

RESEARCH ARTICLE

# Alternative Splicing within and between *Drosophila* Species, Sexes, Tissues, and Developmental Stages

Lauren Gibilisco, Qi Zhou, Shivani Mahajan, Doris Bachtrog\*

Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, United States of America

\* [dbachtrog@berkeley.edu](mailto:dbachtrog@berkeley.edu)



 OPEN ACCESS

**Citation:** Gibilisco L, Zhou Q, Mahajan S, Bachtrog D (2016) Alternative Splicing within and between *Drosophila* Species, Sexes, Tissues, and Developmental Stages. *PLoS Genet* 12(12): e1006464. doi:10.1371/journal.pgen.1006464

**Editor:** Gregory S. Barsh, Stanford University School of Medicine, UNITED STATES

**Received:** June 17, 2016

**Accepted:** November 4, 2016

**Published:** December 9, 2016

**Copyright:** © 2016 Gibilisco et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** 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](http://www.ncbi.nlm.nih.gov/sra)) under BioProject ID PRJNA347569.

**Funding:** This work was funded by NIH grants (R01GM076007, R01GM093182 and R01GM101255) to DB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## 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 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 (“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.

## Author Summary

Alternative pre-mRNA splicing (“AS”) greatly expands the 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 function—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 the proteome diversity within species by creating different combinations of exons from the same genomic loci [1]. The resulting mRNA isoforms are usually expressed in a tissue or developmental-stage specific manner and underlie numerous essential biological processes like sex-determination [2, 3], tissue development [4], and stress response [5]. AS can also greatly increase the proteome diversity between species with similar repertoires of protein-coding genes [6, 7]. Given the correlation of alternative splicing with the evolution of organismal complexity [1], its dynamics across developmental stages, tissues, and species have attracted great attention [6, 7]. 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 alternative splicing events [6, 7]. 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 [6, 7].

AS is also prevalent in *Drosophila*, and has been most extensively studied in *D. melanogaster* [8–16]. 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 [17]. The complexity of AS also differs across developmental stages and tissues [18, 19] and different environmental perturbations [17]. 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 [20], and they diverged from *D. melanogaster* roughly 25 MY ago [21]; *D. nasuta* and *D. albomicans* split only about 0.1 MY ago [22], and diverged from *D. melanogaster* over 60 MY ago [23]. The presence of genomic resources for these species [24–27] 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; [S1 Table](#)). 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 [24]. 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 [Fig 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 alternative splicing events is outlined in [Fig 1](#). We considered skipped exons, alternative 5' and 3' spliced 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 ([Fig 1](#)). We find that between 20 and 37% of all multi-exon genes are alternatively spliced in at least one tissue or stage and

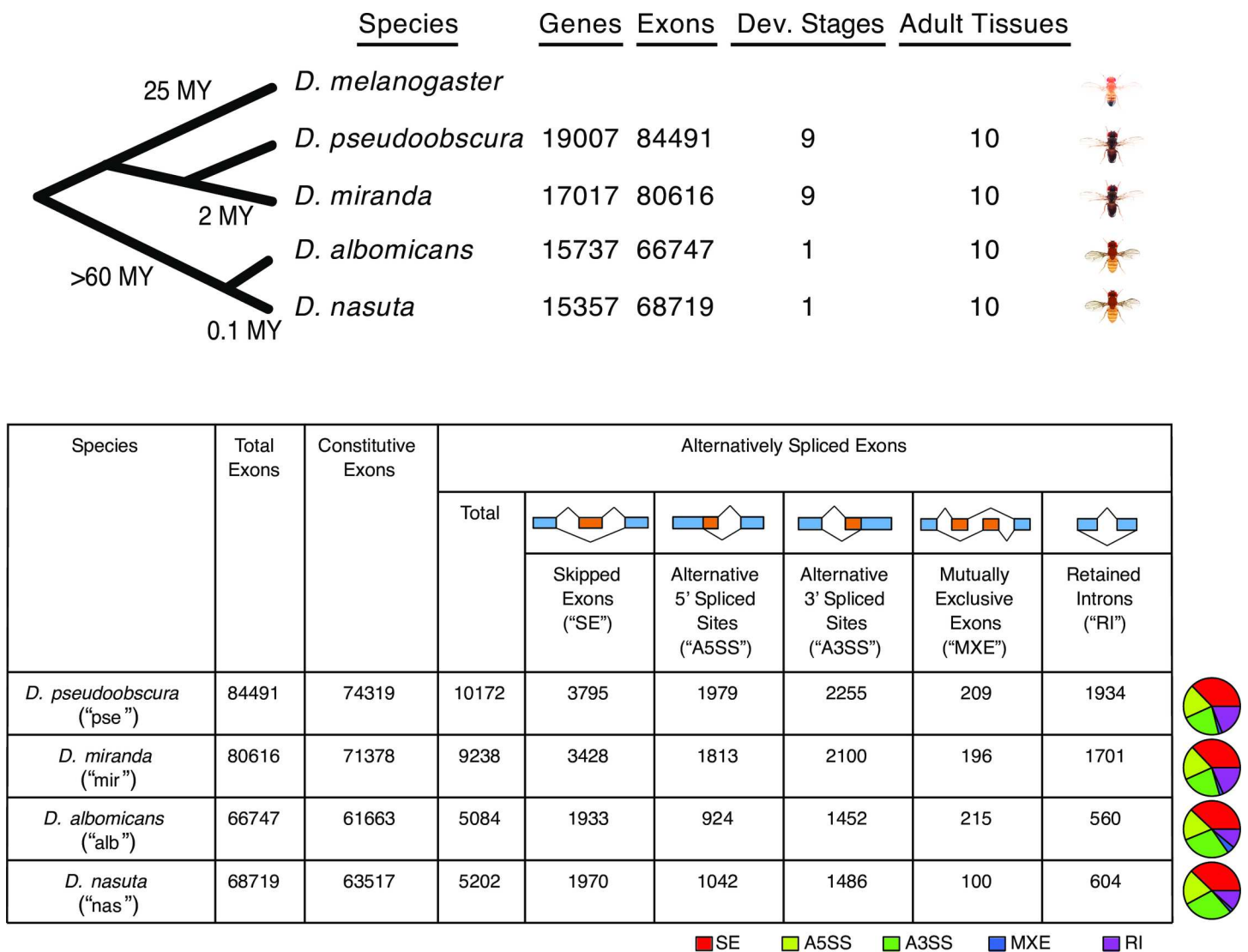


Figure 1

**Fig 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.

doi:10.1371/journal.pgen.1006464.g001

between 5,084 and 10,172 exons (8–12% of all exons) are alternatively spliced (Fig 1). These values are similar to reports in *D. melanogaster* (where 31% of genes were found to be alternatively spliced; [14]), but considerably lower than the degree of AS in mammals, where almost 100% of multi-exon genes are alternatively spliced [7]. Less AS in *Drosophila* is consistent with findings that species further from primates have lower proportions of exons undergoing alternative splicing [6].

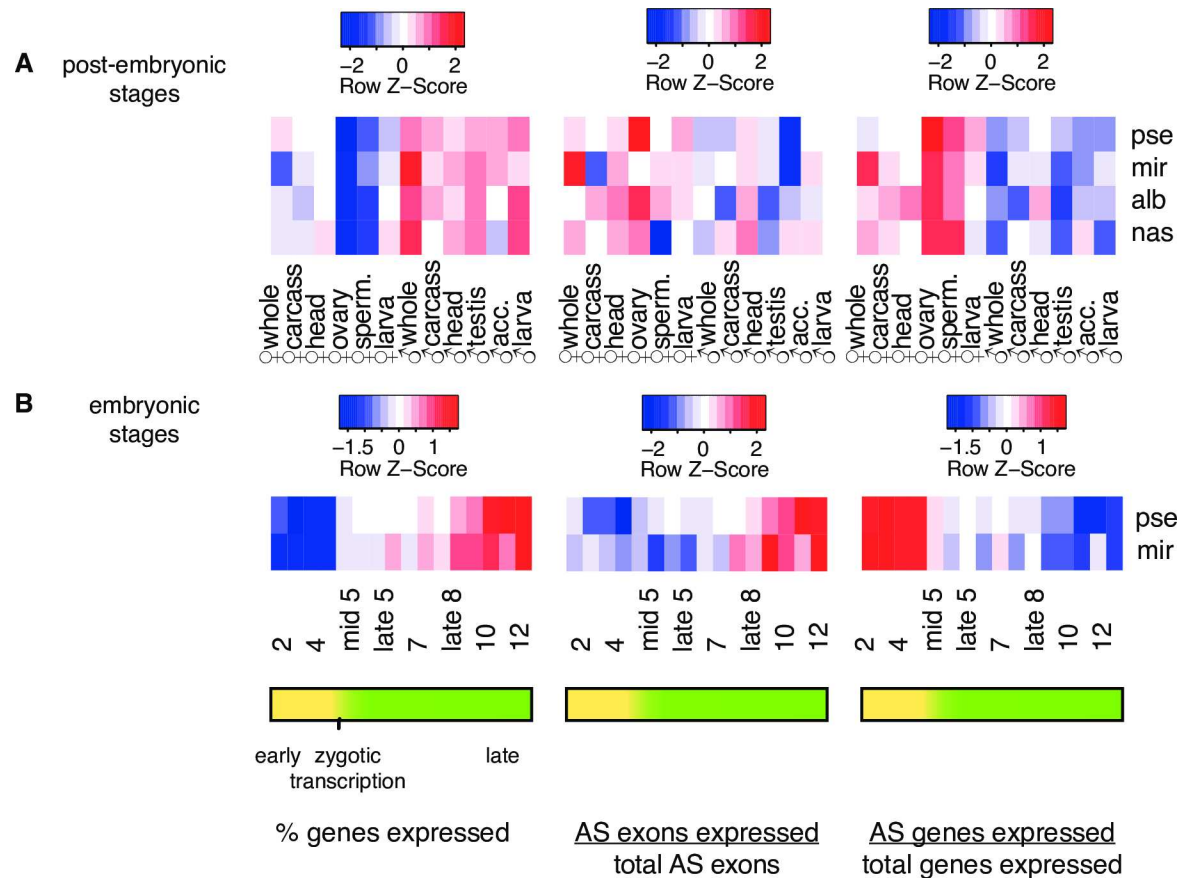
As expected, we detected more genes, exons, and AS events for the two species (*D. pseudoobscura* and *D. miranda*) for which we had both more comprehensive sampling (more samples and RNA-seq reads) and higher quality genome assemblies. Thus, while differences in the numbers of genomic features detected 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 Fig 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 alternative splicing event in all four species, and mutually exclusive exons are the least abundant, consistent with previous studies in *D. melanogaster* [14, 28]. The relative percentage of each type of AS is generally similar between the studied *Drosophila* species (Fig 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 [29], males generally express more genes than females across almost all tissues (Fig 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 (Fig 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; Fig 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 function: 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%; Fig 2A, middle panel). This is consistent with earlier studies in *Drosophila* and humans, which show high transcriptional diversity in brain [30, 31]. 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.



**Fig 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 (A) post-embryonic tissues and (B) embryonic stages. 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’ = 3<sup>rd</sup> instar larva; ‘acc.’ = accessory gland

doi:10.1371/journal.pgen.1006464.g002

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 Fig 2B), when maternal transcripts begin to degrade and widespread zygotic transcription is initiated [24]. This allows us to contrast the landscape of alternative splicing between genes contributed maternally with those transcribed in the early embryo. Over development, both the percentage of total genes expressed (Fig 2B, left panel) and the percentage of alternatively spliced exons expressed (Fig 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 (Fig 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) [32] such as *bbc*, a phosphotransferase, and *Dhc64C*, a gene with ATPase activity, have multiple maternally deposited isoforms, as alternative splicing 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* [32] 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 alternative splicing events of the *Sxl* gene, the master sex determining gene of *Drosophila*, during early embryogenesis in *D. pseudoobscura* and *D. miranda* (see below).

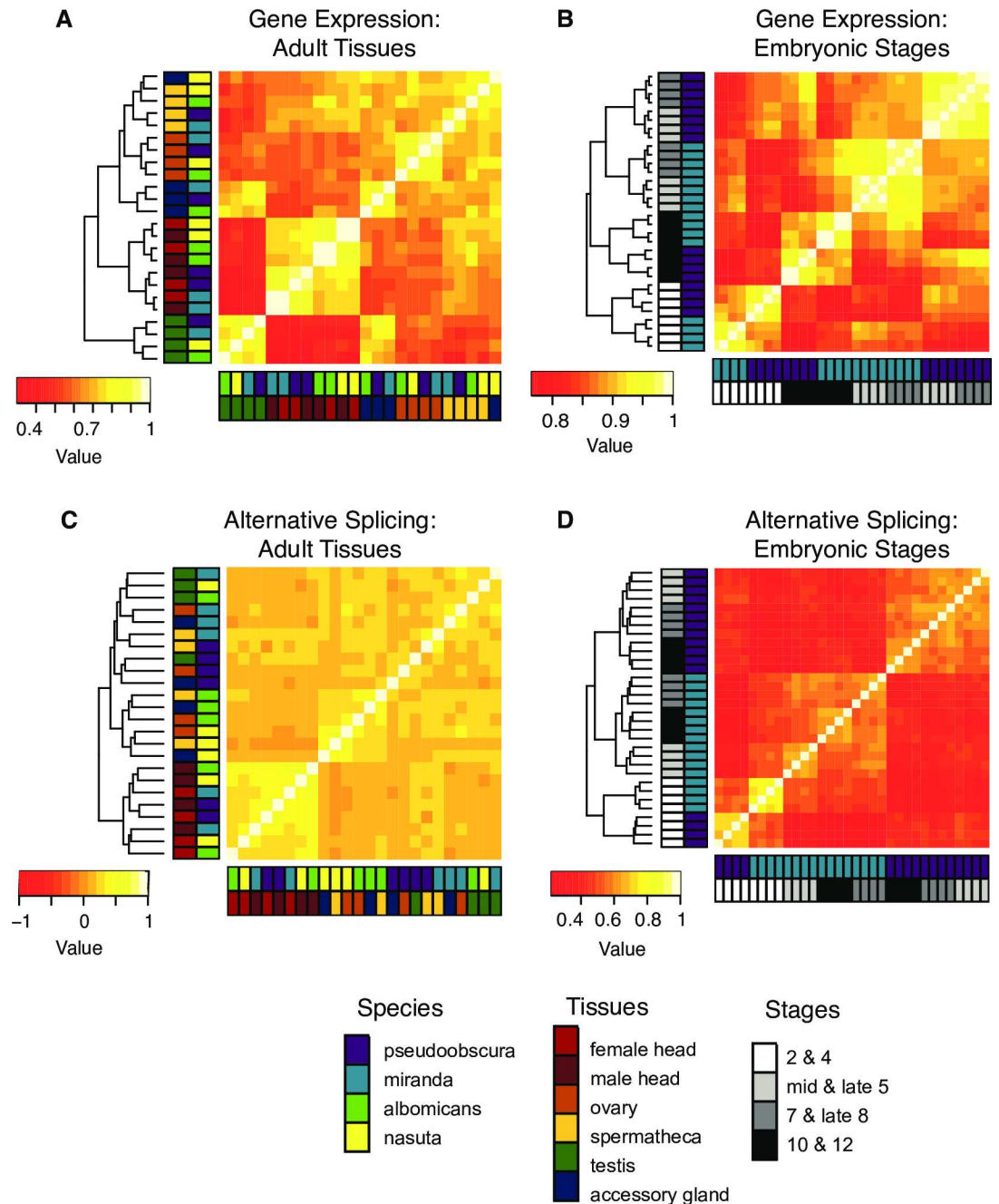
PCA analysis of splicing over embryonic development shows that different embryonic stages have distinct splicing profiles, which form a clock-like pattern in the PCA plot corresponding to the developmental time course (Fig 3; shown is *D. pseudoobscura*, and similar trends are seen for the other species; see S1–S3 Figs). This differs from gene expression, which does not differentiate the different embryonic stages to the same degree as splicing (Fig 3). Remarkably, in PCAs based on AS profiles, ovary is the closest of all post-embryonic tissues to embryonic stages while PCAs based on 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 (Fig 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 alternative splicing 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 S1 Table). As a measure of gene expression for a gene in the interspecific comparisons, we used TPM values (Transcripts Per Million), while alternative splicing was quantified using  $\Psi$  (Percent of exons Spliced In/“PSI”), 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 (Fig 4A). Clustering embryonic stages based on gene expression levels (Fig 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.





**Fig 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 alternative splicing ( $\Psi$ ) 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 ( $n = 1,122$ ).

doi:10.1371/journal.pgen.1006464.g004

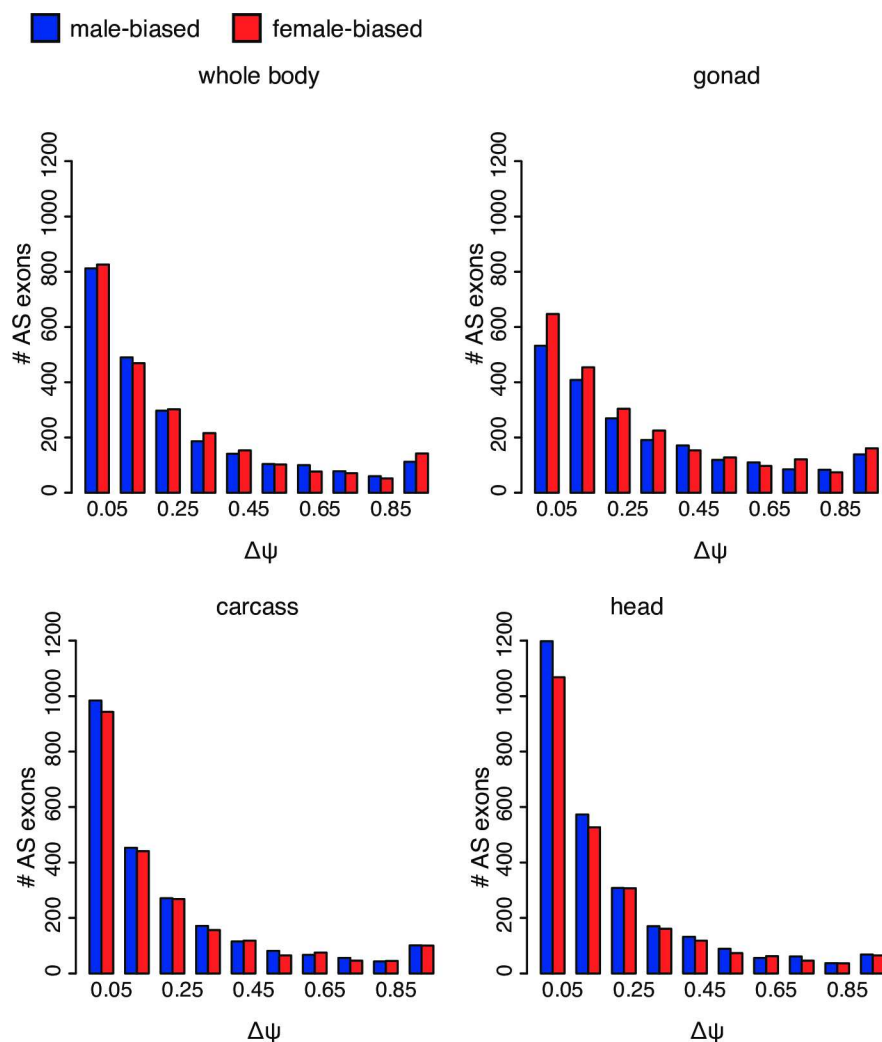
stages 10 & 12 cluster by developmental stage, while only the prezygotic stages 2 and 4 cluster for splicing) is consistent with the strong species-specific clustering observed for splicing and “tissue-dominated clustering” observed for gene expression among vertebrates [7]. A comparison between gene 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

Alternative splicing 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 (S4 Fig). We used  $\Delta\Psi$  values ( $\Delta\Psi = |\Psi_{\text{female}} - \Psi_{\text{male}}|$ ) to assess sex-biased alternative splicing in various male and female tissues and stages (Fig 5), and find sex-biased exons from genes previously observed to have sex-biased isoforms [8], 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



**Fig 5. Sex-biased splicing in *D. pseudoobscura*.** Comparisons of  $\Delta\Psi$  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  $\Delta\Psi$  values and the y-axis shows the number of sex-biased exons. Red bars represent female-biased exons ( $\Psi_{\text{female}} - \Psi_{\text{male}} > 0$ ) and blue bars represent male-biased exons ( $\Psi_{\text{male}} - \Psi_{\text{female}} > 0$ ).

doi:10.1371/journal.pgen.1006464.g005

(Fig 5, S5–S7 Figs). 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 (Fig 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 S5–S7 Figs for  $\Delta\Psi$  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  $\Delta\Psi > 0.10$  (Fig 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%) (S5–S7 Figs). 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 [6, 7, 30, 31], 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 the number of transcripts through AS.

Gene expression levels have been found to cluster by tissue across different mammalian and vertebrate species [6, 7]. 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 [33–37]. 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 [6]) as our *Drosophila* study does (between 0.1 [22] and 60 MY [23]). None of the evolutionary patterns reported differ between distantly- and closely-related species—i.e. species- vs. tissue-specific clustering for alternative splicing vs. gene expression does not differ between closely- vs. distantly-related species (see Fig 4). Thus, divergence time does not appear to affect any of our analysis.

Similar to the findings in mammals and vertebrates, we see strong species-specific clustering for alternative splicing 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 all 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  $\Psi$  values of orthologous exons among species.

While species-specific clustering for alternative splicing is consistent with lineage-specific adaptive evolution [6, 7], it may also support the hypothesis that much splicing is due to erroneous splice site choice, producing non-functional isoforms targeted for degradation/non-sense-mediated decay [38, 39]. 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 [40]. 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 (S8 Fig). 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 alternative splicing. After embryogenesis, gene expression becomes more conserved among tissues and species-specific clustering is less dominant. Alternative splicing profiles cluster samples by species rather than tissue (with the exception of head) into adulthood, suggesting that alternative splicing 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 [41]. The shortness/low complexity of early transcripts is thought to reflect time constraints on producing longer (or spliced) transcripts in quickly dividing early embryos [42, 43] during development [44]. 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 [45] and RNA-seq coverage along transcripts show patterns consistent with intron delay and an inability to fully transcribe long transcripts [46]. 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 [46].

Similar to findings in *D. melanogaster* [45, 46], 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 (S9 Fig).

## Methods

### Data

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

The remaining samples are RNA-seq datasets generated by us from *D. pseudoobscura* (MV25) (male and female 3<sup>rd</sup> instar larva, spermatheca), *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 3<sup>rd</sup> 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 3<sup>rd</sup> 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 (Agilent). 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 (Figs 1–3, Fig 5, S1–S7 Figs, S9A and S9B Fig, S9E and S9F Fig and S4 Table). 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 (S1 Table). 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 [47] to map the reads to each genome (*D. pseudoobscura* Release 3.1 downloaded from <http://www.flybase.org>, *D. miranda* assembly [27], *D. albomicans* assembly [26], *D. nasuta* assembly, unpublished, S3 Table) using the parameters —b2-sensitive,—coverage-search, and—microexon-search and then ran Cufflinks v2.1.1 [48] 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 [49] with this master annotation to obtain expression data for each sample. We use the

“commonly adopted arbitrary inclusion threshold” of FPKM = 1 [50] that is used in other *Drosophila* gene expression studies [18].

We used picard v1.106 (<http://picard.sourceforge.net>) to obtain insert size mean and standard deviation for each sample. We used MATS [51, 52] 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  $\Psi$  (“PSI”, Percent Spliced In) values for each alternatively spliced exon by running pairwise comparisons between all samples within each species and taking an average  $\Psi$  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 [53] to annotate skipped exon alternative splicing events and MISO [4, 54] to compute  $\Psi$  values. **S4 Table** 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 (**S1 Table**) 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 (**Fig 4**, **S8 Fig**, **S9C and S9D Fig** and **S10 and S11 Figs**). Pairwise whole genome alignments for each pair of species were done using the software Mercator (<https://www.biostat.wisc.edu/~cdewey/mercator/>) [55] and MAVID [56]. 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 [57] 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 [57] to normalize TPM values between samples and to compute Jensen-Shannon divergence between samples.

We used bowtie2-build and TopHat v2.0.13 [47] to map the reads to each genome (*D. pseudoobscura* Release 3.1 downloaded from <http://www.flybase.org>, *D. miranda* assembly [27], *D. albomicans* assembly [26], *D. nasuta* assembly (unpublished, S3 Table) using the parameters `—b2-sensitive,—coverage-search,` and `—microexon-search` and then ran Cufflinks v2.1.1 [48] using the `—G` parameter and the reference annotations used for orthology analysis on each sample.

For interspecies splicing analyses, we ran MATS [51, 52] 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  $\Psi$  (“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  $\Psi$  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  $\Psi$  value of 0. If the exon had an FPKM value > 1, the exon was assigned a  $\Psi$  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  $\Psi$  value.

*heatmap.2* was used in R to generate heatmaps, which uses *hclust* to cluster samples. We computed Spearman (Fig 4, S8 Fig) and Pearson (S10 Fig) correlations for GE and AS in adult tissues and embryonic stages in R using the *corr* function. Jensen-Shannon divergence of GE (S11 Fig) was computed using sleuth[57].

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 (S1 Table), 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.

## Supporting Information

**S1 Fig. PCAs based on AS and GE for *D. miranda*.** Alternative splicing (left column) and gene expression (right column) profiles for *D. miranda*. The R function *prcomp* was used to

perform the PCAs. (top) 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). (bottom) 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). “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  
(TIF)

**S2 Fig. PCAs based on AS and GE for *D. albomicans*.** Alternative splicing (left column) and gene expression (right column) profiles for *D. albomicans*. The R function *prcomp* was used to perform the PCAs. (top) 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). (bottom) 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). “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland  
(TIF)

**S3 Fig. PCAs based on AS and GE for *D. nasuta*.** Alternative splicing (left column) and gene expression (right column) profiles for *D. nasuta*. The R function *prcomp* was used to perform the PCAs. (top) 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). (bottom) 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). “carc” = carcass; “ov” = ovary; “sperm” = spermatheca; “larv” = 3rd instar larva; “test” = testis; “acc” = accessory gland  
(TIF)

**S4 Fig. *Sxl* expression in embryonic stages.** Exon 2 is spliced in males (blue) and skipped in females (red). We used IGV[58, 59] to visualize *Sxl* expression. “5.5” = mid stage 5; “5.9” = late stage 5; “8.9” = late stage 8.  
(TIF)

**S5 Fig. Sex-biased splicing in *D. miranda* as described by  $\Delta\Psi$  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  $\Delta\Psi$  values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ( $\Psi_{\text{female}} - \Psi_{\text{male}} > 0$ ) and blue bars represent male-based exons ( $\Psi_{\text{male}} - \Psi_{\text{female}} > 0$ ).  
(TIF)

**S6 Fig. Sex-biased splicing in *D. albomicans* as described by  $\Delta\Psi$  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  $\Delta\Psi$  values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ( $\Psi_{\text{female}} - \Psi_{\text{male}} > 0$ ) and blue bars represent male-based exons ( $\Psi_{\text{male}} - \Psi_{\text{female}} > 0$ ).  
(TIF)

**S7 Fig. Sex-biased splicing in *D. nasuta* as described by  $\Delta\Psi$  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  $\Delta\Psi$  values and the y-axis represents the number of sex-biased exons. Red bars represent female-biased exons ( $\Psi_{\text{female}} - \Psi_{\text{male}} > 0$ ) and blue bars represent male-based exons ( $\Psi_{\text{male}} - \Psi_{\text{female}} > 0$ ).  
(TIF)

**S8 Fig. Spearman correlations based on GE in adult tissues not including head.** Spearman correlations based on gene expression (TPM) for genes orthologous in adult tissues (n = 3005) not including male and female head.

(TIF)

**S9 Fig. 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<sub>2</sub>(length) of genes expressed in early (stage 2–stage 4) and later (mid stage 5–stage 12) embryonic stages from within species analyses in A) *D. pseudoobscura* and B) *D. miranda*. Log<sub>2</sub>(length) of orthologous genes from interspecies analysis expressed in early (stage 2–stage 4) and later (mid stage 5–stage 12) embryonic stages in C) *D. pseudoobscura* and D) *D. miranda*. Number of exons per genes 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.

(TIF)

**S10 Fig. Pearson correlations based on GE and AS.** Pearson correlations based on gene expression (TPM) for genes orthologous in adult tissues (n = 3005) (A) and embryonic stages (n = 6707) (B). Pearson correlations based on alternative splicing ( $\Psi$ ) for exons orthologous and annotated as alternatively spliced in at least one sample in adult tissues (n = 472) (C) and embryonic stages (n = 1122) (D).

(TIF)

**S11 Fig. Heatmaps based on Jensen-Shannon divergence of GE.** Heatmaps based on Jensen-Shannon divergence of gene expression for genes orthologous in adult tissues (n = 3005) (A) and embryonic stages (n = 6707) (B).

(TIF)

**S1 Table. 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

(TIF)

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

(TIF)

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

(TIF)

**S4 Table. Correlation of skipped exon  $\Psi$  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  $\Psi$  values. The second pipeline used AltEventFinder to annotate alternatively spliced exons and MISO to compute their  $\Psi$  values.

(TIF)

## Author Contributions

**Conceptualization:** DB.

**Data curation:** LG QZ SM.

**Formal analysis:** LG SM.  
**Funding acquisition:** DB.  
**Investigation:** LG SM.  
**Methodology:** LG QZ SM DB.  
**Project administration:** DB.  
**Resources:** DB.  
**Supervision:** DB.  
**Validation:** LG QZ SM.  
**Visualization:** LG.  
**Writing – original draft:** DB.  
**Writing – review & editing:** LG QZ SM DB.

## References

1. Nilsen TW, Graveley BR. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457–463. doi: [10.1038/nature08909](https://doi.org/10.1038/nature08909) PMID: [20110989](https://pubmed.ncbi.nlm.nih.gov/20110989/)
2. Cline and TW, Meyer BJ. (1996) Vive la difference: Males vs females in flies vs worms. *Annu Rev Genet* 30: 637–702. doi: [10.1146/annurev.genet.30.1.637](https://doi.org/10.1146/annurev.genet.30.1.637) PMID: [8982468](https://pubmed.ncbi.nlm.nih.gov/8982468/)
3. Schutt C, Nothiger R. (2000) Structure, function and evolution of sex-determining systems in dipteran insects. *Development* 127: 667–677. PMID: [10648226](https://pubmed.ncbi.nlm.nih.gov/10648226/)
4. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476. doi: [10.1038/nature07509](https://doi.org/10.1038/nature07509) PMID: [18978772](https://pubmed.ncbi.nlm.nih.gov/18978772/)
5. Staiger D, Brown JW. (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25: 3640–3656. doi: [10.1105/tpc.113.113803](https://doi.org/10.1105/tpc.113.113803) PMID: [24179132](https://pubmed.ncbi.nlm.nih.gov/24179132/)
6. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338: 1587–1593. doi: [10.1126/science.1230612](https://doi.org/10.1126/science.1230612) PMID: [23258890](https://pubmed.ncbi.nlm.nih.gov/23258890/)
7. Merkin J, Russell C, Chen P, Burge CB. (2012) Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338: 1593–1599. doi: [10.1126/science.1228186](https://doi.org/10.1126/science.1228186) PMID: [23258891](https://pubmed.ncbi.nlm.nih.gov/23258891/)
8. Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. (2011) Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genomics* 12: 364. doi: [10.1186/1471-2164-12-364](https://doi.org/10.1186/1471-2164-12-364) PMID: [21756339](https://pubmed.ncbi.nlm.nih.gov/21756339/)
9. Hartmann B, Castelo R, Minana B, Peden E, Blanchette M, et al. (2011) Distinct regulatory programs establish widespread sex-specific alternative splicing in *Drosophila melanogaster*. *RNA* 17: 453–468. doi: [10.1261/rna.2460411](https://doi.org/10.1261/rna.2460411) PMID: [21233220](https://pubmed.ncbi.nlm.nih.gov/21233220/)
10. Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin SV, McIntyre LM. (2009) Sex-specific splicing in *Drosophila*: Widespread occurrence, tissue specificity and evolutionary conservation. *Genetics* 181: 421–434. doi: [10.1534/genetics.108.096743](https://doi.org/10.1534/genetics.108.096743) PMID: [19015538](https://pubmed.ncbi.nlm.nih.gov/19015538/)
11. Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, et al. (2010) Dynamic regulation of alternative splicing and chromatin structure in *drosophila* gonads revealed by RNA-seq. *Cell Res* 20: 763–783. doi: [10.1038/cr.2010.64](https://doi.org/10.1038/cr.2010.64) PMID: [20440302](https://pubmed.ncbi.nlm.nih.gov/20440302/)
12. Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, et al. (2011) Conservation of an RNA regulatory map between *drosophila* and mammals. *Genome Res* 21: 193–202. doi: [10.1101/gr.108662.110](https://doi.org/10.1101/gr.108662.110) PMID: [20921232](https://pubmed.ncbi.nlm.nih.gov/20921232/)
13. Cherbas L, Willingham A, Zhang D, Yang L, Zou Y, et al. (2011) The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res* 21: 301–314. doi: [10.1101/gr.112961.110](https://doi.org/10.1101/gr.112961.110) PMID: [21177962](https://pubmed.ncbi.nlm.nih.gov/21177962/)
14. Daines B, Wang H, Wang L, Li Y, Han Y, et al. (2011) The *Drosophila melanogaster* transcriptome by paired-end RNA sequencing. *Genome Res* 21: 315–324. doi: [10.1101/gr.107854.110](https://doi.org/10.1101/gr.107854.110) PMID: [21177959](https://pubmed.ncbi.nlm.nih.gov/21177959/)

15. Venables JP, Tazi J, Juge F. (2012) Regulated functional alternative splicing in *Drosophila*. *Nucleic Acids Res* 40: 1–10. doi: [10.1093/nar/gkr648](https://doi.org/10.1093/nar/gkr648) PMID: [21908400](https://pubmed.ncbi.nlm.nih.gov/21908400/)
16. McIntyre LM, Bono LM, Genissel A, Westerman R, Junk D, et al. (2006) Sex-specific expression of alternative transcripts in *Drosophila*. *Genome Biol* 7: R79. doi: [10.1186/gb-2006-7-8-r79](https://doi.org/10.1186/gb-2006-7-8-r79) PMID: [16934145](https://pubmed.ncbi.nlm.nih.gov/16934145/)
17. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, et al. (2014) Diversity and dynamics of the *Drosophila* transcriptome. *Nature* 512: 393–399. doi: [10.1038/nature12962](https://doi.org/10.1038/nature12962) PMID: [24670639](https://pubmed.ncbi.nlm.nih.gov/24670639/)
18. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, et al. (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471: 473–479. doi: [10.1038/nature09715](https://doi.org/10.1038/nature09715) PMID: [21179090](https://pubmed.ncbi.nlm.nih.gov/21179090/)
19. Duff MO, Olson S, Wei X, Garrett SC, Osman A, et al. (2015) Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* 521: 376–379. doi: [10.1038/nature14475](https://doi.org/10.1038/nature14475) PMID: [25970244](https://pubmed.ncbi.nlm.nih.gov/25970244/)
20. Bachtrog D, Charlesworth B. (2002) Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416: 323–326. doi: [10.1038/416323a](https://doi.org/10.1038/416323a) PMID: [11907578](https://pubmed.ncbi.nlm.nih.gov/11907578/)
21. Robinson R. (2012) Natural selection molds genomic insulator elements. *PLoS Biol* 10: e1001421. doi: [10.1371/journal.pbio.1001421](https://doi.org/10.1371/journal.pbio.1001421) PMID: [23139641](https://pubmed.ncbi.nlm.nih.gov/23139641/)
22. Bachtrog D. (2006) A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev* 16: 578–585. doi: [10.1016/j.gde.2006.10.007](https://doi.org/10.1016/j.gde.2006.10.007) PMID: [17055249](https://pubmed.ncbi.nlm.nih.gov/17055249/)
23. Tamura K, Subramanian S, Kumar S. (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21: 36–44. doi: [10.1093/molbev/msg236](https://doi.org/10.1093/molbev/msg236) PMID: [12949132](https://pubmed.ncbi.nlm.nih.gov/12949132/)
24. Lott SE, Villalta JE, Zhou Q, Bachtrog D, Eisen MB. (2014) Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. *PLoS Genet* 10: e1004159. doi: [10.1371/journal.pgen.1004159](https://doi.org/10.1371/journal.pgen.1004159) PMID: [24550743](https://pubmed.ncbi.nlm.nih.gov/24550743/)
25. Zhou Q, Bachtrog D. (2012) Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337: 341–345. doi: [10.1126/science.1225385](https://doi.org/10.1126/science.1225385) PMID: [22822149](https://pubmed.ncbi.nlm.nih.gov/22822149/)
26. Zhou Q, Zhu HM, Huang QF, Zhao L, Zhang GJ, et al. (2012) Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 13: 109. doi: [10.1186/1471-2164-13-109](https://doi.org/10.1186/1471-2164-13-109) PMID: [22439699](https://pubmed.ncbi.nlm.nih.gov/22439699/)
27. Zhou Q, Ellison CE, Kaiser VB, Alekseyenko AA, Gorchakov AA, et al. (2013) The epigenome of evolving *Drosophila* neo-sex chromosomes: Dosage compensation and heterochromatin formation. *PLoS Biol* 11: e1001711. doi: [10.1371/journal.pbio.1001711](https://doi.org/10.1371/journal.pbio.1001711) PMID: [24265597](https://pubmed.ncbi.nlm.nih.gov/24265597/)
28. McManus CJ, Coolon JD, Eipper-Mains J, Wittkopp PJ, Graveley BR. (2014) Evolution of splicing regulatory networks in *Drosophila*. *Genome Res* 24: 786–796. doi: [10.1101/gr.161521.113](https://doi.org/10.1101/gr.161521.113) PMID: [24515119](https://pubmed.ncbi.nlm.nih.gov/24515119/)
29. 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: 233–237. doi: [10.1038/nature06323](https://doi.org/10.1038/nature06323) PMID: [17994089](https://pubmed.ncbi.nlm.nih.gov/17994089/)
30. Mohr C, Hartmann B. (2014) Alternative splicing in *Drosophila* neuronal development. *J Neurogenet* 28: 199–215. doi: [10.3109/01677063.2014.936437](https://doi.org/10.3109/01677063.2014.936437) PMID: [24957133](https://pubmed.ncbi.nlm.nih.gov/24957133/)
31. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476. doi: [10.1038/nature07509](https://doi.org/10.1038/nature07509) PMID: [18978772](https://pubmed.ncbi.nlm.nih.gov/18978772/)
32. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, et al. (2011) Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. *PLoS Biol* 9: e1000590. doi: [10.1371/journal.pbio.1000590](https://doi.org/10.1371/journal.pbio.1000590) PMID: [21346796](https://pubmed.ncbi.nlm.nih.gov/21346796/)
33. Grassa CJ, Kulathinal RJ. (2011) Elevated evolutionary rates among functionally diverged reproductive genes across deep vertebrate lineages. *Int J Evol Biol* 2011: 274975. doi: [10.4061/2011/274975](https://doi.org/10.4061/2011/274975) PMID: [21811675](https://pubmed.ncbi.nlm.nih.gov/21811675/)
34. Prokupek A, Hoffmann F, Eyun S, Moriyama E, Zhou M, et al. (2008) An evolutionary expressed sequence tag analysis of *Drosophila spermatheca* genes. *Evolution* 62: 2936–2947. doi: [10.1111/j.1558-5646.2008.00493.x](https://doi.org/10.1111/j.1558-5646.2008.00493.x) PMID: [18752616](https://pubmed.ncbi.nlm.nih.gov/18752616/)
35. Swanson WJ, Vacquier VD. (2002) The rapid evolution of reproductive proteins. *Nature Reviews Genetics* 3: 137–144. doi: [10.1038/nrg733](https://doi.org/10.1038/nrg733) PMID: [11836507](https://pubmed.ncbi.nlm.nih.gov/11836507/)
36. Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* 156: 1879–1888. PMID: [11102381](https://pubmed.ncbi.nlm.nih.gov/11102381/)
37. Tsaour SC, Wu CI. (1997) Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. *Mol Biol Evol* 14: 544–549. PMID: [9159932](https://pubmed.ncbi.nlm.nih.gov/9159932/)
38. Pickrell JK, Pai AA, Gilad Y, Pritchard JK. (2010) Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* 6: e1001236. doi: [10.1371/journal.pgen.1001236](https://doi.org/10.1371/journal.pgen.1001236) PMID: [21151575](https://pubmed.ncbi.nlm.nih.gov/21151575/)



39. Lareau LF, Brooks AN, Soergel DA, Meng Q, Brenner SE. (2007) The coupling of alternative splicing and nonsense-mediated mRNA decay. *Adv Exp Med Biol* 623:190–211. PMID: [18380348](#)
40. Catalán A, Hutter S, Parsch J. (2012) Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* 13: 654. doi: [10.1186/1471-2164-13-654](#) PMID: [23170910](#)
41. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, et al. (2014) The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Reports* 6: 285–292. doi: [10.1016/j.celrep.2013.12.030](#) PMID: [24440719](#)
42. Mourier T, Jeffares DC. (2005) Eukaryotic intron loss. *Science* 300: 1393. doi: [10.1126/science.1080559](#) PMID: [12775832](#)
43. Jeffares DC, Mourier T, Penny D. (2006) The biology of intron gain and loss. *Trends in Genetics* 22: 16–22. doi: [10.1016/j.tig.2005.10.006](#) PMID: [16290250](#)
44. Foe VE, Odell G, Edgar B. (1993) Mitosis and morphogenesis in the *Drosophila* embryo: Point and counterpoint. *The Development of Drosophila melanogaster* 1: 149–300.
45. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. (2007) Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biol* 5: e117. doi: [10.1371/journal.pbio.0050117](#) PMID: [17456005](#)
46. Artieri CG, Fraser HB. (2014) Transcript length mediates developmental timing of gene expression across *Drosophila*. *Mol Biol Evol* 31: 2879–2889. doi: [10.1093/molbev/msu226](#) PMID: [25069653](#)
47. Trapnell C, Pachter L, Salzberg SL. (2009) TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* 25: 1105–1111. doi: [10.1093/bioinformatics/btp120](#) PMID: [19289445](#)
48. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515. doi: [10.1038/nbt.1621](#) PMID: [20436464](#)
49. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 31: 46–53. doi: [10.1038/nbt.2450](#) PMID: [23222703](#)
50. Younus F, Chertemps T, Pearce SL, Pandey G, Bozzolan F, et al. (2014) Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of *Drosophila melanogaster*. *Insect Biochem Mol Biol* 53: 30–43. doi: [10.1016/j.ibmb.2014.07.003](#) PMID: [25038463](#)
51. Park JW, Tokheim C, Shen S, Xing Y. (2013) Identifying differential alternative splicing events from RNA sequencing data using RNASEq-MATS. *Methods Mol Biol* 1038:171–179. doi: [10.1007/978-1-62703-514-9\\_10](#) PMID: [23872975](#)
52. Shen S, Park JW, Huang J, Dittmar KA, Lu ZX, et al. (2012) MATS: A bayesian framework for flexible detection of differential alternative splicing from RNA-seq data. *Nucleic Acids Res* 40: e61. doi: [10.1093/nar/gkr1291](#) PMID: [22266656](#)
53. Zhou A, Breese MR, Hao Y, Edenberg HJ, Li L, et al. (2012) Alt event finder: A tool for extracting alternative splicing events from RNA-seq data. *BMC Genomics* 13 Suppl 8: S10.
54. Katz Y, Wang ET, Airoidi EM, Burge CB. (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* 7: 1009–1015. doi: [10.1038/nmeth.1528](#) PMID: [21057496](#)
55. Dewey CN (2007) Aligning multiple whole genomes with Mercator and MAVID. *Methods Mol Biol* 395: 221–236. PMID: [17993677](#)
56. Bray N, Pachter L. (2004) MAVID: Constrained ancestral alignment of multiple sequences. *Genome Res* 14: 693–699. doi: [10.1101/gr.1960404](#) PMID: [15060012](#)
57. Bray N, Pimentel H, Melsted P, Pachter L. (2015) Near-optimal RNA-seq quantification. *Nat Biotechnol* 34: 525–527.
58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29: 24–26. doi: [10.1038/nbt.1754](#) PMID: [21221095](#)
59. Thorvaldsdóttir H, Robinson JT, Mesirov JP. (2013) Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178–192. doi: [10.1093/bib/bbs017](#) PMID: [22517427](#)