and **S11**). Pairwise whole genome alignments for each pair of species were done using the software Mercator (https://www.biostat.wisc.edu/~cdewey/mercator/) (Dewey & Pachter, 2006) and MAVID (Bray & Pachter, 2004). First we used Mercator to build an orthology map for each pair of species. Then MAVID was used to perform global whole genome alignments. Finally, for each exon/gene in each species, the coordinates of the corresponding ortholog in the other species were determined using the sliceAlignment program (Mercator distribution).

Using pairwise alignments of *D. pseudoobscura*, *D. miranda*, *D. albomicans*, and *D. nasuta* to *D. melanogaster*, we kept genes/exons aligned with > 0.5 overlap between all pairs. We kept all genes/exons with 1:1 orthology. If there were multiple genes/exons from one species that aligned to the same *D. melanogaster* gene/exon, we kept the pair with the highest overlap score. If the highest overlap score was shared between the *D. melanogaster* gene/exon and more than one gene/exon from the other species, we did not use the gene(s)/exon(s) in our analysis. If there were multiple *D. melanogaster* genes/exons that aligned to the same gene/exon from the other species, we kept the pair with the highest overlap score. If the highest overlap score was shared between one gene/exon from the other species and more than one *D. melanogaster* gene/exon, we did not use the gene(s)/exon(s) in our analysis. For comparisons over embryonic development, this left us 6,707 genes with 1:1 orthologous relationships between *D. pseudoobscura* and *D. miranda* and 1,122 exons with 1:1 orthologous relationships between *D. pseudoobscura* and *D. miranda* that were also annotated as alternatively spliced in at least one sample and expressed with an FPKM > 1 in at least one sample per species. For comparisons of post-embryonic tissues, we recovered 3,005 genes with 1:1 orthologous relationships among *D. pseudoobscura*, *D. miranda*, *D. albomicans*, and *D. nasuta* and 472 exons with 1:1 orthologous relationships among the four species that were also annotated as alternatively spliced in at least one sample and expressed with an FPKM > 1 in at least one sample per species.

We used orthology information to create annotations for each species containing only genes orthologous among all species (*D. pseudoobscura*, *D. miranda*, *D. albomicans*, and *D. nasuta* for comparisons of post-embryonic samples; *D. pseudoobscura* and *D. miranda* for comparisons of embryonic samples) and ran kallisto (Bray, Pimentel, Melsted, & Pachter, 2015) using these annotations. We used TPM (Transcripts Per Million) values from each sample for each gene for our interspecies gene expression analysis. We used sleuth (Bray et al., 2015) to normalize TPM values between samples and to compute Jensen-Shannon divergence between samples.

We used bowtie2-build and TopHat v2.0.13 (Trapnell et al., 2009) to map the reads to each genome (*D. pseudoobscura* Release 3.1 downloaded from http://www.flybase.org, *D. miranda* assembly (Q. Zhou et al., 2013), *D. albomicans* assembly (Q. Zhou et al., 2012), *D. nasuta* assembly (unpublished, **Table S3**) using the parameters --b2-sensitive, --coverage-search, and --microexon-search and then ran Cufflinks v2.1.1 (Trapnell et al., 2010) using the –G parameter and the reference annotations used for orthology analysis on each sample.

For interspecies splicing analyses, we ran MATS (Park et al., 2013; Shen et al., 2012) to on one hand detect and annotate putative AS events from our aligned RNA-seq data (bam files) and orthology annotations. MATS was also used to get Ψ ("PSI", Percent Spliced In) values for each alternatively spliced exon. The program was run using the same parameters as for the "within-species" analysis. MATS compares splicing in samples pairwise, and we compared all samples within species that shared the same read lengths. Spermatheca reads from *D. nasuta* and *D. albomicans* were trimmed to 76 base pairs (the size of all other reads in these species). In each sample, we took an average Ψ value for each exon calculated using reads on target and junction counts, excluding samples for which the exon was not classified as alternatively spliced. For exons alternatively spliced in some samples but not others, we looked at the expression calculated for that exon in Cufflinks. If the exon had an FPKM value < 1 and the upstream or downstream exons had an FPKM value > 1, the exon was assigned a Ψ value of 0. If the exon had an FPKM value > 1, the exon was assigned a Ψ value of 1. If the exon had an FPKM value < 1 and the upstream and downstream exons had an FPKM value < 1, the exon was not assigned a Ψ value.

heatmap.2 was used in R to generate heatmaps, which uses *hclust* to cluster samples. We computed Spearman (**Figure 4** and **S8**) and Pearson (**Figure S10**) correlations for GE and AS in adult tissues and embryonic stages in R using the *corr* function. Jensen-Shannon divergence of GE (**Figure S11**) was computed using sleuth (Bray et al., 2015).

For analyses among species, we directly compare only orthologous genes and exons recovered in all four species. Therefore, our inferences are limited by the least complete genome assembly and annotation. We did not subsample datasets for our interspecies analyses (**Table S1**), but instead required that each orthologous gene or exon be expressed at FPKM > 1 in at least one sample per species to be included in our analyses. We consider our interspecies analysis conservative; we may be underestimating the number of orthologous genes and comparable splicing events. However, due to our approach, we do not expect a considerable number of splicing events or expressed genes to be erroneously identified as species-specific.

Data Availability

All sequencing reads generated in this study are deposited at the National Center for Biotechnology Information Short Reads Archive (www.ncbi.nlm.nih.gov/sra) under the BioProject ID PRJNA347569.

References

Artieri, C. G., & Fraser, H. B. (2014). Transcript length mediates developmental timing of gene expression across drosophila. *Molecular Biology and Evolution, 31*(11), 2879-2889. doi:10.1093/molbev/msu226 [doi]

Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Current Opinion in Genetics & Development, 16*(6), 578-585.

Bachtrog, D., & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature, 416*(6878), 323-326.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science (New York, N.Y.), 338*(6114), 1587-1593. doi:10.1126/science.1230612 [doi]

Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M., & Clark, A. G. (2000). Molecular population genetics of male accessory gland proteins in drosophila. *Genetics, 156*(4), 1879-1888.

Bray, N., Pimentel, H., Melsted, P., & Pachter, L. (2015). Near-optimal RNA-seq quantification. *arXiv Preprint arXiv:1505.02710,*

Bray, N., & Pachter, L. (2004). MAVID: Constrained ancestral alignment of multiple sequences. *Genome Research, 14*(4), 693-699. doi:10.1101/gr.1960404 [doi]

Brooks, A. N., Yang, L., Duff, M. O., Hansen, K. D., Park, J. W., Dudoit, S., et al. (2011). Conservation of an RNA regulatory map between drosophila and mammals. *Genome Research, 21*(2), 193-202. doi:10.1101/gr.108662.110 [doi]

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., et al. (2014). Diversity and dynamics of the drosophila transcriptome. *Nature, 512*(7515), 393-399.

Catalán, A., Hutter, S., & Parsch, J. (2012). Population and sex differences in drosophila melanogaster brain gene expression. *BMC Genomics, 13*(1), 1.

Chang, P. L., Dunham, J. P., Nuzhdin, S. V., & Arbeitman, M. N. (2011). Somatic sex-specific transcriptome differences in drosophila revealed by whole transcriptome sequencing. *BMC Genomics, 12*, 364-2164-12-364. doi:10.1186/1471-2164-12-364 [doi]

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B. D., et al. (2011). The transcriptional diversity of 25 drosophila cell lines. *Genome Research, 21*(2), 301-314. doi:10.1101/gr.112961.110 [doi]

Cline and, T. W., & Meyer, B. J. (1996). Vive la difference: Males vs females in flies vs worms. *Annual Review of Genetics, 30*(1), 637-702.

Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., et al. (2011). The drosophila melanogaster transcriptome by paired-end RNA sequencing. *Genome Research, 21*(2), 315-324. doi:10.1101/gr.107854.110 [doi]

De Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the drosophila embryo. *PLoS Biol, 5*(5), e117.

Dewey, C., & Pachter, L. (2006). MERCATOR: Multiple whole-genome orthology map construction. *Software Available at Http://Bio.Math.Berkeley.Edu/Mercator,*

Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., et al. (2015). Genome-wide identification of zero nucleotide recursive splicing in drosophila. *Nature, 521*(7552), 376-379.

Foe, V. E., Odell, G., & Edgar, B. (1993). Mitosis and morphogenesis in the drosophila embryo: Point and counterpoint. *The Development of Drosophila Melanogaster, 1*, 149-300.

Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K., et al. (2010). Dynamic regulation of alternative splicing and chromatin structure in drosophila gonads revealed by RNA-seq. *Cell Research, 20*(7), 763-783.

Grassa, C. J., & Kulathinal, R. J. (2011). Elevated evolutionary rates among functionally diverged reproductive genes across deep vertebrate lineages. *International Journal of Evolutionary Biology, 2011*, 274975. doi:10.4061/2011/274975 [doi]

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., et al. (2011). The developmental transcriptome of drosophila melanogaster. *Nature, 471*(7339), 473-479.

Hartmann, B., Castelo, R., Minana, B., Peden, E., Blanchette, M., Rio, D. C., et al. (2011). Distinct regulatory programs establish widespread sex-specific alternative splicing in drosophila melanogaster. *RNA (New York, N.Y.), 17*(3), 453-468. doi:10.1261/rna.2460411 [doi]

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., et al. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Reports, 6*(2), 285-292.

Jeffares, D. C., Mourier, T., & Penny, D. (2006). The biology of intron gain and loss. *TRENDS in Genetics, 22*(1), 16-22.

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods, 7*(12), 1009-1015.

Lareau, L. F., Brooks, A. N., Soergel, D. A., Meng, Q., & Brenner, S. E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. *Alternative Splicing in the Postgenomic Era, 623*

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A., & Eisen, M. B. (2011). Noncanonical compensation of zygotic X transcription in early drosophila melanogaster development revealed through single-embryo RNA-seq. *PLoS-Biology, 9*(2), 278.

Lott, S. E., Villalta, J. E., Zhou, Q., Bachtrog, D., & Eisen, M. B. (2014). Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. *PLoS Genetics, 10*(2), e1004159. doi:10.1371/journal.pgen.1004159 [doi]

McIntyre, L. M., Bono, L. M., Genissel, A., Westerman, R., Junk, D., Telonis-Scott, M., et al. (2006). Sex-specific expression of alternative transcripts in drosophila. *Genome Biology, 7*(8), R79.

McManus, C. J., Coolon, J. D., Eipper-Mains, J., Wittkopp, P. J., & Graveley, B. R. (2014). Evolution of splicing regulatory networks in drosophila. *Genome Research, 24*(5), 786-796. doi:10.1101/gr.161521.113 [doi]

Merkin, J., Russell, C., Chen, P., & Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science (New York, N.Y.), 338*(6114), 1593-1599. doi:10.1126/science.1228186 [doi]

Mohr, C., & Hartmann, B. (2014). Alternative splicing in drosophila neuronal development. *Journal of Neurogenetics, 28*(3-4), 199-215.

Mourier, T., & Jeffares, D. C. (2003). Eukaryotic intron loss. *Science (New York, N.Y.), 300*(5624), 1393. doi:10.1126/science.1080559 [doi]

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature, 463*(7280), 457-463.

Park, J. W., Tokheim, C., Shen, S., & Xing, Y. (2013). Identifying differential alternative splicing events from RNA sequencing data using RNASeq-MATS. *Deep sequencing data analysis* (pp. 171-179) Springer.

Pickrell, J. K., Pai, A. A., Gilad, Y., & Pritchard, J. K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet, 6*(12), e1001236.

Prokupek, A., Hoffmann, F., Eyun, S., Moriyama, E., Zhou, M., & Harshman, L. (2008). An evolutionary expressed sequence tag analysis of drosophila spermatheca genes. *Evolution, 62*(11), 2936-2947.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nature Biotechnology, 29*(1), 24-26.

Robinson, R. (2012). Natural selection molds genomic insulator elements.

Schutt, C., & Nothiger, R. (2000). Structure, function and evolution of sex-determining systems in dipteran insects. *Development (Cambridge, England), 127*(4), 667-677.

Shen, S., Park, J. W., Huang, J., Dittmar, K. A., Lu, Z. X., Zhou, Q., et al. (2012). MATS: A bayesian framework for flexible detection of differential alternative splicing from RNA-seq data. *Nucleic Acids Research, 40*(8), e61. doi:10.1093/nar/gkr1291 [doi]

Staiger, D., & Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *The Plant Cell, 25*(10), 3640-3656. doi:10.1105/tpc.113.113803 [doi]

Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature Reviews Genetics, 3*(2), 137-144.

Tamura, K., Subramanian, S., & Kumar, S. (2004). Temporal patterns of fruit fly (drosophila) evolution revealed by mutation clocks. *Molecular Biology and Evolution, 21*(1), 36-44. doi:10.1093/molbev/msg236 [doi]

Telonis-Scott, M., Kopp, A., Wayne, M. L., Nuzhdin, S. V., & McIntyre, L. M. (2009). Sex-specific splicing in drosophila: Widespread occurrence, tissue specificity and evolutionary conservation. *Genetics, 181*(2), 421-434. doi:10.1534/genetics.108.096743 [doi]

Thorvaldsdottir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics, 14*(2), 178-192. doi:10.1093/bib/bbs017 [doi]

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology, 31*(1), 46-53.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology, 28*(5), 511-515.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics (Oxford, England), 25*(9), 1105-1111. doi:10.1093/bioinformatics/btp120 [doi]

Tsaur, S. C., & Wu, C. I. (1997). Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of drosophila. *Molecular Biology and Evolution, 14*(5), 544-549.

Venables, J. P., Tazi, J., & Juge, F. (2012). Regulated functional alternative splicing in drosophila. *Nucleic Acids Research, 40*(1), 1-10. doi:10.1093/nar/gkr648 [doi]

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008a). Alternative isoform regulation in human tissue transcriptomes. *Nature, 456*(7221), 470-476.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008b). Alternative isoform regulation in human tissue transcriptomes. *Nature, 456*(7221), 470-476.

Younus, F., Chertemps, T., Pearce, S. L., Pandey, G., Bozzolan, F., Coppin, C. W., et al. (2014). Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of drosophila melanogaster. *Insect Biochemistry and Molecular Biology, 53*, 30-43.

Zhang, Y., Sturgill, D., Parisi, M., Kumar, S., & Oliver, B. (2007). Constraint and turnover in sex-biased gene expression in the genus drosophila. *Nature, 450*(7167), 233-237.

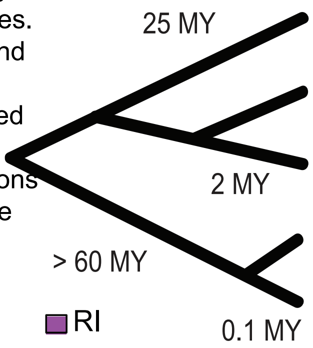
Zhou, Q., Ellison, C. E., Kaiser, V. B., Alekseyenko, A. A., Gorchakov, A. A., & Bachtrog, D. (2013). The epigenome of evolving drosophila neo-sex chromosomes: Dosage compensation and heterochromatin formation.

Zhou, A., Breese, M. R., Hao, Y., Edenberg, H. J., Li, L., Skaar, T. C., et al. (2012). Alt event finder: A tool for extracting alternative splicing events from RNA-seq data. *BMC Genomics, 13 Suppl 8*, S10-2164-13-S8-S10. Epub 2012 Dec 17. doi:10.1186/1471-2164-13-S8-S10 [doi]

Zhou, Q., & Bachtrog, D. (2012). Sex-specific adaptation drives early sex chromosome evolution in drosophila. *Science (New York, N.Y.), 337*(6092), 341-345. doi:10.1126/science.1225385 [doi]

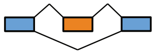
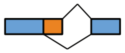
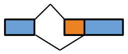
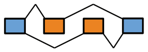
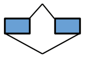
Zhou, Q., Zhu, H. M., Huang, Q. F., Zhao, L., Zhang, G. J., Roy, S. W., et al. (2012). Deciphering neo-sex and B chromosome evolution by the draft genome of drosophila albomicans. *BMC Genomics, 13*, 109-2164-13-109. doi:10.1186/1471-2164-13-109 [doi]

Figure 1. Overview of RNA-seq data used for analyses and alternative splicing in four *Drosophila* species. Top: Numbers of genes and exons detected and numbers of stages/ tissues for each of four *Drosophila* species. Bottom: Numbers of total, constitutive, and alternatively spliced exons, as well as a breakdown of types of alternatively spliced exons, for each species. Proportions of different types of alternatively spliced exons are shown as pie charts to the right of the table.



Species	Genes	Exons	Dev. Stages	Adult Tissues
D. melanogaster				
D. pseudoobscura	19,007	84,491	9	10
D. miranda	17,017	80,616	9	10
D. albomicans	15,737	66,747	1	10
D. nasuta	15,357	68,719	1	10

🟥SE 🟨A5SS 🟩A3SS 🟦MXE 🟪RI

Species	Total Exons	Constitutive Exons	Alternatively Spliced Exons					
			Total					
				Skipped Exons (“SE”)	Alternative 5’ Splice Sites (“A5SS”)	Alternative 3’ Splice Sites (“A3SS”)	Mutually Exclusive Exons (“MXE”)	Retained Introns (“RI”)
D. pseudoobscura (“pse”)	84,491	74,319	10,172	3,795	1,979	2,255	209	1,934
D. miranda (“mir”)	80,616	71,378	9,238	3,428	1,813	2,100	196	1,701
D. albomicans (“alb”)	66,747	61,663	5,084	1,933	924	1,452	215	560
D. nasuta (“nas”)	68,719	63,517	5,202	1,970	1,042	1,486	100	604

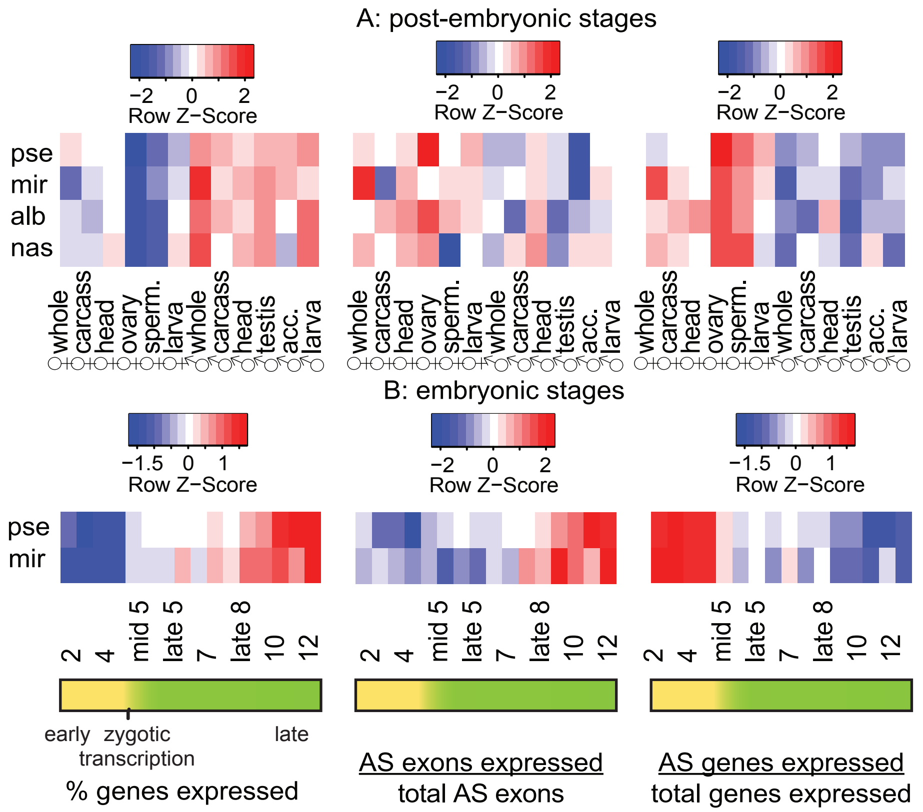


Figure 2. Gene expression and alternative splicing profiles across tissues and development. Comparisons of % of total genes expressed (left panel), % annotated alternatively spliced exons expressed (middle panel), and % of total genes expressed that are annotated as alternatively spliced (right panel) for post-embryonic tissues (A) and embryonic stages (B). Each row of each heatmap is scaled separately by Z-score.
"pse" = *D. pseudoobscura*; "mir" = *D. miranda*; "alb" = *D. albomicans*; "nas" = *D. nasuta*; "sperm." = spermatheca; "larva" = 3rd instar larva; "acc." = accessory gland

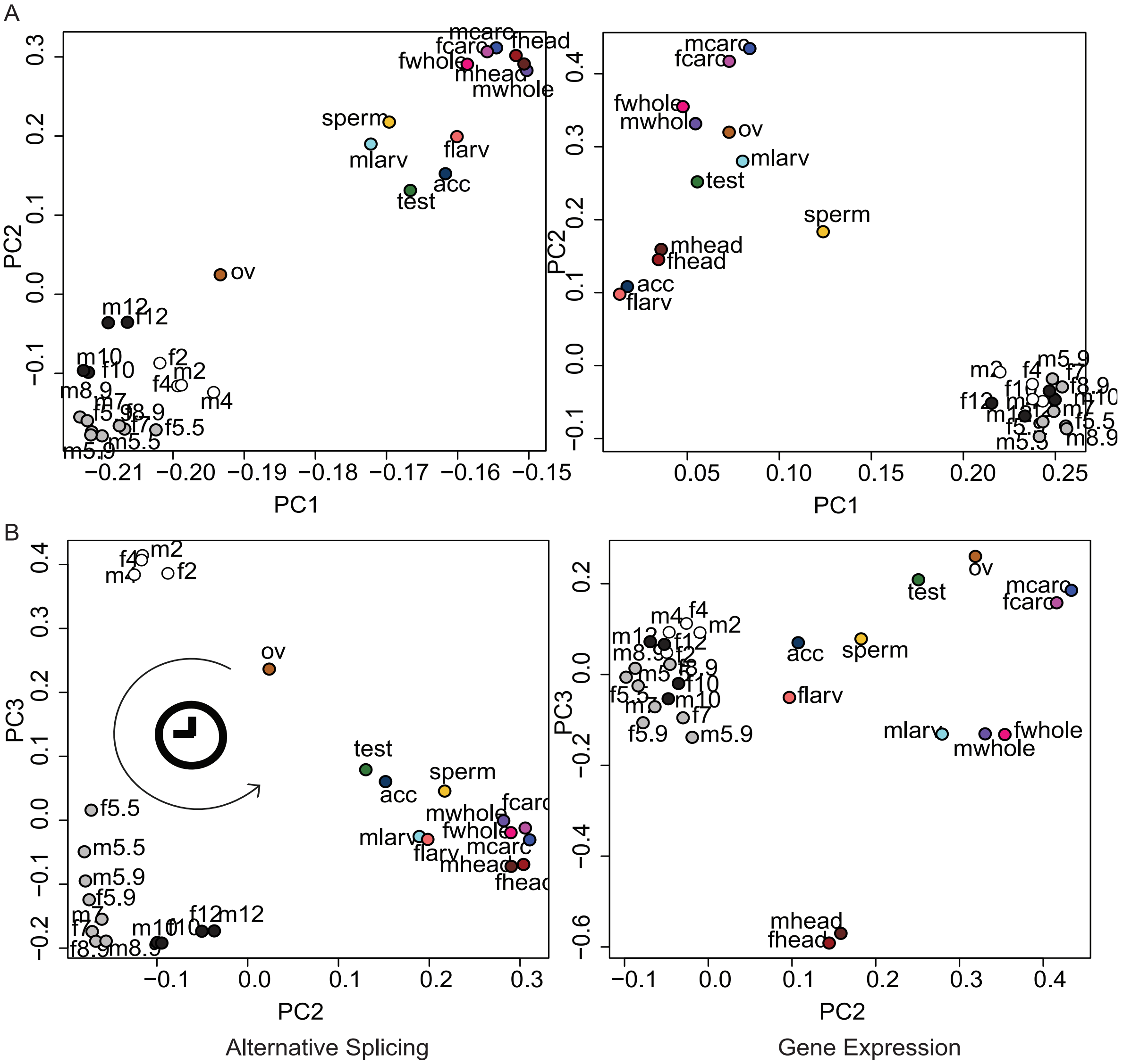


Figure 3. PCAs based on alternative splicing and gene expression. Alternative splicing (left column) and gene expression (right column) profiles for *D. pseudoobscura*. The R function *prcomp* was used to perform the PCAs. PC1 (AS: 57.9% of the variance & GE: 50.1% of the variance) and PC2 (AS: 10.7% of the variance & GE: 12.4% of the variance) (A). PC2 and PC3 (AS: 4.5% of the variance & GE: 7.5% of the variance) (B). “f” = female; “m” = male; “5.5” = mid stage 5; “5.9” = late stage 5; “8.9” = late stage 8; “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland

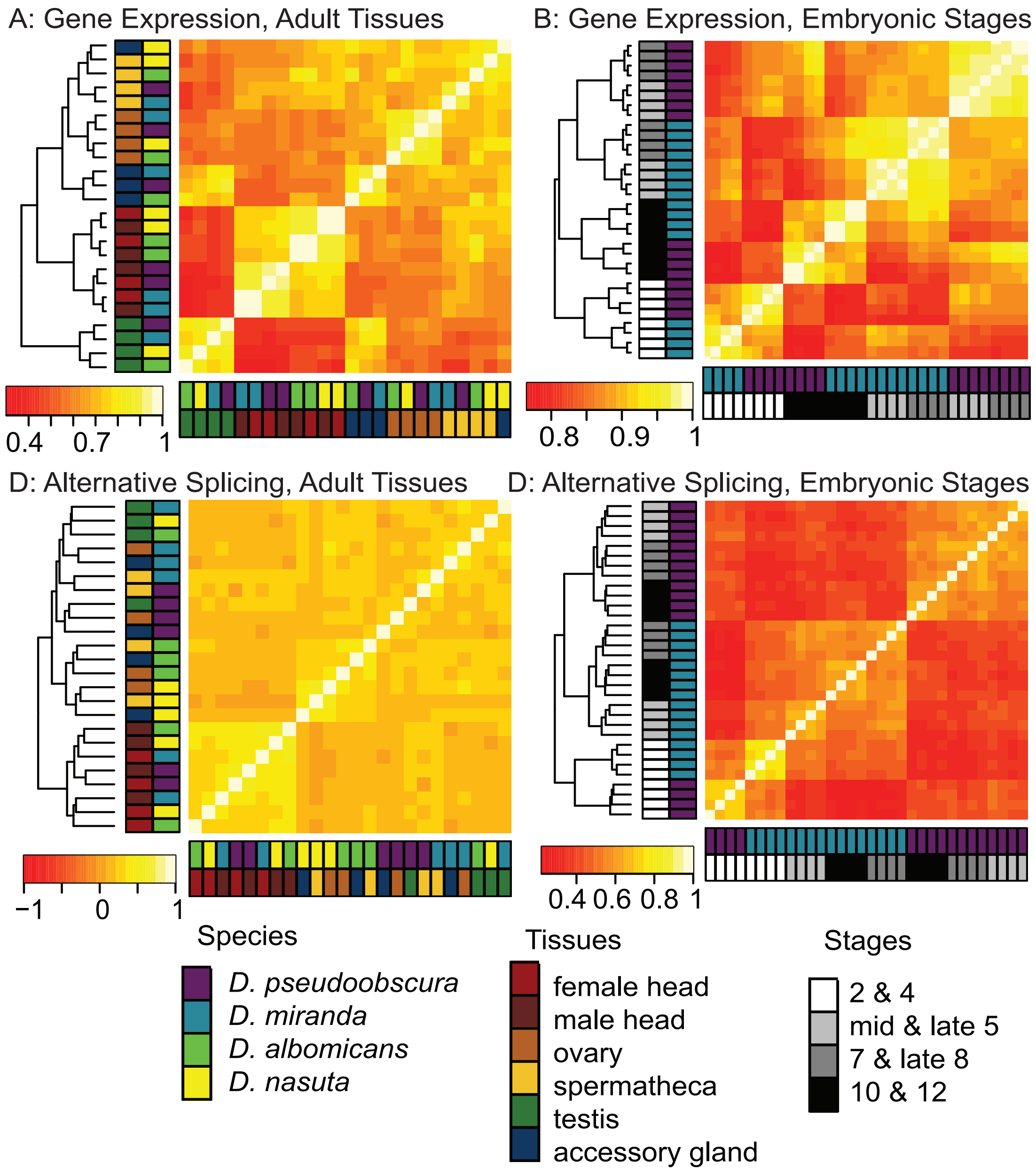


Figure 4. Correlations of gene expression versus alternative splicing. Spearman correlations based on gene expression (TPM) for orthologous genes in (A) adult tissues (n=3,005) and (B) embryonic stages (n=6,707) (mean of three replicates per sex/stage). Spearman correlations based on AS (Ψ) for orthologous exons (C) expressed in all species and annotated as alternatively spliced in at least one sample in adult tissues (n=472) and (D) expressed in both species and annotated as alternatively spliced in at least one sample in embryonic stages (n=1,122).

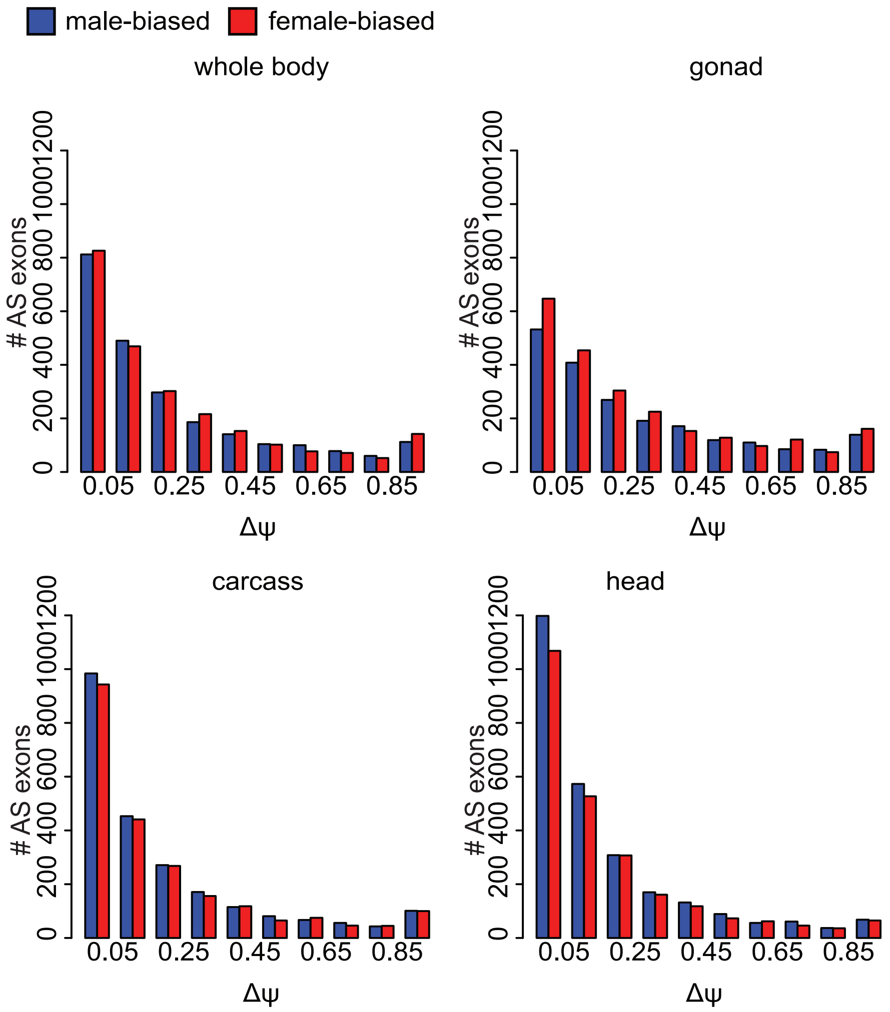
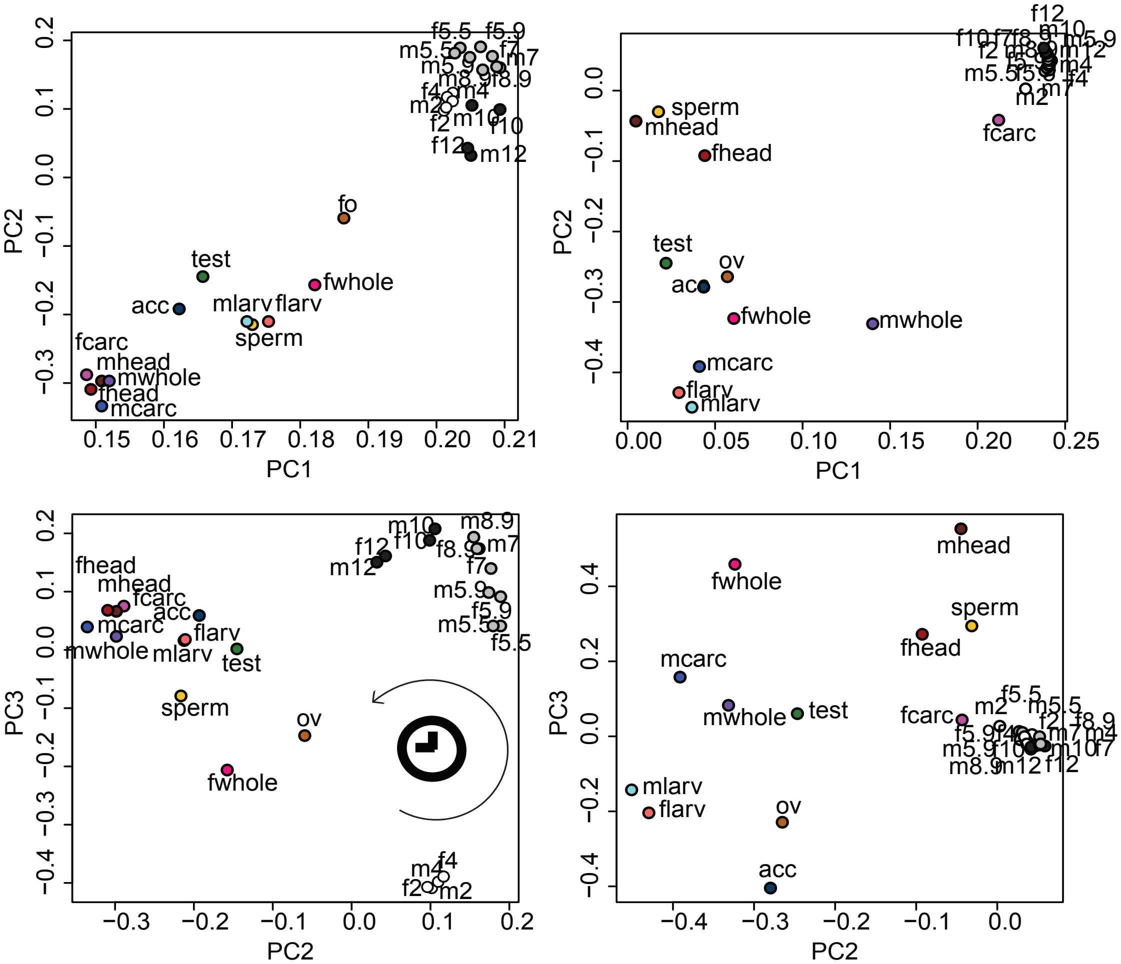


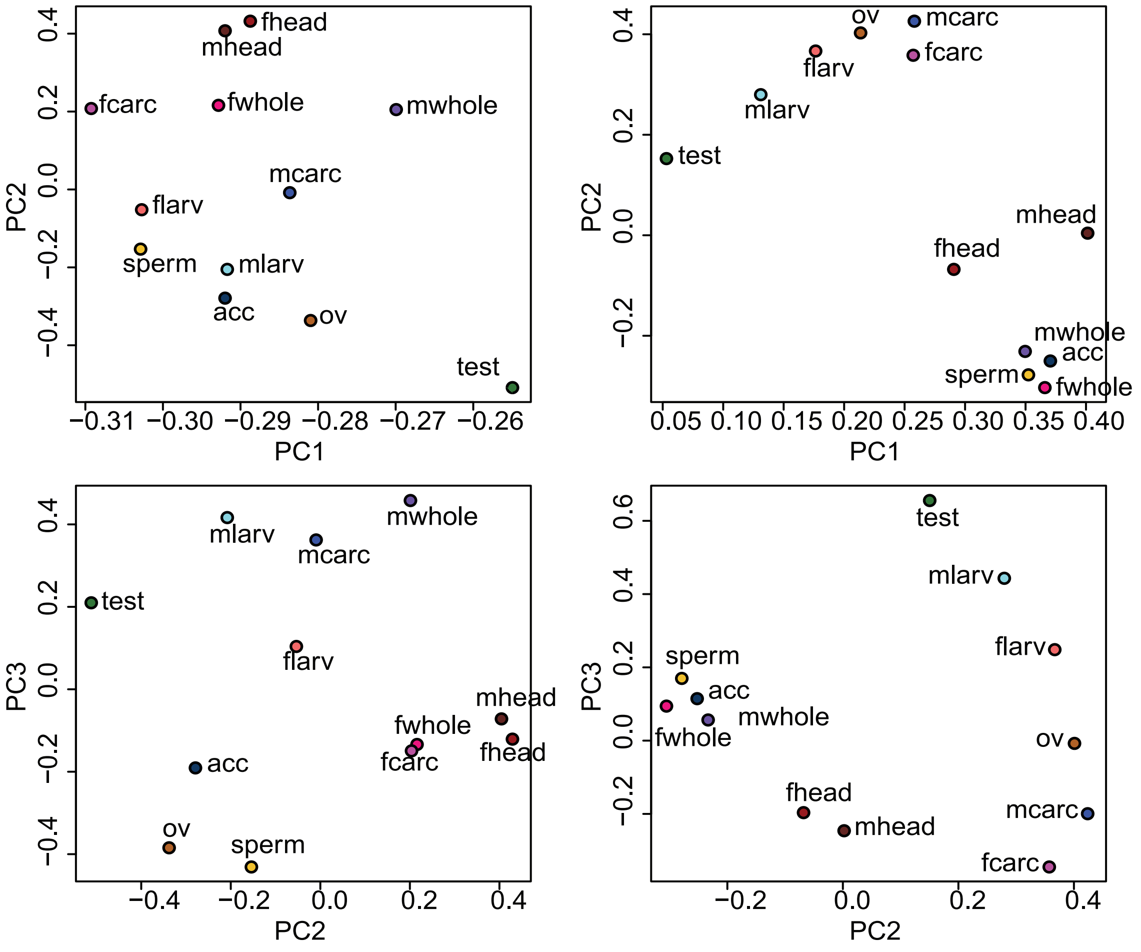
Figure 5. Sex-biased splicing in *D. pseudoobscura.* Comparisons of **Δ**Ψ distributions are between males and females for whole body (top left), gonad (ovary and testis, top right), carcass (bottom left), and head (bottom right). The x-axis gives ΔΨ values and the y-axis shows the number of sex-biased exons. Red bars represent female-biased exons ($\Psi_{female} - \Psi_{male} > 0$) and blue bars represent male-based exons ($\Psi_{male} - \Psi_{female} > 0$).



Alternative Splicing

Gene Expression

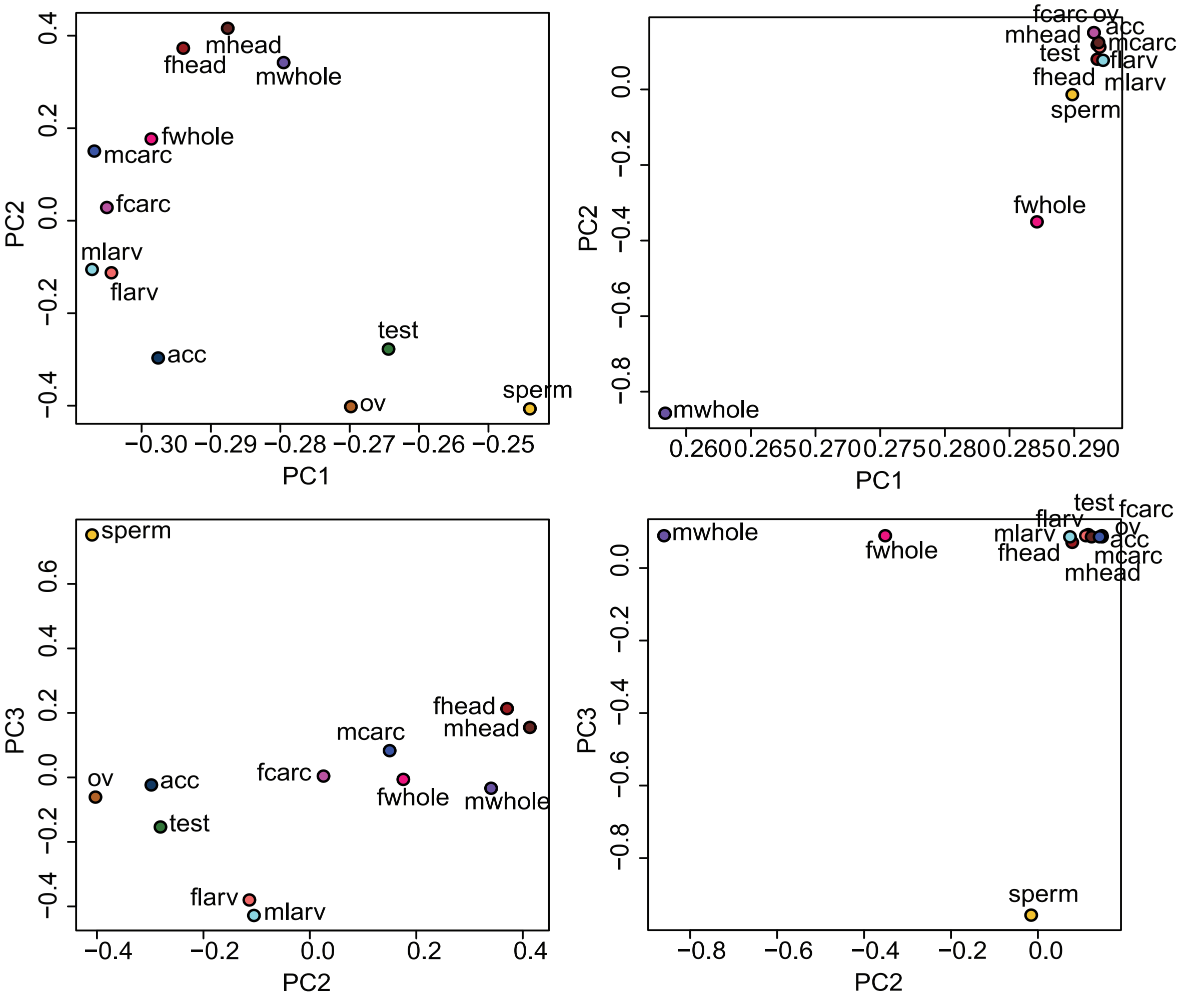
Figure S1. PCAs based on alternative splicing and gene expression for *D. miranda*. AS (left column) and gene expression (right column) profiles for *D. miranda*. The R function *prcomp* was used to perform the PCAs. PC1 (AS: 74.7% of the variance & GE: 98.4% of the variance) and PC2 (AS: 7.4% of the variance & GE: 1.5% of the variance) (top). PC2 (AS: 7.4% of the variance & GE: 1.5% of the variance) and PC3 (AS: 4.3% of the variance & GE: 0.05% of the variance) (bottom). “f” = female; “m” = male; “5.5” = mid stage 5; “5.9” = late stage 5; “8.9” = late stage 8; “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland



Alternative Splicing

Gene Expression

Figure S2. PCAs based on alternative splicing and gene expression for *D. albomicans*. AS (left column) and gene expression (right column) profiles for *D. albomicans*. The R function *prcomp* was used to perform the PCAs. PC1 (AS: 54.7% of the variance & GE: 40.5% of the variance) and PC2 (AS: 7.6% of the variance & GE: 23.7% of the variance) (top). PC2 (AS: 7.6% of the variance & GE: 23.7% of the variance) and PC3 (AS: 6.0% of the variance & GE: 9.2% of the variance) (bottom). “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland



Alternative Splicing

Gene Expression

Figure S3. PCAs based on alternative splicing and gene expression for *D. nasuta*. AS (left column) and gene expression (right column) profiles for *D. nasuta*. The R function *prcomp* was used to perform the PCAs. PC1 (AS: 54.3% of the variance & GE: 97.3% of the variance) and PC2 (AS: 7.7% of the variance & GE: 2.5% of the variance) (top). PC2 (AS: 7.7% of the variance & GE: 2.5% of the variance) and PC3 (AS: 5.3% of the variance & GE: 0.2% of the variance) (bottom). “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland

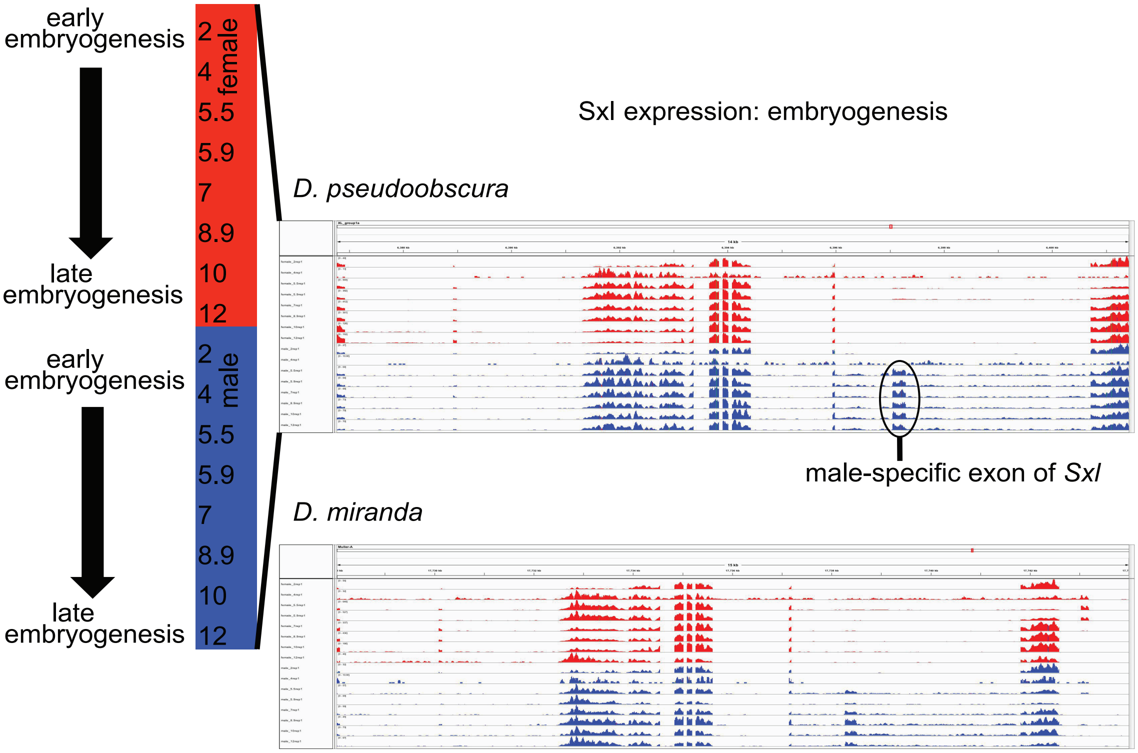


Figure S4. *Sxl* expression in embryonic stages. Exon 2 is spliced in in males (blue) and skipped in females (red). We used IGV (J. T. Robinson et al., 2011; Thorvaldsdottir, Robinson, & Mesirov, 2013) to visualize *Sxl* expression. “5.5” = mid stage 5; “5.9” = late stage 5; “8.9” = late stage 8.

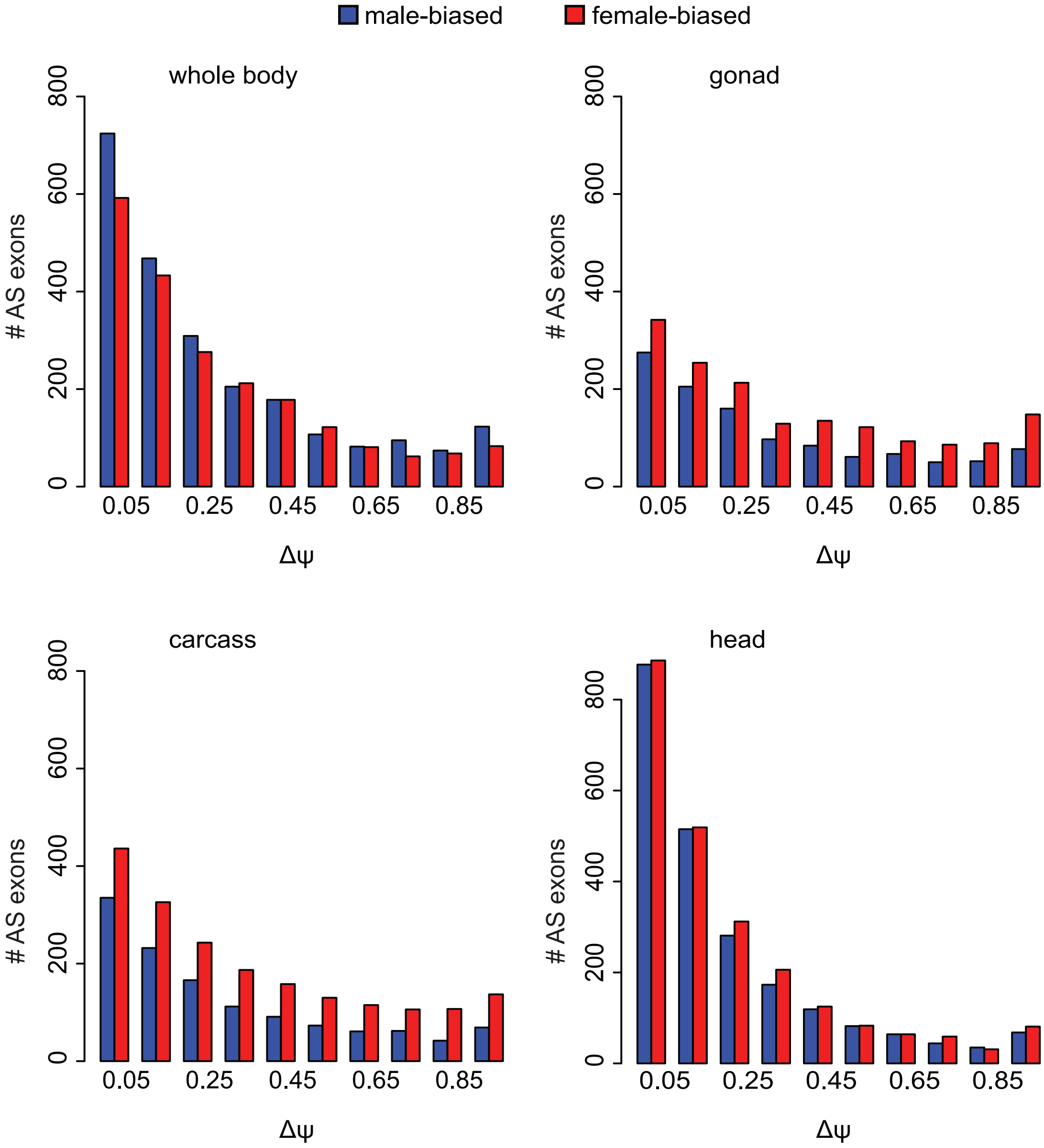


Figure S5. Sex-biased splicing in *D. miranda* as described by ΔΨ distributions. Comparisons are between males and females for whole body (top left), gonad (ovary and testis; top right), carcass (bottom left), and head (bottom right). The x-axis represents ΔΨ values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ($\Psi_{female} - \Psi_{male} > 0$) and blue bars represent male-based exons ($\Psi_{male} - \Psi_{female} > 0$).

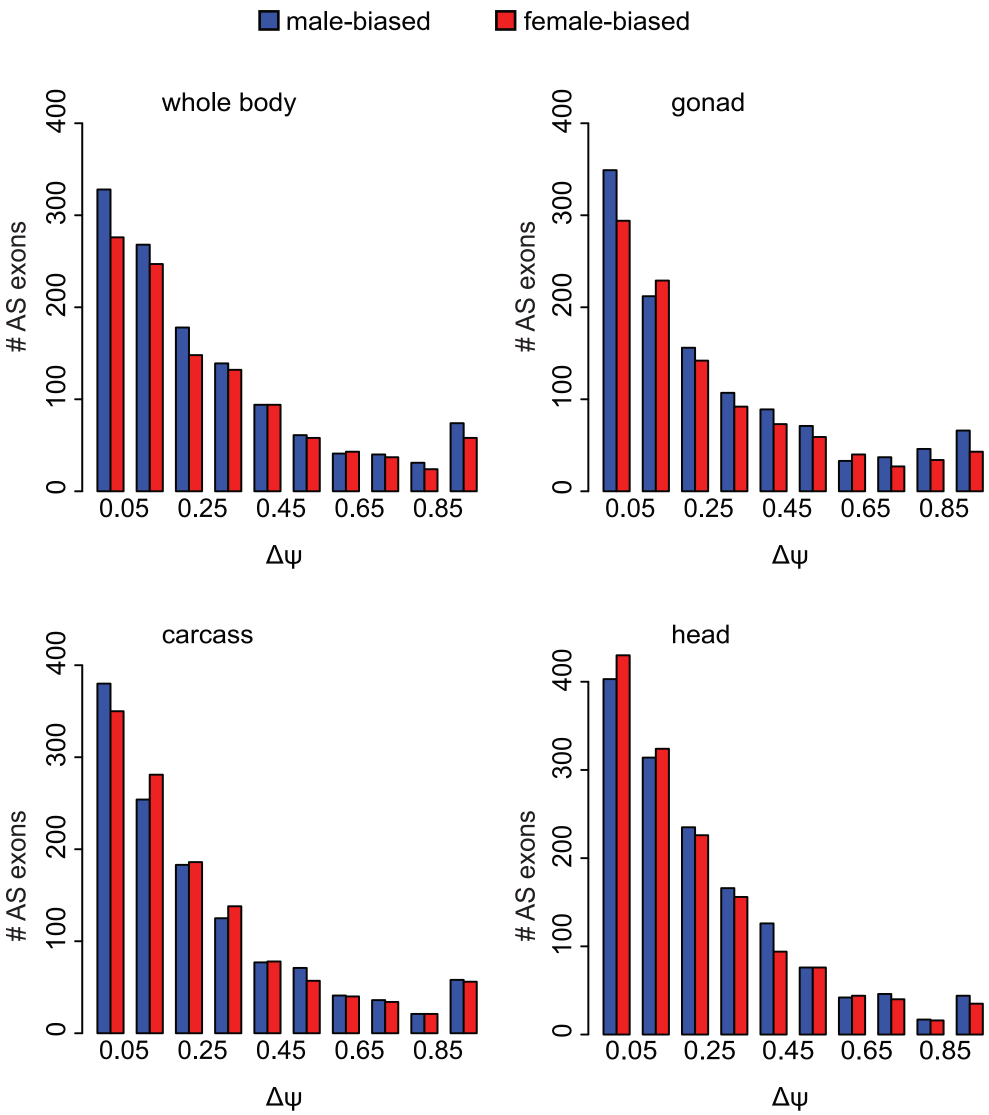


Figure S6. Sex-biased splicing in *D. albomicans* as described by ΔΨ distributions. Comparisons are between males and females for whole body (top left), gonad (ovary and testis; top right), carcass (bottom left), and head (bottom right). The x-axis represents ΔΨ values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ($\Psi_{female} - \Psi_{male} > 0$) and blue bars represent male-based exons ($\Psi_{male} - \Psi_{female} > 0$).

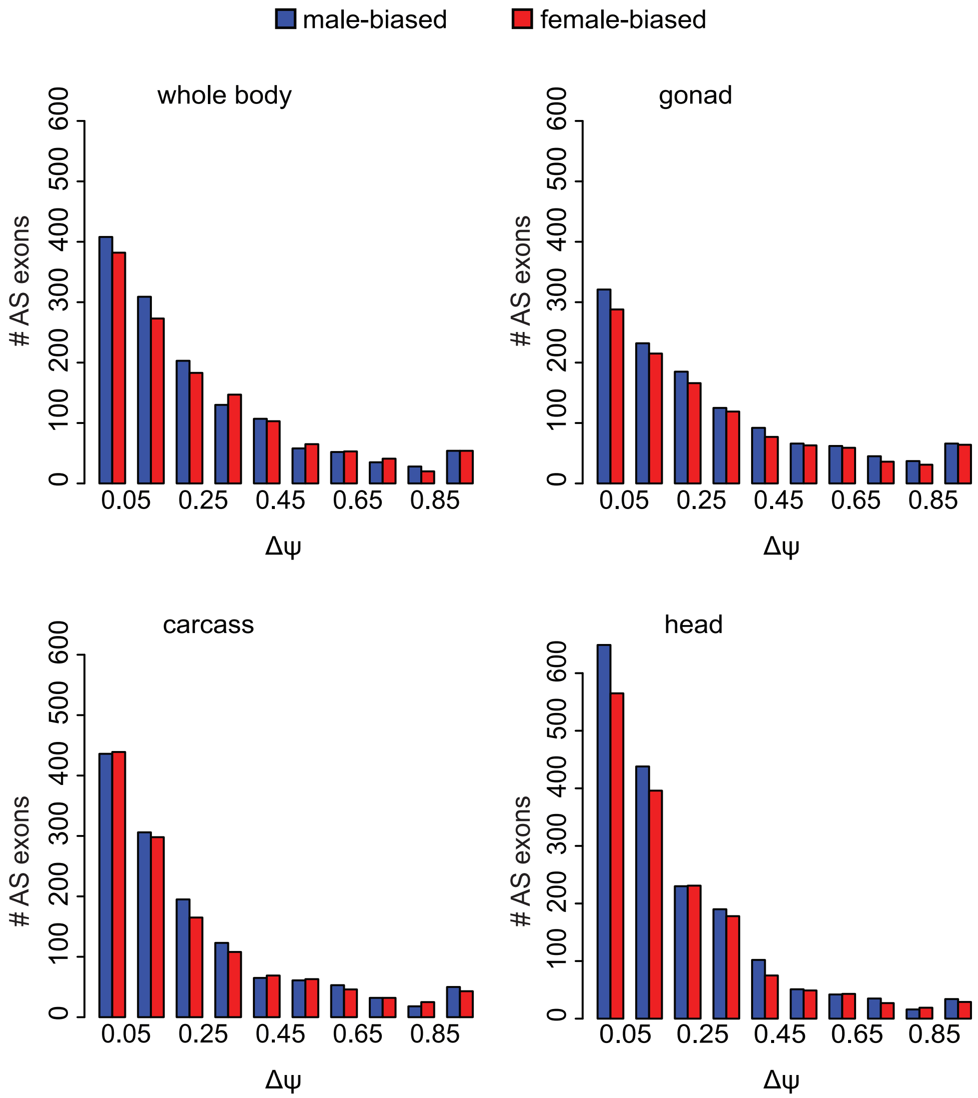


Figure S7. Sex-biased splicing in *D. nasuta* as described by ΔΨ distributions. Comparisons are between males and females for whole body (top left), gonad (ovary and testis; top right), carcass (bottom left), and head (bottom right). The x-axis represents ΔΨ values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ($\Psi_{female} - \Psi_{male} > 0$) and blue bars represent male-based exons ($\Psi_{male} - \Psi_{female} > 0$).

Figure S8. Spearman correlations based on gene expression in adult tissues not including head. Spearman correlations based on gene expression (TPM) for genes orthologous in adult tissues (n=3,005), not including male and female head.

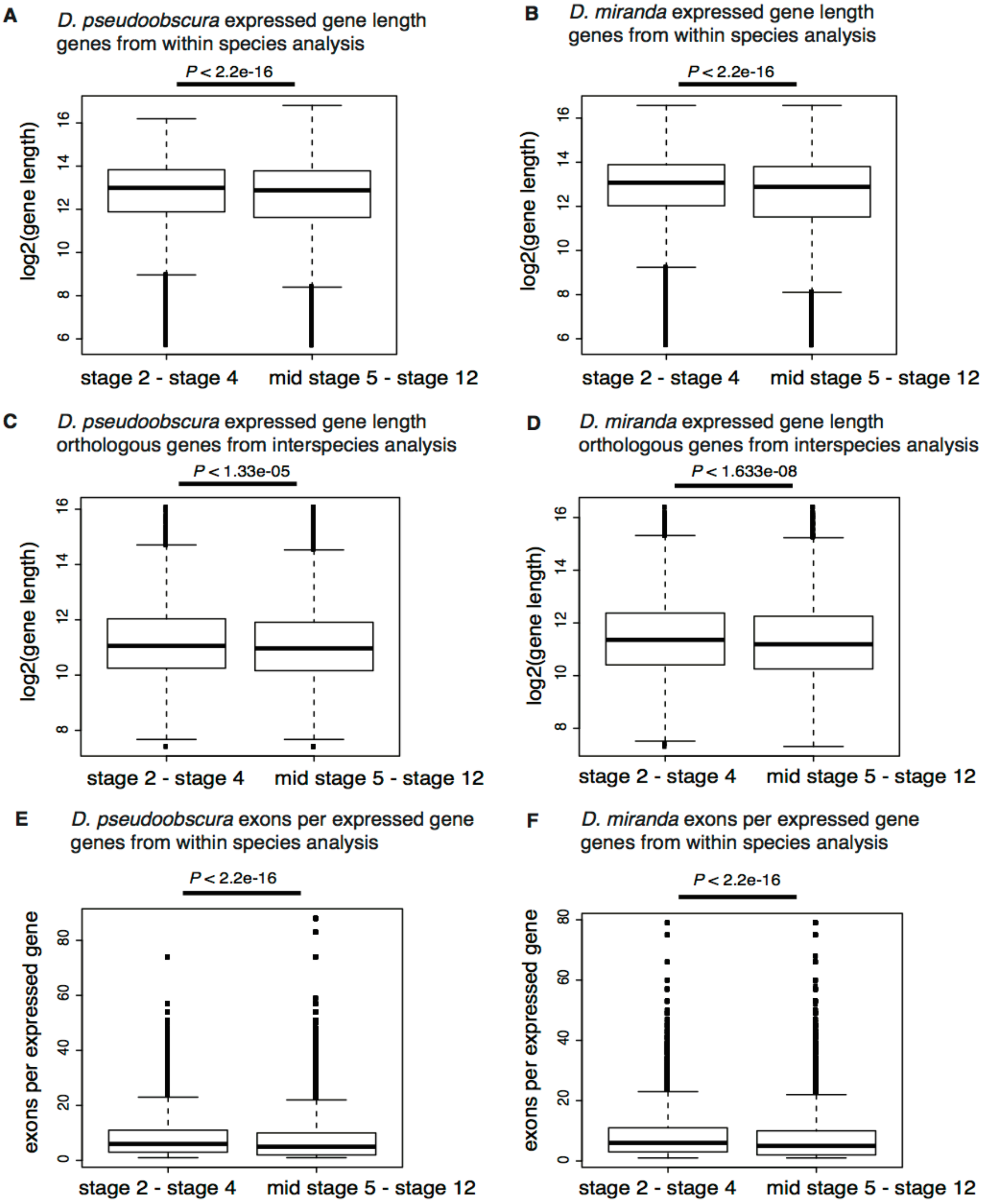


Figure S9. Boxplots comparing length and exon number of expressed genes between early (stage 2 – stage 4) and later (mid stage 5 – stage 12) embryonic stages. Log$_2$(length) of genes expressed in early (stage 2 – stage 4) and later (mid stage 5 – stage 12) embryonic stages from within-species analyses in *D. pseudoobscura* (A) and *D. miranda* (B). Log$_2$(length) of orthologous genes from interspecies analysis expressed in early (stage 2 – stage 4) and later (mid stage 5 – stage 12) embryonic stages in *D. pseudoobscura* (C) and *D. miranda* (D). Number of exons per gene expressed in early (stage 2 – stage 4) and later (mid stage 5 – stage 12) embryonic stages from within-species analyses in E) *D. pseudoobscura* and F) *D. miranda*. *P*-values indicate the results of Wilcoxon rank sum tests.

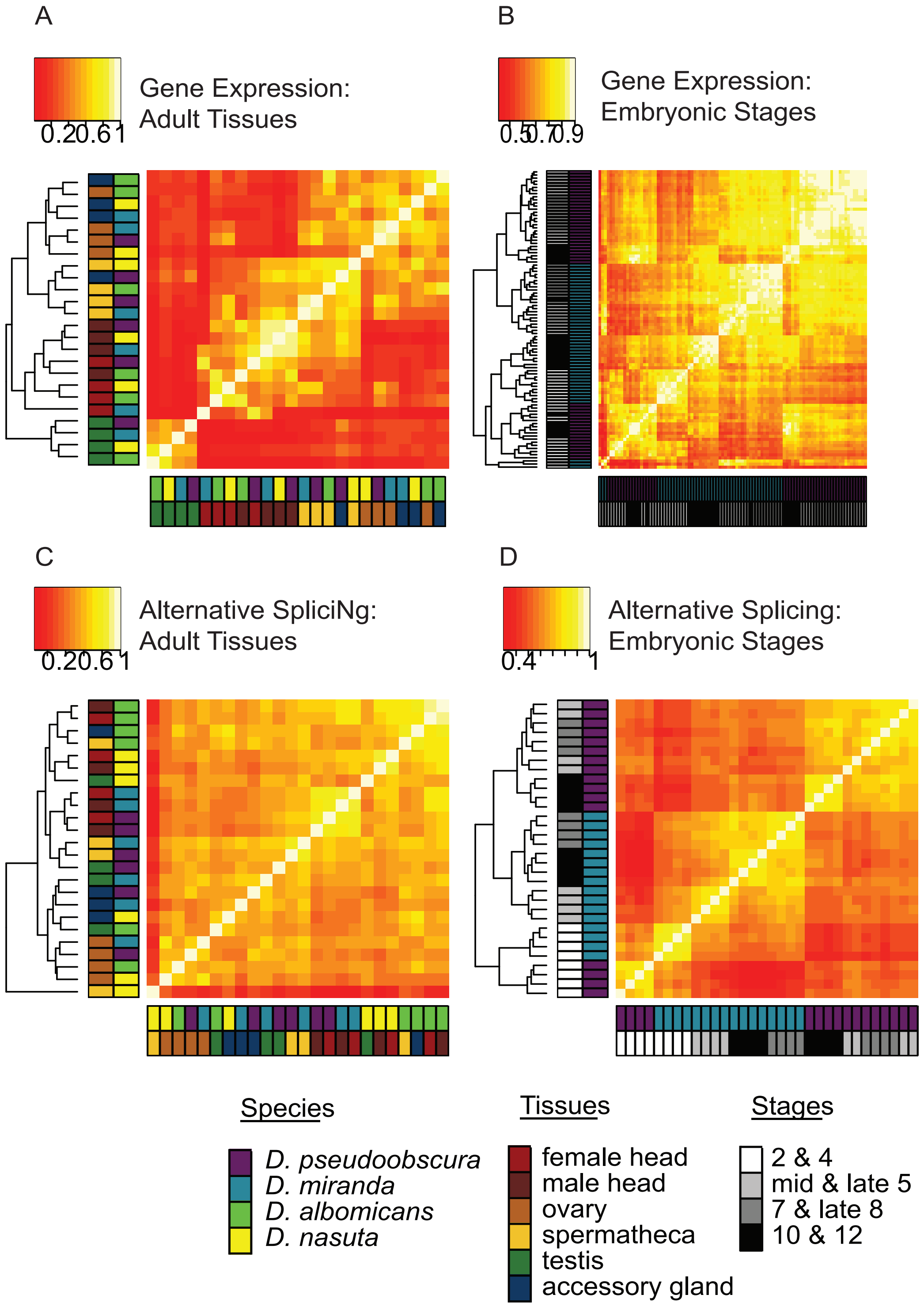


Figure S10. Pearson correlations based on gene expression and alternative splicing. Pearson correlations based on gene expression (TPM) for genes orthologous in adult tissues (n=3,005) (A) and embryonic stages (n=6,707) (B). Pearson correlations based on AS (Ψ) for exons orthologous and annotated as alternatively spliced in at least one sample in adult tissues (n=472) (C) and embryonic stages (n=1,122) (D).

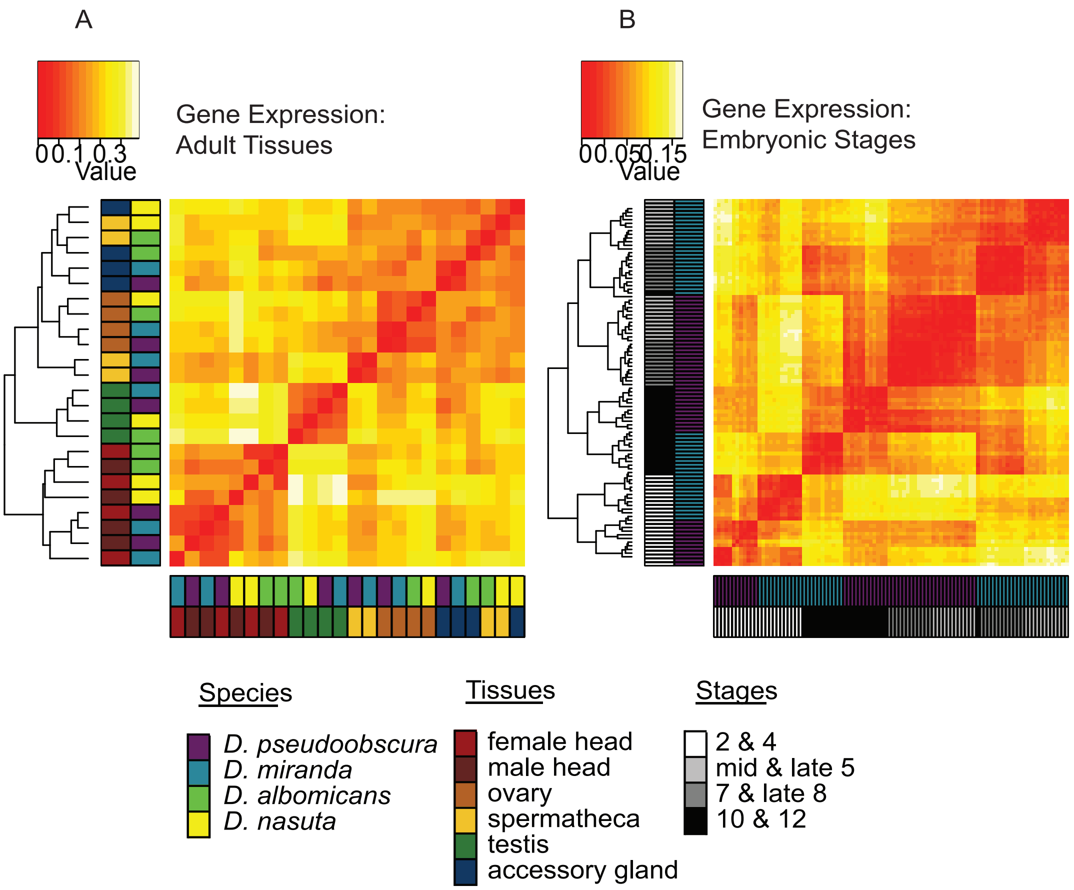


Figure S11. Heatmaps based on Jensen-Shannon divergence of GE. Heatmaps based on Jensen-Shannon divergence of gene expression for genes orthologous in adult tissues (n=3,005) (A) and embryonic stages (n=6,707) (B).

tissue/stage	D. pseudoobscura	D. miranda	D. albomicans	D. nasuta
female whole	467,377,020	51,231,232	26,727,418	23,167,266
female carcass	57,432,046	67,129,846	26,511,478	25,880,266
female head	33,905,716	40,465,784	23,512,022	24,109,998
ovary	54,058,410	65,438,942	25,874,364	26,272,194
spermatheca	58,710,452	57,849,644	23,684,462	10,615,680
male whole	313,334,258	54,544,916	26,203,150	26,610,468
male carcass	53,319,952	27,364,622	25,642,428	24,161,432
male head	44,494,468	45,094,584	23,692,120	23,644,298
testis	53,156,840	27,045,342	26,805,002	24,199,712
accessory gland	51,061,292	57,588,570	24,138,710	27,651,206
female 3rd instar larva	169,557,998	129,680,490	26,190,516	27,066,670
male 3rd instar larva	193,814,674	489,989,564	15,787,642	25,555,764
female embryonic stage 2	83,008,370*	104,506,456*	N/A	N/A
female embryonic stage 4	98,155,834*	96,885,044*	N/A	N/A
female embryonic stage 5.5	86,845,349*	101,672,710*	N/A	N/A
female embryonic stage 5.9	96,666,822*	107,292,918*	N/A	N/A
female embryonic stage 7	129,582,880*	91,276,706*	N/A	N/A
female embryonic stage 8.9	88,024,376*	95,564,504*	N/A	N/A
female embryonic stage 10	104,551,402*	104,479,700*	N/A	N/A
female embryonic stage 12	91,942,430*	114,018,074*	N/A	N/A
male embryonic stage 2	100,450,462*	102,908,968*	N/A	N/A
male embryonic stage 4	96,909,292*	90,476,136*	N/A	N/A
male embryonic stage 5.5	92,415,258*	100,957,974*	N/A	N/A
male embryonic stage 5.9	105,462,372*	136,574,154*	N/A	N/A
male embryonic stage 7	101,848,886*	108,006,670*	N/A	N/A
male embryonic stage 8.9	88,260,010*	94,594,086*	N/A	N/A
male embryonic stage 10	102,957,216*	97,035,706*	N/A	N/A
male embryonic stage 12	94,887,134*	98,228,172*	N/A	N/A
intrapecific comparison	10,357,750**	8,514,094**	5,307,840	5,307,840
*three replicates were used: read number is sum of reads				
**two replicates were used for embryonic samples: read number is reads used per replicate				

Table S1. Data used. The numbers of pairs of paired-end RNA-seq reads used for interspecies analyses, broken down by species and tissue/sex/stage, and intraspecies analysis. “5.5” = mid stage 5; “5.9” = late stage 5; “8.9” = late stage 8

sample	*D. pseudoobscura*	*D. miranda*	*D. albomicans*	*D. nasuta*
female whole	SRP009072	SRP009365	SRP010449	SRP010449
female carcass	SRP009072	SRP009365	*	*
female head	SRP001791	*	*	*
ovary	SRP009072	SRP009365	*	*
spermatheca	*	*	*	*
female 3rd instar larva	*	SRP009365	*	*
male whole	SRP009072	SRP009365	SRP010449	SRP010449
male carcass	SRP009072	SRP009365	*	*
male head	SRP001791	*	*	*
testis	SRP009072	SRP009365	*	*
accessory gland	SRP009072	SRP009365	*	*
male 3rd instar larva	*	SRP009365	*	*
embryonic stages	SRP034605	SRP034605	N/A	N/A

*data not publicly available before publication: see DATA ACCESS

Table S2. SRA identifiers for all RNA-seq datasets publicly available before publication used in this study.

D. nasuta genome assembly

	Length (bp)
Muller-A	31,028,076
Muller-B	23,121,377
Muller-C	23,086,324
Muller-D	29,470,473
Muller-E	30,485,585
Muller-F	1,765,445
Total	138,957,280
N50	29,470,473

Table S3. Information for the unpublished *D. nasuta* genome assembly.

D. miranda tisssue	Correlation (r^2)
female whole	0.84
female carcass	0.74
female third instar larva	0.76
ovary	0.81
spermatheca	0.78
male whole	0.72
male carcass	0.70
testis	0.75
accessory gland	0.78
male third instar larva	0.79
female embryonic stage 2	0.84
female embryonic stage 4	0.81
female embryonic stage 7	0.79
female embryonic stage 8.9	0.78
female embryonic stage 10	0.81
female embryonic stage 12	0.71

Table S4. Correlation of skipped exon Ψ values for alternatively spliced exons in post-embryonic *D. miranda* tissues computed by two pipelines. The first pipeline, described in Materials and Methods, used MATS to annotate alternatively spliced exons and compute their Ψ values. The second pipeline used AltEventFinder to annotate alternatively spliced exons and MISO to compute their Ψ values.

Chapter 2: Heterochromatin sinks create a mutational burden in males

Abstract

Large portions of eukaryotic genomes consist of repetitive elements, and the establishment of transcription-repressing heterochromatin during early development safeguards genome integrity. Heterochromatin formation coincides with genome-wide activation of zygotic expression. Males often contain substantially more heterochromatic DNA than females, due to the presence of a large, repeat-rich Y chromosome. Here, we show that heterochromatin formation is delayed in early male Drosophila embryos. This coincides with an increase in repeat expression in male versus female embryos, and more *de novo* insertions of repeats in males. Thus, the Y chromosome may indirectly create a mutational burden in males by postponing the establishment of silencing chromatin marks.

Introduction

In most animals, the zygotic genome is initially inactive and the first stages of embryonic development are solely controlled by maternal proteins and transcripts (Newport & Kirschner, 1982a; Newport & Kirschner, 1982b). In *Drosophila melanogaster*, zygotic transcription begins about an hour into development (at the preblastoderm stage 2) and gradually increases; by the end of stage 4 (syncytial blastoderm), widespread zygotic transcription is observed (the maternal-to-zygotic transition) (Lécuyer et al., 2007; Pritchard & Schubiger, 1996).

Concordant with genome-wide activation of zygotic expression, the embryo also needs to assure that genomic regions are silenced whose transcription would be harmful (Haig, 2016). In particular, large fractions of eukaryotic genomes consist of repetitive DNA (Cavalier-Smith, 1985; Dover, 1982; John & Miklos, 1988), and transcriptional activation of repeats could result in their mobilization, causing insertional mutations and genomic instability (Craig, 2002; Hedges & Deininger, 2007). Silencing of repeats is achieved through establishment of constitutive heterochromatin in all cells during early development at repetitive DNA at centromeres, telomeres, and along the Y chromosome (Brown & Bachtrog, 2017; Francisco & Lemos, 2014; Henikoff, 1996).

To ensure efficient silencing of repeats in the embryo, the Drosophila mother deposits many components for heterochromatin assembly into the egg, including methyltransferases (such as Su(var)3-9) to induce histone modifications, structural components of heterochromatin (such as HP1a), and small piRNAs which are thought to target genomic regions for heterochromatin formation (Czech & Hannon, 2016; Halic & Moazed, 2010). This allows for the rapid formation of heterochromatin during early development, simultaneously with zygotic genome activation (Li, Harrison, Villalta, Kaplan, & Eisen, 2014; Lu, Ma, & Eissenberg, 1998; Yuan & O'Farrell, 2016).

In species with heteromorphic XY sex chromosomes, male embryos face two challenges that may delay initial establishment of heterochromatin at the onset of zygotic transcription. On one hand, mothers cannot deposit piRNAs to recognize and silence male-specific (i.e. Y-specific) repeats, and silencing of such repeats may depend on spreading of heterochromatin from adjacent loci, or production of Y-linked piRNAs in male embryos. Also, the presence of a large, repeat-rich Y chromosome implies that males may contain substantially more heterochromatic DNA than females. The heterochromatic Y chromosome may create a 'chromatin sink' for repressive chromatin marks (Francisco & Lemos, 2014), delaying the establishment of a repressive chromatin structure in repeat-rich regions in males relative to females. This in turn could result in increased repeat activity during early development in male versus female embryos.

Results & Discussion

Heterochromatin formation differs between sexes

To test for differences in the dynamics of establishing heterochromatin between the sexes, we assayed chromatin and gene expression profiles from single Drosophila embryos during early development (**Figure 1**). We gathered replicate ChIP-seq data for a repressive histone modification typical of heterochromatin (H3K9me3) from precisely staged single male and female *Drosophila pseudoobscura* embryos during the onset of heterochromatin formation. In particular, heterochromatin in *Drosophila melanogaster* begins to form during the short (roughly 8–20 min) post-fertilization nuclear cycles 11–13, but does not mature into a stable and cytologically visible domain until cycle 14, when interphase extends to 1.5 hours (Vlassova, Graphodatsky, Belyaeva, & Zhimulev, 1991; Yuan & O'Farrell, 2016). Chromatin assembly during this period is prior to any significant zygotic transcription, and thus dependent on maternally loaded RNA and proteins. Analysis of an inducible reporter gene has found that silencing occurs at the onset of gastrulation, about 1 hour after heterochromatin is visible cytologically (Lu et al., 1998).

We collected 3-11 embryos from each sex and developmental stage at the beginning of cycle 14 (the onset of cellularization of the blastoderm) and in the middle of cycle 14 (gastrulation of the embryo). These two time points can easily be recognized morphologically, and thus allow for precise staging of embryos, and reflect the transition of initiation of heterochromatin towards maturation into a stable, repressive chromatin mark.

We used a 'spike in' normalization method to compare the genomic distribution of the repressive chromatin mark (H3K9me3) across samples (Bonhoure et al., 2014). Specifically, we 'spiked in' a *D. melanogaster* embryo to each *D. pseudoobscura* embryo sample prior to ChIP and sequencing. *D. melanogaster* chromatin served as an internal standard for the immunoprecipitation experiment.

The relative recovery of *D. pseudoobscura* ChIP signal vs. *D. melanogaster* ChIP signal, normalized by their respective input counts, was estimated using a linear regression model (Bonhoure et al., 2014), and used to quantity the absolute abundance of the heterochromatin chromatin mark in each *D. pseudoobscura* embryo (see Methods for details). We identified the sex of the embryos by mapping sequenced input DNA to autosomal and X-specific sequences (**Figures S1 and S2**).

In wildtype Drosophila, heterochromatin is highly enriched in pericentromeric regions, the small dot chromosome, and along the entire length of the Y chromosome (Bonaccorsi & Lohe, 1991; Gatti & Pimpinelli, 1983; Hannah, 1951). **Figure 2A** shows the genomic distribution of the repressive H3K9me3 mark for single male and female *D. pseudoobscura* embryos across development. Overall, we find that levels of heterochromatin enrichment are similar for replicates within a sex, but differ between developmental stages and among sexes (**Figure 2B**; for individual embryos see **Figures S3 and S4**). As expected, we find lower levels of heterochromatin at the earliest stages of development, and a general increase of H3K9me3 enrichment at repetitive regions through embryonic development for both sexes ($p<2.2e^{-16}$, Wilcoxon test for both males and females; **Figure 2B**). Most interestingly, however, at the beginning of cycle 14, male embryos appear to harbor reduced levels of H3K9me3 along their repeat-rich regions compared to females. Subtraction plots (i.e. subtracting median enrichment of H3K9me3 signal in female minus male embryos across 10 kb windows; **Figure 2A**) and boxplots (**Figure 2B**) both demonstrate that H3K9me3 enrichment is significantly lower along repetitive regions in males relative to females ($p<2.2e^{-16}$, Wilcoxon test; **Figure 2B**). In late cycle 14 when heterochromatin has matured (i.e. gastrulation), H3K9me3 enrichment ceases to differ significantly between sexes (p=0.24, Wilcoxon test; **Figure 2B**). Thus, overall patterns of heterochromatin distribution are consistent with the idea that the increased repeat content of males may delay the establishment of repressive chromatin in male vs. female embryos during the initial establishment of heterochromatin.

Increased repeat expression in male embryos

Heterochromatin formation during early embryogenesis leads to the transcriptional silencing of repetitive DNA and transposable elements (TEs) (Elgin & Reuter, 2013). To establish whether delayed establishment of suppressive heterochromatin in male embryos results in a de-repression of repeats, we analyzed single male and female embryo RNA-seq data during early embryogenesis (Lott, Villalta, Zhou, Bachtrog, & Eisen, 2014). To assess sex-specific expression patterns of repetitive elements, we first built consensus TE sequences for *D. pseudoobscura* (see Methods). Both male and female embryos show similar temporal activation of zygotic genes at the onset of the maternal- to zygotic transition (**Figure S5**) (Gibilisco, Zhou, Mahajan, & Bachtrog, 2016). Most strikingly, however, repeat expression differs dramatically between male and female embryos (**Figure 3**). In particular, females efficiently suppress the majority of their TEs during early development, consistent with heterochromatin being established early on. Males, in contrast, show a dramatic de-repression of repeats in early embryogenesis. In particular, concordant with the global activation of zygotic gene expression at the onset of cellularization of the blastoderm during early cycle 14, we see a global expression of repeats in male embryos but not female embryos (**Figures 3A, 3B**, and **S5**).

Repeats with higher genomic coverage in males vs. females supposedly have copies located on the Y chromosome. Overall, we find repeats that show higher copy number in males relative to females, as estimated by genomic coverage analysis (**Figure 4A**), to show higher expression levels **(Figure 4B**). This suggests that de-repression of repeats in males is especially pronounced for repeats located on the Y chromosome.

The *Drosophila* female deposits many components for heterochromatin assembly into the egg, to ensure efficient silencing of repeats in the embryo (Brennecke et al., 2008; Liang et al., 2008). Small piRNAs are thought to target genomic regions for heterochromatin formation (Brennecke et al., 2008), suggesting that a lack of Y-linked piRNA sequences in female parents could be directly responsible for reduced heterochromatin formation in male embryos. We find that repeat expression during early embryogenesis correlates with abundance of maternally-deposited piRNA (**Figures 4C and 4D**). However, we find no correlation between piRNA abundance of repeats in embryos and male/female genomic coverage, **Figure S7**), arguing against a simple lack of piRNAs being responsible for expression of male-biased repeats

Note that a loss of heterochromatin at pericentromeric regions and the Y chromosome should not necessarily result in increased expression across all TE families present. The vast majority of TEs located in the pericentromere and the Y chromosome are dormant and fragmented, incapable of transposition (Ananiev, Barsky, Ilyin, & Ryzic, 1984; Pimpinelli et al., 1995).

More transposon insertion in males

Increased transposable element activity in male embryos could result in a higher mutational burden in male flies due to *de novo* TE insertions. To test if differences in repeat expression between male and female embryos result in sex-specific differences in TE movement, we mapped novel repeat insertions in male and female fly embryos and larvae by deep sequencing. Insertions were defined by paired-end reads where one read mapped uniquely to the genome and the other to a transposon sequence. To identify *de novo* transposition events in individual full-sib male and female larvae, we compared the genomic positions of transposons between parents and siblings; novel TE insertions in single larvae were identified as not being present in parental ovaries or testes or any siblings. Most interestingly, we detect significantly more autosomal TE insertions in males, relative to females (**Figure 5A)**. Only for *de novo* insertions on the X, where females are diploid and males are haploid, are the numbers of TE insertions similar between sexes.

Note that when comparing only the parental ovaries and testes, we found 6769 *de novo* insertions in testes compared to 2186 *de novo* insertions in ovaries. Larvae already have developed gonads, and tissue-specific effects may contribute to sex-differences in TE insertions. We thus also analyzed input DNA reads from our single embryo ChIP-seq data to identify *de novo* insertions in males and females at the beginning and middle of cycle 14. During these developmental stages, follicle progenitor cells and the genital disc have not developed yet (Campos-Ortega & Hartenstein, 2013), so any *de novo* insertions will not be due to differences in the gonads. At the beginning of cycle 14, right at the onset of zygotic transcription, we see few novel TE insertions in either sex, and no significant differences between sexes (p=0.16, Wilcoxon test; **Figure 5B**). However, later in development in the middle of cycle 14, we detect more novel TE insertions, and find that males have significantly more autosomal *de novo* insertions compared to females (p=0.003, Wilcoxon test; **Figure 5B**). Thus, this indeed suggests that TEs are more active during early embryogenesis in males compared to females, resulting in higher levels of insertional mutations and genomic instability.

To conclude, our data show that heterochromatin formation is delayed in male embryos relative to females, presumably due to the larger repeat content in male relative to female embryos. Delayed heterochromatin formation results in higher expression of TE's, and more genomic insertions of novel TE's in male embryos relative to females. This will result in higher levels of somatic TE mutations in males, and potentially more heritable germline insertions.

Materials & Methods

Fly strains

We used *D. pseudoobscura* strain SS-R2 kept at 18°C and *D. melanogaster* strain Oregon-R kept at 25°C.

Embryo collection for ChIP-seq

Flies were allowed to lay on molasses plates with yeast paste for 15 minutes (*D. melanogaster*) or 30 minutes (*D. pseudoobscura*), and embryos were aged to appropriate stages. After aging, embryos were washed, dechorionated with bleach, and viewed live under a light microscope for a maximum of 10 minutes (*D.* melanogaster) or 20 minutes (*D. pseudoobscura*) . Embryo stages were visually confirmed, and embryos were immediately flash frozen in liquid nitrogen and stored at -80°C.

Chromatin immunoprecipitation and sequencing

We performed ChIP-seq using a protocol adapted from (Brind'Amour et al., 2015). Briefly, single embryos were homogenized with a pipette tip, and chromatin was digested for 7.5 minutes at 21ºC using micrococcal nuclease (MNase) (New England Biolabs). We spiked DNA from *D. pseudoobscura* single embryos with DNA from *D. melanogaster* stage 7 (gastrulation) embryos so that each sample had 20% spike (*D. melanogaster*) DNA (i.e. one *D. melanogaster* embryo was used for four *D. pseudoobscura* embryos). We set aside 10% of each sample as input and incubated the remaining chromatin with Dynabeads Protein G (Invitrogen) for 2-6 hours. The H3K9me3 antibody (Diagenode, 1.65 ug/ul) was incubated for >3 hours with Dynabeads Protein G to bind the antibody to the beads, before adding it to the chromatin (0.25 ul per embryo) for overnight incubation. The chromatin-antibody-bead complexes were washed first with low-salt buffer, then with high-salt buffer. DNA was eluted from the chromatin-antibody-bead complexes by shaking at 65ºC for 1-1.5 hours. We extracted DNA from our ChIP samples and from our input using a phenol/chloroform/isoamyl alcohol mixture, then cleaned the DNA with Agencourt AmpureXP beads before library preparation. Libraries were prepared using the ThruPLEX DNA-seq kit (Rubicon) followed by two rounds of cleaning with AmpureXP beads.

100bp paired-end sequencing of our samples was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

ChIP-seq read mapping and data normalization

Reads were mapped to our unpublished *D. pseudoobscura* genome, which we received courtesy of Steven Schaeffer and updated with Hi-C data and to release 6.12 of the FlyBase *D. melanogaster* genome (Gramates et al., 2017) using bowtie2 v. 2.2.9 (Langmead & Salzberg, 2012). We determined the sexes of the embryos using bedtools to get coverage of chromosomes from the input reads. We then remapped the ChIP and input reads to sex-specific *D. pseudoobscura* genomes: scaffolds with a genomic female:male coverage ratio < 0.5 were classified as Y-linked, scaffolds wih a genomic female:male coverage ratio >= 0.5 and <=0.85 were removed, and all scaffolds with a genomic female:male coverage ratio of >0.85 were classified as autosomal or X-linked. To calculate H3K9me3 enrichment while accounting for ChIP efficiency using our internal spike-in, we used a pipeline adapted from (Bonhoure et al., 2014)The pipeline was modified as follows: we used 10 kb genomic windows over the entire genome (we did not presume to know where H3K9me3 peaks would be). We used total counts mapping to our windows, and did not use partial counts: if a read mapped partially to a window, it was counted. Instead of total reads mapping to the genome, we summed the counts for each window. Lastly, we did not log transform the final enrichment number. Eight embryos for which the ChIP failed (i.e. that showed no H3K9me3 enrichment of their spike at the annotated *D. melanogaster* pericentromeres) were excluded. We also excluded one outlier embryo with considerably lower amounts of H3K9me3 enrichment in *D. pseudoobscura* (compared to the remaining 21 embryos). For one embryo with >10x more input reads than ChIP reads, we subsampled the number of input reads to match the number of ChIP reads. In total, we analyzed 8 female and 3 male *D. pseudoobscura* embryos at the middle of cycle 14 (gastrulation), and 3 female stage and 7 male embryos at the beginning of cycle 14 (onset of cellularization of the blastoderm).

Repeat annotation

We annotated *D. pseudoobscura* repeats using REPdenovo (downloaded November 7, 2016) (Chu, Nielsen, & Wu, 2016). REPdenovo was run on raw sequencing reads using the parameters MIN_REPEAT_FREQ 3, RANGE_ASM_FREQ_DEC 2, RANGE_ASM_FREQ_GAP 0.8, K_MIN 30, K_MAX 50, K_INC 10, K_DFT 30, GENOME_LENGTH 196210428, ASM_NODE_LENGTH_OFFSET -1, MIN_CONTIG_LENGTH 100, IS_DUPLICATE_REPEATS 0.85, COV_DIFF_CUTOFF 0.5, MIN_SUPPORT_PAIRS 20, MIN_FULLY_MAP_RATIO 0.2, TR_SIMILARITY 0.85, and RM_CTN_CUTOFF 0.9. We only kept REPdenovo repeats with blast hits to the Repbase *Drosophila* repeat annotation (downloaded March 22, 2016) (Bao, Kojima, & Kohany, 2015).

We used tblastn (Boratyn et al., 2012) with the parameters -evalue 1e-6, -num_alignments 1, and -num_descriptions 1 to blast translated *D. pseudoobscura* genes (release 3.04) from FlyBase (Gramates et al., 2017) to repeats without hits to Repbase repeats. We eliminated any repeats with blast hits to *D. pseudoobscura* genes.

This resulted in 1,194 repeats with an N50 of 427, totaling 905,416 base pairs.

mRNA-seq data analysis

We used published RNA-seq (Lott et al., 2014) data to look at sex-specific repeat expression during embryogenesis. We ran HISAT2 (Ngwa, Wojciechowski, Zack, Beaty, & Ruczinski, 2017) using the default parameters on the paired end data to align the RNA-seq reads to our repeat annotation.

To look at repeat expression over embryogenesis, we used bedtools (Quinlan & Hall, 2010) to calculate coverage of each repeat in our annotation. We normalized coverage by the average genome coverage of all mapped reads for each sample to our unpublished *D. pseudoobscura* genome.

To look at expression of zygotic genes over embryogenesis, we ran HISAT2 using the default parameters on the paired end data to align the RNA-seq reads to version 3.04 of the FlyBase *D. pseudoobscura* genome (Gramates et al., 2017). We used a list of *D. pseudoobscura* zygotic genes as defined in (Gibilisco et al., 2016). We used cuffdiff (Trapnell et al., 2010) and version 3.04 of the FlyBase gtf annotation for *D. pseudoobscura* to calculate FPKM values.

piRNA-seq data generation and analysis

We generated a piRNA library for 0-1 hour *D. pseudoobscura* embryos which is before the onset of zygotic transcription, i.e. the piRNAs in the library are maternally-deposited. Please see chapter three of this dissertation for more details on piRNA library construction.

Adapters were trimmed from the reads using trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (). Trimmed reads were mapped to our *D. pseudoobscura* repeat annotation with bowtie2 (Langmead & Salzberg, 2012) using the default parameters and bedtools (Quinlan & Hall, 2010)was used to calculate coverage of our repeat library.

Calculating genomic repeat coverage

Paired-end male and female *D. pseudoobscura* genomic reads were mapped to our *D. pseudoobscura* repeat library. We used bedtools (Quinlan & Hall, 2010)to calculate male and female genomic coverage of repeats, and normalized coverage by the average genomic coverage of all mapped reads for each sample to our unpublished *D. pseudoobscura* genome.

D. pseudoobscura pericentromere annotation

We manually annotated pericentromeres in *D. pseudoobscura* based on genome-wide plots of H3K9me3 enrichment using the R program *plotly.*

gDNA-seq of larvae and parental gonads & bioinformatics analysis

We collected third instar larvae from a single female x singe male cross of *D. pseudoobscura* strain SS-R2 kept at 18ºC. Individual larvae and the parental gonads were frozen on dry ice and stored at -80ºC. Genomic DNA was extracted using phenol:choloroform:isoamly alcohol and gDNA libraries were constructed using the Illumina TruSeq Nano DNA Library Prepration Kit with custom dual-indexed barcodes to avoid index swapping. 100bp paired-end sequencing of our samples was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

We mapped our paired-end gDNA data to our unpublished *D. pseudoobscura* genome and to our *D. pseudoobscura* repeat annotation, using bowtie2 with default parameters. We used seqtk (https://github.com/lh3/seqtk) to subsample reads to the number of reads of the smallest sample in our larvae + parental gonads data sets and mapped the subsampled reads as well. Regions of the genome where one read mapped uniquely and its mate mapped to our repeat annotation were identified. For each larva, we used bedtools to compare insertions from the subsampled reads in that larva to insertions in the non-subsampled reads of all other samples (other sibling larva and parental gonads). After merging insertions within 1000 bp of one another, we called *de novo* insertions in a larva as insertions not present in any other sibling larvae, or parental gonads.

Calling *de novo* TE insertions in embryos

We used input DNA from ChIP-seq experiments from 3 male and 8 female embryos at the middle of cycle 14 (gastrulation), and 9 female and 8 male embryos at the beginning of cycle 14 (onset of cellularization of the blastoderm). We used the same pipeline to call *de novo* insertions as we used for larvae, except for each embryo, we used bedtools to compare insertions from the subsampled reads in that embryo to insertions in the non-subsampled reads of all other embryos, larvae, and gonads from the parents of the larvae.

Data visualization

Heat maps in **Figures 3A**, **S5**, and **S6** were made using Morpheus (https://software.broadinstitute.org/morpheus). Box plots in **Figures 2B**, **3B**, and **5** were made using *ggplot2* in R. All other figures except **Figure S1** were made in R.

References

Ananiev, E., Barsky, V., Ilyin, Y. V., & Ryzic, M. (1984). The arrangement of transposable elements in the polytene chromosomes of drosophila melanogaster. *Chromosoma, 90*(5), 366-377.

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA, 6*(1), 11.

Bonaccorsi, S., & Lohe, A. (1991). Fine mapping of satellite DNA sequences along the Y chromosome of drosophila melanogaster: Relationships between satellite sequences and fertility factors. *Genetics, 129*(1), 177-189.

Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., et al. (2014). Quantifying ChIP-seq data: A spiking method providing an internal reference for sample-to-sample normalization. *Genome Research, 24*(7), 1157-1168. doi:10.1101/gr.168260.113 [doi]

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct, 7*(1), 12.

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, N.Y.), 322*(5906), 1387-1392. doi:10.1126/science.1165171 [doi]

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M. M., & Lorincz, M. C. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nature Communications, 6*, 6033.

Brown, E., & Bachtrog, D. (2017). The drosophila Y chromosome affects heterochromatin integrity genome-wide. *bioRxiv,* , 156000.

Campos-Ortega, J. A., & Hartenstein, V. (2013). *The embryonic development of drosophila melanogaster* Springer Science & Business Media.

Cavalier-Smith, T. (1985). Cell volume and the evolution of eukaryote genome size. *The Evolution of Genome Size,* , 105-184.

Chu, C., Nielsen, R., & Wu, Y. (2016). REPdenovo: Inferring de novo repeat motifs from short sequence reads. *PloS One, 11*(3), e0150719.

Craig, N. L. (2002). Mobile DNA: An introduction. *Mobile DNA II* (pp. 3-11) American Society of Microbiology.

Czech, B., & Hannon, G. J. (2016). One loop to rule them all: The ping-pong cycle and piRNA-guided silencing. *Trends in Biochemical Sciences, 41*(4), 324-337.

Dover, G. A. (1982). *Genome evolution* Academic Press.

Elgin, S. C., & Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in drosophila. *Cold Spring Harbor Perspectives in Biology, 5*(8), a017780. doi:10.1101/cshperspect.a017780 [doi]

Francisco, F. O., & Lemos, B. (2014). How do y-chromosomes modulate genome-wide epigenetic states: Genome folding, chromatin sinks, and gene expression. *Journal of Genomics, 2*, 94-103. doi:10.7150/jgen.8043 [doi]

Gatti, M., & Pimpinelli, S. (1983). Cytological and genetic analysis of the Y chromosome of drosophila melanogaster. *Chromosoma, 88*(5), 349-373.

Gibilisco, L., Zhou, Q., Mahajan, S., & Bachtrog, D. (2016). Alternative splicing within and between drosophila species, sexes, tissues, and developmental stages. *PLoS Genetics, 12*(12), e1006464.

Gramates, L. S., Marygold, S. J., Santos, G. d., Urbano, J., Antonazzo, G., Matthews, B. B., et al. (2017). FlyBase at 25: Looking to the future. *Nucleic Acids Research, 45*(D1), D663-D671.

Haig, D. (2016). Transposable elements: Self‐seekers of the germline, team‐players of the soma. *BioEssays, 38*(11), 1158-1166.

Halic, M., & Moazed, D. (2010). Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. *Cell, 140*(4), 504-516.

Hannah, A. (1951). Localization and function of heterochromatin in drosophila melanogaster. *Advances in Genetics, 4*, 87-125.

Hedges, D., & Deininger, P. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 616*(1), 46-59.

Henikoff, S. (1996). Dosage‐dependent modification of position‐effect variegation in drosophla. *Bioessays, 18*(5), 401-409.

John, B., & Miklos, G. 1988The eukaryote genome in development and evolution.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357-359.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., et al. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell, 131*(1), 174-187.

Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014). Establishment of regions of genomic activity during the drosophila maternal to zygotic transition. *eLife, 3*, 10.7554/eLife.03737. doi:10.7554/eLife.03737 [doi]

Liang, H., Nien, C., Liu, H., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein zelda is a key activator of the early zygotic genome in drosophila. *Nature, 456*(7220), 400-403.

Lott, S. E., Villalta, J. E., Zhou, Q., Bachtrog, D., & Eisen, M. B. (2014). Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. *PLoS Genetics, 10*(2), e1004159. doi:10.1371/journal.pgen.1004159 [doi]

Lu, B. Y., Ma, J., & Eissenberg, J. C. (1998). Developmental regulation of heterochromatin-mediated gene silencing in drosophila. *Development (Cambridge, England), 125*(12), 2223-2234.

Newport, J., & Kirschner, M. (1982a). A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell, 30*(3), 675-686.

Newport, J., & Kirschner, M. (1982b). A major developmental transition in early xenopus embryos: II. control of the onset of transcription. *Cell, 30*(3), 687-696.

Ngwa, J., Wojciechowski, R., Zack, D. J., Beaty, T., & Ruczinski, I. (2017). Differential expression analysis of gene and transcript abundance for single cell RNA-seq data using STAR and HISAT aligners. *Investigative Ophthalmology & Visual Science, 58*(8), 1850-1850.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., et al. (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), 3804-3808.

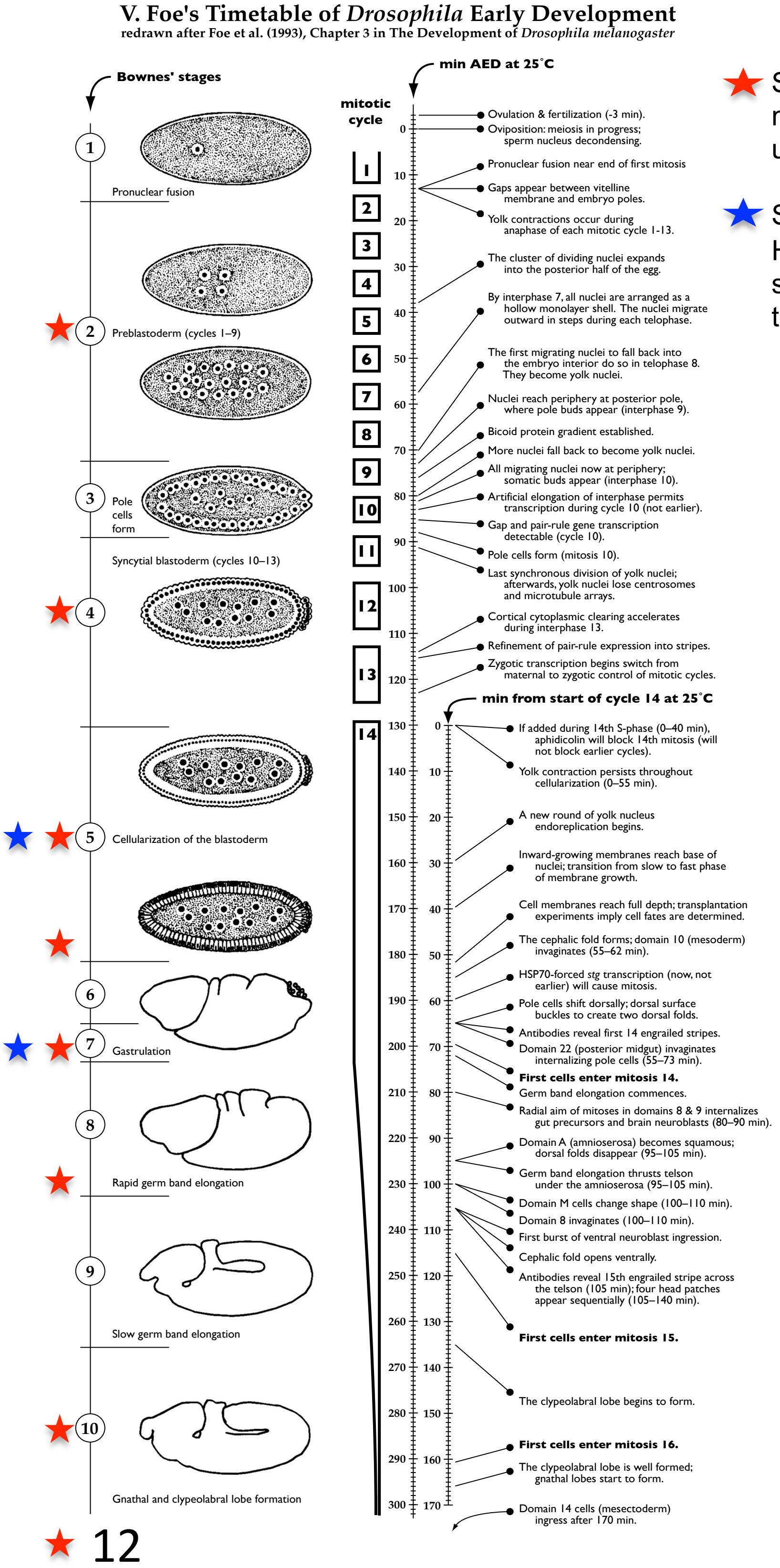
Pritchard, D. K., & Schubiger, G. (1996). Activation of transcription in drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes & Development, 10*(9), 1131-1142.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology, 28*(5), 511-515.

Vlassova, I. E., Graphodatsky, A. S., Belyaeva, E. S., & Zhimulev, I. F. (1991). Constitutive heterochromatin in early embryogenesis of drosophila melanogaster. *Molecular and General Genetics MGG, 229*(2), 316-318.

Yuan, K., & O'Farrell, P. H. (2016). TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in drosophila embryos. *Genes & Development, 30*(5), 579-593. doi:10.1101/gad.272237.115 [doi]



★ Single embryo mRNA-seq data used in this study

★ Single embryo H3K9me3 ChIP-seq data used in this study

Figure 1. Timetable of early Drosophila development downloaded from http://kirschner.med.harvard.edu/files/bionumbers/Timetable%20of%20Drosophila%20Early%20Development.pdf. Blue stars show the stages (5a and 7) we used for ChIP-seq. Red stars show the 8 stages (2, 4, 5.5, 5.9, 7, 8.9, 10, 12) for which we used single embryo mRNA-seq data.

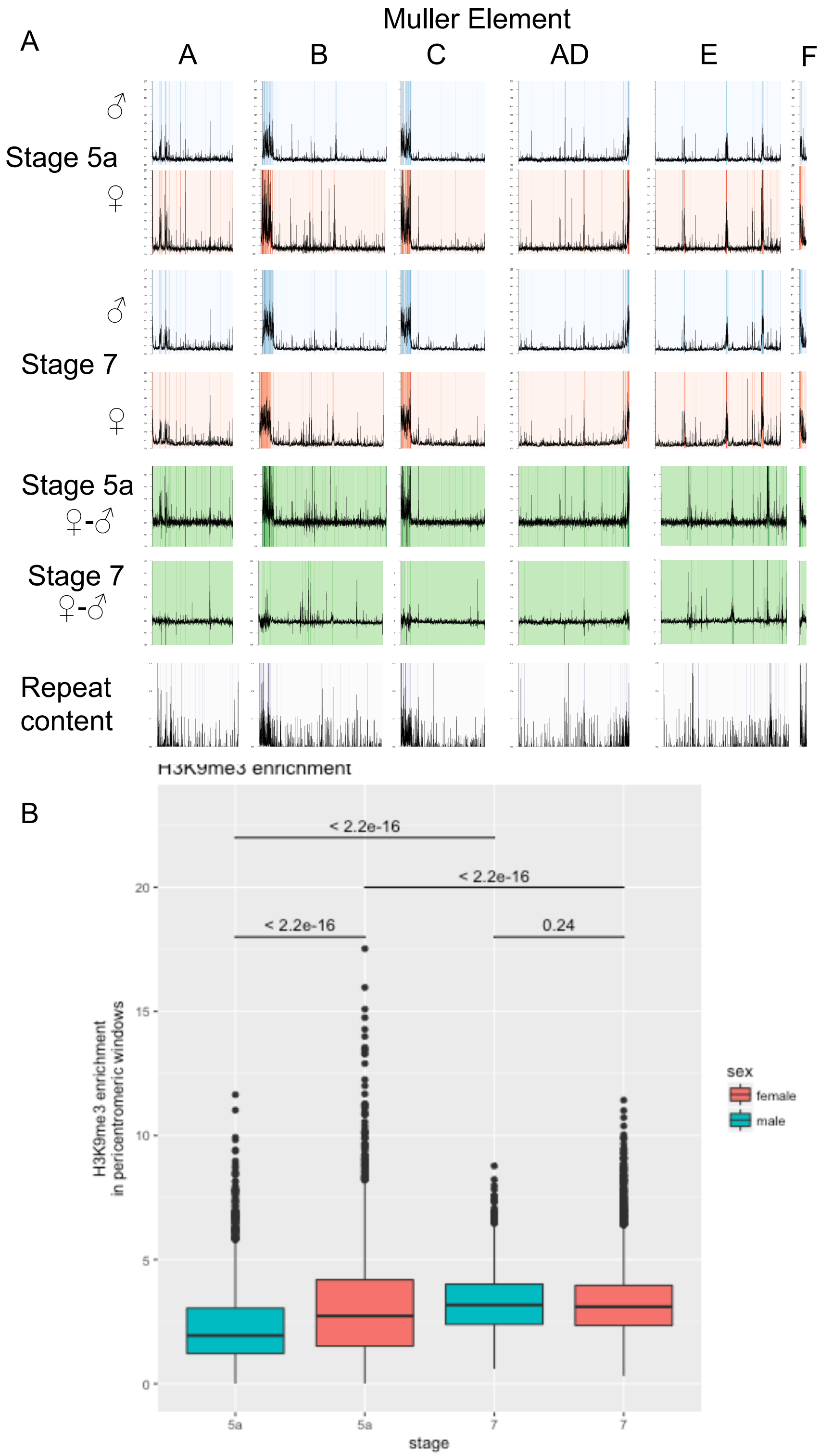
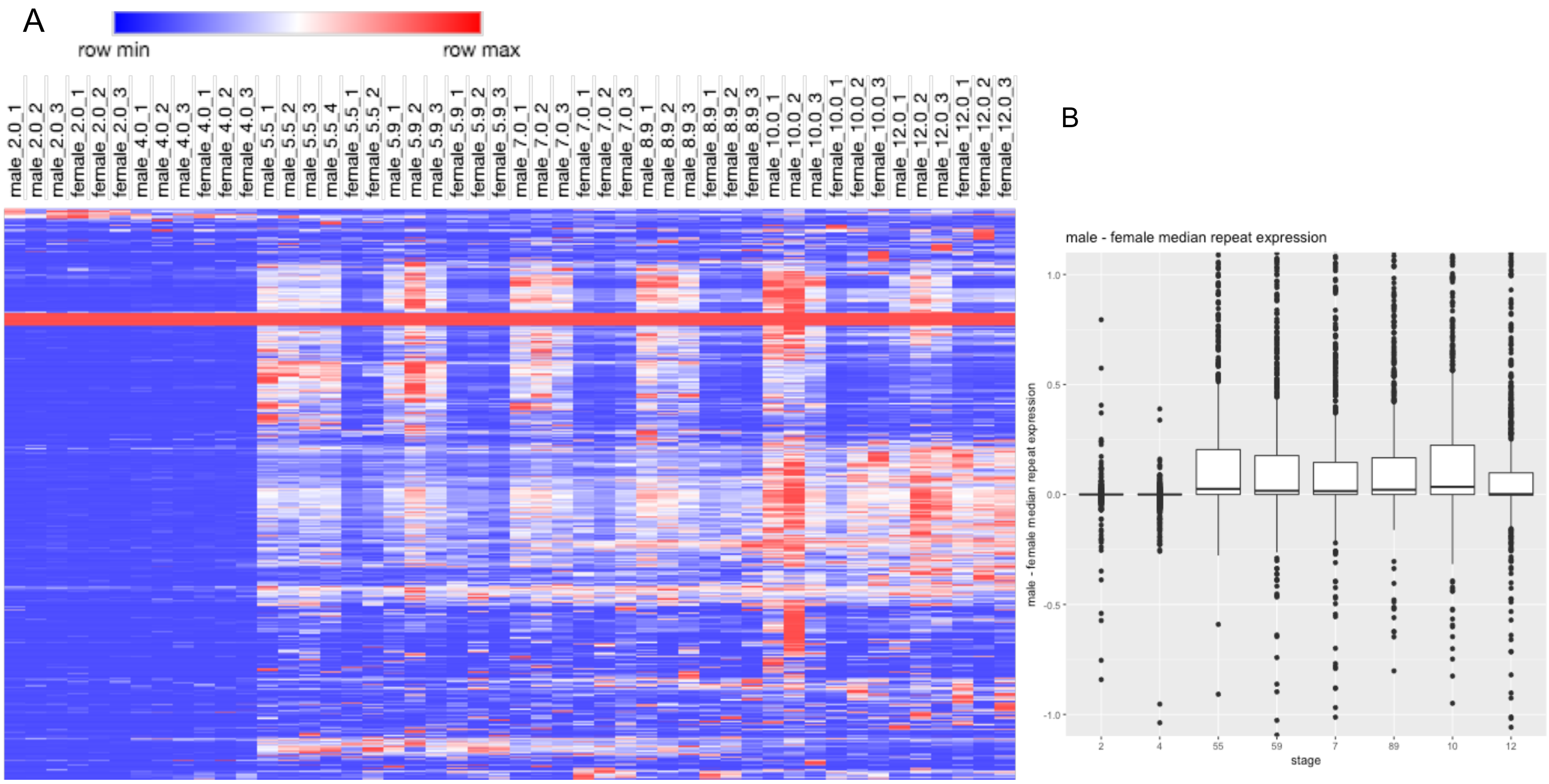


Figure 2. A) Genome-wide median H3K9me3 enrichment in male (blue) and female (red) stage 5a and stage 7 embryos, genome-wide male median – female median H3K9me3 enrichment in stage 5a and stage 7 embryos, and genome-wide repeat content. B) Boxplot for H3K9me3 enrichment in all pericentromeric windows for male (blue) and female (red) stage 5a and stage 7 embryos. P-values were calculated using *wilcox.test* in R.

Figure 3. A) Heat map of repeat expression in single embryos. Rows represent repeats Heat map is scaled by row. B) Box plot showing median male – median female expression of repeats over embryogenesis. Please see Table S1 for p-values.



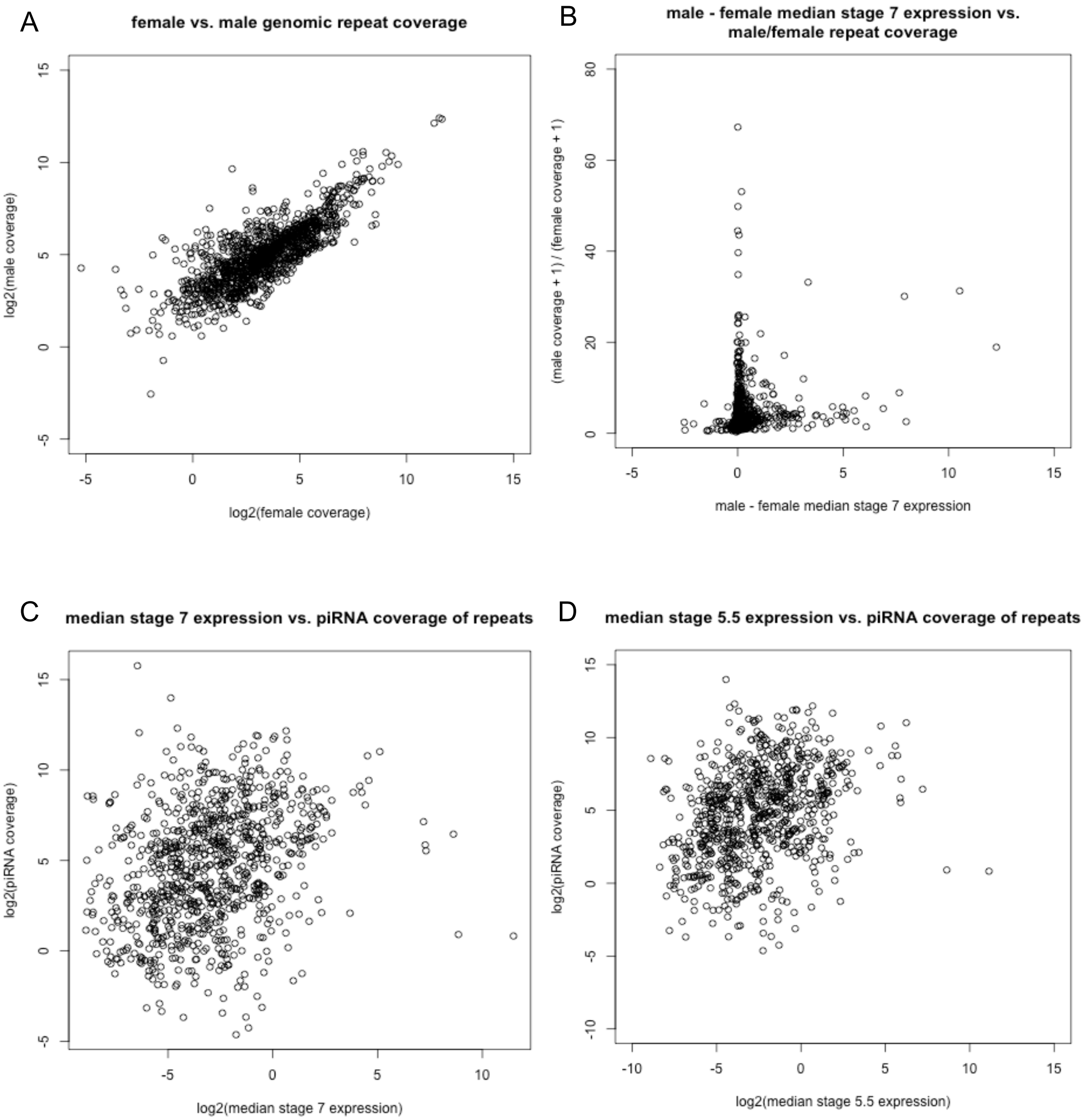


Figure 4. A) log2 of female repeat coverage vs. log2 of male repeat coverage. B) male – female median stage 7 expression of repeats vs. (male genomic coverage + 1) / (female genomic coverage + 1) of repeats. Spearman correlation = 0.34 C) log2 of median stage 7 repeat expression vs. log2 of 0-1 hour embryo (maternally deposited) piRNA coverage. Spearman correlation = 0.46 D) log2 of median stage 5a repeat expression vs. log2 of 0-1 hour embryo (maternally deposited) piRNA coverage. Spearman correlation = 0.46

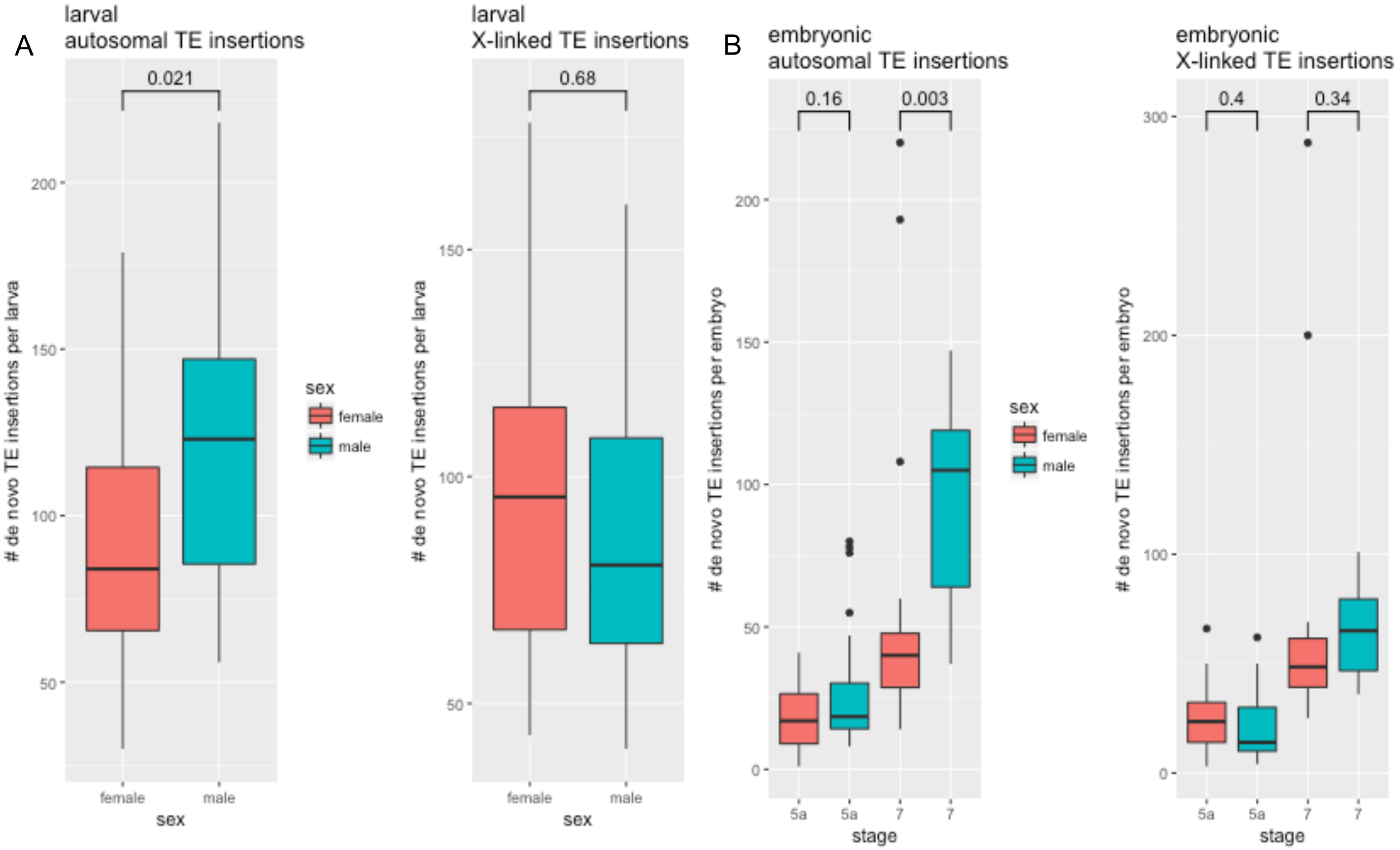


Figure 5. A) Box plots of *de novo* TE insertions in male and female larvae from a single female x single male *D. pseudoobscura* cross. B) Box plots of *de novo* TE insertions in male and female embryos. p-values show results of wilcox tests using *geom_signif* in R.

Stage 5a embryos

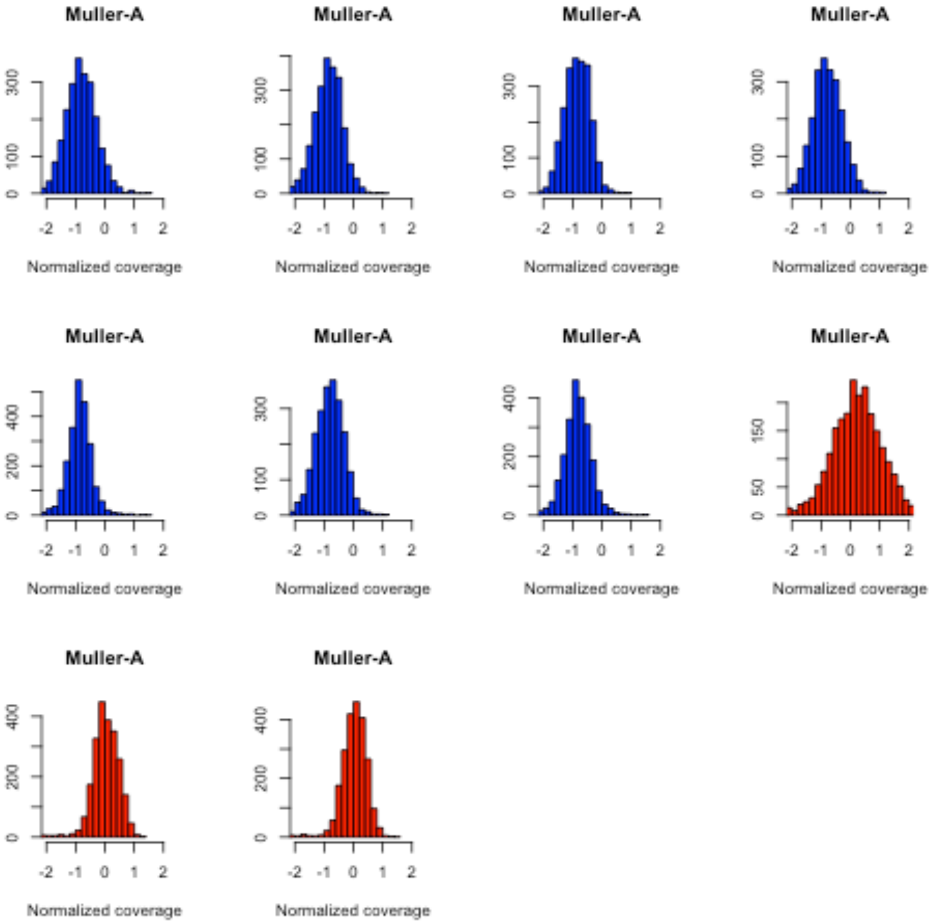


Figure S1. Histograms showing the sexes of stage 5a embryos after sexing embryos based on chromosome coverage then mapping input DNA from ChIP-seq to sex-specific *D. pseudoobscura* genomes. Histograms show log2 of counts mapped to 10 kb windows of Muller-A / counts mapped to 10 kb windows of the three largest autosomes (Muller B, Muller C, and Muller E). Males have half as many X chromosomes as autosomes, so their histograms center around -1. The X:autosome ratio in females is 1:1, so their histograms center around 0. Males are colored blue and females are colored red.

Stage 7 embryos

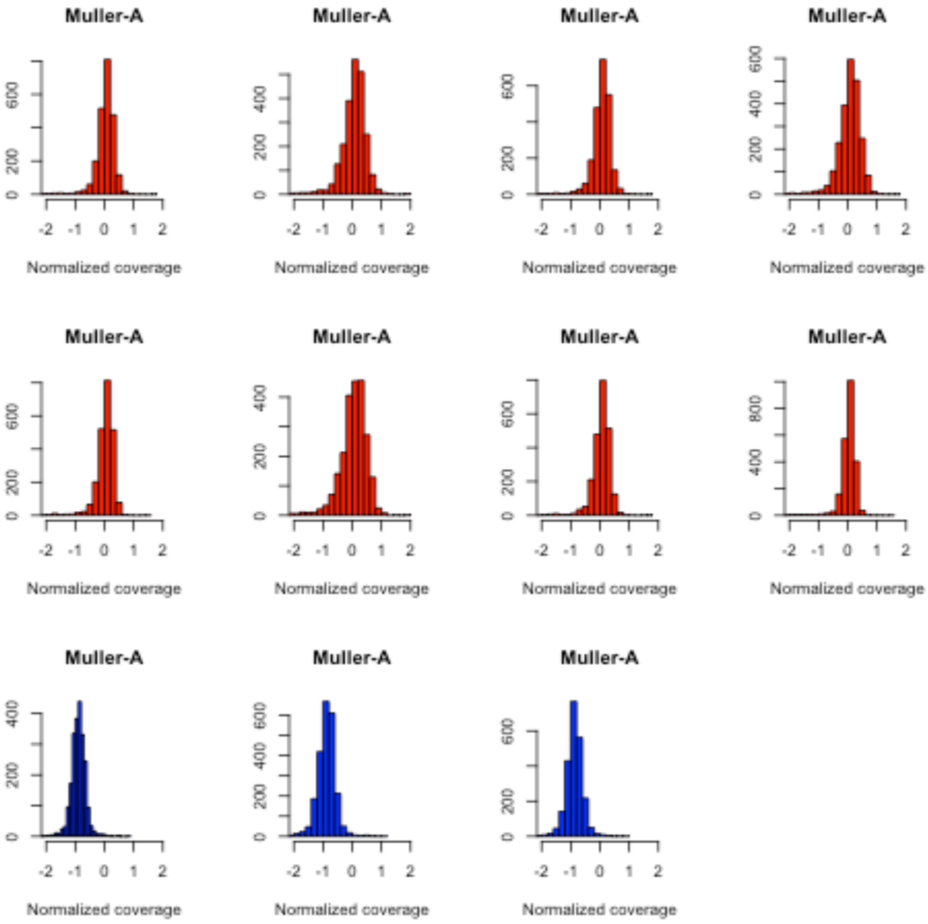


Figure S2. Histograms showing the sexes of stage 7 embryos after sexing embryos based on chromosome coverage then mapping input DNA from ChIP-seq to sex-specific *D. pseudoobscura* genomes. Histograms show log2 of counts mapped to 10 kb windows of Muller-A / counts mapped to 10 kb windows of the three largest autosomes (Muller B, Muller C, and Muller E). Males have half as many X chromosomes as autosomes, so their histograms center around -1. The X:autosome ratio in females is 1:1, so their histograms center around 0. Males are colored blue and females are colored red.

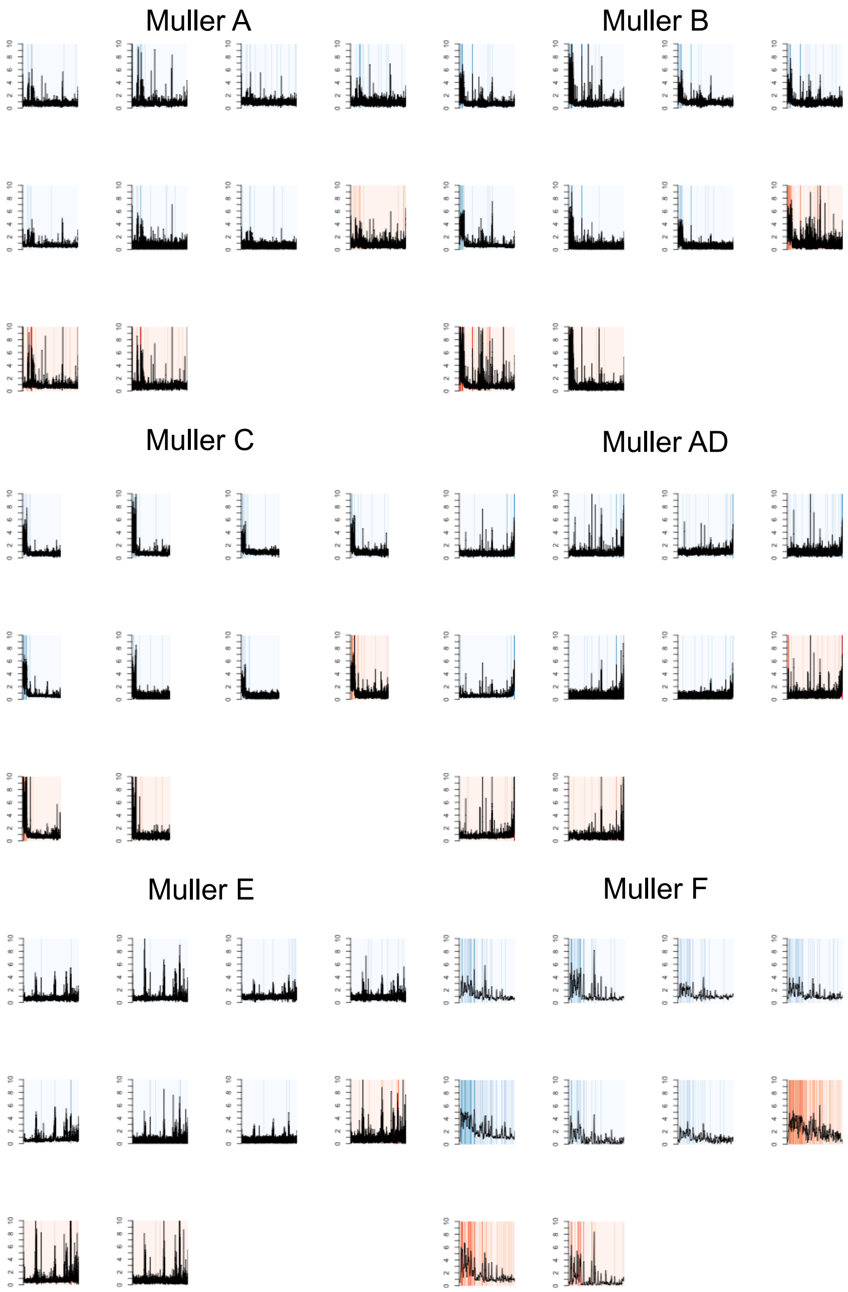


Figure S3. Genome-wide plots of H3K9me3 enrichment in stage 5a single embryos. Males are colored blue and females are colored red.

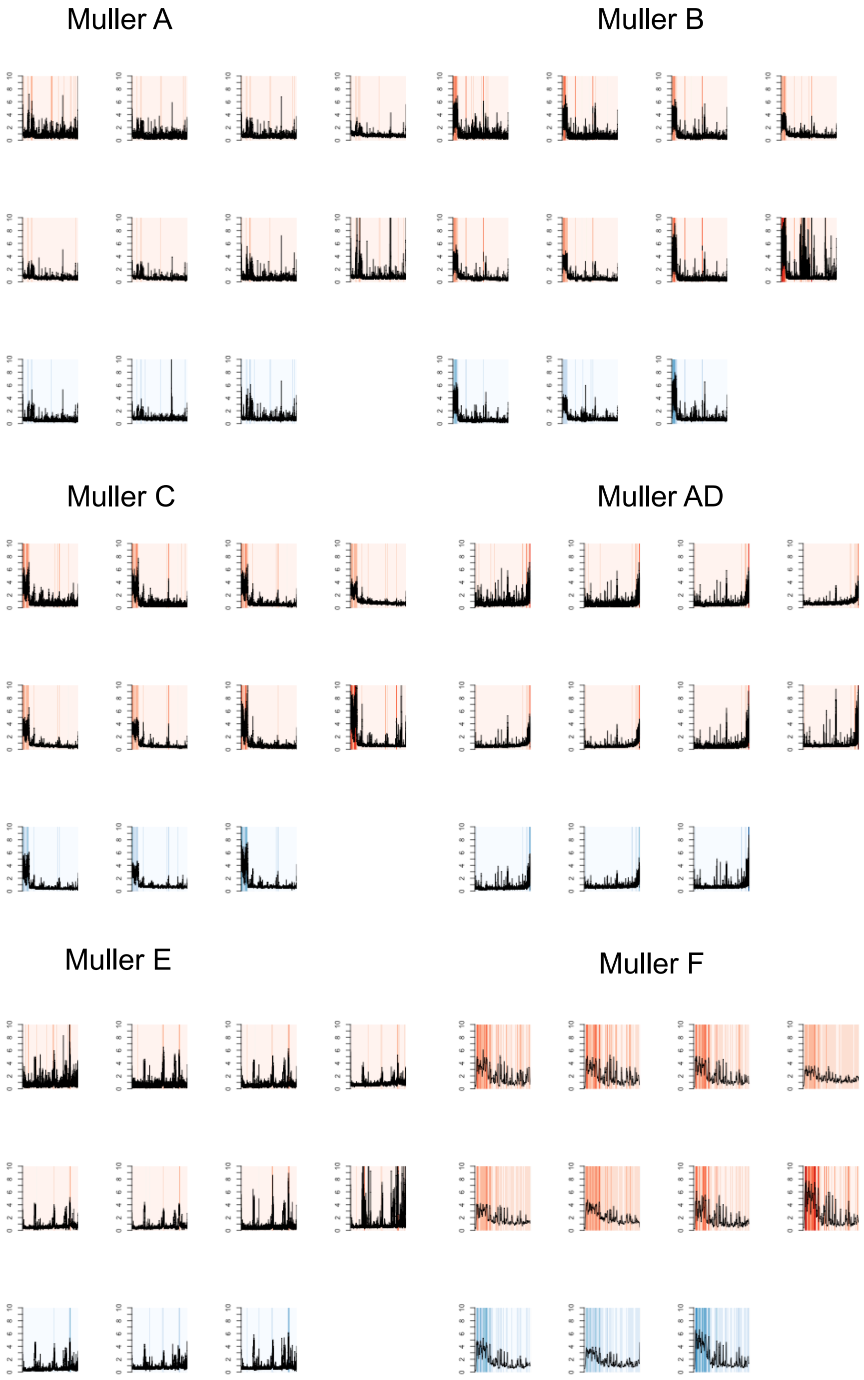


Figure S4. Genome-wide plots of H3K9me3 enrichment in stage 7 single embryos. Males are colored blue and females are colored red.

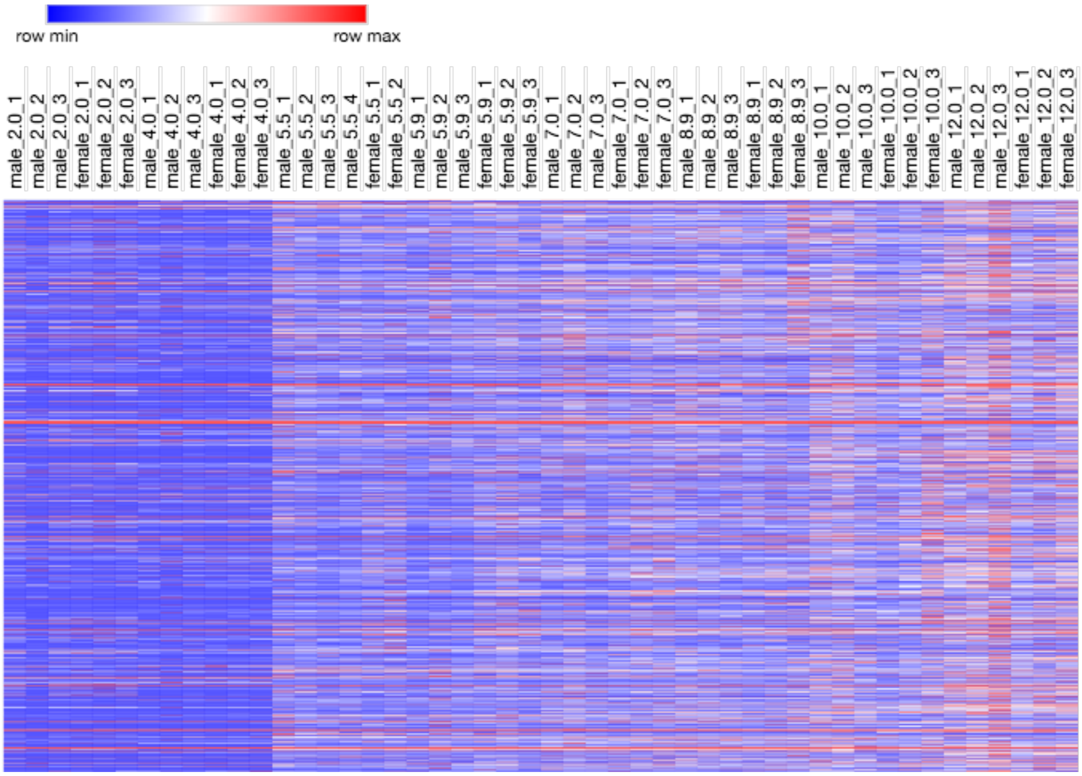


Figure S5. FPKM of zygotic genes in single embryos. Rows represent zygotic genes. Heat map is scaled by row.

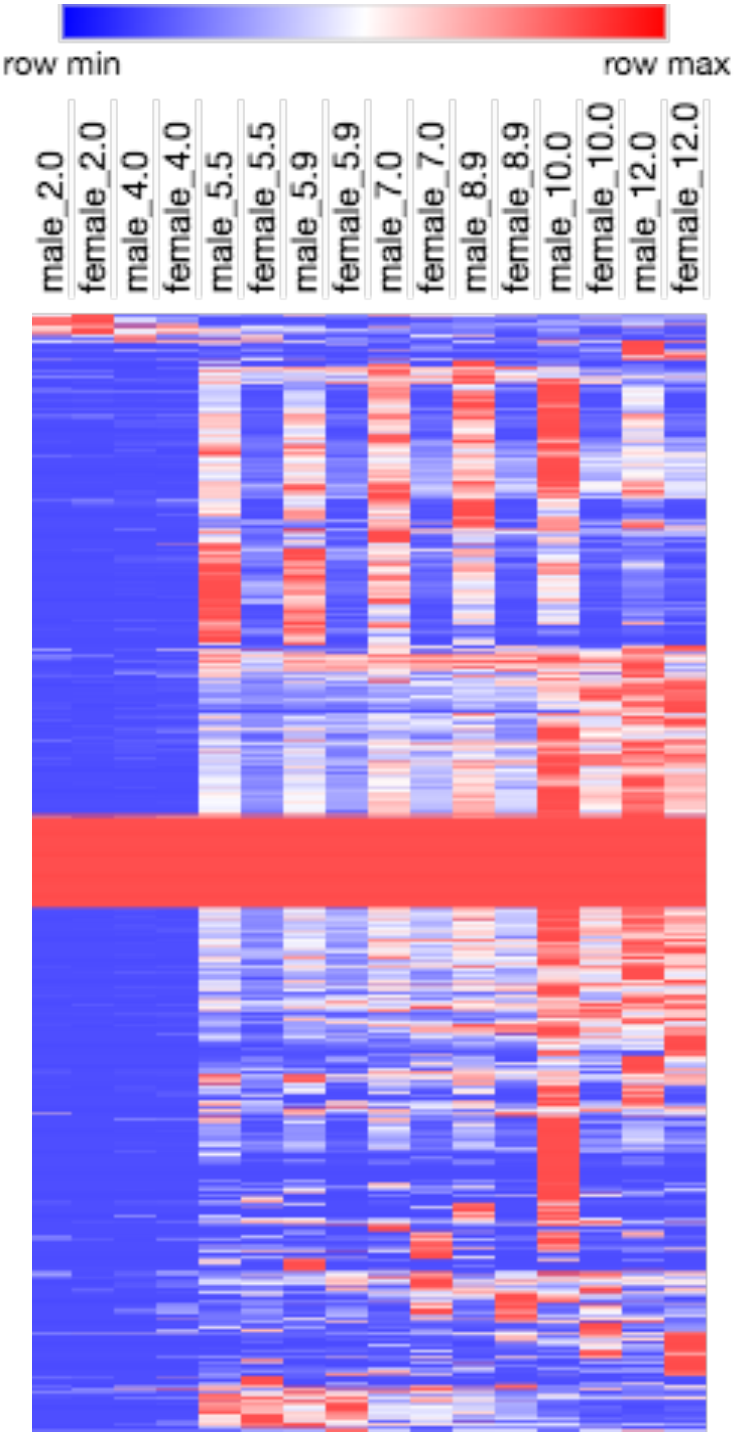


Figure S6. Heat map of median expression of repeats for males and females in stages 2-12 of embryogenesis. Rows represent repeats. Heat map is scaled by row.

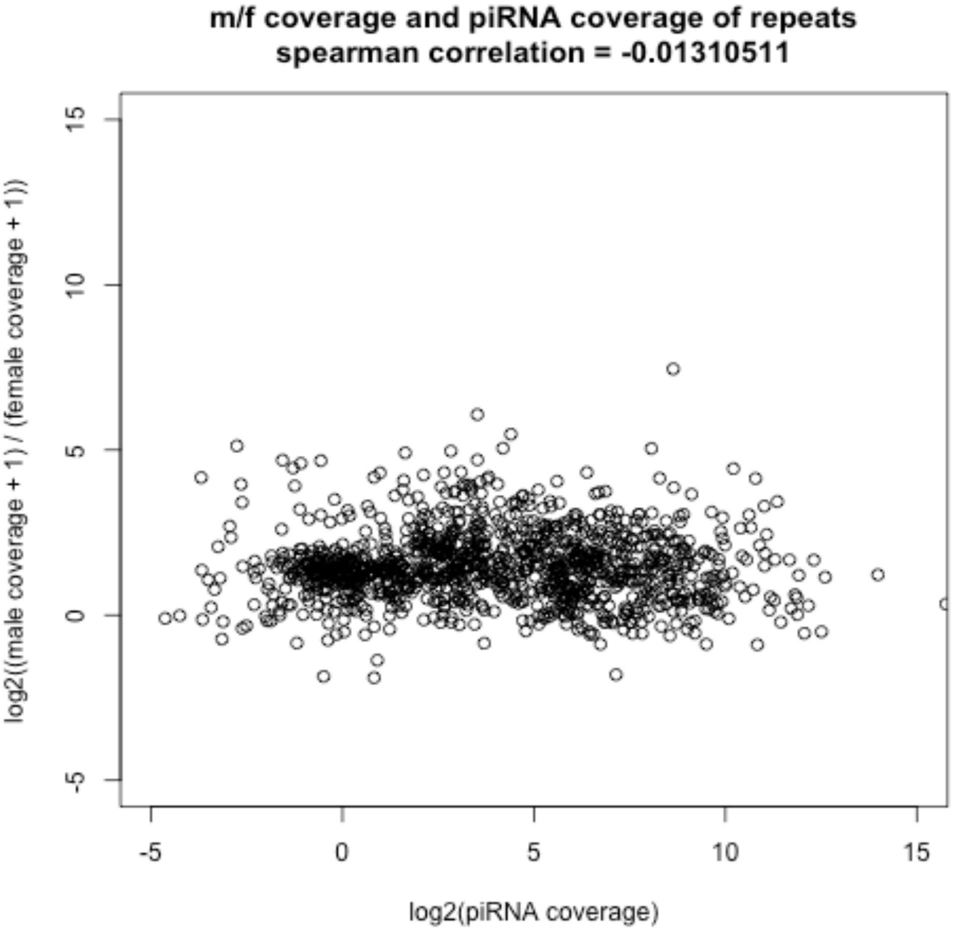


Figure S7. log2 of maternally deposited piRNA coverage of repeats vs. log2 of (male coverage + 1) / (female coverage + 1) of repeats.

Table S1. p-values for Figure 3B. Values were calculated using geom_signif in R.

stage A	stage B	p-value (Wilcoxon test)
2	4	2.24E-05
2	5.5	< 2.2e-16
2	5.9	< 2.2e-16
2	7	< 2.2e-16
2	8.9	< 2.2e-16
2	10	< 2.2e-16
2	12	< 2.2e-16
4	5.5	< 2.2e-16
4	5.9	< 2.2e-16
4	7	< 2.2e-16
4	8.9	< 2.2e-16
4	10	< 2.2e-16
4	12	< 2.2e-16
5.5	5.9	0.01454
5.5	7	0.0001963
5.5	8.9	0.05145
5.5	10	0.1413
5.5	12	1.55E-14
5.9	7	0.206
5.9	8.9	0.5579
5.9	10	0.0001054
5.9	12	2.33E-07
7	8.9	0.07714
7	10	6.87E-07
7	12	0.0001547
8.9	10	0.0007567
8.9	12	1.35E-08
10	12	< 2.2e-16

Chapter 3: Characterization of piRNAs and piRNA clusters in *D. pseudoobscura* and *D. melanogaster*

Abstract

Piwi-interacting RNAs (piRNAs) are short (23-29 nt) RNAs that bind to PIWI proteins and direct post-transcriptional and transcriptional silencing of target transposons. Many piRNAs are sense or antisense to repeats, and are produced by loci called piRNA clusters. These clusters often contain fragments of many different repeats and act as traps for transposable elements, and piRNAs produced from these clusters use sequence homology to target transcripts of transposable elements for degradation, and induce transcriptional silencing of repeats through heterochromatin formation. Repeat content evolves quickly across species, but less is known how piRNA clusters co-evolve. I sequenced piRNAs from *D. melanogaster*, *D. pseudoobscura*, and *D. miranda* ovaries, testes, and 0-1 hour embryos, to identify and characterize piRNAs and piRNA clusters in these three species. Comparison of piRNA clusters reveals a dynamic picture of their evolution. We identify many piRNA clusters that are unique to a species, or that have newly emerged on the repeat-rich neo-Y chromosome of *D. miranda*. We also find several homologous clusters that are syntenic between *D. miranda* and *D. pseudoobscura*, but whose sequence composition and repeat content has dramatically changed between species. Thus, piRNA clusters co-evolve with the rapidly changing repeat content of species, both by *de novo* formation of clusters, and by turn-over of the repeats contained within a cluster.

Introduction

Transposable elements (TEs) comprise large fractions of eukaryotic genomes (Cavalier-Smith, 1985; Dover, 1982; John & Miklos, 1988). Their transposition can have deleterious effects, and host genomes have evolved strategies to repress TE activity, such as silencing repeats through heterochromatin formation (Brown & Bachtrog, 2017; Francisco & Lemos, 2014; Henikoff, 1996) and posttranscriptional repression through RNAi (Chirn et al., 2015; Ernst, Odom, & Kutter, 2017; Zanni et al., 2013). Repression of TE activity is especially important in the germline, where new TE insertions would be heritable (Chirn et al., 2015).

Germline transposon silencing is mediated by short (23-29 nt) Piwi-interacting RNAs (piRNAs), which are bound by PIWI proteins (Piwi, Ago3, and Aub) and direct post-transcriptional and transcriptional silencing of target transposons (Brennecke et al., 2007). Many piRNAs are sense or antisense to repeats, and are produced by loci called piRNA clusters. These clusters often contain fragments of many different repeats and act as “traps” for transposable elements (Brennecke et al., 2007).

Repeat content evolves quickly across species, but less is know how piRNA clusters co-evolve. Are piRNA clusters born *de novo*, or are the same clusters used? Are sequences within piRNA clusters conserved, or are TEs exchanged within them? piRNA cluster conservation has been studied for a handful of clusters annotated in *D. melanogaster* (Chirn et al., 2015), but given the interesting TE activity we observed in *D. pseudoobscura* and *D. miranda* (Chapter 2), and the presence of neo-sex chromosomes in *D. miranda* and their autosomal counterparts in *D. pseudoobscura*, we were interested to characterize piRNAs and piRNA clusters in these two species. Here we compare locations and sequences of piRNAs and piRNA clusters in *D. pseudoobsura* and *D. miranda* to address the question of how piRNA clusters evolve.

Results & Discussion

Repeat content differs among species

We used repdenovo (Chu, Nielsen, & Wu, 2016) to make repeat libraries for both *D. miranda* and *D. pseudoobscura* and RepeatModeler (http://www.repeatmasker.org) to make one repeat library for the two species combined (see **Methods**). **Figure 1** shows the proportion of different types of repeats annotated by RepeatModeler in *D. pseudoobscura* and *D. miranda*. *D. miranda* has a much larger proportion of helitrons, which have been found to play an important role in the *D. miranda* genome: the neo-X (Muller C), which is an autosome in *D. pseudoobscura*, was found to recruit helitrons in order to gain binding sites for the male specific lethal (MSL) complex to achieve dosage compensation in males (Ellison & Bachtrog, 2013).

Figure 2 shows comparisons of the repeat landscape of the *D. pseudoobscura* and *D. miranda* genomes using the RepeatModeler repeat annotation for both species (see **Methods**). *D. miranda* has more total base pairs mapping to most of the individual repeats (**Figure 2A*,*** right). This is also the case for more general classes of repeats (**Figure 2C**). We also calculate a higher total number of instances a specific repeat (**Figure 2B**) or class of repeats (**Figure 2D**) is annotated in the genome in *D. miranda* compared to *D. pseudoobscura*. This difference in repeat content may be due to the presence of the neo-Y chromosome in *D. miranda*, which is an autosome in *D. pseudoobscura*. Y chromosomes accumulate repetitive DNA (Ananiev, Barsky, Ilyin, & Ryzic, 1984; Junakovic, Terrinoni, Di Franco, Vieira, & Loevenbruck, 1998; Pimpinelli et al., 1995), and we see 71,091,108 bp (78%) of the *D. miranda* neo-Y chromosome annotated as repetitive using our RepeatModeler annotation, compared to 1,181,862 bp on its former homolog, the neo-X (5%) or its Muller C ortholog in *D. pseudoobscura*, which has 4,333,269 bp (18%) annotated as repetitive.

Of the five repeats with largest difference in the number of base pairs in *D. miranda* and *D. pseudoobscura*, with more base pairs in *D. miranda*, four are helitrons map to the ISY element, which is a binding site for the dosage compensation complex, and was found to have increased on the neoY in *D. miranda* (Ellison & Bachtrog, 2013). Of the five repeats with largest difference in the number of base pairs in *D. miranda* and *D. pseudoobscura*, with more base pairs in *D. pseudoobscura*, four are of unknown classifications and one is a LINE/CR1 element.

Characteristics of *Drosophila* piRNAs and small RNAs

We annotated our piRNAs and small RNAs (**Figures 3 and 4**) and characterized their length distributions (**Figure 5**).

We do not observe stark differences between species in the types of repeats targeted by piRNAs (**Figure 3**). In ovary, testis, and 0-1 hour embryos, *D. miranda* has the largest proportion of piRNAs mapping to genes, followed by *D. pseudoobscura* (**Figure 4**)*. D. melanogaster* has the smallest proportion of piRNA reads mapping to genes. We also see this trend for small RNAs from ovary and 0-1 hour embryos. However, for testis small RNAs, *D. pseudoobscura* has the largest proportion mapping to genes, followed by *D. miranda,* then *D. melanogaster*.

Ovarian piRNA length is similar among species (**Figure 5**), and 0-1 hour embryonic piRNA length is similar to ovarian piRNA length. This is expected, as piRNAs from 0-1 hour embryos are maternally deposited, and we expect these embryonic piRNAs to be a subset of ovarian piRNAs. Testis piRNA lengths differ the most, with the *D. miranda* distribution skewing right compared to the *D. melanogaster* and *D. pseudoobscura* distributions.

Small RNA distributions follow a similar pattern, with 0-1 hour embryonic small RNA length distributions almost identical in all three species (**Figure 5**). Ovarian small RNA lengths differ slightly more than ovarian piRNA lengths among species, but are still similar and similar to 0-1 hour embryonic small RNA length distributions. Once again, testis small RNA length distributions are less conserved. In this case, the *D. pseudoobscura* length distribution is skewed left compared to *D. melanogaster* and *D. miranda.*

piRNA clusters

We used a pipeline adapted from (Brennecke et al., 2007) to call piRNA clusters in *D. pseudoobscura* and *D. miranda* using piRNA reads from ovary and testis (see **Methods**). We annotated 216 piRNA clusters in *D. pseudoobscura* (**Table 1**) and 227 piRNA clusters in *D. miranda* (**Table 2**).

In *D. pseudoobscura*, the majority (138; 64%) of piRNA clusters are expressed in both ovary and testis. 63 (29%) are expressed as piRNA clusters only in ovary, and 15 (7%) are expressed as piRNA clusters only in testis. This is different from *D. miranda*, where the majority of our annotated clusters are expressed as piRNA clusters only in ovary (129; 57%), while 45 (20%) are expressed as piRNA clusters only in testis and 53 (23%) are expressed as clusters in both tissues.

We used gene synteny to detect homologous clusters between *D. pseudoobscura* and *D. miranda* (see **Methods**). The two species diverged only 2 million years ago (Bachtrog & Charlesworth, 2002), but transposable element content and piRNA clusters are known to turn over quickly between species and even within strains of a single species (Vieira & Biémont, 2004; Zanni et al., 2013). We detected 26 pairs of homologous clusters between the two species (**Table 3**), and of those, 16 with 1:1 homology (**Table 4**).

Clusters were identified as homologous based on gene synteny only, and we wanted to investigate whether the sequences of the clusters are conserved. We used our RepeatModeler repeat annotation and calculated the number of repeats shared by homologous clusters and the number of repeats found only in a cluster in one species (**Table 4**). Interestingly, repeats/repeat fragments within homologous piRNA clusters are not strongly conserved, and even homologous clusters that share many of the same repeats, also have many repeats unique to each species (for example, clusters p109 and m106 contain 106 shared repeats but also 130 and 117 repeats unique to *D. miranda* and *D. pseudoobscura*, respectively).

Figure 6 shows alignment plots for four homologous piRNA clusters between *D. pseudoobscura* and *D. miranda*. While there are clearly sequences that are conserved between species, we also see large differences in the repeat content and order within homologous clusters. Indeed, we also find clusters that we call homologous based on synteny but are so different in sequence that we could not make alignment plots for them.

D. miranda neo-Y piRNA clusters

D. miranda has a recently formed neo-Y chromosome, which has accumulated many repeats (71,091,108 bp; 78%) since diverging from the neo-X, which only has 1,181,862 bp (5%) annotated as repetitive, about 2 million years ago (Bachtrog & Charlesworth, 2002). We thus were interested in identifying potentially novel piRNA clusters on the *D. miranda* neo-Y chromosome compared to the neo-X. Based on alignments of homologous genes on the neo-X and neo-Y chromosome, we indeed find new piRNA clusters on the neo-Y chromosome that appear to be absent in the same syntenic position on the neo-X. **Figure 7** shows an example of a novel piRNA cluster on the neo-Y. This region contains many more repeats around the genes on the neo-Y compared to their counterparts on the neo-X, and we also see insertions of repetitive DNA between genes on the neo-Y that have become a piRNA cluster. Thus, the formation of piRNA clusters on the neo-Y chromosomes provides an interesting aspect of sex chromosome evolution: while it is known that Y chromosomes accumulate repeats and deleterious sequences, it is possible that some of these harmful sequences can also become coopted for genome defense.

Methods

Fly strains

We used *D. pseudoobscura* strain SS-R2 and *D. miranda* strain MSH22 kept at 18°C and *D. melanogaster* strain Oregon-R kept at 25°C.

Tissue collection

We dissected testes and ovaries from 3-8 day old virgin flies of *D. pseudoobscura* (strain SS-R2) and *D. miranda* (strain MSH22) reared at 18°C and *D. melanogaster* strain Oregon-R at 25°C on Bloomington food. Tissues were preserved in RNAlater (Ambion).

Embryo collection

Flies were allowed to lay on molasses plates with yeast paste for one hour and immediately flash frozen in liquid nitrogen and stored at -80°C.

RNA extraction

We used Trizol (Invitrogen) and GlycoBlue (Invitrogen) to extract and isolate total RNA.

piRNA isolation

We developed our own protocol for piRNA isolation based on $NaIO_4$ reaction and beta elimination (Abe et al., 2014; Cao et al., 2009; Gebert, Ketting, Zischler, & Rosenkranz, 2015; Kirino & Mourelatos, 2007; Ohara et al., 2007; Simon et al., 2011). We prepared 5x borate buffer (148 mM borax, 148 mM boric acid, pH 8.6: 0.139 g borax and 0.464 g boric acid in 15 ml H_2O), 1x borax buffer (30 mM borax, 40 mM boric acid, 50 mM NaOH, pH 9.5: 0.572 g borax, 0.093 g boric acid, 0.1 g NaOH in 50 ml H_2O), and 200 mM sodium periodate (0.214 g $NaIO_4$ in 5 ml H_2O).

We started with 20 ug total RNA in 13.5 ul water and added 4 ul 5x borate buffer and 2.5 ul freshly dissolved 200mM sodium periodate. This was incubated at room temperature for 10 minutes, then quenched with 2 ul gylcerol and incubated for an additional 10 minutes at room temperature, during which we poked holes in the tops of the tubes with sterile needles. We vacuum dried the reaction for 1 hour at room temperature, then added 50 ul 1x borax buffer and incubated for 90 minutes at 45°C. We added 1.33 ul GlycoBlue (Invitrogen) and 200 ul ice cold 100% EtOH and precipitated the reaction at -80°C overnight. We spun down the reaction at 4°C, maximum speed, for 15 minutes, removed the supernatant, and let the RNA pellet dry, and resuspended the RNA in 20 ul H_2O.

Library preparation and sequencing

20 ul of RNA prepared as described above (for our piRNA libraries) or 20 ug total RNA (for our small RNA libraries) was resolved on a 15% TBE-urea gel (Invitrogen). Custom 18-nt and 30-nt ladders were used to size select 19-29 nt long RNA, which was purified and used as input for library preparation using the Illumina TruSeq Small RNA Library Preparation Kit to ligate adapters, reverse transcribe and amplify libraries, and purify the cDNA constructs. The protocol included an additional size selection step after library preparation on a 6% TBE gel (Invitrogen).

50bp single-end sequencing of our samples was performed on an Illumina HiSeq 4000 at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Bioinformatics analysis

Adapters were trimmed from the reads using trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were mapped to our *D. pseudoobscura* and *D. miranda* repeat annotations (repdenovo) and our *D. pseudoobscura* + *D. miranda* (RepeatModeler) repeat annotation with bowtie2 (Langmead & Salzberg, 2012)using default parameters. Bedtools [39] was used to calculate coverage of our repeat library. We used the same method to map our reads to the genomes of *D. melanogaster* (FlyBase v. 6.12 (Gramates et al., 2017)), *D. pseudoobscura* (unpublished), and *D. miranda* (unpublished).

piRNA cluster calling

We first called clusters separately, using piRNA reads from ovary and testis. To call piRNA clusters, we identified all 1kb windows with > 99 mapped piRNA reads. We merged windows within 20 kb of each other and kept those with > 99 uniquely mapped piRNA reads. Of these intervals, we kept regions with > 5 different repeats (from our repdenovo repeat annotations) mapped as piRNA clusters. For a final annotation of piRNA clusters, we combined clusters annotated from ovary piRNA reads and clusters annotated from testis piRNA reads in each species. We used bedtools to merge any clusters within 20kb of each other.

While we detected 431 clusters in *D. pseudoobscura* (**Table 1**), the 216 we describe in the this chapter are the ones mapping to the longest scaffolds (Muller elements and two Y-linked scaffolds), as we only used the Muller elements, neo-Y scaffolds, and longest Y-linked scaffolds in our *D. miranda* analysis.

piRNA cluster calling pipeline validation

142 piRNA clusters were annotated from *D. melanogaster* ovary piRNA using the Release 5 assembly(Brennecke et al., 2007). We used the Coordinates Converter tool on flybase.org to update the coordinates to the current assembly (v. 6.12) (Gramates et al., 2017). This left us with 96 clusters with coordinates in the current assembly.

We compared these clusters to the 249 clusters called by our pipeline using *D. melanogaster* piRNA. Of the 96 clusters, 13% (12) were 100% covered by our annotated clusters, 28% (27) were covered >= 75% by our annotated clusters, 43% (41) were covered >= 50% by our annotated clusters, 64% (61) were covered >= 25% by our annotated clusters, and 2% (2) did not overlap at all with clusters called using our pipeline.

Repeat annotation

Within-species annotation

We annotated *D. pseudoobscura* and *D. miranda* repeats using REPdenovo (downloaded November 7, 2016) (Brennecke et al., 2008; Chu et al., 2016). REPdenovo was run on raw sequencing reads using the parameters MIN_REPEAT_FREQ 3, RANGE_ASM_FREQ_DEC 2, RANGE_ASM_FREQ_GAP 0.8, K_MIN 30, K_MAX 50, K_INC 10, K_DFT 30, GENOME_LENGTH 196210428, ASM_NODE_LENGTH_OFFSET -1, MIN_CONTIG_LENGTH 100, IS_DUPLICATE_REPEATS 0.85, COV_DIFF_CUTOFF 0.5, MIN_SUPPORT_PAIRS 20, MIN_FULLY_MAP_RATIO 0.2, TR_SIMILARITY 0.85, and RM_CTN_CUTOFF 0.9. We only kept REPdenovo repeats with blast hits to the Repbase *Drosophila* repeat annotation (downloaded March 22, 2016) (Bao, Kojima, & Kohany, 2015).

We used tblastn (Boratyn et al., 2012) with the parameters -evalue 1e-6, -num_alignments 1, and -num_descriptions 1 to blast translated *D. pseudoobscura* genes (release 3.04) from FlyBase (Gramates et al., 2017) to repeats without hits to Repbase repeats. We eliminated any repeats with blast hits to *D. pseudoobscura* genes.

For *D. pseudoobscura*, this resulted in 1,194 repeats with an N50 of 427, totaling 905,416 base pairs. For *D. miranda*, this resulted in 999 repeats with an N50 of 631, totaling 964,435 base pairs.

Between-species repeat annotation

We concatenated our *D. pseudoobscura* and *D. miranda* genomes and ran RepeatModeler version 1.0.5 (http://www.repeatmasker.org) to make a single repeat annotation.

piRNA cluster homology analysis

We determined homologous genes between *D. pseudoobscura* and *D. miranda* and used gene synteny to identify potentially homologous clusters. Using only genes with homologs in the other species, we used bedtools to determine the three upstream and three downstream genes flanking each annotated piRNA cluster that did not overlap with the cluster. We called clusters as homologous if they shared at least one upstream and one downstream gene.

Data visualization

Figures 1, **3**, **4**, and **5** were made using Google Sheets. Genome browser plots in **Figure 7** were made using our lab genome browser. Plots in **Figure 2** were made in R. Alignment plots (**Figure 6**) were made using nucmer with the parameter “-c 200”, delta-filter with the parameters “-r –q –i 50 –l 50”, and mummerplot.

References

Abe, M., Naqvi, A., Hendriks, G. J., Feltzin, V., Zhu, Y., Grigoriev, A., et al. (2014). Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in drosophila. *Genes & Development, 28*(1), 44-57. doi:10.1101/gad.226654.113 [doi]

Ananiev, E., Barsky, V., Ilyin, Y. V., & Ryzic, M. (1984). The arrangement of transposable elements in the polytene chromosomes of drosophila melanogaster. *Chromosoma, 90*(5), 366-377.

Bachtrog, D., & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature, 416*(6878), 323-326.

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA, 6*(1), 11.

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct, 7*(1), 12.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in drosophila. *Cell, 128*(6), 1089-1103.

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, N.Y.), 322*(5906), 1387-1392. doi:10.1126/science.1165171 [doi]

Brown, E., & Bachtrog, D. (2017). The drosophila Y chromosome affects heterochromatin integrity genome-wide. *bioRxiv,* , 156000.

Cao, F., Li, X., Hiew, S., Brady, H., Liu, Y., & Dou, Y. (2009). Dicer independent small RNAs associate with telomeric heterochromatin. *RNA (New York, N.Y.), 15*(7), 1274-1281. doi:10.1261/rna.1423309 [doi]

Cavalier-Smith, T. (1985). Cell volume and the evolution of eukaryote genome size. *The Evolution of Genome Size,* , 105-184.

Chirn, G., Rahman, R., Sytnikova, Y. A., Matts, J. A., Zeng, M., Gerlach, D., et al. (2015). Conserved piRNA expression from a distinct set of piRNA cluster loci in eutherian mammals. *PLoS Genetics, 11*(11), e1005652.

Chu, C., Nielsen, R., & Wu, Y. (2016). REPdenovo: Inferring de novo repeat motifs from short sequence reads. *PloS One, 11*(3), e0150719.

Dover, G. A. (1982). *Genome evolution* Academic Press.

Ellison, C. E., & Bachtrog, D. (2013). Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science (New York, N.Y.), 342*(6160), 846-850. doi:10.1126/science.1239552 [doi]

Ernst, C., Odom, D. T., & Kutter, C. (2017). The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nature Communications, 8*, 1411.

Francisco, F. O., & Lemos, B. (2014). How do y-chromosomes modulate genome-wide epigenetic states: Genome folding, chromatin sinks, and gene expression. *Journal of Genomics, 2*, 94-103. doi:10.7150/jgen.8043 [doi]

Gebert, D., Ketting, R. F., Zischler, H., & Rosenkranz, D. (2015). piRNAs from pig testis provide evidence for a conserved role of the piwi pathway in post-transcriptional gene regulation in mammals. *PloS One, 10*(5), e0124860.

Gramates, L. S., Marygold, S. J., Santos, G. d., Urbano, J., Antonazzo, G., Matthews, B. B., et al. (2017). FlyBase at 25: Looking to the future. *Nucleic Acids Research, 45*(D1), D663-D671.

Henikoff, S. (1996). Dosage‐dependent modification of position‐effect variegation in drosophla. *Bioessays, 18*(5), 401-409.

John, B., & Miklos, G. 1988The eukaryote genome in development and evolution.

Junakovic, N., Terrinoni, A., Di Franco, C., Vieira, C., & Loevenbruck, C. (1998). Accumulation of transposable elements in the heterochromatin and on the Y chromosome of drosophila simulans and drosophila melanogaster. *Journal of Molecular Evolution, 46*(6), 661-668.

Kirino, Y., & Mourelatos, Z. (2007). Mouse piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. *Nature Structural and Molecular Biology, 14*(4), nsmb1218.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357-359.

Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., & Suzuki, T. (2007). The 3'termini of mouse piwi-interacting RNAs are 2'-O-methylated. *Nature Structural and Molecular Biology, 14*(4), 349-351.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., et al. (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), 3804-3808.

Simon, B., Kirkpatrick, J. P., Eckhardt, S., Reuter, M., Rocha, E. A., Andrade-Navarro, M. A., et al. (2011). Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a piwi protein. *Structure, 19*(2), 172-180.

Vieira, C., & Biémont, C. (2004). Transposable element dynamics in two sibling species: Drosophila melanogaster and drosophila simulans. *Drosophila melanogaster, drosophila simulans: So similar, so different* (pp. 115-123) Springer.

Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., et al. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proceedings of the National Academy of Sciences of the United States of America, 110*(49), 19842-19847. doi:10.1073/pnas.1313677110 [doi]

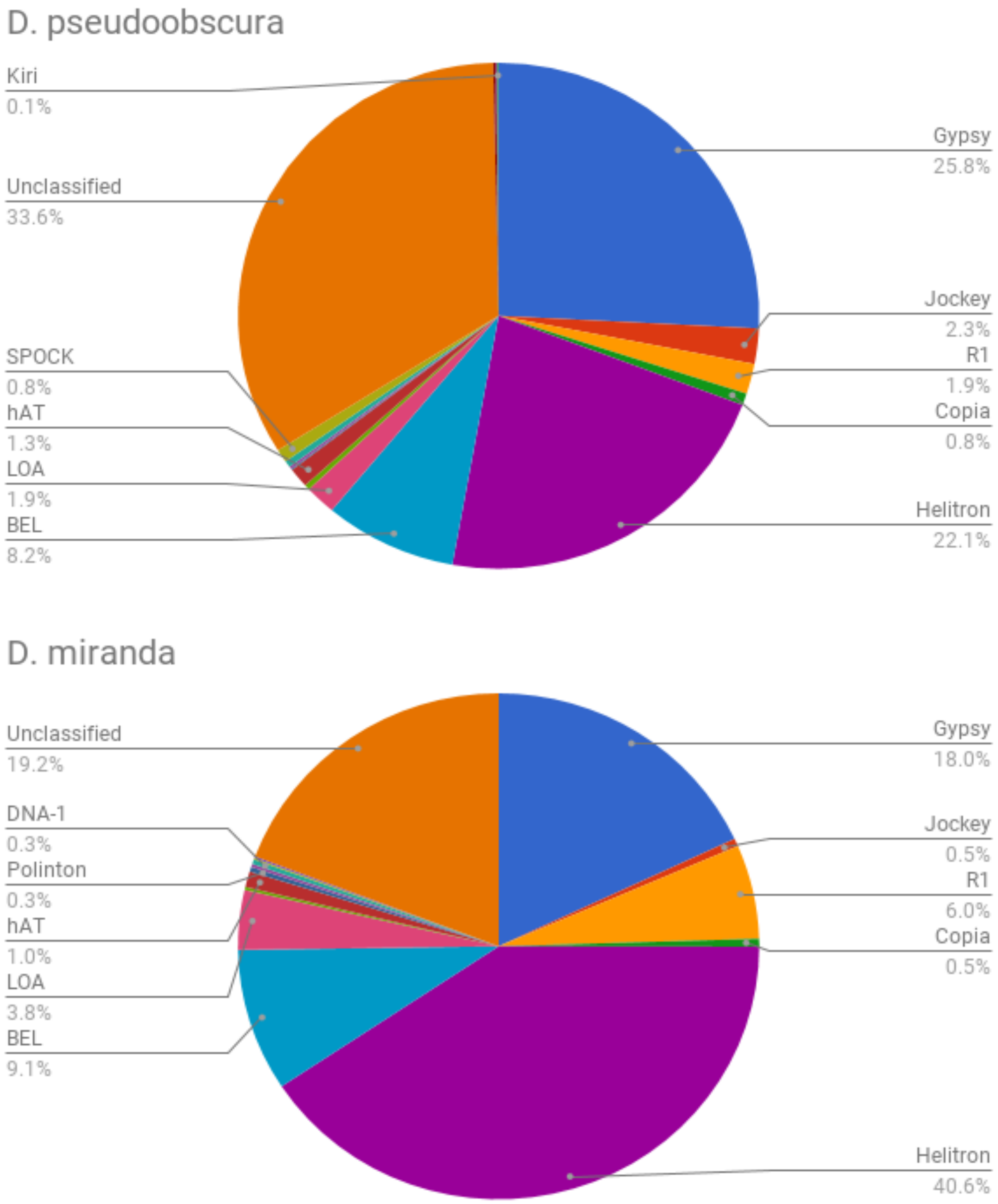


Figure 1. Classification of *D. pseudoobscura* and *D. miranda* annotated repeats using RepeatModeler annotation.

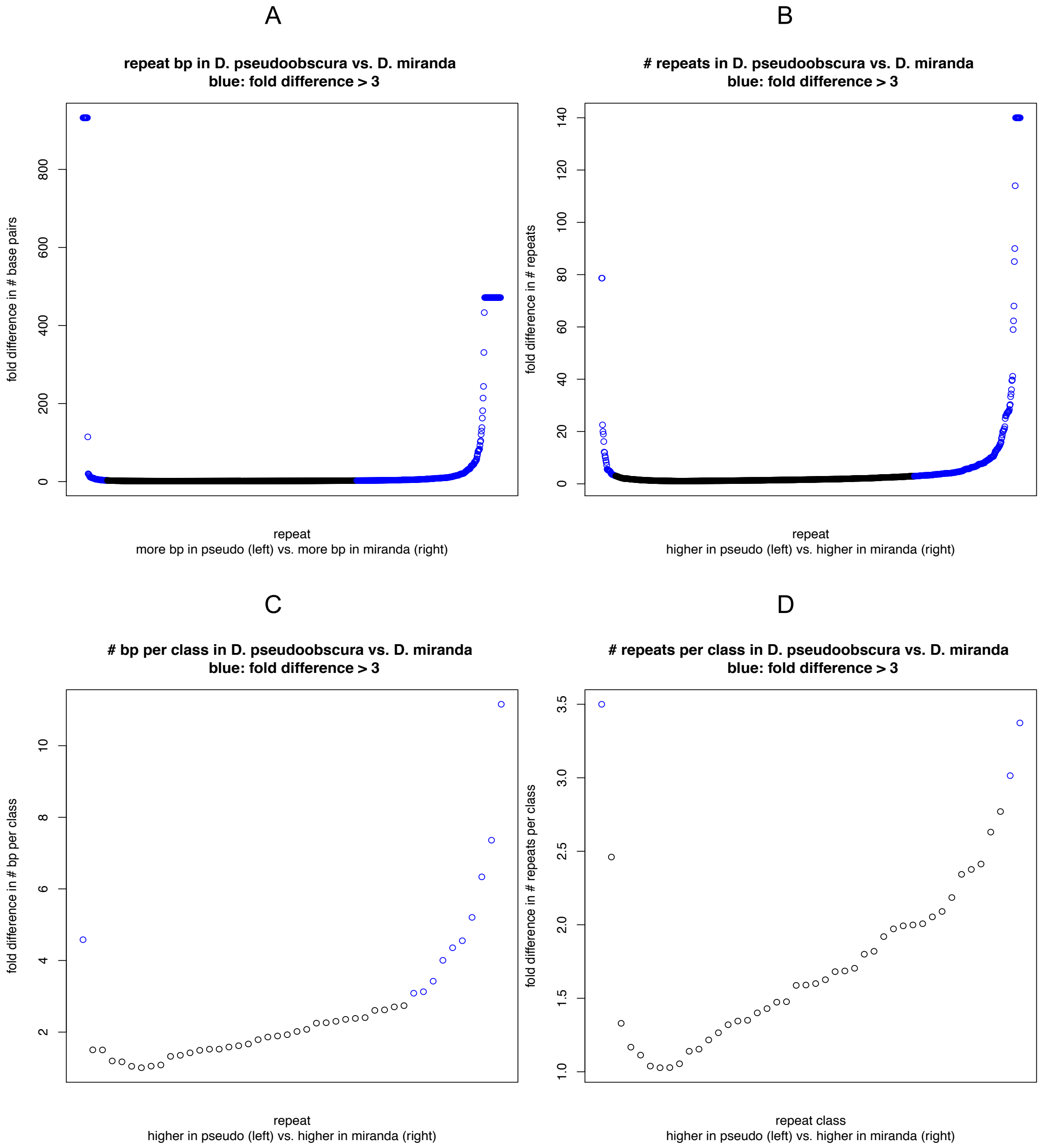


Figure 6. Three alignments of homologous clusters between *D. pseudoobscura* and *D. miranda*. A) *D. pseudoobscura* cluster p110 vs. *D. miranda* cluster m107. B) *D. pseudoobscura* cluster p105 vs. *D. miranda* cluster m105. C) *D. pseudoobscura* cluster p154 vs. *D. miranda* cluster m126.

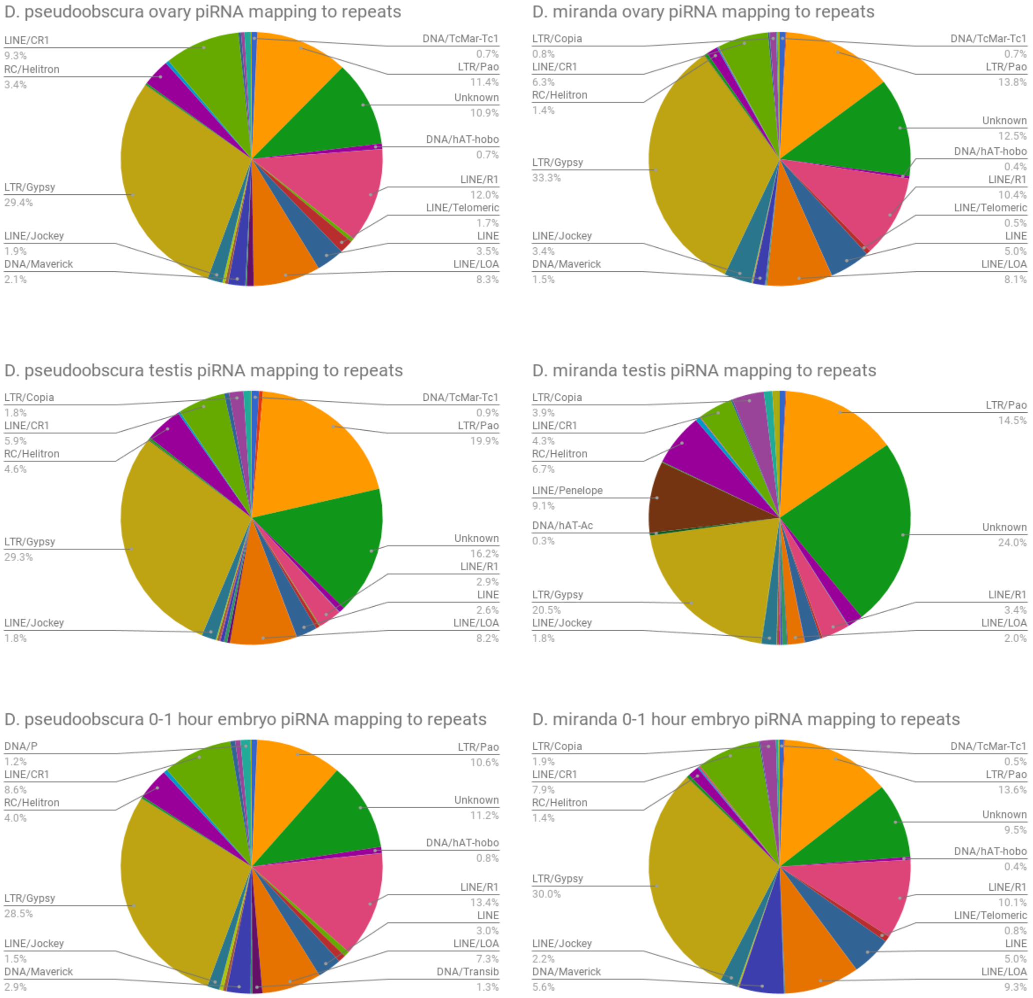
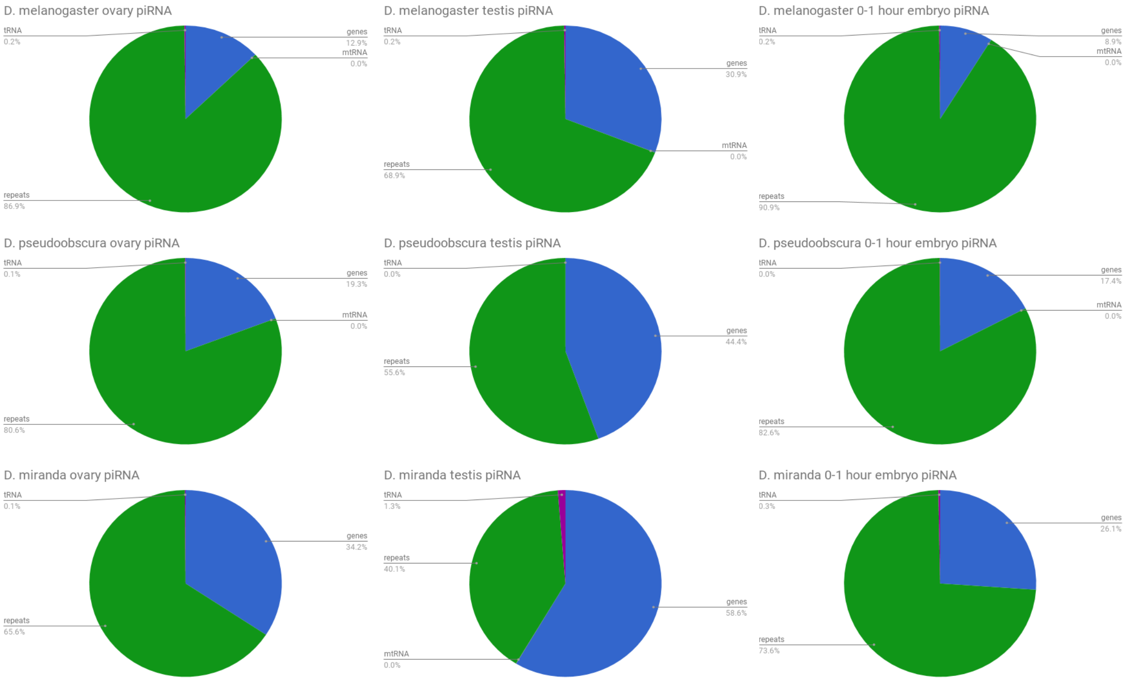


Figure 3. Classification of *D. pseudoobscura* and *D. miranda* ovary, testis, and 0-1 hour embryo piRNAs mapping to repeats using RepeatModeler annotation.

piRNA annotations



smRNA annotations

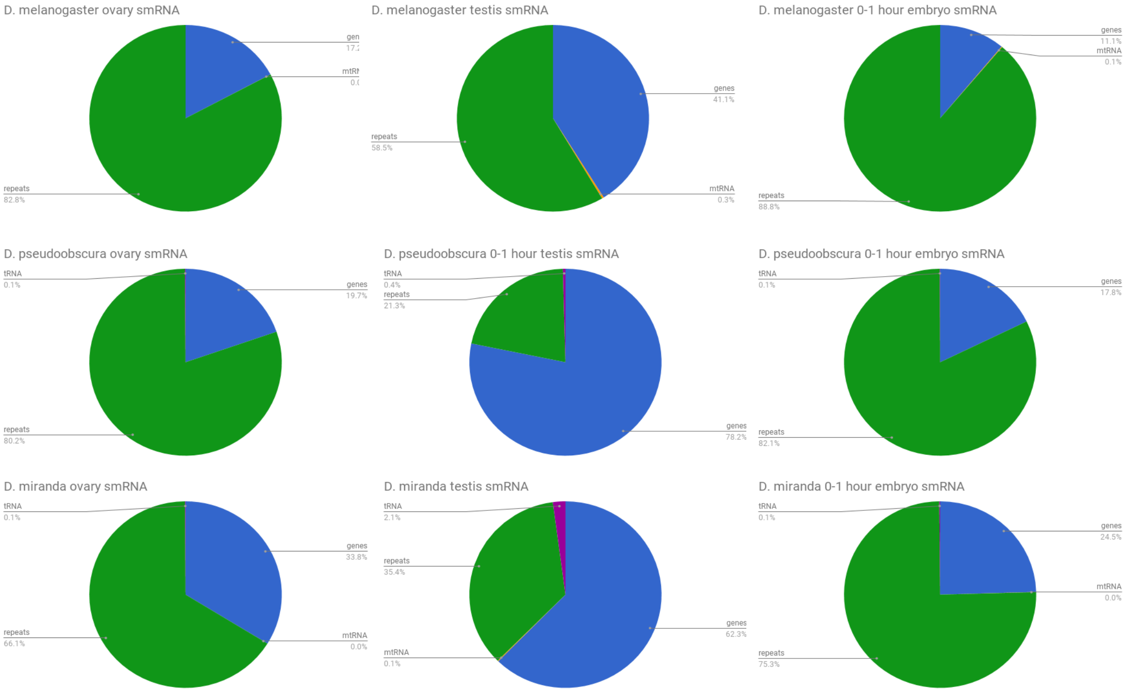
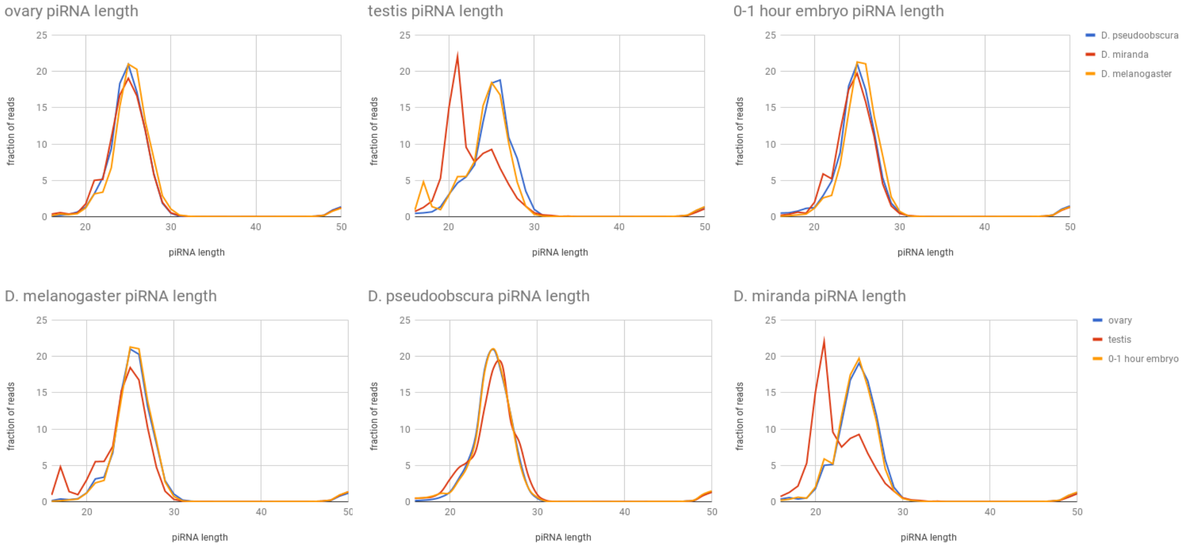


Figure 4. Annotation of our piRNA and small RNA libraries using within-species repeat libraries (repdenovo).

piRNA length distributions



smRNA length distributions

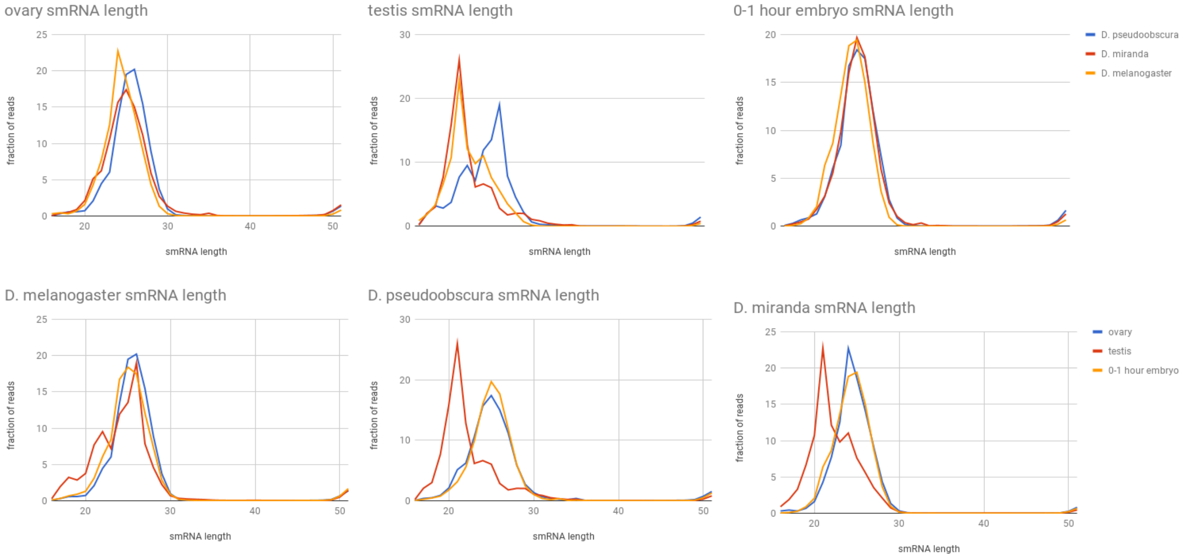
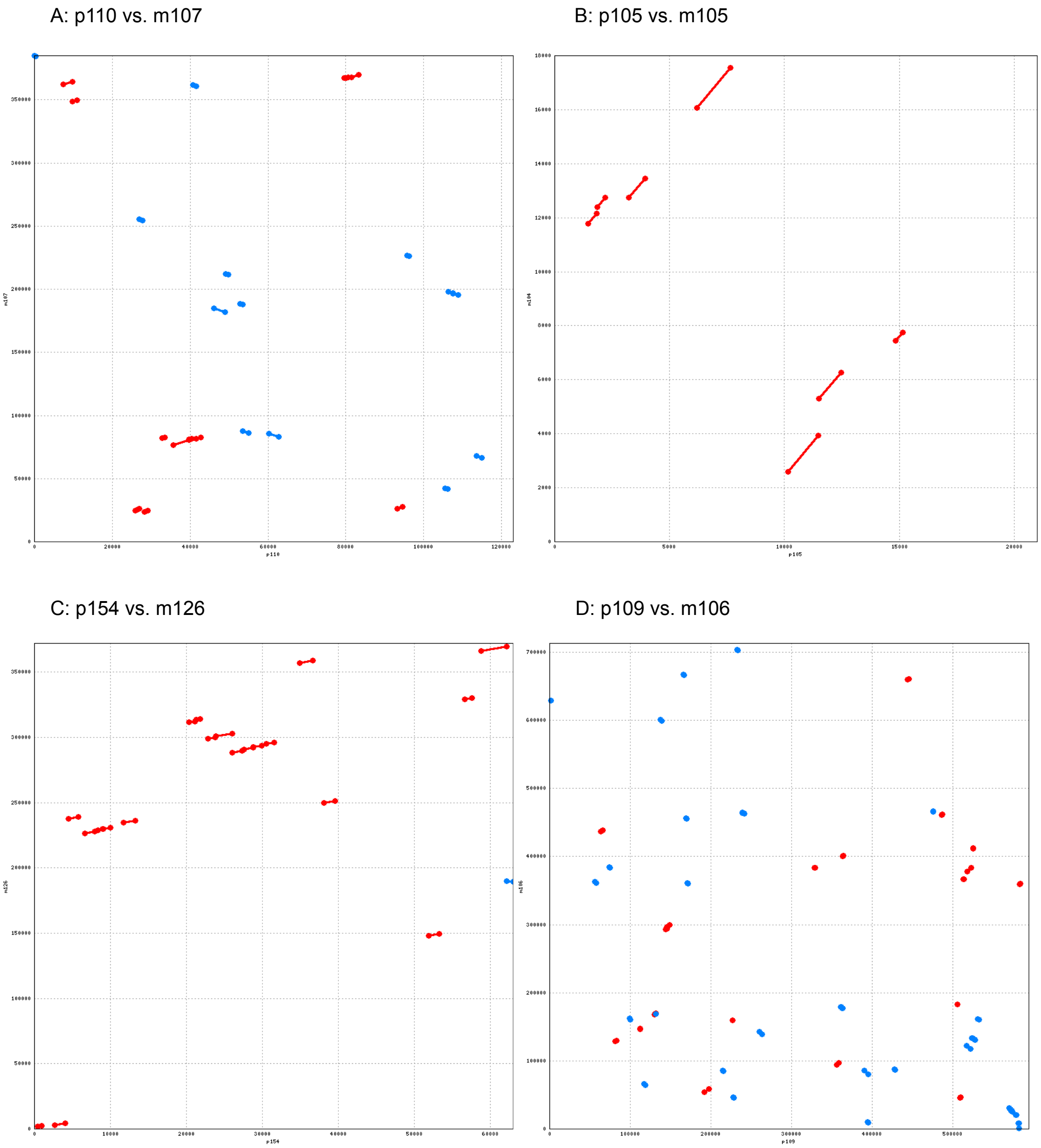


Figure 5. Length distributions of our piRNA and small RNA libraries.



A: p110 vs. m107

B: p105 vs. m105

C: p154 vs. m126

D: p109 vs. m106

Figure 6. Three alignments of homologous clusters between *D. pseudoobscura* and *D. miranda*. A) *D. pseudoobscura* cluster p110 vs. *D. miranda* cluster m107. B) *D. pseudoobscura* cluster p105 vs. *D. miranda* cluster m105. C) *D. pseudoobscura* cluster p154 vs. *D. miranda* cluster m126.

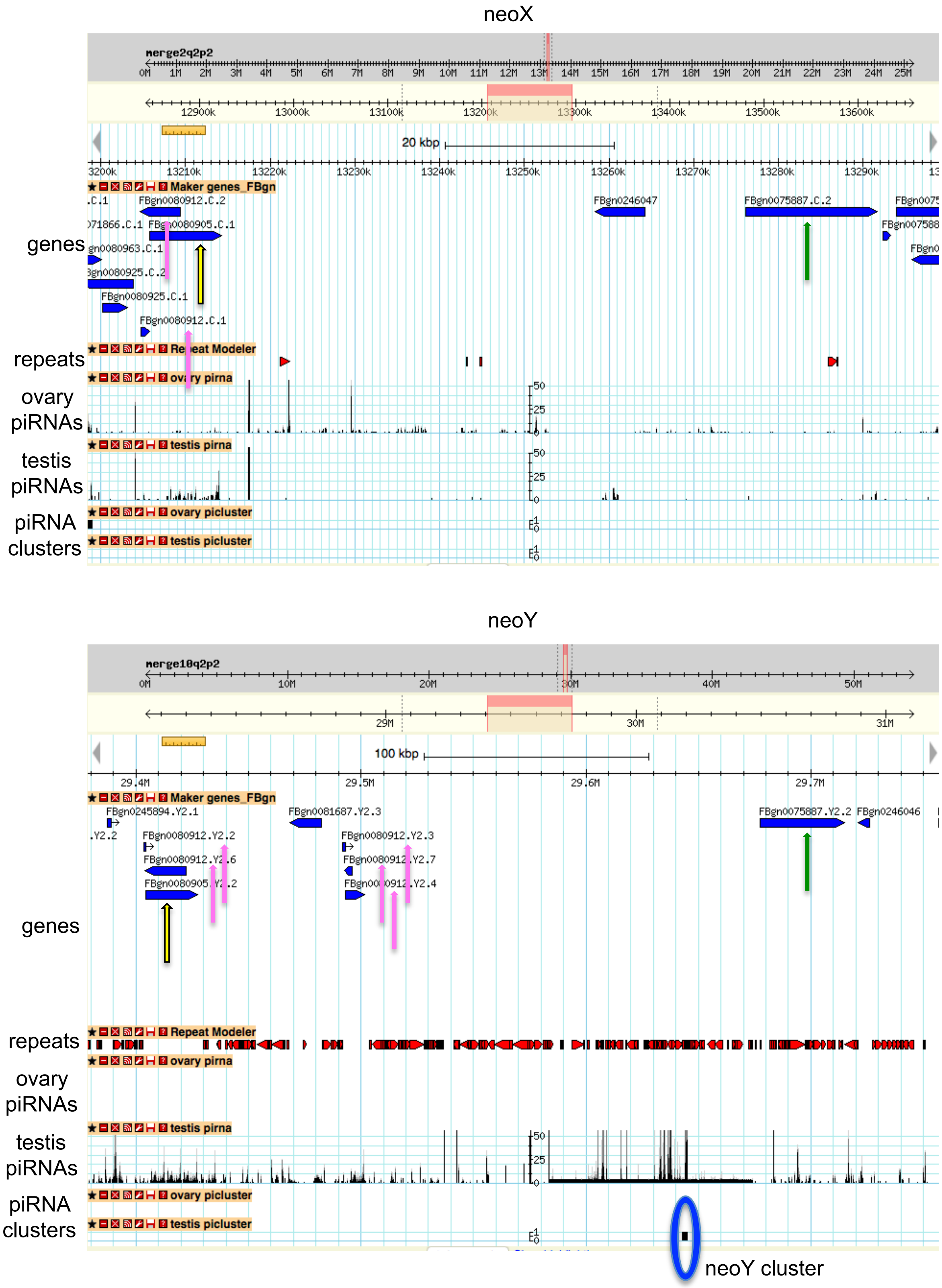


Figure 7. In *D. miranda*, an example of a new neo-Y cluster that not present on the neo-X. Homologous genes between the neo-X and neo-Y are shown using colored arrows.

Table 1. Table of piRNA clusters annotated in *D. pseudoobscura*.

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p1	scaf_merge1	B	1000	6000	5000	both	576
p2	scaf_merge1	B	45000	63000	18000	both	2540
p3	scaf_merge1	B	116000	121000	5000	ovary	1474
p4	scaf_merge1	B	145000	170000	25000	both	2153
p5	scaf_merge1	B	223000	234000	11000	testis	667
p6	scaf_merge1	B	335000	377000	42000	testis	6001
p7	scaf_merge1	B	402000	418000	16000	both	5773
p8	scaf_merge1	B	517000	563000	46000	both	6788
p9	scaf_merge1	B	586000	649000	63000	both	10762
p10	scaf_merge1	B	691000	734000	43000	both	3165
p11	scaf_merge1	B	766000	975000	209000	both	51735
p12	scaf_merge1	B	1002000	1017000	15000	both	2147
p13	scaf_merge1	B	1050000	1098000	48000	both	4597
p14	scaf_merge1	B	1120000	1191000	71000	both	5329
p15	scaf_merge1	B	1279000	1296000	17000	both	2482
p16	scaf_merge1	B	1332000	1408000	76000	both	8950
p17	scaf_merge1	B	1438000	1541000	103000	both	15294
p18	scaf_merge1	B	1580000	1590000	10000	testis	9014
p19	scaf_merge1	B	1622000	1630000	8000	both	1457
p20	scaf_merge1	B	1652000	1710000	58000	both	3558
p21	scaf_merge1	B	1835000	1842000	7000	testis	6124
p22	scaf_merge1	B	1885000	1916000	31000	both	3968
p23	scaf_merge1	B	2017000	2033000	16000	both	1025
p24	scaf_merge1	B	2115000	2146000	31000	both	3257
p25	scaf_merge1	B	2227000	2243000	16000	both	3561

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p26	scaf_merge1	B	2283000	2357000	74000	both	10328
p27	scaf_merge1	B	2421000	2535000	114000	both	38345
p28	scaf_merge1	B	2651000	2692000	41000	both	4716
p29	scaf_merge1	B	2723000	2882000	159000	both	54951
p30	scaf_merge1	B	2903000	2968000	65000	both	7460
p31	scaf_merge1	B	3028000	3100000	72000	both	13697
p32	scaf_merge1	B	3178000	3232000	54000	both	4653
p33	scaf_merge1	B	3256000	3317000	61000	both	12851
p34	scaf_merge1	B	3374000	3380000	6000	testis	191
p35	scaf_merge1	B	5412000	5435000	23000	both	58337
p36	scaf_merge1	B	5943000	5956000	13000	both	2464
p37	scaf_merge1	B	14122000	14135000	13000	both	3886
p38	scaf_merge1	B	14248000	14261000	13000	testis	2661
p39	scaf_merge1	B	20686000	20690000	4000	testis	459
p40	scaf_merge1	B	20788000	20792000	4000	testis	1126
p41	scaf_merge1	B	20839000	20841000	2000	testis	14
p42	scaf_merge1	B	21253000	21257000	4000	testis	191
p43	scaf_merge1	B	21300000	21306000	6000	testis	440
p44	scaf_merge1	B	23358000	23367000	9000	both	584
p45	scaf_merge1	B	24005000	24012000	7000	testis	220
p46	scaf_merge1	B	24496000	24501000	5000	testis	2791
p47	scaf_merge1	B	24862000	24864000	2000	both	396
p48	scaf_merge1	B	27409000	27413000	4000	both	521
p49	scaf_merge1	B	29274000	29278000	4000	both	32
p50	scaf_merge1	B	34000000	34007000	7000	testis	402
p51	scaf_merge1	B	35490000	35507000	17000	both	34581

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p52	scaf_merge2	C	17000	44000	27000	both	7850
p53	scaf_merge2	C	137000	195000	58000	both	15529
p54	scaf_merge2	C	236000	369000	133000	both	17245
p55	scaf_merge2	C	412000	464000	52000	both	4598
p56	scaf_merge2	C	485000	600000	115000	both	29868
p57	scaf_merge2	C	646000	694000	48000	both	11927
p58	scaf_merge2	C	767000	820000	53000	both	9135
p59	scaf_merge2	C	844000	931000	87000	both	21639
p60	scaf_merge2	C	1050000	1104000	54000	both	6873
p61	scaf_merge2	C	1163000	1168000	5000	testis	195
p62	scaf_merge2	C	1202000	1209000	7000	both	555
p63	scaf_merge2	C	1275000	1315000	40000	both	15166
p64	scaf_merge2	C	1363000	1381000	18000	testis	3966
p65	scaf_merge2	C	1406000	1439000	33000	both	9420
p66	scaf_merge2	C	1468000	1616000	148000	both	46573
p67	scaf_merge2	C	1682000	1784000	102000	both	16502
p68	scaf_merge2	C	1839000	1852000	13000	testis	4329
p69	scaf_merge2	C	1904000	1923000	19000	both	2770
p70	scaf_merge2	C	1966000	2087000	121000	both	25182
p71	scaf_merge2	C	2117000	2266000	149000	both	29020
p72	scaf_merge2	C	2297000	2347000	50000	both	6319
p73	scaf_merge2	C	2375000	2686000	311000	both	73420
p74	scaf_merge2	C	2732000	2880000	148000	both	23748
p75	scaf_merge2	C	2995000	3000000	5000	testis	252
p76	scaf_merge2	C	3093000	3100000	7000	both	1882
p77	scaf_merge2	C	3670000	3675000	5000	testis	112

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p78	scaf_merge2	C	3909000	3917000	8000	testis	1524
p79	scaf_merge2	C	4037000	4044000	7000	testis	227
p80	scaf_merge2	C	4385000	4409000	24000	both	7881
p81	scaf_merge2	C	4469000	4516000	47000	both	15046
p82	scaf_merge2	C	4869000	4876000	7000	testis	2603
p83	scaf_merge2	C	4942000	4979000	37000	testis	5931
p84	scaf_merge2	C	5032000	5082000	50000	testis	12476
p85	scaf_merge2	C	5594000	5603000	9000	testis	1643
p86	scaf_merge2	C	6690000	6697000	7000	both	4194
p87	scaf_merge2	C	7992000	7999000	7000	both	2774
p88	scaf_merge2	C	11046000	11054000	8000	both	195
p89	scaf_merge2	C	11194000	11202000	8000	both	270
p90	scaf_merge2	C	12780000	12785000	5000	both	2215
p91	scaf_merge2	C	12841000	12850000	9000	both	3371
p92	scaf_merge2	C	14135000	14138000	3000	testis	18
p93	scaf_merge2	C	15141000	15163000	22000	both	6241
p94	scaf_merge2	C	15254000	15261000	7000	testis	258
p95	scaf_merge2	C	19127000	19136000	9000	both	221
p96	scaf_merge2	C	19404000	19406000	2000	testis	32
p97	scaf_merge2	C	22263000	22265000	2000	both	0
p98	scaf_merge2	C	23814000	23864000	50000	both	254062
p99	scaf_merge3	A	0	154000	154000	both	29554
p100	scaf_merge3	A	195000	253000	58000	both	10202
p101	scaf_merge3	A	936000	1060000	124000	both	32245
p102	scaf_merge3	A	1306000	1307000	1000	both	3480
p103	scaf_merge3	A	1551000	1567000	16000	both	13682

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p104	scaf_merge3	A	1945000	2486000	541000	both	2457103
p105	scaf_merge3	A	2654000	2675000	21000	testis	6664
p106	scaf_merge3	A	2698000	2705000	7000	ovary	685
p107	scaf_merge3	A	2737000	2776000	39000	both	19550
p108	scaf_merge3	A	2879000	2882000	3000	testis	47
p109	scaf_merge3	A	3566000	4160000	594000	both	1219515
p110	scaf_merge3	A	4485000	4608000	123000	both	845055
p111	scaf_merge3	A	4774000	4789000	15000	both	33887
p112	scaf_merge3	A	5460000	5464000	4000	ovary	13213
p113	scaf_merge3	A	5499000	5507000	8000	both	5401
p114	scaf_merge3	A	7500000	7504000	4000	both	422
p115	scaf_merge3	A	8122000	8124000	2000	testis	84
p116	scaf_merge3	A	8183000	8192000	9000	both	91
p117	scaf_merge3	A	12203000	12204000	1000	testis	0
p118	scaf_merge3	A	12908000	12911000	3000	testis	0
p119	scaf_merge3	A	14649000	14654000	5000	both	1319
p120	scaf_merge3	A	14705000	14706000	1000	ovary	183
p121	scaf_merge3	A	18171000	18173000	2000	both	148
p122	scaf_merge3	A	18662000	18668000	6000	testis	4173
p123	scaf_merge3	A	19657000	19659000	2000	ovary	303
p124	scaf_merge3	A	20404000	20405000	1000	both	105
p125	scaf_merge3	A	20956000	20962000	6000	testis	137
p126	scaf_merge3	A	21128000	21144000	16000	both	4176
p127	scaf_merge3	A	22082000	22098000	16000	ovary	3984
p128	scaf_merge3	A	22189000	22192000	3000	both	487
p129	scaf_merge3	A	22800000	22801000	1000	testis	26

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p130	scaf_merge3	A	22992000	23036000	44000	both	94020
p131	scaf_merge4	AD	317000	421000	104000	both	203437
p132	scaf_merge4	AD	1064000	1066000	2000	both	530
p133	scaf_merge4	AD	1684000	1729000	45000	both	3640
p134	scaf_merge4	AD	2007000	2017000	10000	both	3223
p135	scaf_merge4	AD	2047000	2054000	7000	both	1760
p136	scaf_merge4	AD	2639000	2644000	5000	both	1124
p137	scaf_merge4	AD	3140000	3149000	9000	ovary	864
p138	scaf_merge4	AD	4281000	4288000	7000	both	3111
p139	scaf_merge4	AD	17705000	17706000	1000	testis	0
p140	scaf_merge4	AD	17920000	17927000	7000	both	17725
p141	scaf_merge4	AD	18720000	18756000	36000	ovary	32101
p142	scaf_merge4	AD	18801000	18824000	23000	testis	3008
p143	scaf_merge4	AD	18956000	18958000	2000	both	14
p144	scaf_merge4	AD	24205000	24206000	1000	testis	0
p145	scaf_merge4	AD	25133000	25144000	11000	ovary	5681
p146	scaf_merge4	AD	26274000	26285000	11000	both	1100
p147	scaf_merge4	AD	26352000	26361000	9000	testis	190
p148	scaf_merge4	AD	27240000	27260000	20000	both	50328
p149	scaf_merge4	AD	28163000	28169000	6000	ovary	5266
p150	scaf_merge4	AD	28355000	28362000	7000	testis	7210
p151	scaf_merge4	AD	29227000	29238000	11000	both	4183
p152	scaf_merge4	AD	29317000	29319000	2000	testis	27
p153	scaf_merge4	AD	29744000	29917000	173000	both	27423
p154	scaf_merge4	AD	29938000	30001000	63000	both	2836
p155	scaf_merge4	AD	30366000	30649000	283000	both	989247

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p156	scaf_merge4	AD	30832000	30837000	5000	testis	585
p157	scaf_merge4	AD	30885000	30930000	45000	both	10725
p158	scaf_merge4	AD	30977000	31028000	51000	both	14360
p159	scaf_merge4	AD	31049000	31071000	22000	both	673
p160	scaf_merge4	AD	31164000	31203000	39000	both	10049
p161	scaf_merge4	AD	31369000	31387000	18000	both	6110
p162	scaf_merge4	AD	31438000	31473000	35000	both	5193
p163	scaf_merge4	AD	31504000	31520000	16000	both	9499
p164	scaf_merge4	AD	31564000	31579000	15000	both	7610
p165	scaf_merge4	AD	31612000	31663000	51000	both	8160
p166	scaf_merge4	AD	31714000	32114000	400000	both	94982
p167	scaf_merge4	AD	32183000	32199000	16000	testis	2830
p168	scaf_merge4	AD	32306000	32309000	3000	testis	702
p169	scaf_merge4	AD	32332000	32339000	7000	testis	486
p170	scaf_merge4	AD	32402000	32407000	5000	testis	125
p171	scaf_merge4	AD	32465000	32488000	23000	both	3404
p172	scaf_merge4	AD	32554000	32561000	7000	both	758
p173	scaf_merge4	AD	32617000	32619000	2000	testis	23
p174	scaf_merge4	AD	32889000	32893000	4000	both	670
p175	scaf_merge4	AD	32965000	32975000	10000	both	1907
p176	scaf_merge4	AD	33179000	33182000	3000	testis	145
p177	scaf_merge4	AD	33273000	33305000	32000	both	9884
p178	scaf_merge4	AD	33441000	33467000	26000	testis	877
p179	scaf_merge4	AD	33530000	33600000	70000	both	334585
p180	scaf_merge5	E	49000	87000	38000	both	72939
p181	scaf_merge5	E	800000	808000	8000	both	7462

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p182	scaf_merge5	E	1599000	1600000	1000	testis	65
p183	scaf_merge5	E	2083000	2086000	3000	ovary	4900
p184	scaf_merge5	E	3158000	3169000	11000	both	3019
p185	scaf_merge5	E	7644000	7656000	12000	ovary	3295
p186	scaf_merge5	E	7940000	7947000	7000	both	2904
p187	scaf_merge5	E	8175000	8219000	44000	both	8834
p188	scaf_merge5	E	8290000	8301000	11000	both	579
p189	scaf_merge5	E	11201000	11203000	2000	both	41
p190	scaf_merge5	E	14408000	14410000	2000	both	1902
p191	scaf_merge5	E	16291000	16299000	8000	both	329
p192	scaf_merge5	E	19087000	19134000	47000	both	291608
p193	scaf_merge5	E	19171000	19182000	11000	both	1482
p194	scaf_merge5	E	20193000	20212000	19000	testis	204
p195	scaf_merge5	E	20236000	20240000	4000	testis	116
p196	scaf_merge5	E	20372000	20389000	17000	ovary	2395
p197	scaf_merge5	E	20533000	20542000	9000	both	742
p198	scaf_merge5	E	20800000	20806000	6000	both	214
p199	scaf_merge5	E	20874000	20876000	2000	testis	13
p200	scaf_merge5	E	24254000	24256000	2000	testis	1
p201	scaf_merge5	E	24704000	24706000	2000	testis	70
p202	scaf_merge5	E	26212000	26218000	6000	both	12589
p203	scaf_merge5	E	26536000	26538000	2000	both	0
p204	scaf_merge5	E	26705000	26721000	16000	both	12015
p205	scaf_merge5	E	26839000	26840000	1000	testis	24
p206	scaf_merge5	E	27739000	27750000	11000	both	2467
p207	scaf_merge5	E	27791000	27803000	12000	both	3833

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p208	scaf_merge5	E	28379000	28397000	18000	ovary	8713
p209	scaf_merge5	E	28933000	28935000	2000	testis	20
p210	scaf_merge5	E	30386000	30392000	6000	ovary	1786
p211	scaf_merge5	E	30484000	30982000	498000	both	140652
p212	scaf_merge5	E	31008000	31012000	4000	testis	1390
p213	scaf_merge6	Y	78000	89000	11000	testis	0
p214	scaf_merge6	Y	132000	145000	13000	testis	0
p215	scaf_merge6	Y	179000	188000	9000	testis	0
p216	scaf_merge8	Y	20000	41000	21000	testis	0
p217	scf7180000004798	NA	12000	21000	9000	testis	1920
p218	scf7180000004798	NA	53000	133000	80000	both	19872
p219	scf7180000004830	NA	0	17000	17000	testis	764
p220	scf7180000004831	NA	11000	20000	9000	both	8676
p221	scf7180000004923	NA	0	68000	68000	both	16963
p222	scf7180000004923	NA	91000	149000	58000	both	11836
p223	scf7180000004923	NA	193000	208000	15000	both	1842
p224	scf7180000004923	NA	268000	347000	79000	both	12352
p225	scf7180000004950	NA	0	15980	15980	both	612
p226	scf7180000004985	NA	2000	24000	22000	both	2703
p227	scf7180000005021	NA	3000	8000	5000	both	516
p228	scf7180000005028	NA	12000	76000	64000	both	51680
p229	scf7180000005028	NA	145000	179820	34820	both	8451
p230	scf7180000005041	NA	11000	23000	12000	testis	0
p231	scf7180000005077	NA	43000	55000	12000	testis	2814
p232	scf7180000005077	NA	82000	91000	9000	both	960
p233	scf7180000005077	NA	128000	136000	8000	both	692

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p234	scf7180000005077	NA	157000	237000	80000	both	13091
p235	scf7180000005077	NA	279000	293000	14000	both	2666
p236	scf7180000005081	NA	0	23500	23500	both	54816
p237	scf7180000005085	NA	0	5000	5000	testis	0
p238	scf7180000005105	NA	5000	21000	16000	testis	0
p239	scf7180000005118	NA	0	9941	9941	testis	0
p240	scf7180000005120	NA	2000	11000	9000	testis	0
p241	scf7180000005127	NA	1000	7000	6000	testis	0
p242	scf7180000005128	NA	15000	107000	92000	both	22847
p243	scf7180000005129	NA	43000	62000	19000	ovary	4637
p244	scf7180000005129	NA	91000	94000	3000	both	252
p245	scf7180000005135	NA	0	9000	9000	testis	0
p246	scf7180000005194	NA	0	11000	11000	testis	0
p247	scf7180000005243	NA	7000	24000	17000	testis	2092
p248	scf7180000005249	NA	15000	23000	8000	testis	648
p249	scf7180000005297	NA	18000	43000	25000	both	2827
p250	scf7180000005306	NA	23000	32000	9000	testis	0
p251	scf7180000005313	NA	3000	9000	6000	testis	0
p252	scf7180000005348	NA	11000	18000	7000	testis	0
p253	scf7180000005410	NA	0	97000	97000	both	48885
p254	scf7180000005423	NA	2000	5000	3000	testis	0
p255	scf7180000005428	NA	43000	77000	34000	both	7610
p256	scf7180000005432	NA	8000	20000	12000	testis	0
p257	scf7180000005438	NA	17000	25000	8000	testis	0
p258	scf7180000005449	NA	9000	15000	6000	testis	0
p259	scf7180000005457	NA	6000	155000	149000	both	49376

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p260	scf7180000005490	NA	0	12000	12000	testis	0
p261	scf7180000005502	NA	2000	12000	10000	testis	0
p262	scf7180000005506	NA	0	95000	95000	both	24826
p263	scf7180000005530	NA	16000	29875	13875	testis	0
p264	scf7180000005553	NA	0	29000	29000	both	8292
p265	scf7180000005559	NA	0	6000	6000	testis	0
p266	scf7180000005566	NA	0	9562	9562	both	2095
p267	scf7180000005568	NA	1000	7000	6000	testis	2478
p268	scf7180000005568	NA	41000	47900	6900	testis	14677
p269	scf7180000005574	NA	11000	15000	4000	testis	0
p270	scf7180000005593	NA	2000	17000	15000	ovary	28201
p271	scf7180000005598	NA	0	32000	32000	testis	11142
p272	scf7180000005669	NA	0	14835	14835	testis	0
p273	scf7180000005673	NA	0	8000	8000	both	2065
p274	scf7180000005676	NA	0	11475	11475	both	2905
p275	scf7180000005679	NA	3000	12000	9000	testis	0
p276	scf7180000005726	NA	2000	9000	7000	testis	0
p277	scf7180000005736	NA	4000	18000	14000	testis	0
p278	scf7180000005737	NA	37000	40723	3723	testis	0
p279	scf7180000005748	NA	2000	10000	8000	testis	0
p280	scf7180000005750	NA	15000	27000	12000	testis	0
p281	scf7180000005805	NA	8000	36279	28279	testis	1457
p282	scf7180000005813	NA	3000	9000	6000	testis	0
p283	scf7180000005842	NA	12000	13000	1000	testis	0
p284	scf7180000005893	NA	0	8000	8000	testis	0
p285	scf7180000005926	NA	4000	17000	13000	testis	0

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p286	scf7180000005951	NA	1000	8000	7000	testis	0
p287	scf7180000005970	NA	2000	9000	7000	testis	0
p288	scf7180000005979	NA	0	8000	8000	testis	0
p289	scf7180000006026	NA	0	25682	25682	ovary	26371
p290	scf7180000006027	NA	0	62000	62000	both	119039
p291	scf7180000006065	NA	2000	12000	10000	both	1315
p292	scf7180000006098	NA	0	5000	5000	testis	0
p293	scf7180000006108	NA	0	30150	30150	ovary	32131
p294	scf7180000006109	NA	0	17000	17000	both	24788
p295	scf7180000006148	NA	0	153866	153866	both	344220
p296	scf7180000006150	NA	0	39484	39484	both	61547
p297	scf7180000006213	NA	15000	70338	55338	both	25252
p298	scf7180000006219	NA	0	18000	18000	testis	0
p299	scf7180000006233	NA	0	19000	19000	testis	1352
p300	scf7180000006269	NA	2000	14000	12000	testis	0
p301	scf7180000006270	NA	3000	51000	48000	both	15505
p302	scf7180000006274	NA	0	17000	17000	testis	0
p303	scf7180000006295	NA	3000	13798	10798	testis	1856
p304	scf7180000006299	NA	4000	23000	19000	testis	0
p305	scf7180000006306	NA	0	13799	13799	testis	2604
p306	scf7180000006326	NA	20000	34000	14000	both	2007
p307	scf7180000006359	NA	40000	91000	51000	both	2932
p308	scf7180000006359	NA	173000	282000	109000	both	25359
p309	scf7180000006359	NA	330000	335000	5000	both	3744
p310	scf7180000006359	NA	412000	423000	11000	ovary	2115
p311	scf7180000006377	NA	44000	58000	14000	both	614

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p312	scf7180000006377	NA	83000	98000	15000	both	19275
p313	scf7180000006379	NA	15000	24000	9000	ovary	2459
p314	scf7180000006379	NA	47000	59000	12000	both	6949
p315	scf7180000006379	NA	146000	209000	63000	both	12461
p316	scf7180000006379	NA	230000	267000	37000	both	8157
p317	scf7180000006379	NA	296000	337000	41000	both	7728
p318	scf7180000006379	NA	376000	402000	26000	both	4914
p319	scf7180000006379	NA	425000	451000	26000	both	16082
p320	scf7180000006379	NA	481000	491000	10000	testis	3576
p321	scf7180000006379	NA	526000	581000	55000	both	10774
p322	scf7180000006386	NA	0	23941	23941	testis	1408
p323	scf7180000006391	NA	43000	69000	26000	both	4034
p324	scf7180000006391	NA	105000	124000	19000	testis	8489
p325	scf7180000006392	NA	31000	42000	11000	both	1064
p326	scf7180000006392	NA	74000	147000	73000	both	19616
p327	scf7180000006392	NA	188000	201000	13000	both	5208
p328	scf7180000006392	NA	251000	258000	7000	testis	1823
p329	scf7180000006424	NA	6000	16000	10000	testis	0
p330	scf7180000006425	NA	0	47000	47000	both	6304
p331	scf7180000006473	NA	0	9000	9000	testis	0
p332	scf7180000006492	NA	23000	31708	8708	testis	0
p333	scf7180000006498	NA	36000	158000	122000	both	34409
p334	scf7180000006498	NA	179000	229000	50000	both	8857
p335	scf7180000006498	NA	355000	366000	11000	both	6585
p336	scf7180000006499	NA	25000	59000	34000	both	1912
p337	scf7180000006500	NA	9000	24000	15000	testis	8104

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p338	scf718000006534	NA	136000	142000	6000	both	158
p339	scf718000006534	NA	250000	257000	7000	both	1160
p340	scf718000006534	NA	362000	381000	19000	both	2037
p341	scf718000006534	NA	442000	464000	22000	both	13223
p342	scf718000006534	NA	508000	518000	10000	both	2165
p343	scf718000006555	NA	0	16266	16266	ovary	25750
p344	scf718000006586	NA	19000	24000	5000	ovary	2187
p345	scf718000006591	NA	14000	37000	23000	testis	0
p346	scf718000006595	NA	0	12692	12692	both	10629
p347	scf718000006596	NA	0	18000	18000	ovary	5458
p348	scf718000006598	NA	37000	60000	23000	both	7506
p349	scf718000006626	NA	22000	29000	7000	both	869
p350	scf718000006676	NA	10000	20897	10897	testis	0
p351	scf718000006721	NA	0	87302	87302	both	32540
p352	scf718000006727	NA	0	33907	33907	both	18776
p353	scf718000006735	NA	0	8000	8000	testis	0
p354	scf718000006750	NA	8000	17000	9000	both	600
p355	scf718000006750	NA	184000	202000	18000	both	5357
p356	scf718000006750	NA	236000	245000	9000	both	1604
p357	scf718000006750	NA	362000	446000	84000	both	17825
p358	scf718000006750	NA	470000	485000	15000	testis	1144
p359	scf718000006750	NA	521000	529000	8000	testis	322
p360	scf718000006769	NA	16000	36000	20000	both	10717
p361	scf718000006770	NA	69000	90000	21000	both	3634
p362	scf718000006770	NA	114000	186000	72000	both	7514
p363	scf718000006770	NA	216000	239000	23000	both	6903

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p364	scf7180000006770	NA	286000	296000	10000	testis	992
p365	scf7180000006770	NA	362000	370000	8000	testis	187
p366	scf7180000006771	NA	20000	129000	109000	both	12005
p367	scf7180000006812	NA	0	7928	7928	both	966
p368	scf7180000006832	NA	18000	29828	11828	testis	1905
p369	scf7180000006845	NA	0	8000	8000	testis	0
p370	scf7180000006856	NA	0	62992	62992	testis	0
p371	scf7180000006858	NA	56000	109000	53000	both	13727
p372	scf7180000006858	NA	157000	167000	10000	both	658
p373	scf7180000006858	NA	251000	271000	20000	testis	8252
p374	scf7180000006858	NA	324000	350000	26000	both	8344
p375	scf7180000006858	NA	441000	443000	2000	testis	62
p376	scf7180000006858	NA	465000	475000	10000	both	719
p377	scf7180000006858	NA	571000	577000	6000	testis	463
p378	scf7180000006858	NA	614000	615000	1000	testis	11
p379	scf7180000006858	NA	643000	646000	3000	testis	152
p380	scf7180000006859	NA	0	111000	111000	both	21894
p381	scf7180000006859	NA	181000	219000	38000	both	9035
p382	scf7180000006862	NA	0	9000	9000	both	955
p383	scf7180000006863	NA	20000	58000	38000	testis	2678
p384	scf7180000006863	NA	80000	85000	5000	testis	3227
p385	scf7180000006863	NA	146000	209000	63000	both	14315
p386	scf7180000006863	NA	234000	264000	30000	both	4518
p387	scf7180000006902	NA	0	11000	11000	testis	2426
p388	scf7180000006905	NA	34000	61000	27000	both	2139
p389	scf7180000006905	NA	100000	128000	28000	testis	2235

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p390	scf7180000006905	NA	195000	259000	64000	both	5746
p391	scf7180000006905	NA	313000	319000	6000	testis	209
p392	scf7180000006905	NA	342000	346000	4000	both	238
p393	scf7180000006905	NA	421000	516000	95000	both	27452
p394	scf7180000006905	NA	564000	608000	44000	both	19739
p395	scf7180000006929	NA	0	19000	19000	both	1351
p396	scf7180000006929	NA	72000	87000	15000	both	1136
p397	scf7180000006929	NA	123000	143000	20000	both	2472
p398	scf7180000006929	NA	270000	366000	96000	both	20196
p399	scf7180000006941	NA	46000	58000	12000	testis	2070
p400	scf7180000006942	NA	18000	115000	97000	both	23413
p401	scf7180000006943	NA	13000	79000	66000	both	23504
p402	scf7180000006943	NA	215000	225000	10000	both	1448
p403	scf7180000006944	NA	6000	58000	52000	both	84781
p404	scf7180000006944	NA	84000	118000	34000	both	7624
p405	scf7180000006944	NA	167000	205000	38000	both	11705
p406	scf7180000006944	NA	228000	235000	7000	testis	2449
p407	scf7180000006944	NA	266000	348000	82000	both	12909
p408	scf7180000006944	NA	382000	430000	48000	both	68335
p409	scf7180000006957	NA	0	165000	165000	both	51629
p410	scf7180000006957	NA	191000	216000	25000	both	2493
p411	scf7180000006957	NA	246000	256000	10000	testis	659
p412	scf7180000006963	NA	5000	146000	141000	both	17322
p413	scf7180000006963	NA	182000	246000	64000	both	12257
p414	scf7180000006963	NA	284000	310000	26000	both	11863
p415	scf7180000006963	NA	398000	444000	46000	both	5074

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
p416	scf7180000006963	NA	545000	552000	7000	testis	1978
p417	scf7180000006963	NA	573000	595000	22000	both	5266
p418	scf7180000006963	NA	661000	701000	40000	both	4726
p419	scf7180000006963	NA	813000	822000	9000	both	2829
p420	scf7180000006968	NA	34000	49000	15000	both	9774
p421	scf7180000006968	NA	177000	222000	45000	both	12700
p422	scf7180000006968	NA	283000	319000	36000	both	8601
p423	scf7180000006969	NA	0	20000	20000	testis	2681
p424	scf7180000006969	NA	56000	96000	40000	both	9845
p425	scf7180000006970	NA	10000	19000	9000	both	4072
p426	scf7180000006970	NA	159000	168000	9000	testis	677
p427	scf7180000006974	NA	34000	88000	54000	both	10908
p428	scf7180000006975	NA	4000	33000	29000	both	3254
p429	scf7180000006975	NA	102000	115000	13000	both	555
p430	scf7180000006975	NA	142000	188000	46000	both	9238
p431	scf7180000006975	NA	269000	288000	19000	both	3723

Table 1. Tab								
name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p1	342	1687	1543	111	20	669	175	no
p2	1436	4218	5813	68	99	338	1603	no
p3	208	1395	278	144	1	26	3	no
p4	11109	3254	18672	33	166	194	1238	no
p5	3777	1790	2134	10	34	155	65	no
p6	7035	5562	4722	68	209	463	931	no
p7	1275	2031	3117	529	14	317	111	no
p8	9034	7910	12955	182	324	408	584	no
p9	8423	9099	4544	1037	1021	2193	1361	no
p10	3369	5351	9773	436	321	1594	198	no
p11	31045	32240	20458	5094	3306	3208	1827	no
p12	2796	2065	2561	625	987	891	1398	no
p13	17953	10355	12945	546	1815	852	2521	no
p14	18701	7586	11329	1033	2492	1538	2116	no
p15	3255	2008	5670	101	36	341	138	no
p16	29048	9501	10555	892	1447	1431	1326	no
p17	20623	12731	29009	1576	1879	3621	3589	no
p18	470	6481	1626	105	36	563	873	no
p19	1327	651	1688	359	516	167	594	no
p20	19336	5491	14561	748	2754	381	5674	no
p21	1295	609	717	25	4	100	324	no
p22	2018	3049	3271	127	165	726	409	no
p23	5276	1175	3333	102	744	206	809	no
p24	9592	3067	7480	603	599	565	1280	no
p25	4397	1935	2941	201	511	329	400	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p26	14243	31463	39854	590	461	4913	3193	no
p27	26648	15038	18311	1285	5532	2619	2702	no
p28	5519	14157	4639	419	455	923	663	no
p29	28928	26756	27942	5214	1739	8143	5456	no
p30	14119	8390	20549	625	812	2886	4386	no
p31	8588	25397	20113	4352	693	20081	12247	no
p32	4503	18721	25378	425	486	8171	9931	no
p33	23202	38511	31026	394	606	5085	6631	no
p34	398	617	940	42	63	322	663	no
p35	13097	30769	7324	42796	1468	19086	4682	no
p36	378	23876	3524	332	47	11131	802	yes
p37	5081	83009	29137	1166	1037	29979	5079	no
p38	3719	50947	23286	22	49	254	1075	no
p39	3491	2123	30732	48	9	1063	164	no
p40	123	1055	243	14	21	103	81	no
p41	0	230	64	14	0	228	56	no
p42	81	520	4185	68	79	431	4170	no
p43	2256	855	666	5	6	141	35	no
p44	4011	1641	1484318	355	743	1156	246178	yes
p45	1202	674	1393	0	2	1	176	no
p46	168	1306	113	53	10	85	19	no
p47	608	85	108	373	607	81	102	no
p48	1638	1013	146783	299	235	446	10858	no
p49	557	106	9310	8	495	43	9040	no
p50	1372	970	10965	81	8	120	71	no
p51	12116	7703	4812	25	86	14	164	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p52	19301	4067	2464	1084	483	796	233	no
p53	5862	12558	5911	399	838	2678	1716	no
p54	42580	13475	26158	2295	10654	1239	3786	no
p55	5712	6514	6298	454	771	1584	2258	no
p56	32605	12747	14660	1851	1046	895	1245	no
p57	21807	3525	4978	1252	451	376	320	no
p58	4792	8599	8960	1353	268	1564	1932	no
p59	11552	17085	14381	2502	847	3903	3942	no
p60	7067	5725	10140	347	668	1371	2852	no
p61	154	847	283	11	30	340	33	no
p62	707	637	3633	117	22	148	45	no
p63	8569	9332	5804	335	357	2161	791	no
p64	3861	7228	3556	23	27	131	308	no
p65	2063	8577	2311	209	145	1362	627	no
p66	67888	22786	41774	407	705	3405	9786	no
p67	18081	13243	19049	645	1327	2792	5105	no
p68	1348	2302	1889	53	20	64	290	no
p69	1329	1943	3956	98	142	476	410	no
p70	21059	49940	24688	1133	829	7409	4795	no
p71	37421	28366	23042	916	1350	2652	3016	no
p72	3229	7946	4495	2579	928	3642	1890	no
p73	61369	87136	92353	4483	4189	29303	42543	no
p74	21010	36792	43724	3607	3300	18420	22490	no
p75	132	1905	420	42	11	1504	189	no
p76	2566	2192	1326	214	55	805	330	no
p77	431	557	1309	6	30	239	753	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p78	569	1893	1176	3	5	57	92	no
p79	625	1216	649	57	53	768	130	no
p80	83447	40743	250768	2655	2383	8403	2457	no
p81	16788	34568	39555	1423	2449	12220	12320	no
p82	878	5977	1028	47	33	1185	319	no
p83	5446	13934	7889	86	24	584	493	no
p84	11873	9916	7245	61	62	401	174	no
p85	5697	4045	64240	5	100	147	9295	no
p86	300	876366	2025	1184	9	167908	152	no
p87	368	83297	2626	284	20	56557	387	no
p88	2751	2312	126378	111	1896	1439	52924	yes
p89	5664	2171	787764	36	3204	821	683491	yes
p90	1087	1393	1008	283	90	133	95	no
p91	557	201764	1505	350	331	85250	833	no
p92	21	879	133	17	21	682	110	no
p93	1718	3119	1630	843	32	247	18	no
p94	4936	2474	685162	9	50	190	4314	no
p95	4325	1827	690528	44	134	417	50320	no
p96	0	281	0	32	0	277	0	no
p97	155	2	294	0	155	2	294	no
p98	205223	142882	43227	56958	8623	8804	1801	no
p99	82063	49854	78038	2231	4786	4257	9460	no
p100	9948	8767	5317	269	1799	160	253	no
p101	45402	30513	36310	10445	13534	14364	17547	yes
p102	0	1756	0	3480	0	1756	0	no
p103	4798	29991	1071	2315	2622	519	226	yes

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p104	1189949	224050	117531	866653	791887	61200	41929	no
p105	13670	9962	21381	15	110	86	1429	yes
p106	2333	419	418	2	99	20	18	yes
p107	16661	33357	8804	674	10680	5924	1248	yes
p108	3423	1132	3845	16	5	1019	42	no
p109	562986	342643	338753	681590	121373	102887	60158	yes
p110	333062	64699	34248	598485	188131	41807	13112	yes
p111	11450	7259	6068	20097	602	2535	1382	yes
p112	701	1378	66	177	181	30	24	no
p113	588	944	215	354	81	154	45	no
p114	7059	2834	114483	9	6596	25	111218	no
p115	69	156	85	9	19	77	42	no
p116	2347	140	316314	25	1942	84	264272	yes
p117	60	0	155	0	60	0	152	no
p118	4	0	998	0	4	0	991	no
p119	1	21336	0	1316	1	21202	0	no
p120	0	12	0	183	0	12	0	no
p121	0	111	1	148	0	111	1	yes
p122	3792	982	724	26	2	253	58	no
p123	0	43	1	298	0	43	0	no
p124	0	306	7	104	0	297	7	no
p125	92	3195	4634	83	36	171	3259	no
p126	1409	157994	29038	187	52	10285	924	yes
p127	9659	1644	2595	154	308	2	4	no
p128	213	239	165	14	117	56	95	no
p129	7	144	8	21	3	137	1	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p130	83386	43675	42594	44925	37467	26859	21831	no
p131	236020	8816	11997	155299	214495	4143	8922	no
p132	62	183	42	466	58	167	39	no
p133	21602	15428	235582	86	563	522	9196	no
p134	705	55597	3701	131	227	12963	2172	no
p135	1835	285	50322	257	46	45	10	no
p136	558	1742	92	535	72	243	13	no
p137	13581	94	443	425	5725	38	129	no
p138	139	84158	614	1846	19	70845	493	no
p139	0	123	0	0	0	121	0	no
p140	4723	2058	441	17398	428	1991	72	yes
p141	10388	6130	7035	50	105	47	30	no
p142	3616	3034	3239	46	68	106	355	no
p143	163	19	948	0	129	0	911	no
p144	64	0	112	0	64	0	112	no
p145	2620	768	368	94	22	9	7	no
p146	3744	5221	53723	128	239	1479	15673	no
p147	543	1190	1903	39	123	942	1508	yes
p148	4352	47258	11479	34172	2133	32393	3167	no
p149	1119	481	137	1102	77	31	9	no
p150	2290	700	2262	10	26	16	138	no
p151	6488	2561	3769	1333	1319	1230	2221	no
p152	4	309	122	10	3	274	106	no
p153	444014	30008	627953	3831	235359	6723	455601	yes
p154	378783	1832	355361	450	321211	458	301548	yes
p155	123393	148899	53508	127461	6821	22287	8718	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p156	330	970	3048	34	49	197	230	no
p157	3182	15466	6940	2032	682	5533	3596	no
p158	4489	22284	10869	1572	656	11562	4189	no
p159	2005	2474	2216	171	613	807	1006	no
p160	3299	3951	2956	2323	659	1824	1418	no
p161	3702	4583	1889	382	284	763	225	no
p162	4474	7697	3761	2034	345	2051	810	no
p163	2221	11398	1436	2798	59	8685	101	no
p164	1515	3019	752	739	71	212	76	no
p165	8010	5953	5826	1706	555	1387	693	no
p166	147382	44308	47893	37244	33010	6847	6630	no
p167	62006	2079	3353	50	33	184	83	no
p168	45	455	214	1	4	82	81	no
p169	596	3609	1759	57	75	2305	1385	no
p170	11695	1233	1927	38	25	1046	421	no
p171	4889	3947	2931	216	192	966	592	no
p172	3074	832	1841	2	101	38	182	no
p173	406	67	1234	3	45	32	427	no
p174	68	768	2721	227	24	577	2607	no
p175	3269	1978	3567	71	157	1070	1793	no
p176	42	171	206	1	1	84	119	no
p177	2145	11774	3575	3608	841	8834	2572	no
p178	603	1855	2075	32	129	1117	1689	no
p179	103766	140681	49930	43183	11079	16360	7413	no
p180	130605	3237	56055	49664	7943	144	810	no
p181	5497	3555	667	4358	4940	1045	444	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p182	0	105	0	65	0	105	0	no
p183	22996	2573	12651	49	367	0	6	no
p184	6041	5776	61331	811	303	1750	1831	yes
p185	6061	374	950	253	703	27	9	no
p186	720	182919	3493	1335	113	86432	1413	yes
p187	8572	7110	11571	598	481	421	1389	no
p188	2870	1554	3271	109	128	543	266	no
p189	1476	228	4417	0	124	9	1765	no
p190	19	1261	9	1023	0	674	1	no
p191	4467	16610	746154	55	397	1015	98537	no
p192	64609	344210	90342	66863	4163	68801	18165	no
p193	15839	13933	199341	876	2150	6267	89655	yes
p194	158	10625	8295	28	20	7421	7063	no
p195	83	483	520	27	17	366	388	no
p196	38091	794	4053	538	561	144	72	no
p197	210	2500	723	610	15	1956	351	no
p198	711	310	2397	22	211	96	1402	no
p199	105	69	728	3	7	40	441	no
p200	23	16	1281	1	23	16	1269	no
p201	10	364	240	54	0	129	0	no
p202	476	578265	2772	1293	231	18808	1786	yes
p203	2907	0	1598	0	2902	0	1595	no
p204	2426	839629	9202	2819	409	66166	5893	no
p205	9	58	143	0	2	17	131	no
p206	1377	33017	15751	1220	373	28605	13553	no
p207	425	7505	4721	740	109	4800	4356	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p208	3848	1155	594	164	95	17	0	no
p209	0	368	0	20	0	366	0	no
p210	5758	1824	536	35	78	4	2	no
p211	130857	116179	130837	19004	22992	32263	47402	no
p212	297	872	263	68	4	196	19	no
p213	0	1272	1439	0	0	79	44	no
p214	0	2422	1461	0	0	25	86	no
p215	0	2862	1010	0	0	17	135	no
p216	0	2912	3907	0	0	131	89	no
p217	682	1867	4165	27	27	34	81	no
p218	10755	13603	7619	1608	520	1230	577	no
p219	3282	1218	3277	36	67	94	146	no
p220	17601	3692	9939	351	1034	363	654	no
p221	7025	19659	10171	240	232	1446	1104	no
p222	3115	11501	9473	262	183	2493	1387	no
p223	16098	1814	2734	55	129	374	131	no
p224	14338	8830	10464	1416	1434	1711	2556	no
p225	10882	1145	7205	27	541	46	620	no
p226	7781	5219	4516	132	68	463	190	no
p227	929	169	1115	4	139	3	375	no
p228	17397	9155	8022	2701	1454	2315	1389	no
p229	27228	8422	6029	262	652	936	1468	no
p230	0	2951	1345	0	0	133	99	no
p231	3143	1968	1829	18	30	34	99	no
p232	1525	696	1084	40	204	88	672	no
p233	2247	766	1215	32	137	112	140	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p234	11331	9924	10031	1307	905	1599	868	no
p235	818	3822	1114	101	64	1221	230	no
p236	38424	6082	6974	11733	2041	464	185	no
p237	0	1098	281	0	0	158	6	no
p238	0	3079	1530	0	0	122	102	no
p239	0	937	3490	0	0	86	51	no
p240	0	504	2020	0	0	56	481	no
p241	0	1217	578	0	0	146	91	no
p242	14447	13621	11215	1608	2478	1280	1351	no
p243	3664	1937	12397	65	227	91	35	no
p244	64	232	546	137	21	12	119	no
p245	0	1549	1512	0	0	29	73	no
p246	0	2188	751	0	0	427	5	no
p247	14750	4634	8893	9	40	58	113	no
p248	940	1608	2176	59	1	247	2	no
p249	15090	2603	8889	27	150	18	230	no
p250	0	747	1653	0	0	26	107	no
p251	0	707	1348	0	0	292	12	no
p252	0	801	1020	0	0	52	108	no
p253	10588	20319	10091	1689	589	1686	1023	no
p254	0	34	2957	0	0	5	120	no
p255	5635	4561	5634	105	68	125	302	no
p256	0	12236	33772	0	0	35	85	no
p257	0	2609	1119	0	0	3	188	no
p258	0	648	1209	0	0	100	36	no
p259	54014	71971	39210	1593	1137	4466	2680	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p260	0	1976	4506	0	0	108	85	no
p261	0	1340	1028	0	0	24	119	no
p262	17993	14791	11189	1310	641	1412	795	no
p263	0	995	2720	0	0	74	124	no
p264	1943	10794	9888	338	377	3614	3974	no
p265	0	625	973	0	0	92	21	no
p266	4007	3118	3839	1113	419	2461	907	no
p267	442	1421	730	7	0	96	39	no
p268	467	3592	397	7	3	226	8	no
p269	0	394	1063	0	0	0	100	no
p270	1823	3602	512	2971	160	65	4	no
p271	646	8883	1343	21	2	114	12	no
p272	0	1276	4214	0	0	124	76	no
p273	1082	1885	1590	194	75	66	234	no
p274	3979	4107	1213	402	1584	81	186	no
p275	0	1360	1908	0	0	61	386	no
p276	0	1176	950	0	0	58	98	no
p277	0	7695	2695	0	0	462	275	no
p278	0	709	1314	0	0	31	139	no
p279	0	2353	2632	0	0	105	264	no
p280	0	2359	1464	0	0	93	38	no
p281	2179	809	5544	6	27	9	254	no
p282	0	1080	351	0	0	234	16	no
p283	0	457	15	0	0	430	0	no
p284	0	1590	577	0	0	150	28	no
p285	0	893	2247	0	0	33	84	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p286	0	957	859	0	0	2	145	no
p287	0	1921	515	0	0	58	62	no
p288	0	1652	586	0	0	114	100	no
p289	25905	4403	5730	692	309	64	19	no
p290	280724	56468	137516	28158	79781	11613	13235	no
p291	1631	1770	900	77	34	223	95	no
p292	0	172	528	0	0	52	49	no
p293	24556	3230	2531	182	547	11	93	no
p294	6067	3822	2435	741	95	479	75	no
p295	80115	288392	53866	16287	3896	12967	4451	no
p296	19556	75782	12874	612	2632	338	1132	no
p297	12189	17764	11425	533	158	892	343	no
p298	0	1096	6905	0	0	243	241	no
p299	9508	2779	6266	15	31	65	50	no
p300	0	1408	1848	0	0	76	106	no
p301	9202	10057	5836	323	150	1431	300	no
p302	0	1972	1689	0	0	142	22	no
p303	4079	1487	3933	7	61	42	207	no
p304	0	5443	2698	0	0	324	60	no
p305	3895	1126	3410	15	79	21	196	no
p306	2560	1196	1986	137	106	257	276	no
p307	18803	4194	13517	284	5543	640	6914	no
p308	55466	7225	7314	1180	7075	336	879	no
p309	273	2420	152	341	79	141	22	no
p310	995	1303	905	95	23	40	54	no
p311	5670	1127	5478	37	152	42	179	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p312	1814	2636	670	155	167	40	123	no
p313	4269	510	620	80	27	18	7	no
p314	1330	2742	1192	233	31	202	56	no
p315	4794	10773	7408	589	353	1175	945	no
p316	5923	8566	3553	376	742	556	562	no
p317	9913	8094	3570	699	782	1237	428	no
p318	5623	4654	3269	254	57	456	187	no
p319	2203	7285	2646	378	257	373	243	no
p320	428	3233	600	120	55	270	204	no
p321	10197	10829	9681	240	492	667	394	no
p322	13340	1330	15021	4	72	8	170	no
p323	7845	8925	6987	44	177	186	207	no
p324	5134	5773	2886	26	7	114	185	no
p325	4403	1459	2673	73	2065	400	955	no
p326	5183	15474	7673	801	637	2670	1911	no
p327	2676	4270	2244	12	90	54	110	no
p328	298	1417	261	49	50	118	35	no
p329	0	13126	2943	0	0	155	212	no
p330	7373	3993	9341	1114	754	1052	1207	no
p331	0	1152	500	0	0	64	38	no
p332	0	1382	6394	0	0	15	118	no
p333	21986	16339	16108	1058	815	2482	2148	no
p334	4829	8562	4720	663	316	2420	1512	no
p335	993	3393	705	628	51	231	86	no
p336	4994	1626	5368	350	722	698	2354	no
p337	3183	6841	5053	44	8	704	90	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p338	2126	139	692	54	346	50	215	no
p339	123	1043	195	530	44	704	61	no
p340	3089	1929	1635	300	31	811	233	no
p341	11058	3489	17104	73	484	592	734	no
p342	1354	10044	4368	136	32	317	308	no
p343	14616	1237	705	989	168	52	10	no
p344	398	621	220	119	55	10	14	no
p345	0	3302	3127	0	0	126	41	no
p346	3109	3179	1827	55	82	39	72	no
p347	8347	1509	2105	25	150	5	28	no
p348	1534	4481	1784	395	300	232	224	no
p349	2722	607	1375	37	113	47	228	no
p350	0	645	3745	0	0	4	111	no
p351	21008	8825	7780	1454	1698	345	1309	no
p352	2837	19703	2634	263	21	1050	19	no
p353	0	5868	1154	0	0	266	20	no
p354	2248	2761	2673	60	56	30	166	no
p355	1614	3868	1339	527	94	546	46	no
p356	646	1714	1647	111	67	633	162	no
p357	62400	19000	20788	201	681	1109	3801	no
p358	1374	3170	7181	42	30	125	255	no
p359	2176	585	2366	9	59	121	197	no
p360	1242	5281	1860	399	294	294	304	no
p361	11272	3425	25126	42	178	58	239	no
p362	25027	5756	12652	703	860	739	1454	no
p363	5768	2199	3490	149	104	118	387	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p364	482	1237	911	105	28	316	91	no
p365	2027	806	2801	21	28	31	295	no
p366	18525	10246	12818	1119	1553	1413	2199	no
p367	1675	4976	1762	14	93	78	88	no
p368	3107	1019	4416	2	16	5	148	no
p369	0	1521	413	0	0	30	94	no
p370	0	55290	20854	0	0	460	117	no
p371	10519	24412	9903	794	504	4175	2252	no
p372	1181	1617	2350	79	283	1048	1440	no
p373	2853	4272	1926	10	47	65	225	no
p374	9480	3853	3300	129	200	629	218	no
p375	30	241	69	39	2	224	47	no
p376	3911	512	4934	162	332	93	297	no
p377	1005	836	843	54	97	498	69	no
p378	413	27	306	2	30	12	255	no
p379	177	106	309	28	40	45	126	no
p380	29261	18055	14453	1105	895	1947	1459	no
p381	6566	5266	3775	231	618	334	317	no
p382	184	1946	310	126	13	220	16	no
p383	5660	7408	5999	114	127	126	1563	no
p384	137	829	213	28	2	93	35	no
p385	21375	8578	12531	633	1068	807	1163	no
p386	14252	2933	19667	263	540	428	763	no
p387	2297	2432	1599	32	37	83	46	no
p388	12010	4001	5654	175	374	339	522	no
p389	1439	4405	5408	45	53	49	554	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p390	3968	21619	8389	454	306	3686	1607	no
p391	3824	342	842	2	7	74	32	no
p392	1038	350	317	68	37	285	104	no
p393	13861	15679	13349	1233	602	2243	2024	no
p394	5094	7498	3670	248	372	351	486	no
p395	11548	963	5301	2	216	35	76	no
p396	3090	1558	1943	284	225	321	368	no
p397	1713	2483	3086	107	538	231	852	no
p398	47972	12526	13950	624	547	748	917	no
p399	2913	1981	2207	50	64	114	176	no
p400	13815	15554	9576	1421	751	1836	938	no
p401	10056	8580	11942	823	603	1019	1650	no
p402	6737	1020	2242	192	91	217	368	no
p403	25409	26142	17803	1054	372	1318	522	no
p404	7706	1564	6367	80	721	226	2631	no
p405	1915	6876	2881	714	846	3251	1052	no
p406	1032	1526	997	23	34	149	155	no
p407	9796	18899	15506	1058	1271	5390	5304	no
p408	5361	11423	4678	816	649	689	978	no
p409	57890	15946	19760	2104	1197	1285	1425	no
p410	7957	2926	4214	110	187	220	367	no
p411	1993	628	1333	45	71	109	341	no
p412	28962	16085	23942	1112	1611	2168	3040	no
p413	11200	12010	10138	372	177	1882	820	no
p414	6799	5236	4532	56	172	92	434	no
p415	11890	3480	6835	223	933	174	1020	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in miranda?
p416	1050	2570	2050	0	14	19	85	no
p417	23182	2972	3931	168	587	108	1045	no
p418	19068	6882	7565	137	217	256	1383	no
p419	1167	1515	553	63	72	167	15	no
p420	1385	3573	1134	172	9	160	24	no
p421	2999	5454	3190	1231	99	1465	103	no
p422	6294	10596	3224	1771	314	3454	146	no
p423	1294	3354	2593	28	39	127	116	no
p424	7668	4347	3922	138	2442	217	390	no
p425	1420	1532	1284	36	67	31	105	no
p426	3843	2250	1922	11	90	64	101	no
p427	7778	11298	6196	223	336	391	280	no
p428	7920	4576	7475	286	84	479	222	no
p429	2593	820	1557	39	56	163	167	no
p430	12833	5508	11435	290	420	344	942	no
p431	4949	1820	3360	384	389	216	201	no

Table 2. Table of piRNA clusters annotated in *D. miranda*.

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m1	merge10q2p2	neoY	1452000	1461000	9000	testis	0
m2	merge10q2p2	neoY	1779000	1804000	25000	testis	0
m3	merge10q2p2	neoY	5183000	5191000	8000	testis	0
m4	merge10q2p2	neoY	8124000	8129000	5000	testis	0
m5	merge10q2p2	neoY	29643000	29645000	2000	testis	0
m6	merge10q2p2	neoY	36728000	36790000	62000	testis	0
m7	merge10q2p2	neoY	37672000	37681000	9000	testis	0
m8	merge10q2p2	neoY	49169000	49188000	19000	testis	0
m9	merge10q2p2	neoY	50681000	50686000	5000	testis	0
m10	merge10q2p2	neoY	51256000	51257000	1000	testis	0
m11	merge10q2p2	neoY	51403000	51423000	20000	testis	0
m12	merge10q2p2	neoY	52483000	52490000	7000	testis	0
m13	merge10q2p2	neoY	53391000	53399000	8000	testis	0
m14	merge10q2p2	neoY	53625000	53630000	5000	testis	0
m15	merge10q2p2	neoY	54157000	54197000	40000	testis	0
m16	merge11q2p2	Y	406000	413000	7000	testis	0
m17	merge11q2p2	Y	525000	575000	50000	testis	0
m18	merge11q2p2	Y	615000	686000	71000	testis	0
m19	merge11q2p2	Y	869000	995000	126000	testis	0
m20	merge11q2p2	Y	1109000	1170000	61000	testis	0
m21	merge11q2p2	Y	1217000	1291000	74000	testis	0
m22	merge11q2p2	Y	1315000	1340000	25000	testis	0
m23	merge11q2p2	Y	1445000	1464000	19000	testis	0
m24	merge11q2p2	Y	1876000	1918000	42000	testis	0
m25	merge11q2p2	Y	1945000	2080000	135000	testis	0

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m26	merge11q2p2	Y	2383000	2400000	17000	testis	0
m27	merge11q2p2	Y	9655000	9658000	3000	testis	0
m28	merge11q2p2	Y	10040000	10042000	2000	testis	0
m29	merge1q2p2	B	1190000	1196000	6000	ovary	15
m30	merge1q2p2	B	1358000	1376000	18000	ovary	40157
m31	merge1q2p2	B	1838000	1848000	10000	ovary	63727
m32	merge1q2p2	B	1877000	1892000	15000	ovary	937
m33	merge1q2p2	B	1943000	1958000	15000	both	116954
m34	merge1q2p2	B	2006000	2010000	4000	ovary	186
m35	merge1q2p2	B	2120000	2128000	8000	ovary	1944
m36	merge1q2p2	B	2167000	2196000	29000	ovary	10738
m37	merge1q2p2	B	2251000	2258000	7000	ovary	342
m38	merge1q2p2	B	2466000	2472000	6000	testis	16139
m39	merge1q2p2	B	2582000	2593000	11000	ovary	20702
m40	merge1q2p2	B	2664000	2679000	15000	both	60019
m41	merge1q2p2	B	8125000	8133000	8000	ovary	1647
m42	merge1q2p2	B	9687000	9696000	9000	both	2325
m43	merge1q2p2	B	12216000	12228000	12000	both	4504
m44	merge1q2p2	B	12541000	12559000	18000	both	314513
m45	merge1q2p2	B	25958000	25973000	15000	both	5674
m46	merge1q2p2	B	26712000	26727000	15000	both	15677
m47	merge1q2p2	B	28707000	28717000	10000	both	298
m48	merge1q2p2	B	32494000	32539883	45883	both	987472
m49	merge2q2p2	C	1135000	1142000	7000	ovary	11037
m50	merge2q2p2	C	1322000	1328000	6000	ovary	153
m51	merge2q2p2	C	1512000	1538000	26000	ovary	251767

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m52	merge2q2p2	C	1566000	1585000	19000	ovary	11857
m53	merge2q2p2	C	1773000	1780000	7000	ovary	29335
m54	merge2q2p2	C	1804000	1822000	18000	ovary	18619
m55	merge2q2p2	C	2072000	2078000	6000	ovary	742
m56	merge2q2p2	C	2124000	2141000	17000	ovary	10734
m57	merge2q2p2	C	2744000	2765000	21000	ovary	34178
m58	merge2q2p2	C	2988000	2999000	11000	ovary	172
m59	merge2q2p2	C	3172000	3178000	6000	ovary	3425
m60	merge2q2p2	C	4499000	4538000	39000	ovary	33914
m61	merge2q2p2	C	5728000	5739000	11000	ovary	5729
m62	merge2q2p2	C	7393000	7403000	10000	testis	1326
m63	merge2q2p2	C	10884000	10910000	26000	ovary	45096
m64	merge2q2p2	C	20326000	20328000	2000	both	9177
m65	merge2q2p2	C	20421000	20427000	6000	ovary	3
m66	merge2q2p2	C	25239000	25306201	67201	both	908824
m67	merge3q2p2	E	0	58000	58000	ovary	423200
m68	merge3q2p2	E	3191000	3196000	5000	testis	15
m69	merge3q2p2	E	7818000	7834000	16000	testis	5603
m70	merge3q2p2	E	9565000	9582000	17000	both	360
m71	merge3q2p2	E	17899000	17975000	76000	both	956196
m72	merge3q2p2	E	18048000	18061000	13000	both	452
m73	merge3q2p2	E	20308000	20311000	3000	both	21
m74	merge3q2p2	E	20674000	20681000	7000	both	6398
m75	merge3q2p2	E	23526000	23538000	12000	testis	1596
m76	merge3q2p2	E	23579000	23588000	9000	testis	352
m77	merge3q2p2	E	24143000	24153000	10000	testis	138

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m78	merge3q2p2	E	24299000	24304000	5000	testis	124
m79	merge3q2p2	E	24461000	24469000	8000	both	7901
m80	merge3q2p2	E	25551000	25573000	22000	both	12757
m81	merge3q2p2	E	25788000	25790000	2000	both	3286
m82	merge3q2p2	E	25859000	25869000	10000	testis	4986
m83	merge3q2p2	E	26380000	26384000	4000	testis	2605
m84	merge3q2p2	E	33395000	33405000	10000	ovary	1886
m85	merge3q2p2	E	33497000	33522000	25000	ovary	133018
m86	merge3q2p2	E	33556000	33570000	14000	ovary	10892
m87	merge3q2p2	E	33593000	33628000	35000	ovary	10124
m88	merge3q2p2	E	33742000	33820000	78000	both	45337
m89	merge3q2p2	E	33841000	33846000	5000	ovary	1039
m90	merge3q2p2	E	35076000	35164000	88000	both	318888
m91	merge3q2p2	E	35238000	35263000	25000	ovary	112767
m92	merge4q2p2	A	0	167000	167000	both	258792
m93	merge4q2p2	A	226000	470000	244000	ovary	100511
m94	merge4q2p2	A	1263000	1266000	3000	ovary	4123
m95	merge4q2p2	A	1322000	1361000	39000	ovary	31716
m96	merge4q2p2	A	1419000	1476000	57000	ovary	10075
m97	merge4q2p2	A	1981000	2033000	52000	both	176678
m98	merge4q2p2	A	2127000	2133000	6000	ovary	16232
m99	merge4q2p2	A	2449000	2542000	93000	ovary	179823
m100	merge4q2p2	A	2641000	2809000	168000	both	1057517
m101	merge4q2p2	A	2831000	2838000	7000	both	11897
m102	merge4q2p2	A	2864000	3077000	213000	ovary	713144
m103	merge4q2p2	A	3112000	3137000	25000	ovary	10741

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m104	merge4q2p2	A	3436000	3454000	18000	both	8430
m105	merge4q2p2	A	4268000	4282000	14000	both	17344
m106	merge4q2p2	A	4308000	5021000	713000	both	2385031
m107	merge4q2p2	A	5340000	5725000	385000	both	2410657
m108	merge4q2p2	A	5887000	5903000	16000	both	213852
m109	merge4q2p2	A	9810000	9818000	8000	ovary	16631
m110	merge4q2p2	A	14749000	14759000	10000	ovary	2
m111	merge4q2p2	A	18683000	18685000	2000	testis	3
m112	merge4q2p2	A	20134000	20140000	6000	both	12016
m113	merge4q2p2	A	22120000	22139000	19000	both	151124
m114	merge4q2p2	A	23475000	23608000	133000	both	67748
m115	merge4q2p2	A	25282000	25339000	57000	both	369310
m116	merge5q2p2	AD	54000	57000	3000	ovary	310
m117	merge5q2p2	AD	4463000	4468000	5000	testis	381
m118	merge5q2p2	AD	10489000	10492000	3000	ovary	2526
m119	merge5q2p2	AD	17854000	17858000	4000	both	11181
m120	merge5q2p2	AD	26351000	26405000	54000	testis	1733
m121	merge5q2p2	AD	26437000	26448000	11000	ovary	1304
m122	merge5q2p2	AD	27108000	27115000	7000	ovary	28
m123	merge5q2p2	AD	27740000	27749000	9000	both	32284
m124	merge5q2p2	AD	27776000	28029000	253000	both	485085
m125	merge5q2p2	AD	28052000	28103000	51000	ovary	126761
m126	merge5q2p2	AD	30980000	31352000	372000	both	188474
m127	merge5q2p2	AD	31872000	32029000	157000	ovary	654583
m128	merge5q2p2	AD	32542000	32553000	11000	ovary	978
m129	merge5q2p2	AD	32676000	32678000	2000	ovary	24

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m130	merge5q2p2	AD	32783000	32793000	10000	ovary	1295
m131	merge5q2p2	AD	32977000	33011000	34000	ovary	84743
m132	merge5q2p2	AD	33154000	33201000	47000	ovary	11187
m133	merge5q2p2	AD	33370000	33400000	30000	ovary	11380
m134	merge5q2p2	AD	33609000	33660000	51000	ovary	25762
m135	merge5q2p2	AD	33691000	33703000	12000	ovary	268
m136	merge5q2p2	AD	34258000	34308000	50000	ovary	30465
m137	merge5q2p2	AD	34446000	34452000	6000	ovary	254
m138	merge5q2p2	AD	34517000	34528000	11000	ovary	624
m139	merge5q2p2	AD	34575000	34578000	3000	ovary	12
m140	merge5q2p2	AD	34675000	34724000	49000	ovary	9929
m141	merge5q2p2	AD	34795000	34813000	18000	ovary	7660
m142	merge5q2p2	AD	35234000	35246000	12000	ovary	138
m143	merge5q2p2	AD	35290000	35308000	18000	ovary	7973
m144	merge5q2p2	AD	35428000	35450000	22000	ovary	20596
m145	merge5q2p2	AD	35601000	35692000	91000	ovary	42836
m146	merge5q2p2	AD	35723000	35733000	10000	ovary	1559
m147	merge5q2p2	AD	35817000	35828000	11000	ovary	2791
m148	merge5q2p2	AD	35889000	35930000	41000	ovary	60826
m149	merge5q2p2	AD	35983000	35998000	15000	ovary	4846
m150	merge5q2p2	AD	36326000	36335000	9000	ovary	668
m151	merge5q2p2	AD	36443000	36451000	8000	ovary	1899
m152	merge5q2p2	AD	36500000	36518000	18000	ovary	4923
m153	merge5q2p2	AD	36586000	36599000	13000	ovary	2672
m154	merge5q2p2	AD	36629000	36632000	3000	ovary	24
m155	merge5q2p2	AD	36731000	36741000	10000	ovary	261

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m156	merge5q2p2	AD	37276000	37288000	12000	ovary	13867
m157	merge5q2p2	AD	37340000	37385000	45000	ovary	37615
m158	merge5q2p2	AD	37487000	37490000	3000	ovary	89
m159	merge5q2p2	AD	38202000	38225000	23000	ovary	20958
m160	merge5q2p2	AD	38299000	38321000	22000	ovary	59184
m161	merge5q2p2	AD	38981000	38998000	17000	ovary	22343
m162	merge5q2p2	AD	39092000	39096000	4000	both	7456
m163	merge5q2p2	AD	39296000	39319000	23000	ovary	104797
m164	merge5q2p2	AD	39348000	39354000	6000	ovary	2240
m165	merge5q2p2	AD	39702000	39713000	11000	ovary	5210
m166	merge5q2p2	AD	39740000	39746000	6000	ovary	1132
m167	merge5q2p2	AD	40104000	40109000	5000	ovary	292
m168	merge5q2p2	AD	40485000	40503000	18000	ovary	7746
m169	merge5q2p2	AD	40561000	40585000	24000	ovary	2942
m170	merge5q2p2	AD	40649000	40678000	29000	ovary	7175
m171	merge5q2p2	AD	40822000	40860000	38000	ovary	19479
m172	merge5q2p2	AD	40930000	40937000	7000	ovary	1419
m173	merge5q2p2	AD	41152000	41165000	13000	ovary	16262
m174	merge5q2p2	AD	41234000	41253000	19000	ovary	54049
m175	merge5q2p2	AD	41462000	41465000	3000	both	95
m176	merge5q2p2	AD	41596000	41598000	2000	both	161
m177	merge5q2p2	AD	41743000	41760000	17000	both	10946
m178	merge5q2p2	AD	41785000	41792000	7000	ovary	1260
m179	merge5q2p2	AD	42081000	42158000	77000	both	64812
m180	merge5q2p2	AD	42260000	42269000	9000	ovary	1353
m181	merge5q2p2	AD	42549000	42567000	18000	ovary	2247

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m182	merge5q2p2	AD	42673000	42708000	35000	both	5512
m183	merge5q2p2	AD	43035000	43047000	12000	ovary	10974
m184	merge5q2p2	AD	43412000	43420000	8000	both	848
m185	merge5q2p2	AD	43460000	43490000	30000	both	43553
m186	merge5q2p2	AD	43576000	43583000	7000	ovary	42405
m187	merge5q2p2	AD	43650000	43664000	14000	ovary	796
m188	merge5q2p2	AD	43698000	43865000	167000	ovary	477406
m189	merge5q2p2	AD	44072000	44086000	14000	ovary	2460
m190	merge5q2p2	AD	44711000	44733000	22000	ovary	59047
m191	merge5q2p2	AD	44884000	44905000	21000	both	160489
m192	merge5q2p2	AD	44939000	44956000	17000	both	4744
m193	merge5q2p2	AD	45537000	45553000	16000	ovary	135389
m194	merge5q2p2	AD	45920000	45926000	6000	ovary	460
m195	merge5q2p2	AD	46187000	46245000	58000	ovary	86842
m196	merge5q2p2	AD	46266000	46296000	30000	ovary	57400
m197	merge5q2p2	AD	46432000	46469000	37000	both	162490
m198	merge5q2p2	AD	46591000	46593000	2000	ovary	253
m199	merge5q2p2	AD	46860000	46867000	7000	ovary	4371
m200	merge5q2p2	AD	46972000	46989000	17000	ovary	5582
m201	merge5q2p2	AD	47700000	47731000	31000	both	22337
m202	merge5q2p2	AD	47929000	47941000	12000	ovary	50671
m203	merge5q2p2	AD	48170000	48180000	10000	ovary	271
m204	merge5q2p2	AD	48917000	48925000	8000	ovary	1604
m205	merge5q2p2	AD	49335000	49366000	31000	ovary	8760
m206	merge5q2p2	AD	49489000	49527000	38000	ovary	36896
m207	merge5q2p2	AD	49563000	49571000	8000	ovary	2416

name	scaffold	Muller element	start	end	length	tissue expressed	# fwd. ovary piRNA
m208	merge5q2p2	AD	49638000	49662000	24000	ovary	208116
m209	merge5q2p2	AD	49932000	49944000	12000	ovary	2880
m210	merge5q2p2	AD	50507000	50521000	14000	ovary	27901
m211	merge5q2p2	AD	50568000	50573000	5000	ovary	1087
m212	merge5q2p2	AD	50595000	50605000	10000	ovary	3065
m213	merge5q2p2	AD	50629000	50672000	43000	ovary	876
m214	merge5q2p2	AD	51311000	51330000	19000	ovary	3520
m215	merge5q2p2	AD	51475000	51484000	9000	ovary	16277
m216	merge5q2p2	AD	51543000	51552000	9000	ovary	31
m217	merge5q2p2	AD	51616000	51629000	13000	ovary	80974
m218	merge5q2p2	AD	51704000	51714000	10000	both	1218
m219	merge5q2p2	AD	51765000	51783000	18000	ovary	92721
m220	merge5q2p2	AD	51870000	51881000	11000	ovary	17340
m221	merge5q2p2	AD	52252000	52256000	4000	both	1185
m222	merge5q2p2	AD	52352000	52354000	2000	ovary	149
m223	merge7q2p2	Y	195000	215000	20000	testis	0
m224	merge7q2p2	Y	290000	402000	112000	testis	0
m225	merge8q2p2	neoY	29430000	29437000	7000	testis	0
m226	merge8q2p2	neoY	30720000	30737000	17000	testis	0
m227	mpm2q2p2	F	2191000	2329000	138000	both	314901

Table 2. Tab

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m1	0	793	208	0	0	51	80	no
m2	0	3473	378	0	0	256	127	no
m3	0	404	223	0	0	98	137	no
m4	0	256	67	0	0	80	18	no
m5	0	1689	35	0	0	5	0	no
m6	0	5731	5151	0	0	271	682	yes
m7	0	1063	232	0	0	79	163	no
m8	0	3001	3411	0	0	32	126	no
m9	0	3620	5346	0	0	42	14	no
m10	0	142	9	0	0	136	0	no
m11	0	3105	1146	0	0	7	94	no
m12	0	251	135517	0	0	0	2107	no
m13	0	188	100	0	0	4	72	no
m14	0	48	154	0	0	26	106	no
m15	0	4391	7671	0	0	412	1483	no
m16	0	86	2975	0	0	2	546	no
m17	0	4146	3243	0	0	899	361	no
m18	0	4205	3120	0	0	730	177	no
m19	0	13534	11799	0	0	1712	3796	no
m20	0	9910	1125	0	0	2512	230	no
m21	0	23480	9098	0	0	2475	696	no
m22	0	3035	886	0	0	60	295	no
m23	0	1709	3629	0	0	818	1040	no
m24	0	3678	1836	0	0	847	633	no
m25	0	19155	15174	0	0	3905	4091	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m26	0	2309	835	0	0	491	88	no
m27	0	4	136	0	0	0	56	no
m28	0	6	121	0	0	6	108	no
m29	1213	62	72	4	153	30	34	no
m30	22457	383	160	85	138	36	25	no
m31	11946	282	30	113	19	21	18	no
m32	5341	98	252	125	370	36	79	no
m33	2071	5298	221	1810	41	20	196	no
m34	18	14	7	132	6	8	2	no
m35	173	99	71	348	1	14	11	no
m36	8923	273	1295	328	2984	33	201	no
m37	135	32	31	91	22	13	0	no
m38	401	61	147	13	0	39	110	no
m39	575	78	99	195	5	51	22	no
m40	989	1130	189	4421	11	335	11	no
m41	9952	78	209	114	41	2	5	no
m42	92415	1047	10389	118	1454	849	5545	no
m43	85175	5115	22312	201	2044	248	4494	yes
m44	12850	17526	13550	15792	846	325	119	no
m45	14067	5401	91921	312	232	327	15227	no
m46	156	440804	954	3790	8	47621	144	yes
m47	459	1840	3134	29	300	251	1705	no
m48	33916	32268	8478	28787	1193	483	412	no
m49	269	1705	8	254	1	8	0	no
m50	6755	113	21	7	100	5	7	no
m51	5927	16	81	300	31	1	9	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m52	9016	344	102	52	837	3	12	no
m53	458	875	47	1005	87	26	11	no
m54	7172	480	129	340	113	14	42	no
m55	9137	97	141	7	1109	3	87	no
m56	457	463	50	319	57	33	15	no
m57	2166	534	141	686	58	46	50	no
m58	1135	17	98	11	129	0	15	no
m59	161	0	0	184	3	0	0	no
m60	2059	458	59	1225	103	106	9	no
m61	23954	10	23	9	261	0	17	no
m62	262	25670	1803	16	41	4586	22	yes
m63	32652	93	8	15554	119	44	2	no
m64	160	4	510	169	0	3	396	no
m65	540	6	23	2	210	0	7	no
m66	50239	18050	7028	193233	6993	1301	189	no
m67	382102	46045	32003	2065	17371	363	308	no
m68	448	205	2369	9	17	193	689	yes
m69	989	144020	9475	86	11	2727	337	yes
m70	3725	12507	222665	13	55	313	8226	no
m71	417616	108118	70427	207748	9621	9834	7092	no
m72	9221	1340	9354	40	353	705	1138	yes
m73	462	26	293	3	410	9	266	no
m74	2505	8853	19401	0	93	9	2545	no
m75	419	61031	16167	51	1	2482	12	no
m76	629	4588	28334	1	74	177	7760	no
m77	70	861	15308	2	16	41	7847	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m78	14	11699	335	10	5	4057	143	no
m79	723	1142	7434	222	222	133	4955	no
m80	50201	5665	505571	119	603	53	6164	no
m81	2	208	5	3286	2	208	0	no
m82	1008	27964	3578	26	5	2744	78	yes
m83	146	13016	1663	22	5	250	71	no
m84	4158	29	23	68	43	1	1	no
m85	2482	289	74	637	40	5	6	no
m86	8537	245	209	421	202	38	27	no
m87	18640	85	185	259	607	14	45	no
m88	157569	615	2494	154	840	109	220	no
m89	4530	224	1225	12	47	12	29	no
m90	26555	1675	1281	6357	851	197	26	no
m91	17095	857	994	962	336	1	0	no
m92	219949	5630	2128	10995	11820	211	569	no
m93	305524	3214	4609	13203	40415	156	357	no
m94	221	25	16	113	53	2	5	yes
m95	52806	404	1639	1528	19810	4	86	yes
m96	397707	281	1548	4351	33448	26	199	yes
m97	17648	4742	4107	9743	4935	259	87	yes
m98	813	59	19	138	45	0	17	yes
m99	226428	13429	3990	10916	71504	131	253	no
m100	63251	11197	1414	161837	14776	1319	336	no
m101	5526	1949	875	178	412	62	204	no
m102	154409	7480	2529	106423	7163	927	336	no
m103	90413	6007	188	94	155	1	5	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m104	20759	2494	20233	3323	443	170	2290	yes
m105	3184	90316	48322	88	247	1142	293	no
m106	2911599	396567	353819	154413	99673	3253	2838	yes
m107	391857	13690	5692	167738	39163	987	983	yes
m108	343	944	215	126822	82	475	60	yes
m109	1522	200	68	425	146	7	5	yes
m110	22135	16	46	2	22129	16	46	no
m111	23	108	157	0	0	29	94	no
m112	457	58492	205	6113	12	42951	10	yes
m113	26806	31693	3965	5087	739	1378	312	no
m114	145888	301869	114063	69	2095	2666	13104	yes
m115	31969	7530	2431	26602	1568	907	400	no
m116	1647	13	59	263	234	1	5	no
m117	487	581	4823	5	20	32	3277	no
m118	0	338	10	1268	0	84	3	no
m119	2011	154	41	9873	54	99	22	yes
m120	2722	53784	29050	16	10	426	100	yes
m121	214	22	6	35	150	10	4	no
m122	399	9	76	21	188	4	69	no
m123	1933	8973	1193	375	2	418	3	no
m124	247965	35666	56925	19671	5097	1356	2503	no
m125	62442	13304	1293	2013	1002	82	8	no
m126	3070322	24482	186036	4713	1205128	2569	71163	yes
m127	146239	6726	4688	20121	2361	174	207	no
m128	9955	38	38	22	148	6	1	no
m129	914	5	88	2	115	0	66	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m130	11711	189	264	14	1168	6	16	no
m131	36101	3234	1539	270	44	25	19	no
m132	168209	477	2820	165	374	13	18	no
m133	27541	431	217	149	201	24	61	no
m134	27233	836	1359	325	155	99	36	no
m135	980	93	141	25	141	35	78	no
m136	18449	1320	2245	529	386	63	99	no
m137	1056	11	90	3	104	1	4	no
m138	10904	70	855	14	535	1	109	no
m139	417	14	103	1	265	1	71	no
m140	11502	508	212	410	705	19	57	no
m141	78620	7	10	9	273	0	0	no
m142	199	55	26	21	141	0	1	no
m143	784	59	110	77	67	5	60	no
m144	19494	91	73	196	343	7	21	no
m145	43296	5210	4109	491	2664	132	99	no
m146	75	299	37	211	24	37	10	no
m147	6248	117	201	165	125	8	2	no
m148	574	7715	295	1641	85	265	33	no
m149	133	125	81	390	10	59	16	no
m150	477	26	142	10	167	1	38	no
m151	67	230	32	311	32	24	9	no
m152	12644	136	838	29	216	0	1	no
m153	23559	28	133	252	380	16	6	no
m154	102	11	26	5	91	0	3	no
m155	148	47	67	60	53	33	13	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m156	9725	1140	314	677	157	60	22	no
m157	6595	3057	1246	581	629	42	103	no
m158	89	25	47	69	49	0	3	no
m159	30496	3667	349	221	204	7	4	no
m160	1546	1999	893	48	74	3	6	no
m161	2144	818	183	2634	3	151	16	no
m162	34	607	171	75	3	0	2	no
m163	300	5518	88	3375	38	98	0	no
m164	175	230	27	192	1	3	7	no
m165	38851	76	90	24	133	18	5	no
m166	3712	124	69	23	457	0	26	no
m167	393	65	21	5	118	5	7	no
m168	119	522	331	4965	17	163	22	no
m169	51823	357	1536	170	634	68	64	no
m170	1896	3333	222	506	224	81	26	no
m171	31574	501	131	610	218	37	70	no
m172	2427	13	107	9	128	3	1	no
m173	4338	353	84	505	59	29	11	no
m174	2606	54	303	120	136	19	16	no
m175	268	10	542	65	45	7	515	no
m176	106	2083	28	116	1	1968	0	no
m177	4249	991	378	288	848	319	13	no
m178	194	4	15	96	50	0	10	no
m179	37970	5848	10559	2153	528	241	170	no
m180	84755	104	4517	24	131	0	86	no
m181	174596	35	8567	19	104	0	11	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m182	2287	1092	214	491	544	270	14	no
m183	210	816	11	175	6	40	0	no
m184	8124	773	635	0	1925	13	156	no
m185	4806	3633	3193	682	25	155	72	no
m186	552	71	33	218	8	52	18	no
m187	107	30	33	190	8	7	19	no
m188	141867	19497	8173	1152	935	196	57	no
m189	165	34	29	124	6	4	0	no
m190	2888	20	19	232	1	5	0	no
m191	9321	409	2130	120	315	59	83	no
m192	3233	305	4176	27	821	11	1057	no
m193	8359	6948	115	11	119	11	12	no
m194	59130	0	9	0	104	0	2	no
m195	30887	5406	751	290	900	130	112	no
m196	862	4380	132	1274	28	9	2	no
m197	2807	14723	254	16240	23	1736	26	no
m198	6	2	1	251	5	2	0	no
m199	133	119	70	251	13	16	8	no
m200	28715	205	288	47	412	16	23	no
m201	15893	1328	358	1230	409	100	23	no
m202	1374	907	93	240	58	7	21	no
m203	1417	52	113	129	52	0	31	no
m204	202	306	55	517	32	14	2	no
m205	25553	452	480	165	1576	27	68	no
m206	103090	2047	577	97	178	22	35	no
m207	565	57	119	55	114	11	20	no

name	# rev. ovary piRNA	# fwd. testis piRNA	# rev. testis piRNA	# fwd. ovary piRNA uniqmap	# rev. ovary piRNA uniqmap	# fwd. testis piRNA uniqmap	# rev testis piRNA uniqmap	homologous clust. in pseudo?
m208	8393	510	2230	427	120	55	33	no
m209	17374	76	77	225	39	16	4	no
m210	1193	332	153	532	415	44	58	no
m211	252	62	44	186	4	3	0	no
m212	1413	58	54	226	94	0	6	no
m213	39152	232	3065	147	365	29	25	no
m214	48929	93	115	120	1399	45	46	no
m215	1047	15	47	1023	120	5	34	no
m216	3231	31	155	1	98	3	1	no
m217	1216	83	54	640	7	5	1	no
m218	209	199	112	264	117	137	73	no
m219	14969	1442	1631	91	166	8	16	no
m220	68	2052	173	2020	5	174	10	no
m221	24	308	9	284	0	79	0	no
m222	264	13	164	129	0	0	0	no
m223	0	418	4812	0	0	133	1549	no
m224	0	22503	6027	0	0	4202	669	no
m225	0	3005	192	0	0	90	14	no
m226	0	74670	2379	0	0	5622	229	no
m227	130349	12191	13261	10255	21681	888	1610	no

Table 3. Table of homologous piRNA clusters between *D. pseudoobscura* and *D. miranda*.

pseudo_cluster	miranda_cluster
p101	m94
p101	m95
p101	m96
p103	m97
p103	m98
p105	m104
p106	m104
p107	m104
p109	m106
p110	m107
p111	m108
p116	m112
p121	m109
p126	m114
p140	m119
p147	m120
p153	m126
p154	m126
p184	m68
p186	m69
p193	m72
p202	m82
p36	m43
p44	m46
p88	m62
p89	m6

Table 4. Table showing overlap of repeats annotated by RepeatModeler for 1:1 homologous clusters between *D. pseudoobscura* and *D. miranda*.

D. pseudoobscura cluster	*D. miranda* cluster	repeats in both clusters	repeats only in *D. pseudoobscura* cluster	repeats only in *D. miranda* cluster
p126	m114	5	6	29
p89	m6	2	12	10
p88	m62	2	12	15
p109	m106	130	117	106
p121	m109	0	5	4
p44	m46	4	1	8
p193	m72	2	7	17
p110	m107	42	61	90
p111	m108	6	9	11
p116	m112	2	4	5
p36	m43	3	18	13
p184	m68	6	4	0
p186	m69	3	4	14
p140	m119	6	3	0
p202	m82	5	13	24
p147	m120	7	6	6

Table S1. Table showing total base pairs mapped to each specific repeat from the RepeatModeler

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-428	0	116
rnd-1_family-429	188	0
rnd-1_family-420	0	0
rnd-1_family-421	3013	5711
rnd-1_family-422	105453	165419
rnd-1_family-423	40934	183098
rnd-1_family-424	25761	60218
rnd-1_family-425	1834	237650
rnd-1_family-426	1132	1237
rnd-1_family-427	70	0
rnd-5_family-1887	5246	7165
rnd-6_family-558	6568	131239
rnd-5_family-4385	9735	7570
rnd-5_family-545	26105	26196
rnd-5_family-3926	128976	6738
rnd-2_family-195	86	268
rnd-6_family-1973	14125	15508
rnd-1_family-789	3858	1864
rnd-3_family-102	12514	21310
rnd-3_family-40	36245	78002
rnd-5_family-104	30341	25854
rnd-5_family-2225	0	36818
rnd-1_family-589	18567	38267
rnd-1_family-588	727	0
rnd-1_family-587	42822	45696
rnd-1_family-586	42353	125594
rnd-1_family-585	26861	23638
rnd-1_family-584	3827	2859
rnd-1_family-583	16750	5999
rnd-1_family-582	0	334
rnd-1_family-581	4250	4105
rnd-1_family-580	20079	8530
rnd-1_family-635	13329	2536
rnd-1_family-634	2172	542
rnd-1_family-637	83	133
rnd-1_family-636	2575	3541
rnd-1_family-631	739	0
rnd-1_family-630	1749	2314
rnd-1_family-633	1334	3611

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-632	935	976
rnd-1_family-728	31783	5817
rnd-1_family-639	10978	9477
rnd-1_family-638	145	183
rnd-5_family-2157	2618	8419
rnd-6_family-5307	82720	78689
rnd-6_family-4575	19725	29299
rnd-5_family-2156	40163	107921
rnd-5_family-187	29788	30750
rnd-1_family-729	2862	2654
rnd-6_family-1786	2669	899
rnd-5_family-181	9788	10871
rnd-1_family-280	93347	66742
rnd-1_family-281	112	79
rnd-1_family-282	385	485
rnd-1_family-283	483685	219171
rnd-1_family-284	5671	57608
rnd-1_family-285	1350	7023
rnd-1_family-286	61229	464182
rnd-1_family-287	1863	5405
rnd-1_family-288	12115	1375
rnd-1_family-289	2598	2907
rnd-6_family-5391	16172	16773
rnd-4_family-194	88425	132461
rnd-1_family-738	8246	13677
rnd-1_family-721	502	1370
rnd-1_family-775	1581	36841
rnd-1_family-734	12945	13085
rnd-1_family-735	2815	6850
rnd-1_family-736	12588	11538
rnd-1_family-720	264	340
rnd-1_family-730	173	52
rnd-1_family-731	31676	2265
rnd-1_family-732	10846	14953
rnd-1_family-733	9825	37536
rnd-3_family-409	180804	353978
rnd-5_family-1285	21418	24994
rnd-6_family-4611	20268	92741
rnd-6_family-1950	24260	40497
rnd-5_family-1358	12339	121479
rnd-5_family-1210	12056	10293

repeat	*D. pseudoobscura*	*D. miranda*
rnd-3_family-151	2804163	2590673
rnd-1_family-129	882173	140715
rnd-1_family-44	20594	52637
rnd-1_family-47	2334	6095
rnd-1_family-46	4635	10680
rnd-1_family-41	36568	108925
rnd-1_family-40	35388	73032
rnd-1_family-43	4981	42648
rnd-1_family-42	8951	28044
rnd-1_family-121	48015	58702
rnd-1_family-120	888	372
rnd-1_family-123	68839	831505
rnd-1_family-122	282985	193525
rnd-1_family-49	18146	65582
rnd-1_family-48	42714	321714
rnd-1_family-127	92566	815979
rnd-1_family-126	151389	179978
rnd-4_family-1419	4575	5972
rnd-1_family-349	137148	63362
rnd-1_family-348	9203	9132
rnd-5_family-4349	216868	323607
rnd-1_family-345	1176	18196
rnd-1_family-344	52612	140490
rnd-1_family-347	22058	42117
rnd-1_family-346	12596	37905
rnd-1_family-341	25687	114557
rnd-1_family-340	35815	117106
rnd-1_family-343	6917	4371
rnd-1_family-342	8722	8314
rnd-5_family-7432	6805	13252
rnd-5_family-248	86229	50839
rnd-5_family-679	11339	231784
rnd-1_family-751	16329	9406
rnd-5_family-522	31	612
rnd-5_family-525	41352	53277
rnd-5_family-524	83846	203986
rnd-5_family-526	26427	40419
rnd-5_family-1308	121226	38973
rnd-6_family-906	3135	184307
rnd-6_family-3235	797	52412
rnd-6_family-9368	12893	12154

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-773	5663	35446
rnd-1_family-798	1024	1118
rnd-5_family-2623	20549	18203
rnd-1_family-799	4558	3145
rnd-6_family-734	11182	36973
rnd-5_family-6343	25836	11789
rnd-1_family-231	0	0
rnd-1_family-230	63	68
rnd-1_family-233	0	1126
rnd-1_family-232	28051	36559
rnd-1_family-235	25284	26055
rnd-1_family-234	26159	77433
rnd-1_family-237	58065	89821
rnd-1_family-236	15993	19352
rnd-1_family-239	0	0
rnd-1_family-238	2339	9017
rnd-1_family-541	34056	96620
rnd-1_family-540	10682	16869
rnd-1_family-547	25652	30716
rnd-1_family-546	37659	28012
rnd-1_family-545	15326	49333
rnd-1_family-544	11934	60427
rnd-1_family-464	9225	9903
rnd-1_family-465	28812	24597
rnd-1_family-466	55513	154587
rnd-1_family-467	156394	241683
rnd-1_family-460	74	176
rnd-1_family-461	9926	11469
rnd-1_family-462	6938	37973
rnd-1_family-463	432	1425
rnd-1_family-785	8746	19838
rnd-1_family-784	707	38864
rnd-1_family-787	3182	5968
rnd-1_family-786	0	107
rnd-1_family-468	2142	2603
rnd-1_family-469	2962	148544
rnd-1_family-783	7922	11157
rnd-1_family-782	6591	15737
rnd-5_family-2530	21541	79710
rnd-5_family-3472	49563	50056
rnd-5_family-1948	4884	17983

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-986	8341	51098
rnd-5_family-358	7928	19292
rnd-6_family-6	33646	40355
rnd-5_family-4902	14643	98276
rnd-1_family-101	0	288
rnd-1_family-338	0	0
rnd-1_family-339	69419	10771
rnd-1_family-100	47156	56054
rnd-6_family-8049	889227	1081978
rnd-1_family-330	9156	53648
rnd-1_family-331	133943	212091
rnd-1_family-332	0	0
rnd-1_family-333	4135	3188
rnd-1_family-334	0	73
rnd-1_family-335	76332	17910
rnd-1_family-336	3878	33646
rnd-1_family-337	379	2198
rnd-5_family-7004	21787	15264
rnd-6_family-2793	13780	55282
rnd-5_family-3283	13213	27113
rnd-6_family-4544	27202	56772
rnd-5_family-256	12490	35931
rnd-5_family-11416	68215	47564
rnd-6_family-2941	190065	353864
rnd-1_family-109	420841	990527
rnd-6_family-1305	9551	17159
rnd-1_family-108	120174	429157
rnd-6_family-2856	7067	13541
rnd-1_family-563	941	4197
rnd-1_family-562	0	1560
rnd-5_family-5662	13828	18927
rnd-1_family-165	190976	814651
rnd-1_family-164	8913	21499
rnd-1_family-167	66477	37570
rnd-1_family-166	7661	39837
rnd-1_family-161	206169	188009
rnd-1_family-160	154	12400
rnd-1_family-163	2277	9477
rnd-1_family-162	391222	321433
rnd-5_family-557	15980	28604
rnd-5_family-4402	24377	84143

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-4543	15995	74255
rnd-6_family-5135	27701	115509
rnd-1_family-169	424	0
rnd-1_family-168	10852	18022
rnd-1_family-569	5325	10416
rnd-4_family-121	310682	2654826
rnd-1_family-568	3516	6150
rnd-6_family-4578	2936	6372
rnd-3_family-8	324794	871145
rnd-1_family-653	12490	39030
rnd-4_family-124	299148	1818522
rnd-1_family-652	6962	13944
rnd-6_family-513	12862	16633
rnd-1_family-651	0	0
rnd-5_family-144	28603	48663
rnd-1_family-763	53207	25165
rnd-1_family-762	2471	77116
rnd-1_family-761	57651	55940
rnd-1_family-650	2559	10111
rnd-1_family-608	4506	19600
rnd-1_family-609	0	0
rnd-1_family-765	20563	83218
rnd-1_family-764	166	69
rnd-1_family-604	0	0
rnd-1_family-56	1763	2291
rnd-1_family-606	608	386
rnd-1_family-607	0	0
rnd-1_family-600	7927	3780
rnd-1_family-601	0	0
rnd-1_family-602	69696	129960
rnd-1_family-57	72272	428607
rnd-5_family-891	16602	39102
rnd-6_family-386	8104	10064
rnd-5_family-2338	55133	46152
rnd-1_family-655	135765	55714
rnd-5_family-2588	4871	12553
rnd-5_family-2589	36877	58345
rnd-1_family-654	4413	3191
rnd-5_family-76	11698	17545
rnd-1_family-700	1961	11931
rnd-6_family-6776	22989	26841

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-2824	152860	74641
rnd-5_family-50	9977	17819
rnd-1_family-77	53144	102425
rnd-1_family-510	2130	8294
rnd-1_family-58	363514	424946
rnd-5_family-71	10414	72493
rnd-5_family-3617	36	4990
rnd-1_family-157	61985	628320
rnd-5_family-1088	23500	35133
rnd-4_family-490	41922	54057
rnd-5_family-660	21276	424840
rnd-6_family-6071	3455	12958
rnd-4_family-498	46028	98997
rnd-1_family-110	204637	308454
rnd-1_family-111	905	1492
rnd-1_family-112	104746	428518
rnd-1_family-113	79748	7575
rnd-1_family-114	145302	432127
rnd-1_family-115	28260	125046
rnd-1_family-18	29032	162857
rnd-1_family-117	0	0
rnd-1_family-118	40135	65706
rnd-1_family-119	7553	37797
rnd-1_family-14	2042	5340
rnd-1_family-15	953532	2536768
rnd-1_family-12	4454	13551
rnd-1_family-13	30109	137094
rnd-1_family-10	4747	9134
rnd-1_family-11	49256	243301
rnd-1_family-374	1792	2371
rnd-1_family-375	148856	406946
rnd-1_family-376	1948	40043
rnd-1_family-377	6871	67062
rnd-1_family-279	76590	402702
rnd-1_family-278	0	0
rnd-1_family-372	1909	3396
rnd-1_family-373	181479	275258
rnd-1_family-275	0	0
rnd-1_family-274	276326	457895
rnd-1_family-277	43831	369910
rnd-1_family-276	0	416

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-271	53455	454632
rnd-1_family-270	23030	13380
rnd-1_family-273	82204	231526
rnd-1_family-272	27551	114716
rnd-1_family-507	21886	16113
rnd-1_family-506	129294	327529
rnd-1_family-505	380	61865
rnd-1_family-504	65388	221638
rnd-1_family-503	0	189708
rnd-1_family-502	0	0
rnd-1_family-501	48567	76614
rnd-1_family-500	7293	18962
rnd-1_family-509	1014	16445
rnd-1_family-508	4001	13873
rnd-5_family-564	9686	11326
rnd-5_family-508	12619	17248
rnd-6_family-4734	4731	1167
rnd-6_family-250	48733	51161
rnd-3_family-10	206274	470782
rnd-1_family-536	15929	72656
rnd-5_family-3784	631708	824247
rnd-5_family-1935	174222	752229
rnd-1_family-760	2491	22494
rnd-5_family-6330	30443	112904
rnd-1_family-767	3562	5464
rnd-4_family-111	41578	43347
rnd-1_family-766	13281	12063
rnd-6_family-1689	4095	8888
rnd-1_family-439	3064	27579
rnd-1_family-438	0	284
rnd-5_family-1614	20321	27726
rnd-1_family-433	0	0
rnd-1_family-432	2787	1696
rnd-1_family-431	0	194
rnd-1_family-430	0	0
rnd-1_family-437	4547	29269
rnd-1_family-436	0	0
rnd-1_family-435	21289	7498
rnd-1_family-434	34555	83130
rnd-5_family-4887	17013	35312
rnd-6_family-4876	1718	55633

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-605	1081	1189
rnd-4_family-1031	5468	7842
rnd-3_family-374	465593	82195
rnd-5_family-6727	4925	41971
rnd-1_family-769	422	786
rnd-1_family-628	42778	42258
rnd-1_family-768	5512	120931
rnd-6_family-19	522035	50252
rnd-6_family-2196	191014	393477
rnd-1_family-208	99	261
rnd-1_family-209	27988	129542
rnd-6_family-8134	923991	1579416
rnd-5_family-10308	621380	716781
rnd-1_family-200	51931	651684
rnd-1_family-201	148	191
rnd-1_family-202	0	0
rnd-1_family-203	154	344
rnd-1_family-204	901	9121
rnd-1_family-205	0	70
rnd-1_family-206	3760	3917
rnd-1_family-207	86086	120470
rnd-1_family-660	0	1081
rnd-5_family-1930	35524	32398
rnd-6_family-407	12866	29552
rnd-3_family-99	21738	34292
rnd-4_family-232	126100	657389
rnd-4_family-236	1700224	14826
rnd-5_family-556	35112	106270
rnd-1_family-293	18384	51468
rnd-1_family-292	337318	56620
rnd-1_family-291	53747	450701
rnd-1_family-290	0	1105
rnd-1_family-297	57167	126776
rnd-1_family-296	107003	107990
rnd-1_family-295	1678	8071
rnd-1_family-294	15664	23085
rnd-1_family-299	152434	152317
rnd-1_family-298	6535	2455
rnd-6_family-5068	7375	12314
rnd-1_family-486	18868	80246
rnd-1_family-487	442	191490

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-484	6025	492613
rnd-1_family-485	1211	3076
rnd-1_family-482	45497	43492
rnd-1_family-483	8105	262839
rnd-1_family-480	24674	107471
rnd-1_family-481	50089	48283
rnd-1_family-727	36123	53994
rnd-1_family-726	23357	60909
rnd-1_family-725	1373	2968
rnd-1_family-724	9048	101123
rnd-1_family-723	2112	1755
rnd-1_family-722	7814	19385
rnd-1_family-488	3678	3019
rnd-1_family-489	0	286
rnd-1_family-27	16536	38061
rnd-1_family-26	1511	3287
rnd-1_family-25	0	0
rnd-1_family-24	705	773
rnd-1_family-23	22301	82175
rnd-1_family-22	12489	116231
rnd-1_family-21	2558	4208
rnd-1_family-20	233947	3715240
rnd-6_family-2942	1475433	945402
rnd-1_family-29	80004	154665
rnd-1_family-28	76315	8511
rnd-5_family-4822	31270	12155
rnd-5_family-1330	166252	38365
rnd-5_family-1207	11999	24792
rnd-1_family-737	260	1410
rnd-5_family-4888	31484	98491
rnd-4_family-862	19837	22679
rnd-5_family-119	0	44
rnd-1_family-52	29990	17876
rnd-1_family-53	4335	4811
rnd-1_family-50	8602	34419
rnd-1_family-51	164	388
rnd-1_family-158	23491	109473
rnd-1_family-159	8772	16761
rnd-1_family-54	3951	6473
rnd-1_family-55	8820	13454
rnd-1_family-154	93767	91791

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-155	11106	164085
rnd-1_family-156	1032	4727
rnd-1_family-59	224539	1648177
rnd-1_family-150	67845	32032
rnd-1_family-151	40140	65684
rnd-1_family-152	16357	23297
rnd-1_family-153	38479	59042
rnd-5_family-1387	11874	12515
rnd-5_family-2243	24699	30191
rnd-1_family-691	1921	3034
rnd-5_family-1381	20161	23306
rnd-6_family-6311	3060	9768
rnd-6_family-6088	10779	60929
rnd-5_family-250	36637	68527
rnd-5_family-17	87896	218896
rnd-6_family-2438	8462	21078
rnd-5_family-18	554867	649166
rnd-5_family-19	12026	100251
rnd-6_family-25	1577	5299
rnd-5_family-6789	41403	38393
rnd-1_family-793	0	0
rnd-3_family-164	488	3275
rnd-5_family-1981	12419	13614
rnd-5_family-828	3985	2895
rnd-5_family-829	94164	52959
rnd-6_family-803	1933	3234
rnd-6_family-22	1989	6873
rnd-1_family-641	33785	58009
rnd-1_family-244	0	436
rnd-1_family-245	143072	376389
rnd-1_family-246	9035	13192
rnd-1_family-247	6603	10823
rnd-1_family-240	30517	38460
rnd-1_family-241	65440	140427
rnd-1_family-242	0	0
rnd-1_family-243	0	220
rnd-1_family-576	187	8316
rnd-1_family-577	0	0
rnd-1_family-574	24519	186062
rnd-1_family-575	73399	143192
rnd-1_family-248	73779	194109

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-249	25221	312356
rnd-1_family-570	3739	186
rnd-1_family-571	4901	5406
rnd-1_family-477	3390	3512
rnd-1_family-476	23247	30017
rnd-1_family-475	7542	34499
rnd-1_family-474	6152	28670
rnd-1_family-473	3403	2742
rnd-1_family-472	54946	61351
rnd-1_family-471	2710	3221
rnd-1_family-470	32117	53613
rnd-1_family-479	11049	217781
rnd-1_family-478	0	0
rnd-5_family-2523	28735	18103
rnd-6_family-645	21294	38406
rnd-5_family-3973	4037	28783
rnd-5_family-69	10371	89076
rnd-1_family-329	8594	12504
rnd-1_family-328	69365	391353
rnd-1_family-323	336127	157683
rnd-1_family-322	241987	303981
rnd-1_family-321	56277	51918
rnd-1_family-320	284669	434836
rnd-1_family-327	21222	21190
rnd-1_family-326	7618	8553
rnd-1_family-325	213397	349509
rnd-1_family-324	33001	96877
rnd-6_family-1826	51093	7469
rnd-1_family-684	9026	85487
rnd-1_family-685	6377	11009
rnd-1_family-686	0	0
rnd-1_family-687	197	205
rnd-1_family-680	0	0
rnd-1_family-681	7514	6960
rnd-1_family-682	19589	25207
rnd-1_family-683	18136	42720
rnd-1_family-688	0	0
rnd-1_family-689	34563	141627
rnd-1_family-396	11917	34975
rnd-1_family-397	0	0
rnd-1_family-394	20231	50827

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-395	92467	55698
rnd-1_family-392	0	0
rnd-1_family-393	234694	26497
rnd-1_family-390	160220	351620
rnd-1_family-391	14366	9987
rnd-1_family-579	155869	98094
rnd-5_family-4496	11030	21589
rnd-1_family-398	26311	20200
rnd-1_family-399	0	1257
rnd-1_family-750	5813	109130
rnd-1_family-198	248496	285535
rnd-1_family-199	233043	624285
rnd-4_family-555	5690	12014
rnd-6_family-5829	11646	21661
rnd-1_family-116	36852	91427
rnd-5_family-1832	37870	117460
rnd-1_family-190	11730	54839
rnd-1_family-191	2013	29372
rnd-1_family-192	0	106
rnd-1_family-193	0	516
rnd-1_family-194	0	0
rnd-1_family-195	740	0
rnd-1_family-196	124306	85425
rnd-1_family-197	2902	4970
rnd-1_family-16	3602	2408
rnd-4_family-349	54578	163113
rnd-1_family-17	1034	125034
rnd-6_family-1700	7396	16141
rnd-3_family-119	33336	85282
rnd-6_family-6455	3802	209696
rnd-6_family-4577	10096	20018
rnd-6_family-4558	33268	35674
rnd-5_family-1618	6621	266954
rnd-1_family-756	132527	33592
rnd-1_family-757	286	1146
rnd-1_family-754	0	3009
rnd-1_family-755	2663	8503
rnd-1_family-752	944	809
rnd-1_family-753	7085	10953
rnd-1_family-619	1106	236868
rnd-1_family-618	8336	15328

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-617	617	1621
rnd-1_family-616	12408	11751
rnd-1_family-615	7499	10868
rnd-1_family-614	189	13327
rnd-1_family-613	37886	47973
rnd-1_family-612	482	3345
rnd-1_family-611	0	0
rnd-1_family-610	0	0
rnd-6_family-6194	1657	86634
rnd-5_family-969	3765	78384
rnd-6_family-144	22619	16361
rnd-5_family-492	18805	13264
rnd-1_family-620	8289	35691
rnd-6_family-398	35452	74899
rnd-5_family-183	70176	53876
rnd-1_family-621	37	37
rnd-5_family-4059	13856	5880
rnd-5_family-219	57606	35218
rnd-1_family-370	27658	112794
rnd-5_family-603	2020	1682
rnd-4_family-137	60065	107131
rnd-1_family-371	0	58
rnd-1_family-45	186436	58099
rnd-5_family-777	137548	127042
rnd-5_family-8	39011	63272
rnd-1_family-128	12791	25635
rnd-5_family-675	7586	9639
rnd-6_family-4568	1677	5099
rnd-5_family-778	5154	4714
rnd-1_family-642	1193	1849
rnd-5_family-498	189454	78162
rnd-1_family-103	57771	49887
rnd-1_family-102	63408	201096
rnd-1_family-69	176042	333559
rnd-1_family-68	32944	62375
rnd-1_family-107	65272	188434
rnd-1_family-106	342178	341940
rnd-1_family-105	1457	2495
rnd-1_family-104	26760	75604
rnd-1_family-63	134646	258469
rnd-1_family-62	12780	42003

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-61	53	150
rnd-1_family-60	345795	1088517
rnd-1_family-67	710977	2060281
rnd-1_family-66	14129	25396
rnd-1_family-65	37931	283649
rnd-1_family-64	14344	351058
rnd-1_family-367	61557	299360
rnd-1_family-366	0	0
rnd-1_family-365	89083	152571
rnd-1_family-364	25863	42902
rnd-1_family-363	0	814
rnd-1_family-362	345	690
rnd-1_family-361	6320	6591
rnd-1_family-360	3194	3274
rnd-1_family-378	786	574
rnd-5_family-30	26692	36101
rnd-1_family-647	83987	109900
rnd-5_family-911	158098	173897
rnd-1_family-369	115227	157284
rnd-1_family-368	805	568
rnd-1_family-648	0	0
rnd-1_family-538	3875	218
rnd-1_family-539	1509	62199
rnd-1_family-649	0	0
rnd-1_family-532	28440	42589
rnd-1_family-533	2001	1970
rnd-1_family-530	37767	56220
rnd-1_family-531	138	442
rnd-1_family-379	39261	56742
rnd-1_family-537	23945	39630
rnd-1_family-534	7786	80224
rnd-1_family-535	25246	9793
rnd-5_family-599	24860	45172
rnd-5_family-1181	6766	337448
rnd-5_family-2078	12786	12201
rnd-1_family-125	1762	1466
rnd-4_family-1526	6621	16869
rnd-1_family-124	0	0
rnd-5_family-11386	60552	10161
rnd-6_family-221	4292	15047
rnd-5_family-3787	28761	77717

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-138	15201	14375
rnd-1_family-408	1325	20990
rnd-1_family-409	54230	54607
rnd-1_family-406	183	2478
rnd-1_family-407	7818	9137
rnd-1_family-404	0	3922
rnd-1_family-405	38582	53517
rnd-1_family-402	12588	12469
rnd-1_family-403	0	0
rnd-1_family-400	20695	21740
rnd-1_family-401	9796	32425
rnd-6_family-6947	44895	29041
rnd-5_family-1082	15202	37654
rnd-1_family-770	0	0
rnd-1_family-771	18221	55149
rnd-5_family-1528	14666	16118
rnd-6_family-1694	7747	31076
rnd-1_family-219	96580	380442
rnd-1_family-218	1911	12044
rnd-1_family-657	51871	97600
rnd-6_family-2270	1233476	1323
rnd-1_family-213	47283	46198
rnd-1_family-212	82361	261845
rnd-1_family-211	7886	19959
rnd-1_family-210	212	0
rnd-1_family-217	70076	397780
rnd-1_family-216	76637	508475
rnd-1_family-215	0	0
rnd-1_family-214	0	0
rnd-1_family-712	0	427
rnd-1_family-713	5517	42132
rnd-1_family-710	29916	2618
rnd-1_family-711	11881	16711
rnd-1_family-716	7183	5291
rnd-1_family-717	0	0
rnd-1_family-714	21609	42786
rnd-1_family-715	2722	3175
rnd-1_family-718	4711	37154
rnd-1_family-719	5221	6598
rnd-1_family-778	44191	67287
rnd-6_family-9834	15988	29545

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-656	2505	26991
rnd-4_family-229	93776	141671
rnd-1_family-779	180437	33457
rnd-6_family-4094	16658	3524
rnd-1_family-672	916	9494
rnd-3_family-364	0	1166096
rnd-5_family-1853	0	0
rnd-6_family-1574	14128	14610
rnd-6_family-4857	9449	29044
rnd-1_family-499	19202	44941
rnd-1_family-498	11870	13389
rnd-6_family-2576	12133	44568
rnd-5_family-382	44293	48911
rnd-6_family-2575	11178	68611
rnd-1_family-491	20464	30296
rnd-1_family-490	1095	1322
rnd-1_family-493	1425	1802
rnd-1_family-492	1020	78455
rnd-1_family-495	10144	27325
rnd-1_family-494	3747	6614
rnd-1_family-497	50144	159839
rnd-1_family-496	1184	19545
rnd-1_family-34	1420	3617
rnd-1_family-35	24434	49467
rnd-1_family-36	5379	7377
rnd-1_family-37	1970	3658
rnd-1_family-30	138	223
rnd-1_family-31	210904	303784
rnd-1_family-32	2749	2927
rnd-1_family-33	24354	44716
rnd-5_family-2778	10380	9716
rnd-1_family-38	712	866
rnd-1_family-39	31033	105704
rnd-5_family-2954	38344	38319
rnd-1_family-578	0	0
rnd-5_family-1309	66198	141314
rnd-5_family-1663	46605	61169
rnd-5_family-1307	142901	73886
rnd-5_family-488	76664	49461
rnd-5_family-720	4350	5391
rnd-5_family-523	15565	67113

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-437	45919	86284
rnd-1_family-748	4040	4857
rnd-5_family-3526	17619	23051
rnd-6_family-1379	11395	19083
rnd-6_family-6418	19594	27848
rnd-5_family-439	33051	68798
rnd-5_family-438	61632	13600
rnd-1_family-666	17539	47731
rnd-1_family-667	9212	42645
rnd-5_family-1868	32121	44558
rnd-1_family-665	13163	21515
rnd-1_family-662	0	1694
rnd-1_family-663	381	834
rnd-1_family-149	66092	106109
rnd-1_family-148	23901	46903
rnd-1_family-147	55096	88177
rnd-1_family-146	63910	186086
rnd-1_family-145	23887	556981
rnd-1_family-144	42984	41110
rnd-1_family-143	335770	314490
rnd-1_family-142	54264	441003
rnd-1_family-141	0	0
rnd-1_family-140	0	624
rnd-5_family-794	16789	50723
rnd-5_family-117	33739	13607
rnd-5_family-791	400	5806
rnd-5_family-4142	19414	24071
rnd-4_family-903	186425	299638
rnd-4_family-1635	22887	30181
rnd-1_family-89	44328	80131
rnd-1_family-88	56493	144102
rnd-6_family-7085	37526	44496
rnd-5_family-1543	32698	60806
rnd-1_family-81	198801	267852
rnd-1_family-80	143133	474442
rnd-1_family-83	197968	813149
rnd-1_family-82	407966	115246
rnd-1_family-85	262299	354850
rnd-1_family-84	876097	1596109
rnd-1_family-87	605576	717508
rnd-1_family-86	322903	1362371

repeat	*D. pseudoobscura*	*D. miranda*
rnd-4_family-2304	10713	27528
rnd-5_family-371	60961	133109
rnd-5_family-1508	11654	22952
rnd-1_family-659	0	0
rnd-5_family-372	53395	102135
rnd-6_family-1197	21544	21401
rnd-6_family-756	31935	198105
rnd-5_family-373	6091	1845
rnd-5_family-535	5693	16586
rnd-6_family-4338	4041	147435
rnd-6_family-2695	54953	81257
rnd-1_family-257	410	976
rnd-1_family-256	52334	101164
rnd-1_family-255	100844	467577
rnd-1_family-254	204225	183192
rnd-1_family-253	62062	278530
rnd-1_family-252	1819	3581
rnd-1_family-251	29336	64892
rnd-1_family-250	1524	277068
rnd-5_family-1647	17725	19282
rnd-5_family-1003	1186	15141
rnd-6_family-3186	360140	274112
rnd-1_family-259	70098	644676
rnd-1_family-258	35936	92247
rnd-1_family-561	842	929
rnd-1_family-560	30290	55579
rnd-1_family-448	6454	5538
rnd-1_family-449	35325	22011
rnd-1_family-565	0	0
rnd-1_family-564	1241	5227
rnd-1_family-567	27311	103327
rnd-1_family-566	19985	29507
rnd-1_family-442	1780	4747
rnd-1_family-443	370	0
rnd-1_family-440	93472	138189
rnd-1_family-441	5364	497680
rnd-1_family-446	1740	3161
rnd-1_family-447	0	111
rnd-1_family-444	0	0
rnd-1_family-445	26613	16934
rnd-1_family-4	29437	69000

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-5	219385	510287
rnd-1_family-6	42064	58182
rnd-1_family-7	142856	433426
rnd-1_family-0	2144937	5769017
rnd-1_family-1	12997	22728
rnd-1_family-2	262391	189445
rnd-1_family-3	1983906	8827155
rnd-5_family-6719	139807	178020
rnd-3_family-74	445043	435697
rnd-3_family-75	96766	109063
rnd-1_family-8	43220	65865
rnd-1_family-9	18308	32524
rnd-6_family-1994	41011	71778
rnd-1_family-318	0	4321
rnd-1_family-319	0	0
rnd-1_family-316	322884	36223
rnd-1_family-317	557	2242
rnd-1_family-314	1574	21260
rnd-1_family-315	0	0
rnd-1_family-312	90337	402614
rnd-1_family-313	99534	454121
rnd-1_family-310	6265	90433
rnd-1_family-311	1188	2131
rnd-5_family-1	3778	74965
rnd-5_family-4	2559	75499
rnd-5_family-5	5604	21876
rnd-1_family-19	18270	56613
rnd-1_family-572	178	0
rnd-6_family-10458	2558	114903
rnd-6_family-2160	36495	25142
rnd-1_family-697	19263	43375
rnd-1_family-696	32391	56335
rnd-1_family-695	17370	25953
rnd-1_family-694	1221	730
rnd-1_family-693	1239	1896
rnd-1_family-692	50043	4407
rnd-1_family-645	16109	5040
rnd-1_family-690	19892	48955
rnd-6_family-3081	54573	56929
rnd-1_family-699	0	0
rnd-1_family-698	22451	11166

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-389	22368	72306
rnd-1_family-388	0	444
rnd-5_family-1850	22589	34709
rnd-1_family-573	0	0
rnd-1_family-381	0	0
rnd-1_family-380	334	1005
rnd-1_family-383	75285	732682
rnd-1_family-382	24033	76613
rnd-1_family-385	8484	22221
rnd-1_family-384	12269	36497
rnd-1_family-387	22214	108512
rnd-1_family-386	60515	64396
rnd-1_family-658	390	1168
rnd-1_family-598	126799	267649
rnd-1_family-599	3954	14570
rnd-6_family-2539	48272	100321
rnd-1_family-594	7501	4029
rnd-1_family-595	26	0
rnd-1_family-596	39701	65767
rnd-1_family-597	196	205
rnd-1_family-590	0	154
rnd-1_family-591	3165	23634
rnd-1_family-592	11897	143145
rnd-1_family-593	0	0
rnd-1_family-622	4164	7844
rnd-1_family-623	2064	3206
rnd-1_family-189	81	0
rnd-1_family-188	46511	226841
rnd-1_family-626	2405	283
rnd-1_family-627	2454	17446
rnd-1_family-624	3299	2905
rnd-1_family-625	1339	105578
rnd-1_family-183	3584	9412
rnd-1_family-182	6831	14088
rnd-1_family-181	16353	75258
rnd-1_family-180	29142	1203543
rnd-1_family-187	25141	55391
rnd-1_family-186	7000	463
rnd-1_family-185	43703	142820
rnd-1_family-184	71500	213574
rnd-6_family-5644	198	5292

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-548	0	114
rnd-1_family-661	0	0
rnd-6_family-5786	10945	19161
rnd-5_family-689	67146	76326
rnd-3_family-191	9190	6615
rnd-1_family-749	27632	25007
rnd-1_family-543	546	806
rnd-1_family-542	6871	7684
rnd-1_family-741	0	111
rnd-1_family-740	133242	25594
rnd-1_family-743	6750	6699
rnd-1_family-742	2077	24464
rnd-1_family-745	31759	105091
rnd-1_family-744	7013	7934
rnd-1_family-747	1950	1589
rnd-1_family-746	1380	4791
rnd-2_family-17	83339	171110
rnd-5_family-1748	7244	39011
rnd-6_family-6162	102765	11229
rnd-6_family-1921	139105	834441
rnd-6_family-1922	476025	222031
rnd-6_family-1923	85162	293158
rnd-5_family-612	141655	29935
rnd-6_family-4748	9693	29969
rnd-5_family-3570	2823	13780
rnd-5_family-2088	5983	7005
rnd-5_family-1267	37519	67808
rnd-5_family-1268	15349	49004
rnd-6_family-2840	2149	1517
rnd-6_family-3577	4293	32712
rnd-6_family-2845	31137	29988
rnd-1_family-78	216433	442496
rnd-1_family-79	11783	16163
rnd-1_family-134	205517	502991
rnd-1_family-135	64321	120927
rnd-1_family-132	27278	281585
rnd-1_family-133	212	68
rnd-1_family-130	61354	419683
rnd-1_family-131	433926	1860819
rnd-1_family-70	235918	236716
rnd-1_family-71	331090	572052

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-72	369733	860480
rnd-1_family-73	492825	1784417
rnd-1_family-74	163934	247590
rnd-1_family-75	85693	37553
rnd-1_family-138	426766	767568
rnd-1_family-139	90256	445683
rnd-1_family-358	6522	217180
rnd-1_family-359	2850	126760
rnd-1_family-629	584	524
rnd-1_family-352	13854	107978
rnd-1_family-353	45597	17660
rnd-1_family-350	413	4008
rnd-1_family-351	81789	141240
rnd-1_family-356	3855	20538
rnd-1_family-303	73303	313426
rnd-1_family-354	51618	49633
rnd-1_family-355	13951	4530
rnd-5_family-38	29507	169262
rnd-5_family-39	11926	50814
rnd-1_family-269	57492	84427
rnd-1_family-529	8981	27741
rnd-1_family-528	16257	38506
rnd-5_family-530	8058	34094
rnd-1_family-525	24221	36289
rnd-1_family-524	21804	40263
rnd-1_family-527	29322	25624
rnd-1_family-526	10205	267298
rnd-1_family-521	21186	30027
rnd-1_family-520	5701	478121
rnd-1_family-523	130040	426092
rnd-1_family-522	0	23883
rnd-1_family-136	35942	356421
rnd-5_family-1195	9441	261140
rnd-1_family-137	11701	20525
rnd-1_family-781	1690	645
rnd-1_family-640	211301	153751
rnd-5_family-585	13369	14206
rnd-1_family-780	5830	30376
rnd-1_family-262	7420	12429
rnd-1_family-603	328	4958
rnd-1_family-263	389	3363

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-5207	25693	17333
rnd-1_family-739	7075	3797
rnd-1_family-558	1075	33127
rnd-1_family-559	10183	11011
rnd-1_family-550	29817	74602
rnd-1_family-551	5197	4826
rnd-1_family-552	26284	311204
rnd-1_family-553	14010	52326
rnd-1_family-554	14043	34763
rnd-1_family-555	11107	48225
rnd-1_family-556	415	1635
rnd-1_family-557	1124	2035
rnd-1_family-411	72209	368277
rnd-1_family-410	118	171
rnd-1_family-413	148083	106903
rnd-1_family-412	452	0
rnd-1_family-415	33018	147753
rnd-1_family-414	145727	280798
rnd-1_family-417	18409	26673
rnd-1_family-416	7560	37239
rnd-1_family-419	0	0
rnd-1_family-418	70412	11605
rnd-1_family-790	3145	7770
rnd-1_family-791	11405	28209
rnd-1_family-796	1300	1715
rnd-1_family-797	816	2248
rnd-1_family-794	10743	10424
rnd-6_family-832	7837	18160
rnd-6_family-428	13774	24439
rnd-5_family-1672	105908	150920
rnd-5_family-84	23211	26978
rnd-4_family-141	9788	12310
rnd-1_family-76	186235	292040
rnd-1_family-228	20316	67546
rnd-1_family-229	95477	65245
rnd-1_family-226	174723	57332
rnd-1_family-227	108832	296813
rnd-1_family-224	21581	21488
rnd-1_family-225	32935	59489
rnd-1_family-222	105366	93586
rnd-1_family-223	3071	7937

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-220	88	911
rnd-1_family-221	71398	661110
rnd-1_family-705	2659	111341
rnd-1_family-704	356	1027
rnd-1_family-707	810	3686
rnd-1_family-706	1294	628
rnd-1_family-701	294	980
rnd-1_family-643	791	892
rnd-1_family-703	0	44
rnd-1_family-702	5744	9077
rnd-1_family-549	4648	77836
rnd-1_family-709	0	102
rnd-1_family-708	8661	24526
rnd-5_family-2851	7111	131882
rnd-1_family-511	733	1344
rnd-6_family-1042	9424	15284
rnd-1_family-644	220	72822
rnd-1_family-518	1338	393
rnd-1_family-519	1583	166003
rnd-5_family-27	4938	20630
rnd-5_family-6280	46765	11788
rnd-5_family-6123	35398	127271
rnd-6_family-3094	3674	45732
rnd-5_family-120	0	0
rnd-1_family-172	405	16893
rnd-1_family-173	33601	45344
rnd-1_family-170	807	12742
rnd-1_family-171	16722	26609
rnd-1_family-176	449	2912
rnd-1_family-177	36952	405057
rnd-1_family-174	1099	23934
rnd-1_family-175	0	6746
rnd-5_family-1516	1282	130950
rnd-1_family-178	469854	457875
rnd-1_family-179	314	9653
rnd-1_family-357	1160	1278
rnd-5_family-1166	13735	153234
rnd-6_family-4626	2229	18447
rnd-5_family-496	15997	22877
rnd-5_family-491	9494	11786
rnd-4_family-655	21249	39669

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-1515	13947	20123
rnd-4_family-368	26184	50202
rnd-4_family-656	135948	1182543
rnd-1_family-646	956	2067
rnd-6_family-4933	38574	46803
rnd-5_family-151	11292	24623
rnd-1_family-679	4420	5394
rnd-1_family-678	32479	21705
rnd-1_family-772	6322	10509
rnd-1_family-773	2353	6088
rnd-1_family-774	4523	18021
rnd-1_family-668	30308	22021
rnd-1_family-776	1332	13346
rnd-1_family-777	5360	8371
rnd-1_family-671	0	71
rnd-1_family-670	0	0
rnd-1_family-673	57	334
rnd-1_family-669	46708	36994
rnd-1_family-675	1161	1849
rnd-1_family-674	12506	1158
rnd-1_family-677	1321	571
rnd-1_family-676	2554	9025
rnd-5_family-1796	795	29366
rnd-5_family-1797	752	12061
rnd-6_family-5533	7364	204105
rnd-6_family-6779	1629	82783
rnd-5_family-4177	11640	27789
rnd-1_family-96	57703	95702
rnd-1_family-97	199545	271319
rnd-1_family-94	166639	434874
rnd-1_family-95	156168	1470558
rnd-1_family-92	8095	10069
rnd-1_family-93	212265	1441314
rnd-1_family-90	83972	128207
rnd-1_family-91	33928	55438
rnd-5_family-32	35268	57606
rnd-1_family-98	7229	14721
rnd-1_family-99	56414	381803
rnd-6_family-1860	18085	80601
rnd-5_family-4480	131298	83507
rnd-5_family-37	6291	112496

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-12152	505607	1881711
rnd-5_family-2837	52859	12119
rnd-6_family-8317	4917	22433
rnd-1_family-301	0	0
rnd-1_family-300	55149	52780
rnd-1_family-268	2313	2509
rnd-1_family-302	49056	54761
rnd-1_family-305	0	92
rnd-1_family-304	30460	47881
rnd-1_family-307	196	335
rnd-1_family-306	64782	174420
rnd-1_family-309	12746	4792
rnd-1_family-308	16406	46725
rnd-1_family-260	277	67621
rnd-1_family-261	7911	107184
rnd-1_family-266	62643	106039
rnd-1_family-267	64313	301172
rnd-1_family-264	145531	317894
rnd-1_family-265	67136	245586
rnd-1_family-514	10799	16593
rnd-1_family-515	2028	990
rnd-1_family-516	34997	4686
rnd-1_family-517	72878	219621
rnd-1_family-459	681	22200
rnd-1_family-458	25371	50732
rnd-1_family-512	4523	7234
rnd-1_family-513	848	400056
rnd-1_family-455	5758	22095
rnd-1_family-454	61214	81417
rnd-1_family-457	1075	3266
rnd-1_family-456	63225	312284
rnd-1_family-451	31704	144332
rnd-1_family-450	53367	215555
rnd-1_family-453	2851	3098
rnd-1_family-452	33384	26859
rnd-5_family-43	65979	34140
rnd-5_family-2316	6067	11447
rnd-5_family-45	3935	4425
rnd-5_family-575	2883	12684
rnd-5_family-48	51404	76679
rnd-2_family-2	14528	84239

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-1785	14618	14035
rnd-1_family-758	632	185
rnd-5_family-2773	20024	8381
rnd-6_family-4154	28844	19081
rnd-1_family-759	10443	16830

Table S2. Table showing number of times each specific repeat from the RepeatModeler annotation is found in the

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-428	0	1
rnd-1_family-429	2	0
rnd-1_family-420	0	0
rnd-1_family-421	24	51
rnd-1_family-422	456	586
rnd-1_family-423	53	211
rnd-1_family-424	34	72
rnd-1_family-425	10	187
rnd-1_family-426	16	17
rnd-1_family-427	1	0
rnd-5_family-1887	81	108
rnd-6_family-558	19	127
rnd-5_family-4385	68	53
rnd-5_family-545	341	333
rnd-5_family-3926	85	7
rnd-2_family-195	2	6
rnd-6_family-1973	114	128
rnd-1_family-789	26	13
rnd-3_family-102	88	102
rnd-3_family-40	262	564
rnd-5_family-104	115	119
rnd-5_family-2225	0	119
rnd-1_family-589	55	110
rnd-1_family-588	8	0
rnd-1_family-587	27	39
rnd-1_family-586	145	317
rnd-1_family-585	72	61
rnd-1_family-584	16	9
rnd-1_family-583	54	14
rnd-1_family-582	0	2
rnd-1_family-581	12	13
rnd-1_family-580	105	50
rnd-1_family-635	28	5
rnd-1_family-634	10	2
rnd-1_family-637	1	1
rnd-1_family-636	8	15
rnd-1_family-631	3	0
rnd-1_family-630	25	33
rnd-1_family-633	4	6

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-632	22	23
rnd-1_family-728	97	6
rnd-1_family-639	164	142
rnd-1_family-638	2	1
rnd-5_family-2157	25	82
rnd-6_family-5307	176	178
rnd-6_family-4575	77	103
rnd-5_family-2156	77	163
rnd-5_family-187	155	168
rnd-1_family-729	62	55
rnd-6_family-1786	75	24
rnd-5_family-181	158	177
rnd-1_family-280	143	165
rnd-1_family-281	1	1
rnd-1_family-282	4	4
rnd-1_family-283	196	173
rnd-1_family-284	10	64
rnd-1_family-285	21	115
rnd-1_family-286	114	579
rnd-1_family-287	5	27
rnd-1_family-288	38	2
rnd-1_family-289	25	22
rnd-6_family-5391	194	199
rnd-4_family-194	187	257
rnd-1_family-738	112	189
rnd-1_family-721	6	16
rnd-1_family-775	1	21
rnd-1_family-734	80	81
rnd-1_family-735	33	55
rnd-1_family-736	87	80
rnd-1_family-720	6	8
rnd-1_family-730	3	1
rnd-1_family-731	36	4
rnd-1_family-732	47	72
rnd-1_family-733	29	106
rnd-3_family-409	470	863
rnd-5_family-1285	321	377
rnd-6_family-4611	30	97
rnd-6_family-1950	150	223
rnd-5_family-1358	31	83
rnd-5_family-1210	204	172

repeat	*D. pseudoobscura*	*D. miranda*
rnd-3_family-151	331	223
rnd-1_family-129	4089	759
rnd-1_family-44	345	952
rnd-1_family-47	30	63
rnd-1_family-46	39	72
rnd-1_family-41	269	662
rnd-1_family-40	199	554
rnd-1_family-43	43	122
rnd-1_family-42	62	176
rnd-1_family-121	274	329
rnd-1_family-120	4	4
rnd-1_family-123	39	337
rnd-1_family-122	96	128
rnd-1_family-49	260	951
rnd-1_family-48	327	1119
rnd-1_family-127	37	283
rnd-1_family-126	78	78
rnd-4_family-1419	30	40
rnd-1_family-349	368	298
rnd-1_family-348	12	9
rnd-5_family-4349	188	257
rnd-1_family-345	2	26
rnd-1_family-344	249	360
rnd-1_family-347	117	167
rnd-1_family-346	92	282
rnd-1_family-341	48	162
rnd-1_family-340	47	136
rnd-1_family-343	11	6
rnd-1_family-342	28	23
rnd-5_family-7432	67	126
rnd-5_family-248	116	86
rnd-5_family-679	25	214
rnd-1_family-751	24	15
rnd-5_family-522	1	8
rnd-5_family-525	72	81
rnd-5_family-524	180	251
rnd-5_family-526	115	136
rnd-5_family-1308	73	40
rnd-6_family-906	13	39
rnd-6_family-3235	2	79
rnd-6_family-9368	18	19

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-773	19	109
rnd-1_family-798	9	4
rnd-5_family-2623	119	107
rnd-1_family-799	7	5
rnd-6_family-734	22	72
rnd-5_family-6343	95	63
rnd-1_family-231	0	0
rnd-1_family-230	1	1
rnd-1_family-233	0	4
rnd-1_family-232	264	333
rnd-1_family-235	24	30
rnd-1_family-234	36	126
rnd-1_family-237	444	686
rnd-1_family-236	122	172
rnd-1_family-239	0	0
rnd-1_family-238	24	137
rnd-1_family-541	38	66
rnd-1_family-540	25	58
rnd-1_family-547	81	93
rnd-1_family-546	58	44
rnd-1_family-545	42	123
rnd-1_family-544	25	107
rnd-1_family-464	90	107
rnd-1_family-465	273	238
rnd-1_family-466	72	188
rnd-1_family-467	67	101
rnd-1_family-460	2	1
rnd-1_family-461	180	207
rnd-1_family-462	24	131
rnd-1_family-463	8	25
rnd-1_family-785	18	58
rnd-1_family-784	1	68
rnd-1_family-787	43	60
rnd-1_family-786	0	1
rnd-1_family-468	38	43
rnd-1_family-469	14	243
rnd-1_family-783	144	205
rnd-1_family-782	24	45
rnd-5_family-2530	47	140
rnd-5_family-3472	220	219
rnd-5_family-1948	60	214

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-986	15	72
rnd-5_family-358	27	72
rnd-6_family-6	251	303
rnd-5_family-4902	61	176
rnd-1_family-101	0	4
rnd-1_family-338	0	0
rnd-1_family-339	93	49
rnd-1_family-100	39	55
rnd-6_family-8049	398	443
rnd-1_family-330	6	57
rnd-1_family-331	116	181
rnd-1_family-332	0	0
rnd-1_family-333	9	14
rnd-1_family-334	0	1
rnd-1_family-335	116	77
rnd-1_family-336	12	40
rnd-1_family-337	1	3
rnd-5_family-7004	41	38
rnd-6_family-2793	23	87
rnd-5_family-3283	62	100
rnd-6_family-4544	32	37
rnd-5_family-256	28	79
rnd-5_family-11416	106	81
rnd-6_family-2941	264	326
rnd-1_family-109	258	449
rnd-6_family-1305	81	145
rnd-1_family-108	62	185
rnd-6_family-2856	41	81
rnd-1_family-563	8	24
rnd-1_family-562	0	2
rnd-5_family-5662	168	233
rnd-1_family-165	112	436
rnd-1_family-164	127	281
rnd-1_family-167	33	32
rnd-1_family-166	10	40
rnd-1_family-161	471	457
rnd-1_family-160	1	8
rnd-1_family-163	6	16
rnd-1_family-162	111	122
rnd-5_family-557	93	198
rnd-5_family-4402	32	74

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-4543	31	103
rnd-6_family-5135	45	198
rnd-1_family-169	3	0
rnd-1_family-168	39	69
rnd-1_family-569	45	92
rnd-4_family-121	128	776
rnd-1_family-568	12	16
rnd-6_family-4578	14	30
rnd-3_family-8	177	382
rnd-1_family-653	39	115
rnd-4_family-124	386	965
rnd-1_family-652	99	191
rnd-6_family-513	156	177
rnd-1_family-651	0	0
rnd-5_family-144	121	186
rnd-1_family-763	41	45
rnd-1_family-762	4	159
rnd-1_family-761	98	101
rnd-1_family-650	4	23
rnd-1_family-608	20	69
rnd-1_family-609	0	0
rnd-1_family-765	62	125
rnd-1_family-764	2	1
rnd-1_family-604	0	0
rnd-1_family-56	8	7
rnd-1_family-606	7	3
rnd-1_family-607	0	0
rnd-1_family-600	64	67
rnd-1_family-601	0	0
rnd-1_family-602	138	224
rnd-1_family-57	85	243
rnd-5_family-891	40	87
rnd-6_family-386	144	189
rnd-5_family-2338	237	203
rnd-1_family-655	71	29
rnd-5_family-2588	21	38
rnd-5_family-2589	61	83
rnd-1_family-654	67	49
rnd-5_family-76	73	95
rnd-1_family-700	3	23
rnd-6_family-6776	56	59

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-2824	203	122
rnd-5_family-50	84	141
rnd-1_family-77	145	272
rnd-1_family-510	6	20
rnd-1_family-58	166	183
rnd-5_family-71	17	50
rnd-5_family-3617	1	140
rnd-1_family-157	29	219
rnd-5_family-1088	223	342
rnd-4_family-490	331	426
rnd-5_family-660	106	366
rnd-6_family-6071	7	23
rnd-4_family-498	117	176
rnd-1_family-110	74	127
rnd-1_family-111	13	19
rnd-1_family-112	74	222
rnd-1_family-113	74	15
rnd-1_family-114	127	336
rnd-1_family-115	26	103
rnd-1_family-18	136	414
rnd-1_family-117	0	0
rnd-1_family-118	48	94
rnd-1_family-119	5	32
rnd-1_family-14	20	38
rnd-1_family-15	2316	5639
rnd-1_family-12	28	73
rnd-1_family-13	359	1170
rnd-1_family-10	16	31
rnd-1_family-11	242	502
rnd-1_family-374	40	56
rnd-1_family-375	120	464
rnd-1_family-376	7	71
rnd-1_family-377	18	106
rnd-1_family-279	76	310
rnd-1_family-278	0	0
rnd-1_family-372	7	15
rnd-1_family-373	1330	1627
rnd-1_family-275	0	0
rnd-1_family-274	243	281
rnd-1_family-277	47	266
rnd-1_family-276	0	4

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-271	48	421
rnd-1_family-270	80	56
rnd-1_family-273	76	199
rnd-1_family-272	38	118
rnd-1_family-507	71	36
rnd-1_family-506	60	112
rnd-1_family-505	3	79
rnd-1_family-504	45	152
rnd-1_family-503	0	128
rnd-1_family-502	0	0
rnd-1_family-501	82	122
rnd-1_family-500	51	126
rnd-1_family-509	6	60
rnd-1_family-508	16	35
rnd-5_family-564	110	126
rnd-5_family-508	82	93
rnd-6_family-4734	68	19
rnd-6_family-250	436	448
rnd-3_family-10	305	594
rnd-1_family-536	15	58
rnd-5_family-3784	807	1437
rnd-5_family-1935	127	286
rnd-1_family-760	10	101
rnd-5_family-6330	83	175
rnd-1_family-767	2	9
rnd-4_family-111	495	521
rnd-1_family-766	35	25
rnd-6_family-1689	33	68
rnd-1_family-439	23	43
rnd-1_family-438	0	2
rnd-5_family-1614	88	105
rnd-1_family-433	0	0
rnd-1_family-432	17	20
rnd-1_family-431	0	1
rnd-1_family-430	0	0
rnd-1_family-437	7	44
rnd-1_family-436	0	0
rnd-1_family-435	53	26
rnd-1_family-434	75	137
rnd-5_family-4887	46	100
rnd-6_family-4876	6	92

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-605	18	21
rnd-4_family-1031	22	38
rnd-3_family-374	3748	715
rnd-5_family-6727	30	229
rnd-1_family-769	5	8
rnd-1_family-628	71	80
rnd-1_family-768	23	113
rnd-6_family-19	208	49
rnd-6_family-2196	143	175
rnd-1_family-208	1	3
rnd-1_family-209	17	61
rnd-6_family-8134	317	529
rnd-5_family-10308	379	497
rnd-1_family-200	61	224
rnd-1_family-201	2	2
rnd-1_family-202	0	0
rnd-1_family-203	1	2
rnd-1_family-204	2	23
rnd-1_family-205	0	1
rnd-1_family-206	25	42
rnd-1_family-207	70	113
rnd-1_family-660	0	5
rnd-5_family-1930	241	218
rnd-6_family-407	59	117
rnd-3_family-99	47	69
rnd-4_family-232	36	146
rnd-4_family-236	496	22
rnd-5_family-556	79	141
rnd-1_family-293	249	273
rnd-1_family-292	248	47
rnd-1_family-291	68	149
rnd-1_family-290	0	4
rnd-1_family-297	86	192
rnd-1_family-296	109	87
rnd-1_family-295	15	51
rnd-1_family-294	113	157
rnd-1_family-299	95	83
rnd-1_family-298	7	4
rnd-6_family-5068	103	143
rnd-1_family-486	53	115
rnd-1_family-487	2	180

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-484	5	180
rnd-1_family-485	3	6
rnd-1_family-482	597	663
rnd-1_family-483	19	159
rnd-1_family-480	93	221
rnd-1_family-481	130	116
rnd-1_family-727	83	78
rnd-1_family-726	71	140
rnd-1_family-725	6	11
rnd-1_family-724	26	106
rnd-1_family-723	8	9
rnd-1_family-722	26	54
rnd-1_family-488	25	18
rnd-1_family-489	0	2
rnd-1_family-27	97	256
rnd-1_family-26	16	30
rnd-1_family-25	0	0
rnd-1_family-24	7	9
rnd-1_family-23	71	250
rnd-1_family-22	92	184
rnd-1_family-21	27	49
rnd-1_family-20	504	6697
rnd-6_family-2942	1045	862
rnd-1_family-29	200	382
rnd-1_family-28	112	68
rnd-5_family-4822	82	76
rnd-5_family-1330	218	42
rnd-5_family-1207	25	64
rnd-1_family-737	2	8
rnd-5_family-4888	277	337
rnd-4_family-862	225	250
rnd-5_family-119	0	1
rnd-1_family-52	116	168
rnd-1_family-53	58	68
rnd-1_family-50	179	722
rnd-1_family-51	2	5
rnd-1_family-158	33	106
rnd-1_family-159	60	73
rnd-1_family-54	25	36
rnd-1_family-55	114	155
rnd-1_family-154	101	128

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-155	12	151
rnd-1_family-156	2	3
rnd-1_family-59	270	972
rnd-1_family-150	141	101
rnd-1_family-151	34	83
rnd-1_family-152	56	64
rnd-1_family-153	70	102
rnd-5_family-1387	177	184
rnd-5_family-2243	75	65
rnd-1_family-691	16	20
rnd-5_family-1381	321	362
rnd-6_family-6311	22	41
rnd-6_family-6088	15	61
rnd-5_family-250	174	274
rnd-5_family-17	317	573
rnd-6_family-2438	69	82
rnd-5_family-18	728	890
rnd-5_family-19	18	167
rnd-6_family-25	19	64
rnd-5_family-6789	63	61
rnd-1_family-793	0	0
rnd-3_family-164	2	7
rnd-5_family-1981	126	139
rnd-5_family-828	93	66
rnd-5_family-829	136	86
rnd-6_family-803	38	64
rnd-6_family-22	25	86
rnd-1_family-641	154	242
rnd-1_family-244	0	4
rnd-1_family-245	350	630
rnd-1_family-246	11	13
rnd-1_family-247	43	51
rnd-1_family-240	176	305
rnd-1_family-241	164	331
rnd-1_family-242	0	0
rnd-1_family-243	0	1
rnd-1_family-576	2	42
rnd-1_family-577	0	0
rnd-1_family-574	18	71
rnd-1_family-575	116	232
rnd-1_family-248	132	279

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-249	16	113
rnd-1_family-570	12	1
rnd-1_family-571	54	64
rnd-1_family-477	12	7
rnd-1_family-476	176	227
rnd-1_family-475	12	44
rnd-1_family-474	13	59
rnd-1_family-473	19	24
rnd-1_family-472	108	117
rnd-1_family-471	9	11
rnd-1_family-470	204	345
rnd-1_family-479	188	339
rnd-1_family-478	0	0
rnd-5_family-2523	50	35
rnd-6_family-645	121	143
rnd-5_family-3973	26	82
rnd-5_family-69	74	188
rnd-1_family-329	29	33
rnd-1_family-328	163	348
rnd-1_family-323	429	238
rnd-1_family-322	202	151
rnd-1_family-321	131	120
rnd-1_family-320	139	214
rnd-1_family-327	374	375
rnd-1_family-326	95	127
rnd-1_family-325	172	275
rnd-1_family-324	64	173
rnd-6_family-1826	85	31
rnd-1_family-684	31	248
rnd-1_family-685	22	29
rnd-1_family-686	0	0
rnd-1_family-687	1	1
rnd-1_family-680	0	0
rnd-1_family-681	119	120
rnd-1_family-682	125	157
rnd-1_family-683	83	151
rnd-1_family-688	0	0
rnd-1_family-689	67	282
rnd-1_family-396	68	171
rnd-1_family-397	0	0
rnd-1_family-394	77	163

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-395	249	185
rnd-1_family-392	0	0
rnd-1_family-393	210	56
rnd-1_family-390	102	152
rnd-1_family-391	44	49
rnd-1_family-579	244	108
rnd-5_family-4496	50	88
rnd-1_family-398	83	74
rnd-1_family-399	0	5
rnd-1_family-750	9	103
rnd-1_family-198	102	128
rnd-1_family-199	88	220
rnd-4_family-555	82	179
rnd-6_family-5829	33	58
rnd-1_family-116	35	150
rnd-5_family-1832	60	127
rnd-1_family-190	29	105
rnd-1_family-191	4	42
rnd-1_family-192	0	1
rnd-1_family-193	0	1
rnd-1_family-194	0	0
rnd-1_family-195	1	0
rnd-1_family-196	61	38
rnd-1_family-197	12	20
rnd-1_family-16	25	22
rnd-4_family-349	401	1071
rnd-1_family-17	8	222
rnd-6_family-1700	46	83
rnd-3_family-119	41	79
rnd-6_family-6455	6	85
rnd-6_family-4577	38	49
rnd-6_family-4558	198	192
rnd-5_family-1618	15	120
rnd-1_family-756	131	50
rnd-1_family-757	6	14
rnd-1_family-754	0	5
rnd-1_family-755	2	12
rnd-1_family-752	5	4
rnd-1_family-753	28	35
rnd-1_family-619	6	200
rnd-1_family-618	137	233

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-617	2	13
rnd-1_family-616	187	181
rnd-1_family-615	26	36
rnd-1_family-614	2	28
rnd-1_family-613	89	124
rnd-1_family-612	1	9
rnd-1_family-611	0	0
rnd-1_family-610	0	0
rnd-6_family-6194	11	115
rnd-5_family-969	24	98
rnd-6_family-144	147	140
rnd-5_family-492	44	39
rnd-1_family-620	20	71
rnd-6_family-398	33	75
rnd-5_family-183	139	158
rnd-1_family-621	1	1
rnd-5_family-4059	60	30
rnd-5_family-219	184	134
rnd-1_family-370	80	267
rnd-5_family-603	50	41
rnd-4_family-137	193	284
rnd-1_family-371	0	1
rnd-1_family-45	387	851
rnd-5_family-777	136	140
rnd-5_family-8	99	151
rnd-1_family-128	17	37
rnd-5_family-675	87	111
rnd-6_family-4568	27	84
rnd-5_family-778	103	93
rnd-1_family-642	4	5
rnd-5_family-498	108	64
rnd-1_family-103	254	227
rnd-1_family-102	41	105
rnd-1_family-69	126	281
rnd-1_family-68	150	239
rnd-1_family-107	27	91
rnd-1_family-106	74	146
rnd-1_family-105	20	35
rnd-1_family-104	90	249
rnd-1_family-63	344	648
rnd-1_family-62	302	1025

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-61	1	3
rnd-1_family-60	87	563
rnd-1_family-67	444	765
rnd-1_family-66	18	32
rnd-1_family-65	371	1044
rnd-1_family-64	270	1181
rnd-1_family-367	227	399
rnd-1_family-366	0	0
rnd-1_family-365	331	406
rnd-1_family-364	28	38
rnd-1_family-363	0	3
rnd-1_family-362	2	4
rnd-1_family-361	81	83
rnd-1_family-360	39	35
rnd-1_family-378	4	4
rnd-5_family-30	142	191
rnd-1_family-647	754	900
rnd-5_family-911	857	929
rnd-1_family-369	1073	1279
rnd-1_family-368	8	7
rnd-1_family-648	0	0
rnd-1_family-538	20	1
rnd-1_family-539	2	55
rnd-1_family-649	0	0
rnd-1_family-532	50	81
rnd-1_family-533	6	5
rnd-1_family-530	174	247
rnd-1_family-531	2	5
rnd-1_family-379	23	56
rnd-1_family-537	225	369
rnd-1_family-534	23	216
rnd-1_family-535	29	9
rnd-5_family-599	74	85
rnd-5_family-1181	6	156
rnd-5_family-2078	146	142
rnd-1_family-125	7	5
rnd-4_family-1526	88	236
rnd-1_family-124	0	0
rnd-5_family-11386	368	80
rnd-6_family-221	15	56
rnd-5_family-3787	62	130

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-138	197	186
rnd-1_family-408	2	40
rnd-1_family-409	177	189
rnd-1_family-406	1	10
rnd-1_family-407	38	34
rnd-1_family-404	0	3
rnd-1_family-405	286	380
rnd-1_family-402	45	45
rnd-1_family-403	0	0
rnd-1_family-400	391	404
rnd-1_family-401	30	59
rnd-6_family-6947	38	37
rnd-5_family-1082	52	120
rnd-1_family-770	0	0
rnd-1_family-771	83	221
rnd-5_family-1528	97	112
rnd-6_family-1694	41	106
rnd-1_family-219	94	190
rnd-1_family-218	6	31
rnd-1_family-657	86	156
rnd-6_family-2270	236	3
rnd-1_family-213	708	708
rnd-1_family-212	76	252
rnd-1_family-211	33	46
rnd-1_family-210	1	0
rnd-1_family-217	75	314
rnd-1_family-216	104	447
rnd-1_family-215	0	0
rnd-1_family-214	0	0
rnd-1_family-712	0	2
rnd-1_family-713	8	55
rnd-1_family-710	38	8
rnd-1_family-711	53	67
rnd-1_family-716	9	12
rnd-1_family-717	0	0
rnd-1_family-714	114	151
rnd-1_family-715	26	34
rnd-1_family-718	16	123
rnd-1_family-719	109	136
rnd-1_family-778	21	6
rnd-6_family-9834	88	86

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-656	22	139
rnd-4_family-229	138	225
rnd-1_family-779	88	25
rnd-6_family-4094	89	35
rnd-1_family-672	5	52
rnd-3_family-364	0	246
rnd-5_family-1853	0	0
rnd-6_family-1574	169	176
rnd-6_family-4857	33	92
rnd-1_family-499	134	226
rnd-1_family-498	171	194
rnd-6_family-2576	30	50
rnd-5_family-382	303	335
rnd-6_family-2575	24	60
rnd-1_family-491	244	367
rnd-1_family-490	4	4
rnd-1_family-493	16	21
rnd-1_family-492	5	89
rnd-1_family-495	24	64
rnd-1_family-494	25	31
rnd-1_family-497	60	164
rnd-1_family-496	3	22
rnd-1_family-34	23	67
rnd-1_family-35	133	245
rnd-1_family-36	40	65
rnd-1_family-37	19	32
rnd-1_family-30	2	3
rnd-1_family-31	303	569
rnd-1_family-32	26	27
rnd-1_family-33	149	263
rnd-5_family-2778	78	69
rnd-1_family-38	8	12
rnd-1_family-39	179	546
rnd-5_family-2954	87	118
rnd-1_family-578	0	0
rnd-5_family-1309	69	161
rnd-5_family-1663	124	204
rnd-5_family-1307	90	74
rnd-5_family-488	121	111
rnd-5_family-720	129	159
rnd-5_family-523	24	94

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-437	31	67
rnd-1_family-748	26	31
rnd-5_family-3526	47	60
rnd-6_family-1379	57	76
rnd-6_family-6418	128	183
rnd-5_family-439	24	55
rnd-5_family-438	255	211
rnd-1_family-666	26	72
rnd-1_family-667	32	57
rnd-5_family-1868	174	223
rnd-1_family-665	57	87
rnd-1_family-662	0	3
rnd-1_family-663	2	5
rnd-1_family-149	74	121
rnd-1_family-148	40	51
rnd-1_family-147	104	129
rnd-1_family-146	61	153
rnd-1_family-145	14	187
rnd-1_family-144	53	55
rnd-1_family-143	197	201
rnd-1_family-142	67	329
rnd-1_family-141	0	0
rnd-1_family-140	0	1
rnd-5_family-794	29	87
rnd-5_family-117	110	49
rnd-5_family-791	8	117
rnd-5_family-4142	39	51
rnd-4_family-903	181	279
rnd-4_family-1635	52	67
rnd-1_family-89	251	441
rnd-1_family-88	154	286
rnd-6_family-7085	226	256
rnd-5_family-1543	77	143
rnd-1_family-81	672	878
rnd-1_family-80	99	295
rnd-1_family-83	137	456
rnd-1_family-82	122	82
rnd-1_family-85	167	220
rnd-1_family-84	600	908
rnd-1_family-87	239	334
rnd-1_family-86	151	532

repeat	*D. pseudoobscura*	*D. miranda*
rnd-4_family-2304	11	38
rnd-5_family-371	106	177
rnd-5_family-1508	36	59
rnd-1_family-659	0	0
rnd-5_family-372	86	178
rnd-6_family-1197	422	421
rnd-6_family-756	53	135
rnd-5_family-373	194	60
rnd-5_family-535	31	74
rnd-6_family-4338	7	77
rnd-6_family-2695	231	306
rnd-1_family-257	6	18
rnd-1_family-256	232	309
rnd-1_family-255	105	422
rnd-1_family-254	179	195
rnd-1_family-253	109	303
rnd-1_family-252	8	14
rnd-1_family-251	118	267
rnd-1_family-250	9	183
rnd-5_family-1647	93	121
rnd-5_family-1003	9	70
rnd-6_family-3186	266	350
rnd-1_family-259	28	185
rnd-1_family-258	96	215
rnd-1_family-561	16	17
rnd-1_family-560	174	280
rnd-1_family-448	35	37
rnd-1_family-449	126	100
rnd-1_family-565	0	0
rnd-1_family-564	5	16
rnd-1_family-567	37	112
rnd-1_family-566	39	76
rnd-1_family-442	2	6
rnd-1_family-443	2	0
rnd-1_family-440	58	91
rnd-1_family-441	2	170
rnd-1_family-446	9	13
rnd-1_family-447	0	1
rnd-1_family-444	0	0
rnd-1_family-445	74	63
rnd-1_family-4	205	375

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-5	711	1695
rnd-1_family-6	197	242
rnd-1_family-7	630	1584
rnd-1_family-0	3213	5938
rnd-1_family-1	83	148
rnd-1_family-2	697	844
rnd-1_family-3	2517	11148
rnd-5_family-6719	572	746
rnd-3_family-74	326	374
rnd-3_family-75	255	321
rnd-1_family-8	181	244
rnd-1_family-9	67	99
rnd-6_family-1994	70	105
rnd-1_family-318	0	9
rnd-1_family-319	0	0
rnd-1_family-316	93	13
rnd-1_family-317	4	17
rnd-1_family-314	1	27
rnd-1_family-315	0	0
rnd-1_family-312	119	331
rnd-1_family-313	86	219
rnd-1_family-310	30	172
rnd-1_family-311	2	5
rnd-5_family-1	15	150
rnd-5_family-4	7	153
rnd-5_family-5	8	32
rnd-1_family-19	153	375
rnd-1_family-572	1	0
rnd-6_family-10458	4	104
rnd-6_family-2160	100	116
rnd-1_family-697	56	128
rnd-1_family-696	228	288
rnd-1_family-695	198	239
rnd-1_family-694	12	7
rnd-1_family-693	19	22
rnd-1_family-692	107	10
rnd-1_family-645	74	22
rnd-1_family-690	43	98
rnd-6_family-3081	306	311
rnd-1_family-699	0	0
rnd-1_family-698	30	18

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-389	18	63
rnd-1_family-388	0	4
rnd-5_family-1850	260	390
rnd-1_family-573	0	0
rnd-1_family-381	0	0
rnd-1_family-380	4	6
rnd-1_family-383	31	204
rnd-1_family-382	49	117
rnd-1_family-385	8	23
rnd-1_family-384	24	38
rnd-1_family-387	15	70
rnd-1_family-386	112	113
rnd-1_family-658	3	4
rnd-1_family-598	74	153
rnd-1_family-599	13	32
rnd-6_family-2539	105	225
rnd-1_family-594	11	13
rnd-1_family-595	2	0
rnd-1_family-596	72	120
rnd-1_family-597	3	4
rnd-1_family-590	0	2
rnd-1_family-591	22	154
rnd-1_family-592	30	256
rnd-1_family-593	0	0
rnd-1_family-622	67	132
rnd-1_family-623	13	22
rnd-1_family-189	1	0
rnd-1_family-188	53	247
rnd-1_family-626	6	3
rnd-1_family-627	19	68
rnd-1_family-624	7	19
rnd-1_family-625	3	82
rnd-1_family-183	13	34
rnd-1_family-182	71	135
rnd-1_family-181	170	319
rnd-1_family-180	37	350
rnd-1_family-187	149	237
rnd-1_family-186	10	1
rnd-1_family-185	134	269
rnd-1_family-184	85	183
rnd-6_family-5644	2	54

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-548	0	1
rnd-1_family-661	0	0
rnd-6_family-5786	41	38
rnd-5_family-689	343	399
rnd-3_family-191	82	90
rnd-1_family-749	27	38
rnd-1_family-543	3	5
rnd-1_family-542	73	63
rnd-1_family-741	0	1
rnd-1_family-740	118	51
rnd-1_family-743	3	3
rnd-1_family-742	8	29
rnd-1_family-745	43	116
rnd-1_family-744	84	88
rnd-1_family-747	24	19
rnd-1_family-746	12	31
rnd-2_family-17	309	467
rnd-5_family-1748	20	128
rnd-6_family-6162	157	52
rnd-6_family-1921	149	504
rnd-6_family-1922	426	362
rnd-6_family-1923	290	622
rnd-5_family-612	84	41
rnd-6_family-4748	21	58
rnd-5_family-3570	15	29
rnd-5_family-2088	72	81
rnd-5_family-1267	73	105
rnd-5_family-1268	19	66
rnd-6_family-2840	12	14
rnd-6_family-3577	30	139
rnd-6_family-2845	89	90
rnd-1_family-78	294	493
rnd-1_family-79	65	86
rnd-1_family-134	79	174
rnd-1_family-135	49	78
rnd-1_family-132	17	133
rnd-1_family-133	1	1
rnd-1_family-130	49	180
rnd-1_family-131	146	708
rnd-1_family-70	524	595
rnd-1_family-71	351	482

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-72	297	478
rnd-1_family-73	286	921
rnd-1_family-74	488	741
rnd-1_family-75	223	131
rnd-1_family-138	281	465
rnd-1_family-139	77	311
rnd-1_family-358	6	120
rnd-1_family-359	7	288
rnd-1_family-629	8	6
rnd-1_family-352	59	373
rnd-1_family-353	49	31
rnd-1_family-350	3	19
rnd-1_family-351	821	950
rnd-1_family-356	35	90
rnd-1_family-303	38	148
rnd-1_family-354	91	119
rnd-1_family-355	82	48
rnd-5_family-38	58	180
rnd-5_family-39	19	68
rnd-1_family-269	139	210
rnd-1_family-529	19	31
rnd-1_family-528	81	121
rnd-5_family-530	70	106
rnd-1_family-525	41	75
rnd-1_family-524	161	271
rnd-1_family-527	207	191
rnd-1_family-526	10	125
rnd-1_family-521	45	45
rnd-1_family-520	7	199
rnd-1_family-523	74	269
rnd-1_family-522	0	57
rnd-1_family-136	34	158
rnd-5_family-1195	11	146
rnd-1_family-137	78	133
rnd-1_family-781	6	2
rnd-1_family-640	173	132
rnd-5_family-585	214	221
rnd-1_family-780	12	17
rnd-1_family-262	83	114
rnd-1_family-603	2	15
rnd-1_family-263	1	7

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-5207	71	66
rnd-1_family-739	46	24
rnd-1_family-558	7	63
rnd-1_family-559	35	36
rnd-1_family-550	121	189
rnd-1_family-551	20	16
rnd-1_family-552	50	335
rnd-1_family-553	17	74
rnd-1_family-554	11	34
rnd-1_family-555	19	62
rnd-1_family-556	7	16
rnd-1_family-557	16	30
rnd-1_family-411	47	216
rnd-1_family-410	2	2
rnd-1_family-413	40	26
rnd-1_family-412	2	0
rnd-1_family-415	115	358
rnd-1_family-414	75	136
rnd-1_family-417	237	344
rnd-1_family-416	39	131
rnd-1_family-419	0	0
rnd-1_family-418	92	53
rnd-1_family-790	6	9
rnd-1_family-791	34	58
rnd-1_family-796	17	20
rnd-1_family-797	3	6
rnd-1_family-794	206	195
rnd-6_family-832	41	92
rnd-6_family-428	86	126
rnd-5_family-1672	97	177
rnd-5_family-84	230	254
rnd-4_family-141	14	14
rnd-1_family-76	823	1259
rnd-1_family-228	43	120
rnd-1_family-229	118	74
rnd-1_family-226	99	38
rnd-1_family-227	159	352
rnd-1_family-224	23	23
rnd-1_family-225	31	64
rnd-1_family-222	477	552
rnd-1_family-223	27	57

repeat	*D. pseudoobscura*	*D. miranda*
rnd-1_family-220	1	4
rnd-1_family-221	46	249
rnd-1_family-705	13	98
rnd-1_family-704	2	10
rnd-1_family-707	9	36
rnd-1_family-706	5	2
rnd-1_family-701	5	12
rnd-1_family-643	5	3
rnd-1_family-703	0	1
rnd-1_family-702	71	96
rnd-1_family-549	14	80
rnd-1_family-709	0	1
rnd-1_family-708	58	146
rnd-5_family-2851	17	98
rnd-1_family-511	10	12
rnd-6_family-1042	115	191
rnd-1_family-644	3	91
rnd-1_family-518	11	4
rnd-1_family-519	5	172
rnd-5_family-27	11	41
rnd-5_family-6280	60	28
rnd-5_family-6123	20	115
rnd-6_family-3094	3	28
rnd-5_family-120	0	0
rnd-1_family-172	1	16
rnd-1_family-173	344	449
rnd-1_family-170	3	14
rnd-1_family-171	12	18
rnd-1_family-176	2	8
rnd-1_family-177	28	133
rnd-1_family-174	7	87
rnd-1_family-175	0	10
rnd-5_family-1516	3	187
rnd-1_family-178	315	207
rnd-1_family-179	2	31
rnd-1_family-357	3	4
rnd-5_family-1166	69	208
rnd-6_family-4626	16	121
rnd-5_family-496	78	91
rnd-5_family-491	90	105
rnd-4_family-655	118	176

repeat	*D. pseudoobscura*	*D. miranda*
rnd-5_family-1515	169	205
rnd-4_family-368	145	267
rnd-4_family-656	96	240
rnd-1_family-646	8	14
rnd-6_family-4933	284	322
rnd-5_family-151	69	105
rnd-1_family-679	12	19
rnd-1_family-678	91	62
rnd-1_family-772	19	19
rnd-1_family-773	10	28
rnd-1_family-774	16	54
rnd-1_family-668	95	75
rnd-1_family-776	11	115
rnd-1_family-777	54	50
rnd-1_family-671	0	1
rnd-1_family-670	0	0
rnd-1_family-673	1	1
rnd-1_family-669	209	175
rnd-1_family-675	18	28
rnd-1_family-674	41	5
rnd-1_family-677	4	2
rnd-1_family-676	5	23
rnd-5_family-1796	4	111
rnd-5_family-1797	10	147
rnd-6_family-5533	19	170
rnd-6_family-6779	3	75
rnd-5_family-4177	69	173
rnd-1_family-96	389	559
rnd-1_family-97	89	124
rnd-1_family-94	43	91
rnd-1_family-95	49	377
rnd-1_family-92	50	54
rnd-1_family-93	93	539
rnd-1_family-90	444	597
rnd-1_family-91	236	397
rnd-5_family-32	87	115
rnd-1_family-98	30	52
rnd-1_family-99	48	200
rnd-6_family-1860	32	138
rnd-5_family-4480	86	61
rnd-5_family-37	10	137

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-12152	176	548
rnd-5_family-2837	209	61
rnd-6_family-8317	7	34
rnd-1_family-301	0	0
rnd-1_family-300	113	125
rnd-1_family-268	15	13
rnd-1_family-302	116	134
rnd-1_family-305	0	1
rnd-1_family-304	88	164
rnd-1_family-307	1	4
rnd-1_family-306	152	340
rnd-1_family-309	20	7
rnd-1_family-308	22	52
rnd-1_family-260	1	59
rnd-1_family-261	38	216
rnd-1_family-266	75	130
rnd-1_family-267	23	88
rnd-1_family-264	141	301
rnd-1_family-265	49	107
rnd-1_family-514	190	295
rnd-1_family-515	25	23
rnd-1_family-516	90	31
rnd-1_family-517	544	799
rnd-1_family-459	1	30
rnd-1_family-458	152	300
rnd-1_family-512	89	132
rnd-1_family-513	2	228
rnd-1_family-455	87	108
rnd-1_family-454	736	1057
rnd-1_family-457	6	22
rnd-1_family-456	87	223
rnd-1_family-451	167	403
rnd-1_family-450	43	122
rnd-1_family-453	23	27
rnd-1_family-452	294	231
rnd-5_family-43	134	77
rnd-5_family-2316	59	96
rnd-5_family-45	33	50
rnd-5_family-575	18	42
rnd-5_family-48	177	224
rnd-2_family-2	83	217

repeat	*D. pseudoobscura*	*D. miranda*
rnd-6_family-1785	56	57
rnd-1_family-758	4	2
rnd-5_family-2773	90	45
rnd-6_family-4154	76	61
rnd-1_family-759	32	41

Table S3. Table showing total base pairs mapped to each repeat class from the RepeatModeler annotation.

class	*D. pseudoobscura*	*D. miranda*
Unknown	15985714	26677602
Simple_repeat	576992	1327807
RC/Helitron	6681271	22864576
LTR/Gypsy	15996296	38128590
DNA/CMC-Transib	168934	440461
LINE/LOA	3191461	6151529
LTR/Copia	994362	2391223
DNA/hAT-hobo	419181	1293785
LTR/Pao	9064368	17149695
buffer	0	0
LINE/Jockey	853068	2306467
LINE/R1	1378907	7175748
DNA/P	197197	443810
LINE/Telomeric	381887	900028
LINE/CR1	2996857	6222923
DNA/hAT-Tip100	47482	300808
LINE	1080162	2178902
DNA/Harbinger	86443	236781
DNA/TcMar-Tc1	400945	610968
DNA/Maverick	509279	1152333
DNA/Transib	71293	127562
LINE/Penelope	84829	624299
DNA	96187	96730
DNA/CMC-EnSpm	27748	43944
DNA/hAT-Blackjack	31704	144332
DNA/hAT-Ac	126820	332427
LINE/I	196153	785630
DNA/PIF-Harbinger	24674	107471
DNA/Academ	10183	11011
SINE/tRNA-Glu	1749	2314
SINE?	10291	8795
LINE/R2	45189	30021
rRNA	620848	135508
LINE/L2	230114	153153
DNA/hAT-hAT5	422	786
SINE/5S	44191	67287
DNA/TcMar-Mariner	51404	76679
LINE/L1	13735	153234
DNA/TcMar-Tigger	26692	36101

class	*D. pseudoobscura*	*D. miranda*
DNA/hAT-hATm	41403	67059
DNA/MULE-MuDR	55133	46152
DNA/MuDR	31484	98491
DNA/hAT	111515	106652
DNA/Novosib	48733	51161
DNA/Sola	19594	27848

Table S4. Table showing number of times each repeat class from the RepeatModeler annotation is found in the genome.

class	*D. pseudoobscura*	*D. miranda*
Unknown	51024	68899
Simple_repeat	3251	5162
RC/Helitron	13165	36469
LTR/Gypsy	12876	26442
DNA/CMC-Transib	714	1408
LINE/LOA	2980	4391
LTR/Copia	701	1397
DNA/hAT-hobo	2467	5391
LTR/Pao	7755	13037
buffer	0	0
LINE/Jockey	2152	3422
LINE/R1	3358	7020
DNA/P	1569	2824
LINE/Telomeric	589	993
LINE/CR1	3477	5655
DNA/hAT-Tip100	452	1189
LINE	939	1877
DNA/Harbinger	470	694
DNA/TcMar-Tc1	1036	1451
DNA/Maverick	323	620
DNA/Transib	205	373
LINE/Penelope	306	717
DNA	330	348
DNA/CMC-EnSpm	156	180
DNA/hAT-Blackjack	167	403
DNA/hAT-Ac	709	1423
LINE/I	201	678
DNA/PIF-Harbinger	93	221
DNA/Academ	35	36
SINE/tRNA-Glu	25	33
SINE?	118	106
LINE/R2	43	49
rRNA	534	217
LINE/L2	355	267
DNA/hAT-hAT5	5	8
SINE/5S	21	6
DNA/TcMar-Mariner	177	224
LINE/L1	69	208
DNA/TcMar-Tigger	142	191

class	*D. pseudoobscura*	*D. miranda*
DNA/hAT-hATm	142	242
DNA/MULE-MuDR	237	203
DNA/MuDR	277	337
DNA/hAT	402	387
DNA/Novosib	436	448
DNA/Sola	128	183