NOVEL APPROACHES TO STUDYING ALTERNATIVE SPLICING IN DROSOPHILA MELANOGASTER: INSIGHTS INTO SEX-SPECIFIC GENE EXPRESSION AND THE EVOLUTION OF SEX DETERMINATION

by

Julia Raices

May, 2024

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:
Georgios Katsaros, Chair
Beatriz Vicoso
Daria Siekhaus
Jan Larsson

The thesis of Julia Raices, titled *Novel approaches to studying alternative splicing in Drosophila melanogaster: Insights into sex-specific gene expression and the evolution of sex determination*, is approved by:

Supervisor: Beatriz Vicoso, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Daria Siekhaus, UCLA, Los Angeles, USA

Signature: _____

Committee Member: Jan Larsson, Umeå Universitet, Umeå, Sweden

Signature: _____

Defense Chair: Georgios Katsaros, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Julia Raices

May, 2024

Signed page is on file

Abstract

Males and females exhibit numerous differences, from the initial stages of sex determination to the development of secondary sexual characteristics. In *Drosophila*, these differences have been thoroughly studied. Extensive research has been performed to understand the role and molecular mode of action of central sex in determining switch genes, such as *transformer* (*tra*) and *Sex-lethal* (*Sxl*). Furthermore, studies have highlighted differential gene expression as an essential mechanism to create sexual dimorphism. An alternative path to sexual dimorphism that has been less explored is alternative splicing, the mechanism through which genes can produce multiple transcripts with distinct properties and functions. The primary switch sex-determining gene *Sxl* is a good example of the role of alternative splicing for sex-specific functions: the inclusion of a specific exon determines the male or female form of the protein, which in turn switches on either the male or female developmental pathway. The genes that act upstream of *Sxl* and determine which form is expressed - the counter genes - have received less attention. This thesis addresses two critical questions about the molecular encoding of sexes in the *Drosophila melanogaster* genome: First, the use of splice forms in male and female tissues in *D. melanogaster* is examined, inferring the molecular and evolutionary parameters shaping the diversity of the splicing landscape. Second, the behaviour of counter genes in *Drosophila*-related species is investigated, shedding light on potential changes leading to their incorporation into the sex-determination pathway.

For the alternative splicing analyses, long-read RNA sequencing of testes, ovaries, female and male midguts, heads, and whole bodies was performed. A novel pipeline was developed to assign unique transcript identifiers for each sequence of exons and introns in the read, enabling detailed comparisons of splicing variants in each tissue/sex. Alternative splicing was found to be more pervasive in females than males (22,201 exclusive splice forms in females versus 12,631 in males), especially when comparing ovaries to other tissues. The ovaries alone displayed 15,299 exclusive splice forms, suggesting most female exclusive splice forms originate there. Genome location and gene age were also correlated with the number of splice forms per gene. In particular, the X and 4th chromosomes (Muller elements A and F) showed more splice forms per gene than other chromosomes. Additionally, genes older than 63 million years exhibited more splice forms per gene than younger genes. Our results suggest that alternative splicing is more prevalent than previously believed, with numerous female-exclusive forms, age, and location playing significant roles in shaping its prevalence.

For the counter genes analyses, we combined published gene expression, genomic, and gene interaction data from various clades (*Bactrocera jarvisi*, *B. oleae*, *Ceratitis capitata*, *Mus musculus*, *Caenorhabditis elegans*, *Homo sapiens*, and *D. melanogaster*). The counter genes *scute* (*sc*), *extra macrochaetae* (*emc*), *groucho* (*gro*), *deadpan* (*dpn*), *daughterless* (*da*), *runt* (*run*), *Sxl*, *hermaphrodite* (*her*), and *tra* maintain conserved Muller element locations between *C. capitata* and *D. melanogaster*, which are most of the counter genes identified in the *C. capitata* genome. Their expression patterns during early embryogenesis in *B. jarvisi* and *D. melanogaster* are also similar for counter genes *dpn*, *gro*, *da*, and *emc*. However, *Sxl* and *sc* are also found to have more extreme expression ratios between the species. Lastly, gene interactions within the counter genes are conserved, with *da-sc* and *gro-dpn* interactions occurring in *Drosophila*, worms, humans, and mice. Interactions such as *dpn-sc*, *dpn-da*, *da-emc*, and *gro-run* are present in *Drosophila*, mice, and humans, suggesting these genes were recruited by ancestral characteristics, primarily during embryogenesis. The conserved expression, location, and interactions of counter genes suggest serendipitous recruitment of such genes instead of a change in those characteristics as they were recruited for this function.

Acknowledgements

To Baku: thanks for being the first one to love me here.

I'm extremely grateful to ISTA for offering me the financial support and resources necessary to complete this work. In this aspect, I'm also extremely grateful to the MSC grant foundation for funding the first years of my PhD studies and the ERC grant under the European Union's Horizon 2020 research and innovation program, number 715257, for the funding of most of my PhD studies. It's essential to note that this endeavour would not have been possible without Beatriz's mentoring and expert feedback throughout all these years.

Words cannot express my gratitude to my partner Elias' emotional support, enduring all my crazies, following me here, and sharing our lives and pets. I want to express my deepest gratitude to Luisa, Lorena, Steffi, and Gemma's support at the workplace and for being the best venting buddies a PhD student could ask for. I am deeply indebted to my committee members Daria Siekhaus and Jan Larsson's constructive criticism, feedback, guidance, and understanding. I also must express my deepest appreciation for Marcela Berti, Barbara Paes, Carolina Barata, and Joana Melo's support in all aspects, help in editing, proofreading, advice, and suggestions during the writing process. It would be impossible to complete this without Juliana's help in maintaining my sanity. From ISTA, many people were essential for this work, but I would like to thank particularly Hilde Janssens, Christine Ostermann, Hitome Tsukuda, and the entire Scientific Computing team – particularly Alois Schlögl - for their help and support during the development and finalisation of this thesis. You were crucial in this process, and I greatly appreciate your contributions. Finally, words are not enough to thank Chico Camargo for helping me improve the text and for revising it.

About the Author

Julia Raices completed a teaching degree and a BSc in Biological Sciences at the University of Sao Paulo, as well as an MSc in Genetics and Evolution under Dr Maria Vibranovski at the same university. In September 2017, Raices joined ISTA for her PhD studies under Dr Beatriz Vicoso. Her main research interests include the evolution of sex and asexuality, the evolution of sexual dimorphism, and reproductive behaviours and systems. She presented her current work at the ESEB Meeting in 2019, at the Molecular Basis and Evolution of Sexual Dimorphism conference in 2020, and at the Drosophila meeting in 2021.

Table of Contents

List of Figures

List of Tables

1 Introduction

Creation myths worldwide discuss the differences between human males and females (Willis, 2006), indicating a long-standing curiosity about the sexes in human history. Decades of research have greatly expanded our understanding of the mechanisms and evolutionary pressures shaping sexual dimorphism, with the science of sexual differences encompassing multiple fields of research, including morphology, physiology, and evolution. Here, we add to this growing body of knowledge by studying how alternative splicing differs between the sexes in *Drosophila melanogaster* and how the upstream regulators of sex determination evolved in this species.

1.1 What constitutes males and females

Countless species exhibit significant morphological and physiological differences between sexes (Hedrick & Temeles, 1989; Karp et al., 2017; Ludwig et al., 2023; Wu et al., 2018; Burtis, 1993; Chang et al., 2011; Millington & Rideout, 2018). These specificities prompt the question of how two distinct organisms, phenotypes, or systems develop from a genome that is, in most cases, largely identical (Punzalan & Hosken, 2010). Differences can be substantial, such as body size in *Drosophila* or feather colour in certain birds. Even in tissues without noticeable morphological differences between the sexes, subtle but consistent phenotypic variations distinguish males from females (Anderson & FitzGerald, 2020; Camara et al., 2008). In *Drosophila,* for example, many head genes show alternative splicing differences in males and females (Telonis-Scott et al., 2009). Brains also show differences in gene expression between the sexes (Watanabe, 2019), as does gut metabolism (Hudry et al., 2019). Even lifespan is influenced by sex (as reviewed by Tower & Arbeitman, 2009). However, the most pronounced differences are observed in the reproductive system. Although the end goal of male and female meiosis is similar - the production of gametes - the morphology and function of the cells produced by female and male gametogenesis are distinctly different.

This study relates to different but related concepts: sex determination, sexual differentiation, and sexual dimorphism. Sex determination refers to the species-wide system leading to the development of male or female reproductive characteristics. This can occur through various mechanisms, including genetic, environmental, social, or epigenetic factors (Bachtrog et al., 2014). Sexual differentiation is the process through which the sex of an individual is set up at the cellular or whole-body level as a result of sexual determination. It includes the development of reproductive organs, secondary sexual characteristics, and sex-specific behaviours. Sexual differentiation is influenced by genetic, hormonal, and environmental factors (Smith & Sinclair, 2004) since it is related to sexual determination. Lastly, sexual dimorphism occurs when males and females display different phenotypes that themselves are a result of sexual differentiation. These can be visible, such as sex combs in *Drosophila melanogaster* or terminalia in several insects, but may also be less obvious, such as pheromones or specific protein products in tissues.

Understanding sex determination and sexual differentiation is crucial for comprehending the development of males and females and their impact on the conservation of sexual species or the study of sex-specific diseases. Sex determination begins early during embryogenesis and can continue until adulthood, when sex organs mature, and reproduction is possible. In Diptera, the alternative splicing of master sex determining genes initiates sex determination

early in embryogenesis, with the process being cell-specific (Salz & Erickson, 2010). From an evolutionary perspective, studying the evolution of sex determination mechanisms can help better understand the diversity of strategies, necessary selective pressures, and their role in maintaining genetic diversity, speciation, and adaptation. From economic and conservation perspectives, understanding these mechanisms can aid in developing or improving methods to influence or preserve the sex ratio in species of interest.

1.2 The Evolution of Sex Determination

Sex determination is a process with a variety of potential underlying mechanisms. Males are organisms that produce many small mobile gametes, while females produce few stationary large gametes. The most studied animal in any field is often humans, therefore a large body of knowledge exists about humans and mammals in general, and the same is true about sex determination. In therian mammals sex determination has been extensively studied, and in many cases a male determining gene is present on the Y chromosome. The discovery of the role of the Y chromosome in sex determination in mealworms overturned previous hypotheses and established the dependence of sex determination on this chromosome (Stevens, 1905). Painter (1922, 1923) proposed that in mammals, the Y chromosome and its centrality for sex determination (for more historical information on sex chromosomes in humans and other mammals please check Wilhelm et al., 2007 or Graves, 2016). The evolution of therian mammals was proposed to have occurred after the evolution of the SRY gene, which defined a novel XY sex chromosome pair, creating a reproductive barrier with ancestral populations of synapsid reptiles (Graves, 2016). The pre-eminent role of the Y chromosome in mammalian sex determination has been a fundamental genetic rule, with observations on mammals with aberrant sex chromosome constitutions providing insights into this process (Graves, 1996).

In 1910, Morgan published a now seminal paper on the sex ratio of *Drosophila* offspring showing that it is consistent with mendelian rules for an XY system. Soon, Bridges (1921,1925) showed the importance of the ratio of X-chromosomes to autosomes in the sex determination of fruit flies, stablishing its difference from what was found in mammals. These two XY systems show two different strategies in sex determination: dominant gene and dosage-dependent sex determination. In mammals a dominant gene in the Y chromosome is the switch to start male sex determination, which means that any mammal with the Y chromosome will be a male, and anyone without it will be female (e.g., XO individuals). Meanwhile in *Drosophila*, the X-chromosome dosage is the key factor for sex determination, so individuals that are XXY and have two sets of autosomes (one from each parent) will be females, and XO individual in this situation will be males (Bridges, 1921).

Despite both being XY systems, the *Drosophila* and mammal sex determination systems show great differences and studies have demonstrated variations in sex determination mechanisms even among closely related species, prompting inquiries into the factors influencing these differences (Hansson & Olsson, 2018). Sex chromosomes also play a crucial role in population divergence and speciation (Hill et al., 2021).

1.3 The Molecular Basis of Sexual Differentiation

Various mechanisms contribute to the phenotypic differences between males and females. Figure 1 lists three of those mechanisms, namely gene location, gene expression, and alternative splicing. As shown in Figure 1, one way to have different phenotypes between males and females is to locate genes controlling male traits on the Y chromosome, which is exclusive to males. However, Y chromosomes are often degenerated and contain few genes (Carvalho, 2002; Koerich et al., 2008). In particular, in *Drosophila*, the Y chromosome is primarily associated with male fertility (Hafezi et al., 2020; Carvalho et al., 2001). The scarcity of Y-linked genes suggests that Y-linkage alone is insufficient to account for most sexual dimorphism. A more prevalent source of variation between sexes, also shown in Figure 1, is sex-specific gene expression (Parisi et al., 2004; Ellegren & Parsch, 2007). Genes predominantly expressed in one sex, known as sex-biased genes, are widespread, with studies estimating that about 58% of *Drosophila* genes are sex-biased (Telonis-Scott et al., 2008). Most tissues exhibit some degree of sex-specific expression (Yang, 2006), particularly in the germline (Parisi et al., 2004).

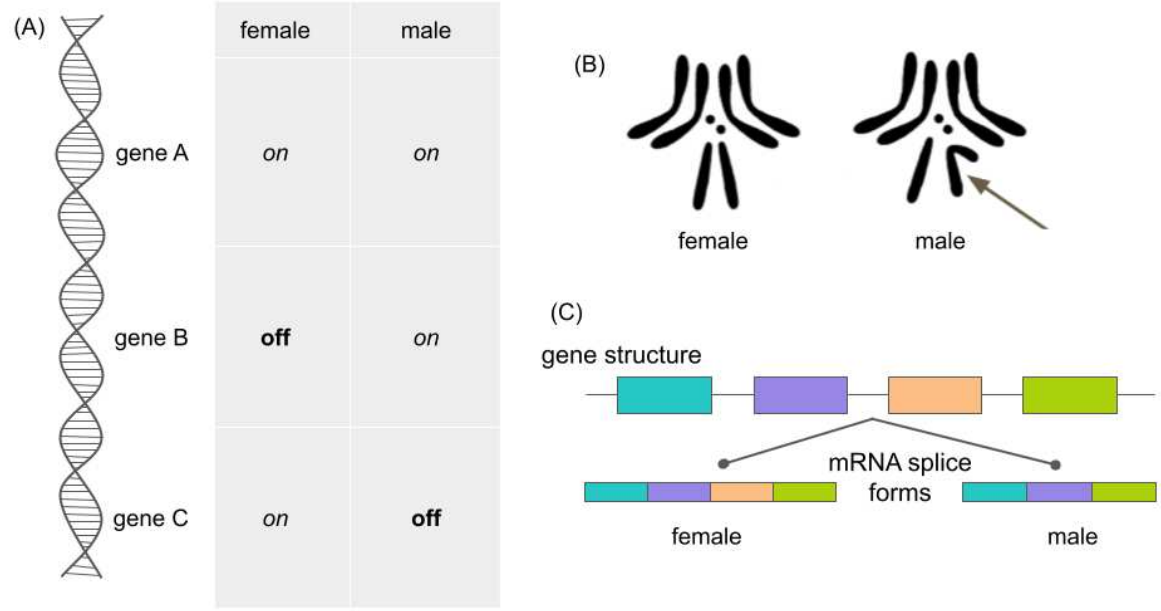

Figure 1. Different strategies for the genetics of sexual dimorphic traits. *In organisms with sex chromosomes, there are different ways to obtain different phenotypes for males and females from what is mostly the same genome: (A) genes that have different expression levels or are expressed in different tissues in males and females; (B) genes present in the sex-specific chromosome (Y-chromosome in the case of Drosophila and mammals) are exclusively expressed in the sex that possesses that chromosome (males in the case of Drosophila and mammals); (C) genes that express different splice forms (alternative splicing) in males and females. Both (A) and (B) have been extensively investigated in the literature, while (C) shows a less studied possibility.*

Another possibility for creating distinct phenotypes that is not yet as well understood is the presence of different splice forms in males and females, as shown in Figure1. Varied splice forms can generate a broader range of proteins from the same gene (Breitbart et al., 1987), which would not be achievable with mere abundance changes of the same protein form. Alternative splicing can also affect the number of proteins produced by a gene, as changes to

the 5' UTR of an mRNA can influence the recruitment of ribosomes and other cellular machinery, thereby affecting the translation efficiency of that splice form (Matera & Wang, 2014; Fiszbein et al., 2019; Gnan et al., 2022; Jabre et al., 2021). Genes can undergo tissue-specific as well as sex-specific splicing (Telonis-Scott et al., 2008; Graveley et al., 2011), with gonad-specific splicing being responsible for a large amount of sex-specific splicing (Gibilisco et al., 2016). Gan et al. (2010) identified 614 genes with sex-specific splice forms in a comparison of testes versus ovaries in *D. melanogaster* RNAseq data. Mohr et al. (2017) detected at least 47 genes with sex-specific splicing in adult fruitfly heads. In somatic tissues, such as heads, 1,370 splice forms exhibited unique prevalences in males and females of the same species, suggesting the significance of sex-specific isoforms for sexual differentiation and dimorphism (Demir & Dickson, 2005; Chang et al., 2011). Identifying these isoforms is crucial for understanding sexual differentiation and identifying key genes and isoforms necessary for its proper establishment in *Drosophila*.

Sex-specific splicing plays a crucial role in *Drosophila*, where male and female splice forms of several sex determining genes regulate aspects of the sex determination pathway (e.g., *Sxl*, *transformer*, and *fruitless*) (Sawanth et al., 2016). For instance, Figure 13 illustrates the functional consequences of male and female forms of *Sxl* in producing *Drosophila* phenotypes.

Alternative splicing remains vital throughout the life of *Drosophila*, as genes such as *fru*, essential for mating behaviours, present sex-specific splice forms in brains (Demir & Dickson, 2005; Kimura et al., 2005). The male form of *fru*, for example, is responsible for male courtship behaviours (Demir & Dickson, 2005; Arbuthnott et al., 2017). Sexual dimorphism has been thoroughly studied through differential gene expression in males and females and how this leads to observable sex differences. The role of the few Y-linked genes of *Drosophila* is also fairly well studied, and has established the Y as being primarily involved in male fertility, but not in establishing sexual dimorphism *per se*. Our study focuses on the lesser-explored aspect of sexual dimorphism: the presence of alternative splice forms in male and female tissues. The extent to which alternative splicing contributes to morphological differences between sexes is relatively unexplored, partly due to the predominance of short reads in RNA sequencing, which complicates accurate splice form detection.

1.4 Studying Alternative Splicing

Alternative splicing represents not only a different expression form from the same gene but also plays a direct role in regulating gene expression (Baralle & Giudice, 2017). Analogous to the importance of studying gene expression for understanding cellular function, the investigation of alternative splicing is crucial for shedding light on the normal operation of cells and tissues, as well as for identifying issues in cells and tissues. As reviewed by Salz (2011), Salz & Erickson (2010), Sawanth et al. (2016), MacDougall et al. (1995), and others (e.g., Telonis-Scott et al., 2008), alternative splicing is vital for sex determination and maintenance in dipterans, a topic explored in this thesis.

Traditionally, studies on quantitative alternative splicing have utilized short reads, focusing on the number of exons or introns used by each gene. While short reads provide a broad perspective on the use of each gene part, they fall short in explaining how these parts are arranged. This research introduces a new pipeline that identifies individual splice forms from long reads, allowing for the simultaneous identification of any gene exon sequence and the

specific splice forms that it is included in. This approach enables the quantification of usage and the number of exons and introns per gene. Figure 2 shows how long reads can offer more precision than short reads in identifying alternative splicing due to being able to sequence the entire mRNA at once, skipping the need to reassemble it form fragments, as occurs with short read technologies.

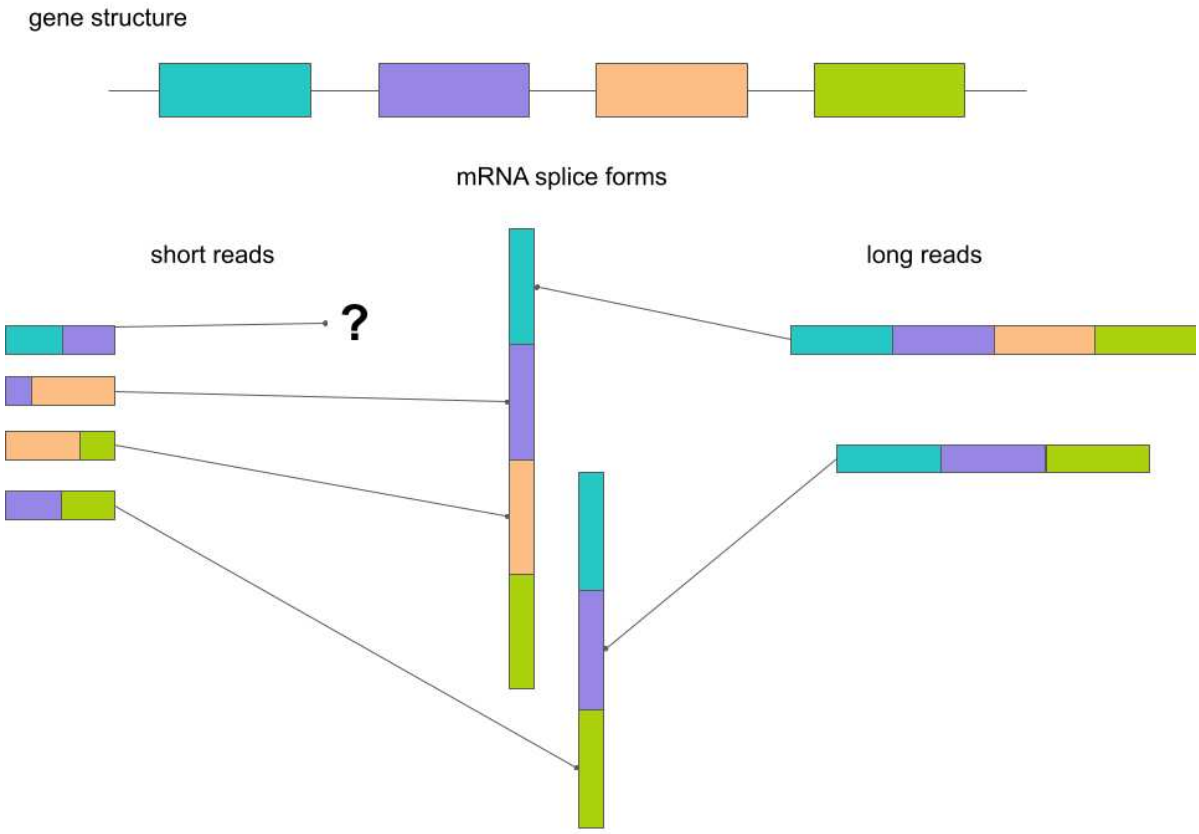

Figure 2. Differences in the use of short and long reads concerning alternative splicing. *In the figure, we can see how short reads and long reads can be used to access alternative splicing. On the top part, we see the gene structure, under it in the middle we can see two mRNA splice forms for that gene. At the left side we see reads obtained from short reads and how they can be assigned (or not) to each mRNA. At the right side we see the same process but for long reads. With this simple example we can observe that some short reads cannot be correctly assigned to any mRNA, while long reads can assess the splice forms without such issues. Adapted from Park et al. (2018).*

Gibilisco et al. (2016) found that, in gonads, female drosophilids express a broader variety of splice forms, while males express a greater number of genes. They suggested that this may correspond to two different strategies to achieve the required transcript diversity. The use of distinct splice strategies raises questions about the diversity of the transcriptome in tissues other than gonads, and whether the splice forms found in each sex are entirely distinct or subsets of each other.

1.5 Identification and Quantification of Different Types of Alternative Splicing

A single gene can transcribe several distinct mature RNAs through alternative splicing, which involves using different combinations of exons from the same gene. The recognized forms of alternative splicing include mutually exclusive exons, skipped exons, retained introns, alternative 5' spliced sites, and alternative 3' spliced sites. This process contributes to a more diverse transcriptome and proteome. Alternative splicing is common, occurring in approximately 74% of multiexon human genes (Yeo et al., 2004). In *Drosophila*, the proportion is smaller, estimated between 30% to 58% of genes producing more than one mRNA splice form (Gibilisco et al., 2016; Brown et al., 2014).

As illustrated in Figure 3, there are various types of alternative splicing. Alternative First Exon (AFE) and Alternative Last Exon (ALE) refer to situations where a transcript's first or last exon differs between splice forms. AFE and ALE are distinct from Alternative 3' Splice Site or Alternative 5' Splice Site, where either the 3' or 5' end of an exon shifts significantly in each splice form (using a threshold of 20 bp, as per the method in our study, more than double that used in Gibilisco et al., 2016). Skipped Exons occur when an exon is included or excluded from its consecutive constitutive exons, leading to multiple splice forms. Mutually Exclusive Exons are similar to Skipped Exons, but two consecutive exons are rarely present in the same splice forms. Finally, Retained Introns (or Intron Retention) involve introns remaining in the mature splice form.

In humans, the brain, liver, and testis exhibit the highest proportion of alternatively spliced genes (Yeo et al., 2004). In *Drosophila*, findings have been inconsistent across species. An early study in *D. melanogaster* found that the nervous system shows the highest transcriptome diversity (Brown et al., 2014). On the other hand, Gibilisco et al. (2016) found that in *D. pseudoobscura*, *D. miranda*, *D. nasuta*, and *D. albomicans*, ovaries and spermatheca have the highest proportion of alternatively spliced genes, while testes show the lowest proportion. Gibilisco et al. (2016) suggested that ovaries and testes employ unique strategies for transcriptome diversity. However, methodological differences across studies and tissues complicate any possible comparisons. For example, much of the tissue-specific data from Brown et al. (2014) originated from neural tissues, which may skew the findings towards higher transcriptome diversity in these tissues.

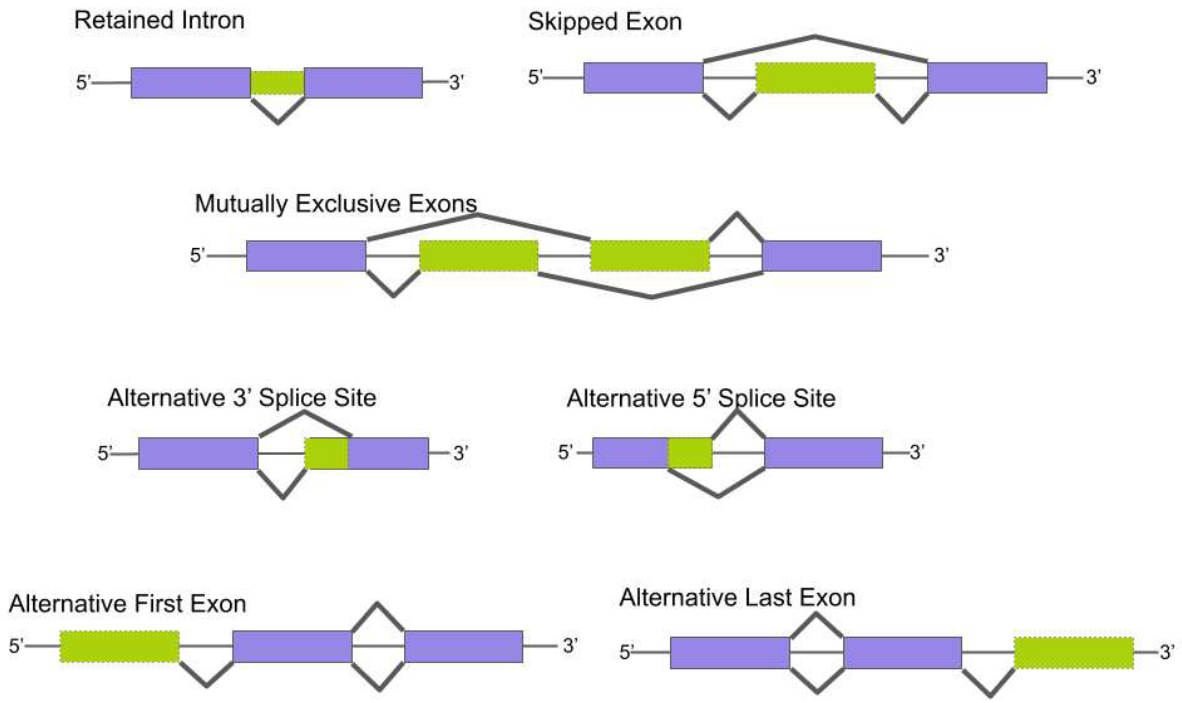

Figure 3. Types of alternative splicing. *Different types of splicing are summarised in the figure. From top left to bottom right, the alternative splicing mechanisms are: Retained Introns (RI), in which an intron is not spliced out of the mature mRNA; Skipped Exon (SE), in which an exon in the sequence is skipped and removed from the mature mRNA along with the bordering introns; Mutually Exclusive Exons (MEE), in which mRNAs tend to have either one or the other exon with the rest of the exon sequence unchanged; Alternative 3' Splice Site (A3SS), where an exon is cut short on its 3' end; Alternative 5' Splice Site (A5SS), where an exon is cut short on its 5' end; Alternative First Exon (AF), in which the first exon in the gene can be skipped so that the mature mRNA actually starts from another exon while the rest of the mRNA is the same; and Alternative Last Exon (AL), where the last exon differs between splice forms while the rest of the mRNA is the same. Inspired by Park et al. (2018).*

The discrepancies observed across *Drosophila* species might be due to variations in the number of genes undergoing alternative splicing or in the frequency of splicing per gene. Mammalian transcriptomes, being more thoroughly characterised, might also contribute to detection biases between these groups. In short, both groups exhibit tissue-varied frequencies of alternative splicing (Telonis-Scott et al., 2008), but whether this reflects the true prevalence of splicing or biases in sampling and/or methodology has yet to be systematically tested.

1.6 Determinants of Alternative Splicing

Alternative splicing is vital for cell functioning and gene expression, and understanding factors that impact the extent of alternative splicing a gene undergoes is crucial. One such factor is chromatin structure. For instance, the presence of a Y chromosome in *Drosophila* impacts the frequency of intron retention in genes throughout the genome, possibly because the presence of a large block of heterochromatin can of itself influence splicing (Wang et al., 2018).

Chromatin state and splicing factors can also impact alternative splicing directly, based on the chromatin configuration of each cell, highlighting a complex regulatory network for splicing (Naftelberg et al., 2015). Variations in histone patterns as well as post-translational modifications of the pre-mRNA can influence gene expression and alternative splicing outcomes depending on their effects on chromatin structure and the recruitment of specific factors, since introns and exons have different chromatin markers and introns are commonly excised from the pre-mRNA during transcription (Shukla & Oberdoerffer, 2012; Moreno et al., 2015). Histone modifications are proposed to act as a mechanism for fine-tuning alternative splicing, particular during co-transcription and altering rates of elongation (Luco et al., 2010; Agirre et al., 2021). Additionally, increased chromatin accessibility has been linked to facilitating intron retention in specific cell states (Petrova et al., 2022). Different chromatin histone modifications influence the RNA polymerase II elongation rates as well as chromatin accessibility, which can lead to intron retention or the usage of specific splice sites depending on the gene (Brown et al., 2012; Luco et al., 2010; Guo et al., 2014).

The sex and tissue that genes are expressed in have also been shown to play critical roles in determining which splice forms are used (Gibilisco et al., 2016). The differential splicing between males and females of key genes of the sex determination pathway of *Drosophila* (as shown in Figure 13) has been found to influence the expression of downstream genes, supporting the idea that alternative splicing can lead to important functional differences in the two sexes (Telonis-Scott et al., 2008; Salvemini et al., 2011). While the most extensive analyses have been performed in mammals and model organisms such as *Drosophila*, similar patterns have been observed in other clades, suggesting that sex-specific splicing likely plays an important role for sexual differentiation throughout the animal kingdom. In particular, Rogers et al. (2020) found extensive sex-specific alternative splicing in birds, specially in gonads. Investigations in the Chinese horseshoe bat revealed strong tissue effects on alternative splicing between sexes in brains and livers (Chen et al., 2023). Chen et al. (2023) found a correlation between gene expression and alternative splicing of the same genes in brains, but not in livers, and proposed that more complex tissues may be more reliant on complementary mechanisms to create sexual dimorphism. Moreover, sex-specific expression, alternative splicing, and genome methylation have been identified in a true bug *(Planococcus citri*, order Hemiptera), where males and females not only occupy different niches, but also exhibit different body plans, since females are neotenous and males go through full metamorphosis (Bain et al., 2021). In humans, alternative splicing is especially prevalent in the brain and testes compared to other tissues, indicating once again that there are important tissue-specific differences in splicing patterns (Grange et al., 2010).

Finally, the age of genes is also expected to influence the amount of alternative splicing that they undergo for two main reasons. First, many new genes arise through retropositions events, i.e., an mRNA molecule is reverse transcribed into DNA and re-inserted into the genome, creating a new intronless gene which cannot undergo many types of alternative splicing, such as exon skipping (Keren et al., 2010; Koterniak et al., 2020; Tikhonov et al., 2018; Olthof et al., 2021). Genes that arise *de novo* from non-coding DNA also initially do not have introns (or only a few), again limiting the different transcripts that can be produced. Second, if selection shapes alternative splicing (see next section) young genes may not have had enough time to undergo such functional diversification. While various mechanisms modulating splicing have been identified, they have typically not been systematically

quantified within a single dataset, such that their relative contributions are still to be fully understood.

1.7 Evolutionary Pressures Shaping Sex-Specific Alternative Splicing

As we have seen in the previous sections, alternative splicing is prevalent and tissue/sex-specific, suggesting that it may play a key role in functional diversification. This raises the question of what neutral and/or selective dynamics have allowed for such a diversification to arise. On the one hand, there is evidence that some splicing variants appears to be spurious and more prevalent in populations with small effective population sizes, suggesting that not all splicing variation have been selectively favored (Bénitière et al., 2024). On the other hand, new splicing variants are expected to be selected for under some circumstances. In particular, antagonistic pleiotropy occurs whenever a gene has more than one function and the two interfere with each other and may be relieved by optimizing different transcripts for each molecular function (Guillaume & Otto, 2024; Mank et al., 2008). Males and females also face distinct reproductive needs, leading to different selective pressures acting on the two sexes. At the genetic level, this can result in intra-locus sexual antagonism, the situation where an allele is advantageous for one sex but detrimental for the other. Alternative splice forms can mitigate the negative effects of sexually antagonistic alleles, providing each sex with optimal protein forms. Studies indicate that sexual antagonism is common (Connallon & Clark, 2011), which may contribute to the prevalence of sex-specific splice forms. However, it is important to consider that mechanistic modulation of splicing could lead to differences in male and female splice forms, as suggested by Wang et al. (2018), who found that the presence of a Y chromosome influences splice form production, possibly due to chromatin structure changes.

Given this expectation that sex-specific selective pressures may promote sex-specific splicing, much work has focused on characterizing sexual dimorphism in splicing, e.g. Blekhman et al. (2009) and Rogers et al. (2020). Blekhman et al. (2009) found that alternative splicing exhibits sex-specific and lineage-specific changes in splice form expression in primates. Rogers et al. (2020) found increased rates of protein sequence evolution at genes with sex-specific splicing in birds and suggested that sex-specific selection influences the evolution of alternative splicing in this group, resulting in the creation of distinct sex-specific protein isoforms. Overall, these studies underscore the intricate relationship between alternative splicing, tissue- and sex-specific differences, emphasizing how these factors collectively contribute to the diversity of gene expression patterns in different sexes and lineages.

Further indirect evidence suggests that sexual antagonism may influence the evolution of sex-specific splicing. First, sexual antagonism is believed to be most prevalent in gonads (Kirkpatrick & Guerrero, 2014; Innocenti & Morrow, 2010), implying that most sex-specific alternative splicing should occur in these tissues if it arises to resolve sexual conflict. The gonad is indeed the tissue with the most differentiated splicing landscape between males and females (Gibilisco et al., 2016; Lang et al., 2019; Xu et al., 2022), although the fact that testes and ovaries are essentially different tissues can also contribute to this pattern. Second, under some evolutionary scenarios, the X-chromosome is expected to accumulate more sexually antagonistic alleles than the autosomes (Rice, 1984; Gibson et al., 2002; Kirkpatrick & Guerrero, 2014). If sexual antagonism is a key driver of sex-specific splicing, we expect X-linked genes to have more sex-specific splice forms than autosomal genes. FlyBase data (Larkin et al., 2021) shows that X-linked genes have more annotated transcripts per gene than

autosomes, and Karlebach et al. (2020) found the same pattern of more alternative splice forms per gene in the X-chromosome of humans, suggesting this pattern is consistent in different taxa. Figure 4 shows the distribution of the number of annotated splice forms per gene for genes on the different chromosomes using FlyBase data (Larkin et al., 2021), and uncovers an excess of splicing variants for both the X and the 4th chromosomes. The fourth chromosome (dot chromosome/ Muller element F) is currently an autosome in *D. melanogaster* but was identified as a sex chromosome in more basal species of the group (Vicoso & Bachtrog, 2013), having undergone similar evolutionary pressures as the X-chromosome for most of its existence, which may explain its high prevalence of splice forms per gene along with the X-chromosome. Whether these annotated transcripts represent shared splice forms or sex-specific variants remains unclear, but if sex-specific, this could support the role of sexual antagonism in their evolution.

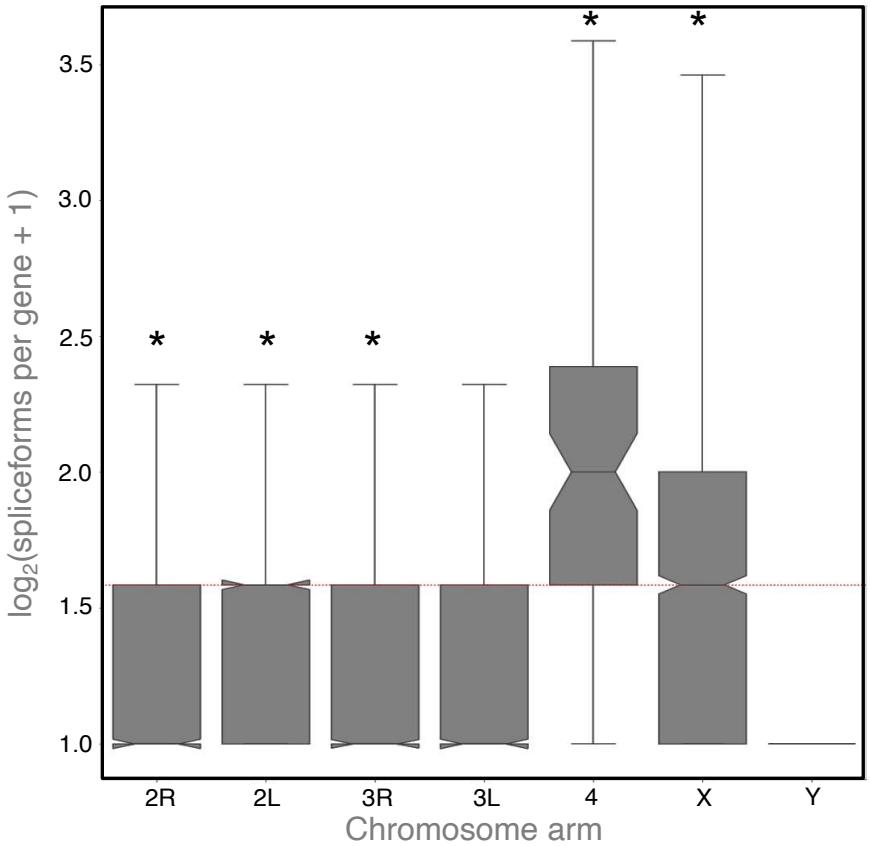

Figure 4. FlyBase data shows that gene location impacts the number of splice forms per gene. *Here we see how, the chromosome arm (equivalent to Muller elements) impacts the number of unique splice forms per gene. The X and 4th chromosomes have more transcripts per gene on average than other chromosomes, it is expected that the X has more splice forms and as the dot chromosome has been theorised to have been an X-chromosome before in this groups' evolution (Vicoso & Bachtrog, 2013), it is in line that the dot chromosome also has more splice forms per gene. Gene data from FlyBase database (Larkin et al., 2021). Asterisk (*) indicates p-value < 0.05 in a Wilcoxon test between the chromosome it is above, and the set of all other chromosomes combined.*

1.8 Summary

The study of alternative splicing and its role in the complex mechanisms of sex determination represents a pivotal area of research within the field of genetics, particularly in model organisms like *Drosophila melanogaster*. Alternative splicing allows the production of diverse protein isoforms from a single gene and can play a crucial role in expanding the functional repertoire of the genome. It is especially significant in the context of sex determination, a tightly regulated process that is essential for the development and fertility of an organism. In *D. melanogaster*, this process is initiated by counter genes that initiate the sexual differentiation pathways. This thesis aims to dissect the intricate patterns of alternative splicing across different tissues and investigate the recruitment of counter genes in the sex determination process of *D. melanogaster*. By addressing these aims, we seek to enhance our understanding of the genetic underpinnings that govern sex determination, providing insights into the broader implications for developmental biology and genetics.

In conclusion, this thesis has embarked on a comprehensive exploration of the patterns of alternative splicing in different tissues and the recruitment of counter genes for sex determination in *D. melanogaster*. Chapter 2 delves into the complexities of alternative splicing, illustrating its variable patterns across tissues and shows a new approach to identifying and quantifying splice forms usage. Chapter 3 focused on the recruitment of counter genes involved in sex determination, shedding light on the delicate balance of genetic factors that dictate sexual development. In Chapter 4, we aimed to synthesize these findings, drawing connections between the mechanisms of alternative splicing and the recruitment of counter genes in sex determination. This endeavour not only contributes to our understanding of genetic regulation in *Drosophila* but also highlights the broader implications of alternative splicing and gene regulation in developmental biology. Through this work, we pave the way for future research to further unravel the complexities of genetic control mechanisms, with the hope of uncovering novel insights into the molecular orchestration of life.

2 Alternative Splicing in Drosophila melanogaster

2.1 Summary

Alternative splicing plays a crucial role in sex determination in *Drosophila melanogaster*, as exemplified by the fact that many genes of the sex determination cascade, such as *Sex-lethal* (*Sxl*), exhibits different male and female splice forms (Park et al., 2004). However, it is still unclear how much of a role sex-specific splicing plays in the downstream processes controlling sexual dimorphism. A reason for this is methodological, as until recently the detection of splicing variants was limited by the prevalent use of short reads for RNA-sequencing. The advent of long read RNA sequencing provided the opportunity to systematically characterize the various existing splice forms and quantify their usage in different tissues/cell types. For instance, since long reads provide more complete sequences of exons and introns compared to short reads, they have allowed for the improved detection of splicing variants in humans (Park et al., 2018). However, detailed comparisons of male and female splicing variants in Drosophila have only been performed using short read sequencing (McIntyre et al., 2006; Gan et al., 2010; Gibilisco et al., 2016). Here, we use PacBio long read RNA sequencing to produce a thorough atlas of male and female splice forms in gonad and two somatic tissues, allowing us to uncover more or distinct splice forms than previously reported. Additionally, in this chapter we develop a novel approach for assessing splicing incidence, using specific notations for each splice form instead of the traditional PSI (Percentage Spliced In). This splice-specific notation reveals new information and reduces the likelihood of misrepresenting the impact of any one exon or intron in constructing specific splice forms.

2.2 Introduction

Alternative splicing is essential in diverse groups, including plants (Syed et al., 2012; Seo et al., 2013) and mammals (Huang et al., 2022), particularly in humans, where it is implicated in diseases such as autism, cancer, and muscular dystrophy (Fu & Ares, 2014; Oltean & Bates, 2013; Soto et al., 2019). The fact that many of these diseases have a different prevalence in the two sexes highlights the importance of understanding how this process operates differently between males and females, a factor often overlooked in medicine (Almqvist et al., 2007; Peters et al., 2019). Alternative splicing can also be closely tied to regulatory functions, such as during sex determination in *Drosophila* (Cornelius et al., 2021; Modrek & Lee, 2002). Over 20% of multi-exon genes in *Drosophila* species show alternative splicing, according to Gibilisco et al. (2016), while in humans, this percentage exceeds 90% (Pan et al., 2008; Wang et al., 2008). Alternative splicing is also prevalent in birds, particularly in gonads (Rogers et al., 2020), and in plants, as evidenced in *A. thaliana* (Jabre et al., 2021; Martín et al., 2021; English et al., 2010).

Different methods for studying alternative splicing can yield varied results regarding splice form identification and prevalence (Li et al., 2017). Microarray was the initial method used to identify alternative splicing in gene expression (Berget et al., 1977; Chow et al., 1977), but it has limitations, including only detecting known genes, alleles, or splice forms. For instance, a *Drosophila* study using microarrays identified alternative splicing in just 828 out of 2,479 known multi-transcript genes, greatly underestimating its prevalence (McIntyre et al., 2006). The advent of next-generation sequencing (NGS), e.g., RNAseq, has facilitated the identification of new genes, alleles, and splice forms (Pollard et al., 2018; Mane et al., 2009).

RNA sequencing does not rely on complementarity to known sequences for identifying new transcript variants in samples. RNA-seq short reads enable new gene and allele identification, but splice form identification is limited by read length. Computational methods have made the quantification of differential exon usage and alternative 5'/3' UTRs from short reads easier and more reliable (Adamopoulos et al., 2018), but identifying the specific combination of exons found in each transcript remains a challenge (as shown in Figure 2). Conversely, SMRTseq yields longer reads that typically encompass the full transcript, simplifying the identification of alternative splicing forms. In Figure 2, we can see how short read sequencing can lead to doubts when quantifying or recovering splice forms due to their only being fragments of the full mRNA, while long reads provide a sounder way to identify it. Li et al. (2017) showed that although RNAseq provides deeper sequencing, SMRTseq can identify more alternative splicing forms in strawberries (33.48% vs. 57.67%). In humans, different splice forms can lead to disease susceptibilities, and certain variants are typical in specific populations (Park etz al., 2018), underscoring the importance of using long reads for accurate splice form identification.

Drosophila melanogaster, a widely used model organism, has been the subject of many alternative splicing studies (Telonis-Scott et al., 2008; Breitbart et al., 1987; Wang et al., 2018; Park et al., 2018; Li et al., 2017; Stegeman et al., 2018). In this species, alternative splicing is vital for sex determination, with sex-specific splicing patterns observed in various genes (Telonis-Scott et al., 2008). The primary sex determining gene, *Sxl*, has distinct splice forms in males and females that initiate the sex determination and dosage compensation genetic cascade (Ilik et al., 2013; Lucchesi & Kuroda, 2015). Alternative splicing also plays a role in gene regulation and expression (Smith et al., 1989; Zhou et al., 2021; Zheng et al., 2019; Kufel et al., 2022), highlighting its critical nature to sex determination and dosage compensation. However, these studies predominantly used RNAseq with short reads, leaving room for improvement through long-read technology. Long reads can identify splice forms used in different tissues and discover new forms for known genes.

The role of alternative splicing in tissue regulation and diversification between organisms therefore remains a subject of debate. In this chapter, we propose to expose differences in splice usage between tissues and sexes in *Drosophila melano*gaster, which can serve as a starting point to better understand its relation to evolution and tissue differentiation. Specifically, long reads are used to annotate each identified splice form, allowing for the detailed observation of the splice forms utilized in various tissues of the two sexes, in particular male and female gonads, heads, and midguts. We recover a higher prevalence of splice forms in ovaries, similar to what was found in other Drosophilid species by Gibilisco et al. (2016). Importantly, we show that this is due to the presence of many ovary-specific transcripts, something that could not be detected with previous datasets. Furthermore, we compare the prevalence of splice forms per gene in different age groups and chromosomal location and find that older genes as well as genes present on the X-chromosome have more splice forms per gene.

2.3 Methods

2.3.1 Fly Care

Flies were maintained in chambers at a constant temperature of 25˚C under a 12-hour light-dark cycle. Canton-S *Drosophila melanogaster* specimens were kept in these conditions and transferred to new vials weekly. Canton-S flies were kindly provided by Dr. Daria Siekhaus at ISTA. Virgin flies were isolated for dissections by removing all adults from the vial and waiting for up to 5 hours for newly emerged flies.

Vials contained a medium supplied by the Vienna BioCenter Facility, enhanced with dry yeast prior to transferring flies into fresh vials. The medium included a blend of cornmeal and soy flour with agar, molasses, malt extract, and preservatives (solutions of Nipagin 15% and propionic and phosphoric acids added during medium preparation).

The parental generation was housed in vials at a low population density (approximately 10-15 mixed-sex flies per vial) for a week before their removal. Virgin flies were then selected from these vials, approximately two weeks after the initial addition of flies. To ensure selection of virgin flies, adult flies were first removed from the vials and vials were left for 5 hours. Newly emerged flies were then collected and sexed under CO_2 anaesthesia, with males identified by their sex combs and females by the absence of these combs. About ten same-sex individuals collected simultaneously were housed in the same vial for 4 to 6 days until dissection.

2.3.2 Dissections

For the creation of a long-read splice form database of *D. melanogaster*, 180 flies were dissected for each replicate, with two replicates per tissue (heads, midguts, and gonads) and sex (males and females). Male and female whole-body samples comprised 6 pooled individuals each. All samples were obtained from virgin flies aged 4 to 6 days post-eclosion from Canton-S *D. melanogaster* lineage.

Heads and gonads were dissected from the same flies on the same day, alternating between batches of 10 males and 10 females. This alternation aimed to normalize any potential time-of-day effects during dissection sessions. Midguts were dissected in a similar fashion. Dissections were conducted in a 10% PBS solution, and tissues were stored in Trizol at -80°C until RNA extraction and subsequent sequencing at the Vienna BioCenter.

2.3.3 RNA Extraction

Ariana Macon, a lab technician at the Vicoso Lab, conducted the RNA extractions. Tissue homogenization for RNA extraction was performed using the Qiagen Tissue Lyser II at a frequency of 30Hz for 2 minutes, with 3mm beads in 1.5mL Eppendorf Safe-lock LoBind tubes as per intructions from the manufacturer.

Total RNA extraction was carried out using the Bioline Isolate II RNA Mini Kit. Homogenized tissue was incubated in ISOL Reagent for five minutes at room temperature for complete dissociation of nucleoprotein complexes. The Phase Lock Gel-Heavy tubes were briefly pre-spun, and the homogenate was added, followed by a 5-minute room temperature incubation. Chloroform was then added to facilitate phase separation.

After centrifugation, the upper aqueous phase containing RNA was transferred to new RNase-free tubes, and RNA was precipitated with ethanol. The RNA was then bound to a mini-spin column and centrifuged.

RNA quality was assessed using the Bioanalyzer RNA Nano before sequencing. Samples were stored at -80°C after being eluted with water, following the protocols of the respective kits and instruments used.

2.3.4 Long Read Sequencing

2.3.4.1 PacBio SMRTseq

Long-read sequencing of whole-body, heads, gonads, and midguts utilized PacBio technology at the Vienna BioCenter. The raw files were processed, demultiplexed, and prepared for sequencing. PacBio RNA IsoSeq sequencing, which sequences full RNA transcripts, was employed to study transcriptome variations, including alternative splicing, isoform diversity, and gene expression levels.

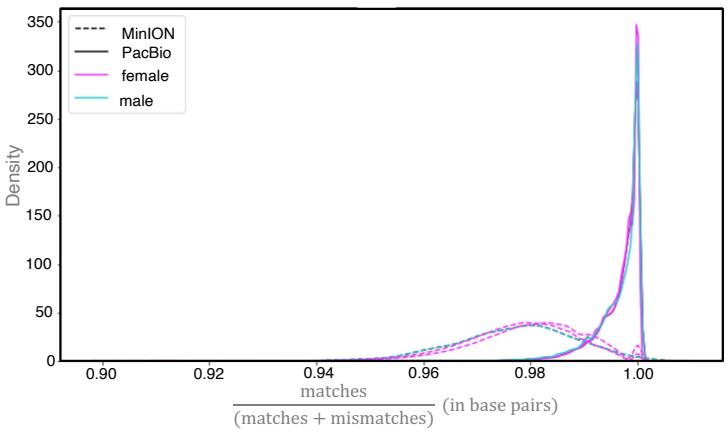

Figure 5. Alignment match to D. melanogaster genes. *Proportion of reads that matched D. melanogaster genes' longest CDS from FlyBase related to total read length. We can see the alignments are quite high, being higher than 0.95 for PacBio, which is one of the reasons we chose this method for use in this study.*

2.3.4.2 MinION Oxford Nanopore

Whole-body samples were also sequenced using the Oxford Nanopore MinION platform. Ariana Macon performed the MinION sequencing. MinION Direct RNA Sequencing directly sequences native RNA molecules using nanopore technology. The RNA sample is loaded onto a flow cell containing nanopores, with the passage of RNA molecules through these pores disrupting the electric current, allowing for real-time RNA sequence identification.

2.3.5 Computational Methods

All scripts and pipelines used in this project are available at github.com/juliaraices/dmel_analysis/ .

2.3.5.1 Long Read Qualities

Whole-body replicates were analysed using both PacBio® and MinION® sequencing to assess the quality of long reads. In order to access this, we compared how our reads aligned to the longest CDS from FlyBase of every gene. Figure 5 shows the proportion of matched base pairs

over the ratio of matched and mismatched base pairs in the read. In Figure 5 we can see that PacBio shows a higher proportion of reads with a ratio of over 0.98 for both female and male samples. Furthermore, Table 1 shows that PacBio has overall longer reads than MinION, which most likely is due to PacBio recovering on average longer length reads, while MinION might be either missing the longest mRNAs or not sequencing the full length of the mRNA. Based on the comparison of read lengths and matches to FlyBase genes, PacBio was chosen for further tissue sequencing.

Table 1. MinION and PacBio read quality comparison. *A notable difference is observed in the number of reads between males and females in MinION, attributed to the sequencing order in the same nanopore cell: males were sequenced first for 24h, followed by females for 40h. PacBio displays consistent read numbers between sexes and larger mean and median read lengths. 1- reads were aligned against CDS database for D. melanogaster from FlyBase; 2- where similarity is calculated as [number of read bases matched to reference]/[length of read in base pairs].*

	MinION Females	*MinION Males*	*PacBio Females*	*PacBio Males*
Mean read length	877	834	2 098	1 850
Median read length	754	707	1 885	1 546
Maximum read length	8 375	17 039	17 398	21 455
Number of reads mapped to FlyBase[1]	1 456 969	702 332	426 366	183 435
Total number of reads	1 934 569	906 318	512 226	225 354
Median similarity to FlyBase genes[2]	0.97864	0.98096	0.99821	0.99834

2.3.5.2 Splice Form Identification Script

Raw sequencing data were demultiplexed and filtered, with BLAT used to align reads to gene exons and introns. A new script pipeline processed this output, generating unique annotations for each identified splice form. This method allowed for the detailed identification of different splice forms in each sample, although it lacked the ability to identify minor modifications in 3' and 5' splice site events.

Figure 6 shows a representation of our pipeline as a flow chart. Following Figure 6 we can see that first all exons or introns matches to the same read are selected and sorted according to their location in the read. From that the name of each exon or intron is written sequentially after the gene id that was attributed to that read. The presence of more than one gene assigned to the same read (chimeric splice forms) or of an intron raise flags that are added to the annotation of each read. For analyses comparing splice forms those transcripts were used, but for analyses comparing gene usage chimeric splice forms were removed to be more conservative and precise in our analyses. For all analyses only splice forms that appeared more than once in the set of all replicates and samples were used to avoid possible artifacts.

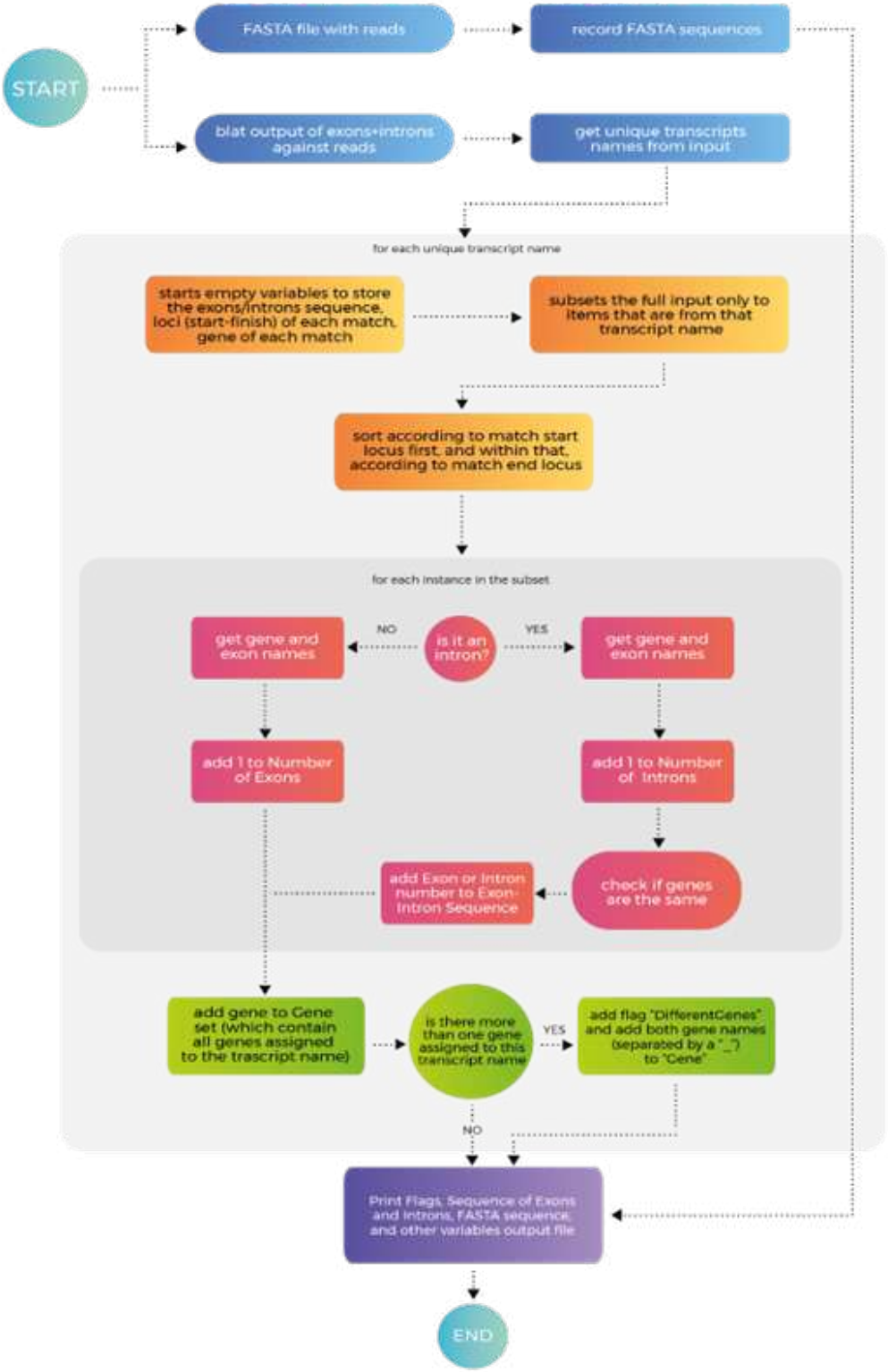

Figure 6. General algorithm for script that creates an identifier for each transcript. *General algorithm for our script to create an alphanumeric string for each transcript consisting of the exon number sequence of each transcript. From BLAT outputs of our reads against exons and introns from FlyBase our script selects all matches for a given read and sorts the matched exons and introns so that they are in the order they appear in the read. From that a string/word is created with the gene to which the read was matched followed by the exon/intron sequence. Flags are raised if more than one gene is matched to the same read or if there is an intron.*

2.3.5.3 Gene Age

Gene ages were determined based on shared gene groupings from Zhang et al. (2010), with genes shared between *D. melanogaster* and the *Drosophila* subgenus considered old (over 63 million years) and those shared with closer groups or exclusive to *D. melanogaster* were deemed new (under 63 million years). Figure 7(A) shows a cladogram where we see different *Drosophila* species, with *D. melanogaster* highlighted in darker colour. In Figure 7(A) we see the dichotomy of *Sophophora* and *Drosophila* subgenus, which is where we set the distinction of old and new genes, with new genes being those that arose after this diversion and old genes those that arose before it.

2.3.5.4 Gene Location

We considered the effects of chromatin state and chromosomal location on alternative splicing per gene, using data from Milon et al. (2014) and Kharchenko et al. (2011) for chromatin analyses.

2.3.5.5 Gene Ontology

Genes with splice forms exclusive to each tissue we selected for Gene Ontology (GO) enrichment analyses, and chimeric genes were discarded. Gene Ontology (https://geneontology.org/; Thomas et al., 2022) was used for GO analyses using the default setting of the platform for *D. melanogaster* as both input and analysis species.

2.3.6 Statistics

Statistical analyses between groups utilized the Wilcoxon test, seeking a p-value of 0.05 or lower to indicate significant differences in data distributions. Python scripts performed the tests and are accessible on GitHub.

2.4 Results

We first identified from the PacBio RNA long reads which splice forms are present in different *Drosophila melanogaster* tissues. A total of 340,610 unique transcript annotations were detected from 13,595 genes; however, only a third (111,659) of splicing variants were represented by at least two transcripts and kept for further analyses (to avoid contamination of the analyses by spurious splice forms). Because the RNA sequences were not stranded, some forms might have been miscalculated due to different annotations for the forward and reverse strands. Many genes (2,992) were only represented by a single transcript, such that we had no power to detect alternative splicing. To obtain a better understanding of how limited we were by expression to quantify variants, we investigated the relationship between gene expression and total variants identified. Figure 7(C) shows that the number of reads (a proxy for expression) are highly correlated to the number of splice forms per gene. There was a correlation between gene expression and the number of splice forms, observed both in whole-body expression and in individual tissue samples. The simple fact of having more reads could lead to more splice forms being found. Some caution should be used when interpreting these expression values, as the reduced number of reads in PacBio experiments makes for noisier estimates of expression. However, recent studies affirm PacBio is reliable for measuring expression (Roy et al., 2021). Furthermore, Figure 20(A) in Appendix 1 shows a scatter plot of the number of reads in whole body versus the expression in TPM (transcripts

per million) from FlyAtlas2 (Leader et al., 2018), where a significant correlation can still be seen. This points to expression being truly correlated to the number of splice forms. Another explanation for this correlation is the higher number of exons found for high expression genes, as we see in the scatterplot in Figure 20.

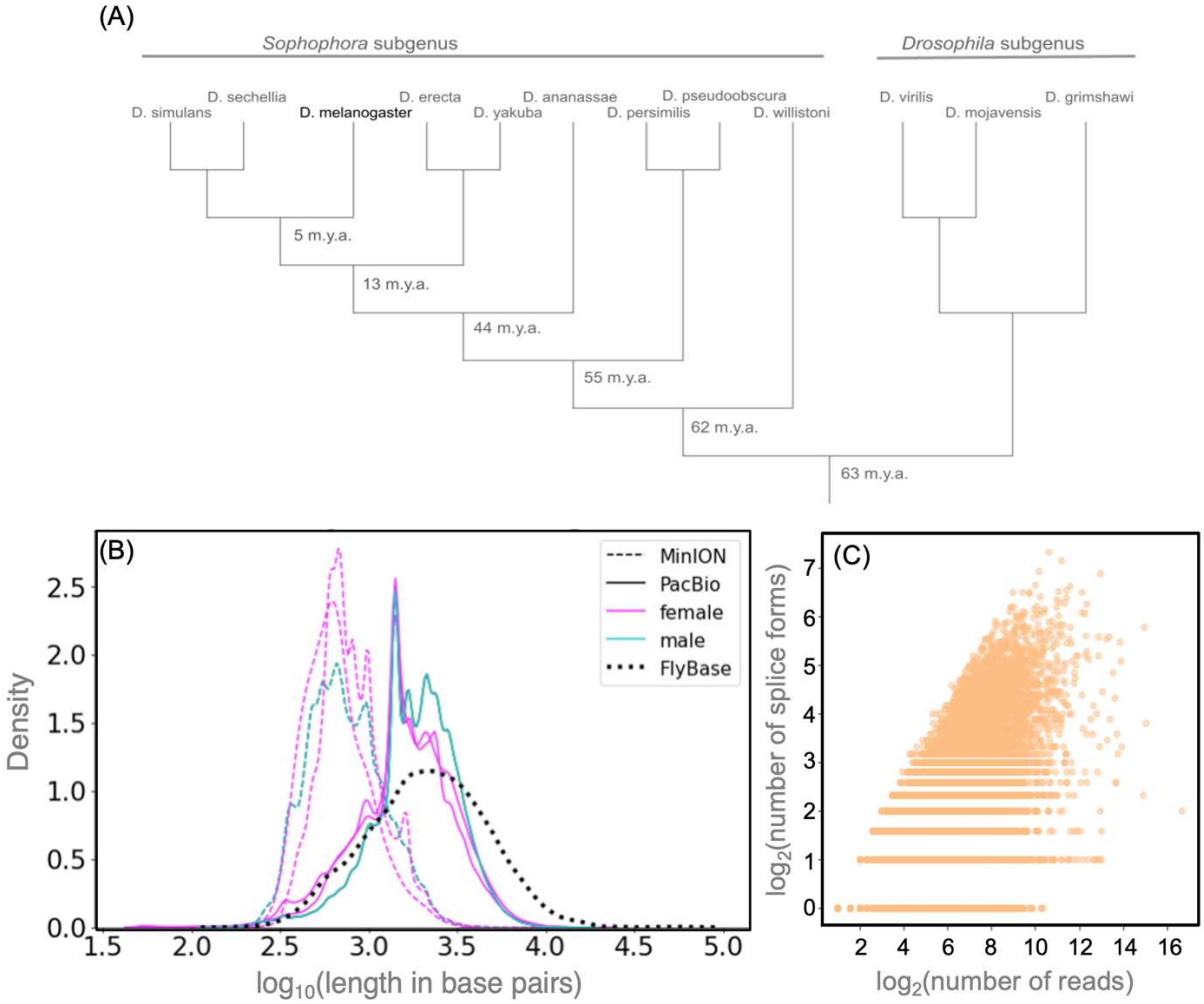

Figure 7. (A) Age determination for protein-coding genes. *"D." stands for Drosophila; and "m.y.a.", million years ago. In darker we can see the position of Drosophila melanogaster in the tree. For our analyses all genes 62 million years or younger were considered young genes. Genes shared between all species in the tree (i.e. 63 million years or older) were considered old genes. Adapted from Zhang et al. (2010).* ***(B) Density curves for male and female long reads using PacBio (full lines) and using MinION (dashe d lines), also shown is the density curve for short read transcripts from FlyBase (dotted lines).*** *All data refers to Drosophila melanogaster. FlyBase database has longer transcripts due to including all currently known splice forms some of which are most likely not represented in our samples. Females have two replicates per technology (in pink), and males have one per technology (in turquoise). PacBio consistently shows higher transcript sizes than MinION.* ***(C) Number of reads vs. the number of splice forms found per gene.*** *Number of reads vs. the number of splice forms found per gene in all tissues. We can see a relatively consistent prevalence of reads for all splice form numbers but as expected the more splice forms found, the more reads were obtained for that gene. It is possible to notice that genes with many reads can fall into any of the splice form number groups showing that probably what we find is not merely an artifact, but a proper signal.*

Most genes (75% or 12,687) had five or more reads, such that common splicing variants should be present in our dataset. The read sizes were close to the annotated transcript lengths of FlyBase (Figure 7(B)), suggesting that whole transcripts were captured, thus allowing for the accurate identification of different transcript forms. We aligned the reads to the genome of *D. melanogaster* and scored what features they overlapped with to check that we had recovered mRNAs (as expected given that the samples were poly-A-selected). Most reads matched coding genes, as seen in Figure 8(A), suggesting minimal artifacts: less than 2,500 reads were assigned to pseudogenes, intergenic regions, or non-coding RNAs in either male or female samples. The comparison of splice forms found in our dataset with the annotated transcripts from FlyBase (Larkin et al., 2021) revealed many novel splice forms, particularly for male and female exclusive splice forms. Figure 8(B) shows the number of unique splice forms found in each tissue and sex, while Figure 8(C) shows the same with a further division between splice forms also present in the FlyBase database and those not. We can see in Figure 8 (B) and (C) most of the splice forms found in our data were not found in FlyBase.

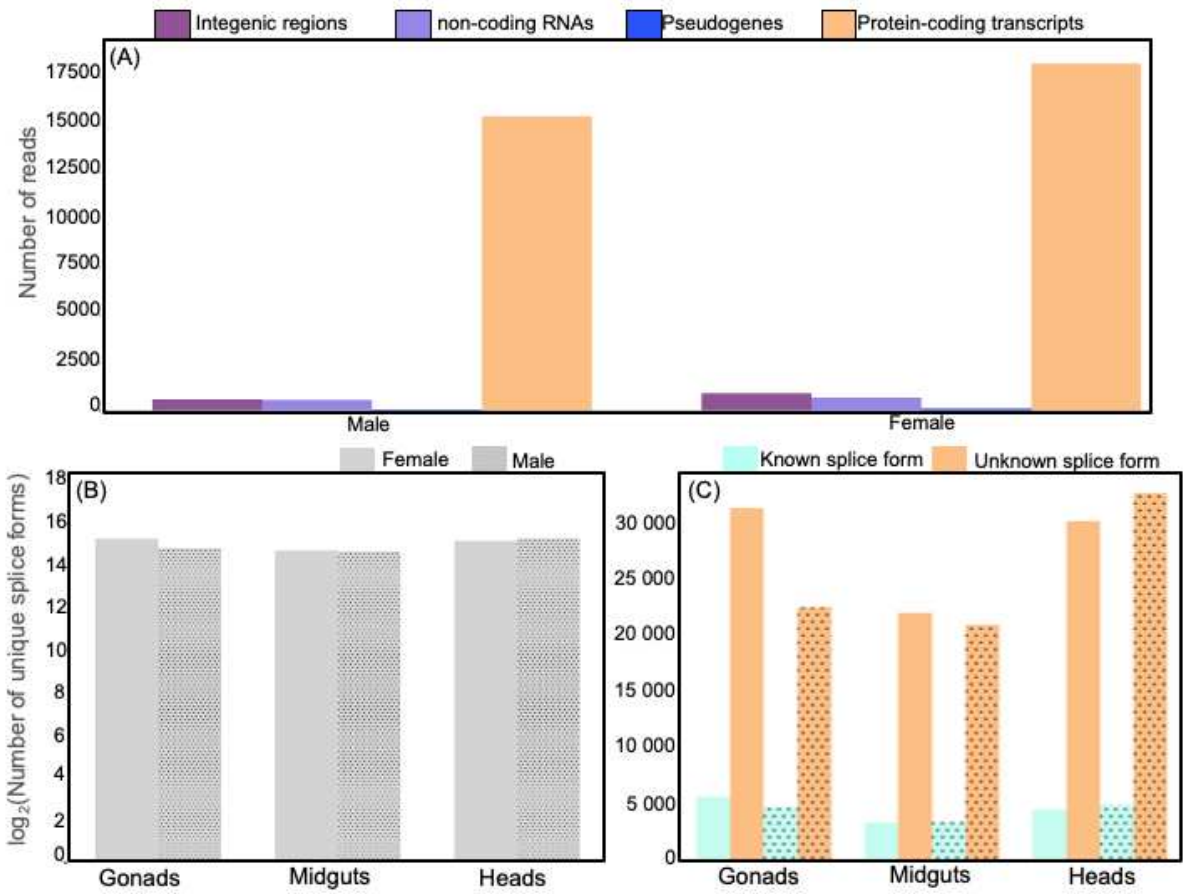

Figure 8. (A) Alignment to coding and non-coding regions from the genome of D. melanogaster. *We can see there is no significant difference in non-coding groups distribution despite the significantly higher number of protein-coding genes. Protein coding genes make most of our sample in both sexes.* ***(B) Unique splice forms found in females and males in each tissue. (C) For male and female splice forms in each tissue the number of splice forms that can be recovered from FlyBase data (known), and those that cannot (unknown).*** *For figures (B) and (C), the same pipeline was used with our long read data and FlyBase transcripts, from which we identified splice forms that were found in ours and in the database from FlyBase, and the ones exclusive to our data set.*

To validate our power to identify male and female-specific splicing variants, the gene *tra*, known to have sex-specific splice forms, we analysed and compared our transcripts with FlyBase annotations. In Figure 9 we see the presence of the female-specific splice form in our female sample, but not in the male one. The presence of known sex-specific splice forms and others in both males and females confirmed the accuracy of the findings. For the comparison of other sex determination genes, see Figures 17 and 18 in Appendix 1.

Figure 10(A) shows that the long read analyses detected over 90,000 unique splice forms, with a majority shared between sexes (62%), and females exhibiting more sex-specific forms. Ovaries showed many tissue-specific forms, aligning with previous studies (Gibilisco et al., 2016). Figure 10(B) shows that most of the female specific forms arise from ovaries, as most ovary specific splice forms are exclusive to that tissue. The heatmap of shared splice forms across tissues and sexes in Figure 11 revealed higher similarities within the same tissue and between male and female samples of the same somatic tissue. The heatmap in Figure 11 is calculated by dividing the number of shared splice forms between column and row tissues divided by the total number of splice forms in the row tissue. The comparison of tissues showed many splice forms expressed across different tissues, despite several being tissue-specific. Figure 19 in Appendix 1 shows a Venn diagram of shared and exclusive splice forms between tissues for females (A) and males (B). Also, in Figure 19 (C, for females) and (D, for males) we see that these numbers of exclusive splice forms has a high correlation to the number of genes expressed, as there are less splice forms where less genes are expressed. Unlike other figures, Figure 19 uses only non-chimeric genes to be more conservative and precise in our analyses.

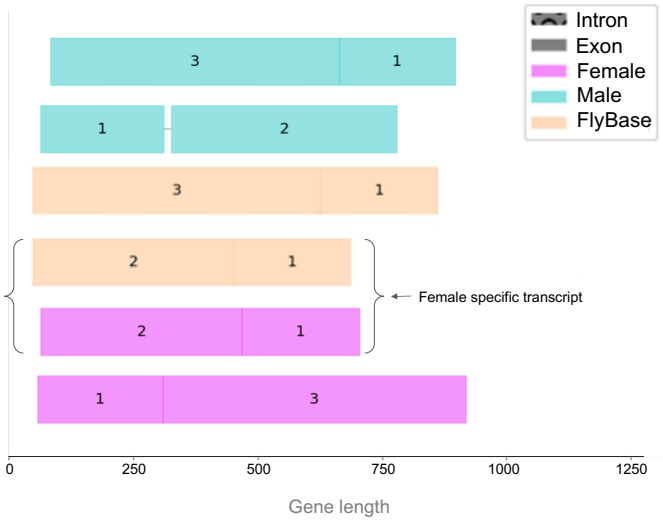

Figure 9. Splice forms of the gene transformer found in males and females. *Alternative splicing of the sex-related genes in male and female Drosophila. Each transcript is labelled with the exon number to which each segment of the transcript was aligned to. Here we can see the known splice forms (orange) and those found in our samples for the gene transformer. Splice forms are aligned to respective gene region.*

Our analyses of FlyBase data had indicated a higher number of splice forms per gene on the 4th (dot) chromosome and the X-chromosome (Figure 4), and we investigated whether this was the case in our data or simply due to a bias in the FlyBase annotation set. Figure 12(C) shows that there is a higher number of splice forms per genes for the X and 4th chromosomes, suggesting that this pattern reflects a biological difference (see discussion). Kharchenko et al. (2011) showed that the chromatin landscape influences gene expression. This may lead to systematic differences in splicing between genes found in different chromatin states, as highly

expressed genes have more chances to undergo alternative splicing. Furthermore, the chromatin landscape itself is known to interact with the spliceosome and shape alternative splicing (Schor et al., 2012; Nissen et al., 2017; Lee & Rio, 2015; Leung et al., 2019; Guo et al., 2014). We therefore combined our annotated splicing variants to chromatin conformation data from Kharchenko et al. (2011) to test for an influence of chromatin state on splicing diversity. The results showed no striking difference in splice form numbers across any single chromatin state from Kharchenko et al.(2011) (Figure 12(A)). However, in Figure 12(B) we see a significant difference between genes that are in open versus close chromatin in the number of splice forms per gene, with genes in open chromatin having significantly more splice forms per gene than those in close chromatin.

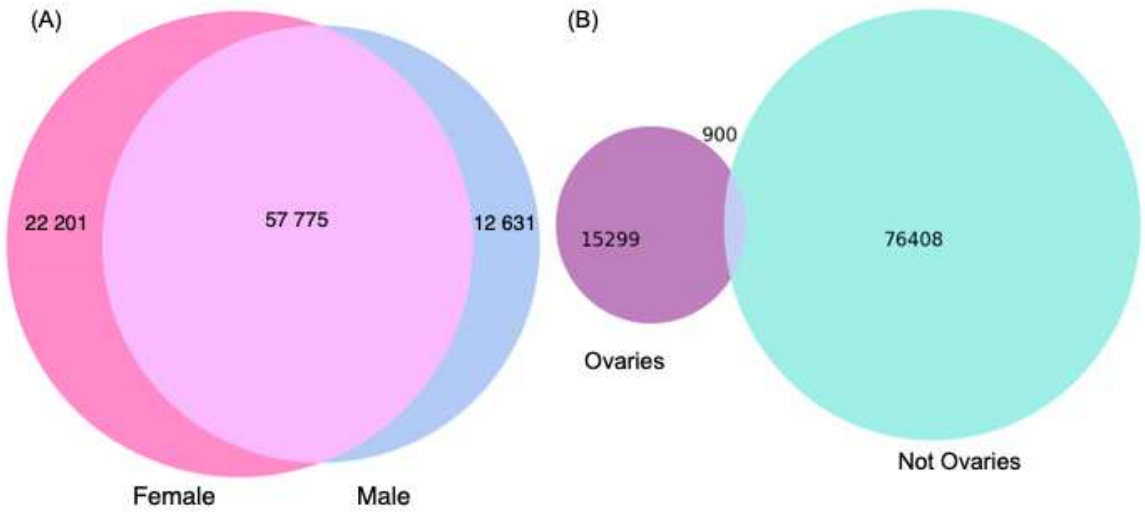

Figure 10. (A) Numbers of shared and sex-specific splice forms found with our method. *The Venn diagram shows that for both males and females most splice forms are shared between the sexes (57775). Females though show a higher proportion of exclusive splice forms (22201) than males (12631). This analyses was done grouping all tissue samples for each sex (heads, gonads, and midguts).* ***(B) Shared and exclusive splice forms between ovaries and all other sampled tissues.*** *When comparing the splice forms present in ovaries with the ones from the combination of all other tissues (male and female heads, and midguts, as well as testis) we see most ovary splice forms are exclusive to that tissue. This shows that most female specific splice forms in our sample are more precisely ovary specific, and only 6902 are female exclusive in the sum of other tissues (all except ovaries, i.e. 22201 female exclusive forms minus 15299 ovary specific ones), versus 12631 in males, as seen in panel (A).*

Older genes exhibited significantly more splice forms per gene than newer genes, as seen in Figure 12(D). This pattern could be attributed to older genes having had more time to evolve the necessary structures for effective splicing. Old genes were defined as those shared between the *Drosophila* and *Sophophora* subgenus, as seen in Figure 7(A).

To gain deeper insights into the possible functional roles and biological significance of the genes with tissue exclusive splice forms, we conducted a Gene Ontology (GO) analyses using GeneOntology (Thomas et al., 2022). We expect somatic tissues to have GO terms related to their function (such as digestion in midguts and behaviour in heads), but ovary GO terms to be related to embryo development, RNA processing, or unknown functions, as such terms would bring weight to our hypothesis of greater alternative splice diversity in ovaries to develop new functions and help embryo development. As expected, the most enriched GO

terms in both male and female heads are for neuron differentiation (GO:0030182), locomotory behaviour (GO:0007626), and cellular component organization (GO:0051128). In Figure 21 we see the overlap of the 20 most significant terms from each sex in heads, and can see most of them have representative genes in both sexes. Female heads show an enrichment of genes related to detection of visible light (GO:0009584), which males do not present. On the other hand, some of the GO terms enriched in males and not females are detection of external stimulus (GO:0009581) and learning (GO:0007612). All those terms in heads are related to brain and eye activity which is consistent with established understanding of those tissues.

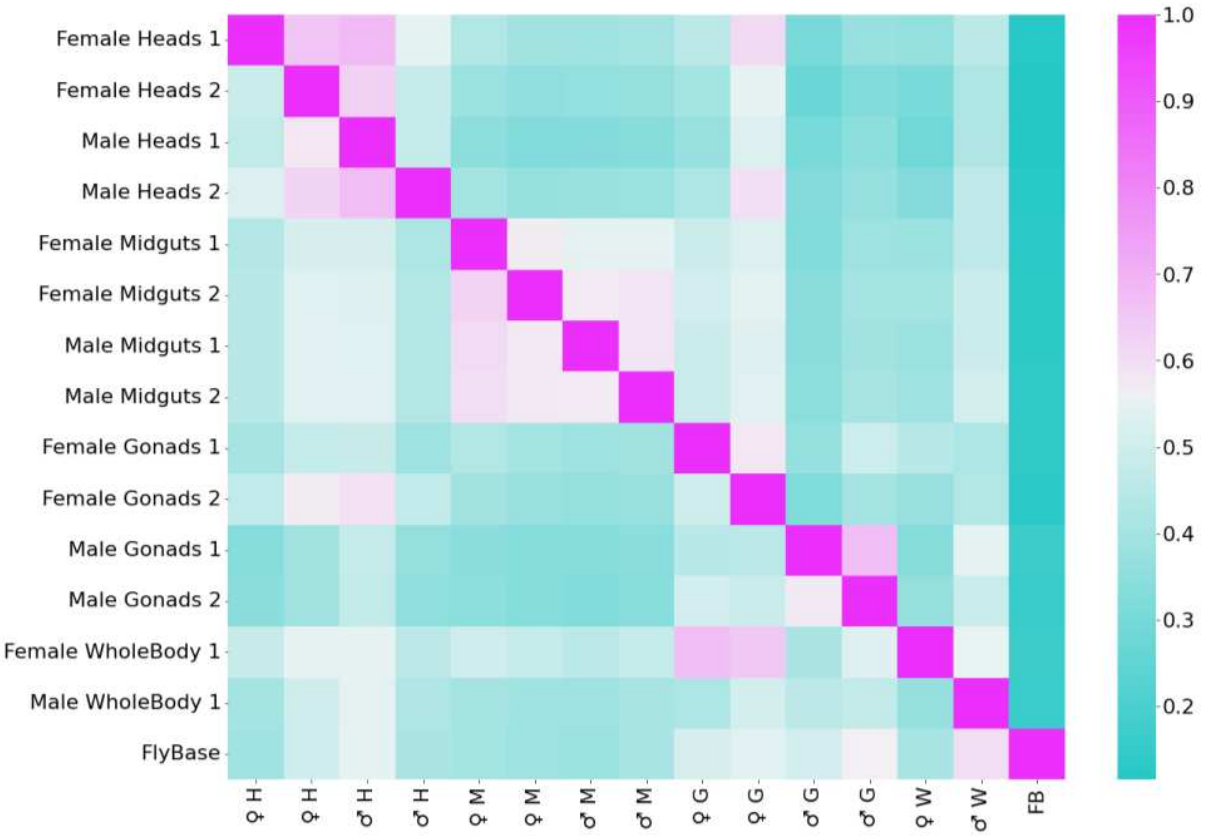

Figure 11. Number of shared splice forms between samples. *Number of shared splice forms between column and row tissues divided by the total number of splice forms in the row tissue. Most splice forms found in each tissue were not retrieved from FlyBase, but around half of the FlyBase splice forms were found in each sample. We can also see higher number of shared forms between samples of the same tissues ("squares" formed between all midgut samples and between all heads samples). Unexpectedly we do not find a high overlap between whole body samples and each individual tissue.*

For midguts, the sex exclusive terms found in each sex are related to cell and tissue maintenance (e.g. supramolecular fiber organisation [GO:1902903], regulation of biogenesis [GO:0044087], or tissue development [GO:0009888] for females; and cytoskeleton organization [GO:0032956], or epithelial cell development [GO:0002064] for males), as are the terms common to both sexes (cell-cell junction assembly [GO:0007043], establishment or maintenance of cell polarity [GO:0007163], epidermal cell differentiation [GO:0009913], positive regulation of epithelial cell migration [GO:0010634], cell-cell junction organization

[GO:0045216], or organic substance transport [GO:0071702]). See Figure 22 for the overlap of the 20 most significant terms from each sex in midguts.

Gonads show the least number of shared significantly enriched GO terms, with just one such term between the 20 terms with more genes in each sex (i.e. phosphate-containing compound metabolic process GO:0006796), and 27 in all terms with significantly enriched GO terms (see Figure 23 for the overlap of the 20 most significant terms from each sex in gonads). Those terms relate to cellular function and division, as expected for gonad cells (such as mitotic cell cycle [GO:0000278], mitotic cell cycle process [GO:1903047], meiotic cell cycle [GO:0051248], meiotic cell cycle process [GO:1903046], microtubule cytoskeleton organization [GO:0000226], regulation of translation [GO:0006417], spindle organization [GO:0007051], chromosome segregation [GO:0007059], or organelle fission [GO:0048285]). And, as expected for gonads, ovaries show an enrichment for GO:0007292 female gamete generation, and testis show an enrichment for GO:0048232 male gamete generation. Genes with ovary exclusive splice forms show -among others- enrichment in embryo development GO terms (such as embryonic axis specification [GO:0000578] and embryonic pattern specification [GO:0009880]), and egg development (like oocyte construction [GO:0009880] and oocyte axis specification [GO:0007309]). Genes with testis exclusive splice forms show an enrichment of genes related to sperm production (as sperm individualization [GO:0007291], spermatid development [GO:0007286]).

2.5 Discussion

Our methodology identified more splice forms than those currently catalogued in FlyBase. FlyBase has been shown to need to increase its database in relation to splice forms in *D. melanogaster* (Daines et al., 2010), suggesting the need for updated methods to include additional splice forms. In this study we use a new methodology for splice form identification, and even under stringent filtering[1] to remove artifacts, and find splice forms that are not present in FlyBase, indicating an opportunity for expansion of the database. It should be noted that our pipeline cannot detect modifications at the 3' or 5' ends of mRNA, which have been a significant focus of splice type identification, such that even this large number of novel transcript variants may be an underestimate.

The high incidence of splice form expression in ovaries aligns with previous findings (Gibilisco et al., 2016), indicating the importance of alternative splicing in ovarian expression. Furthermore, ovaries show many tissue-specific splice forms (Figure 10(B)), which is already known to occur, Gibilisco et al. (2016) shows that the highest fraction of expressed genes that are alternatively spliced is found in the ovary. This suggests that the ovaries have a unique profile of alternative splicing compared to other tissues, which we also find in this study. Additionally, research in humans by Kraaij et al. (1998) shows that in ovaries, the expression levels of alternative transcripts of the FSH (Folicule Stimulating Hormone) receptor show a constant ratio to the expression level of full-length mRNA of the same gene, indicating a specific regulation of alternative splicing in the ovaries also occurs in other groups. Furthermore, pig ovaries exhibit differential expression of clock-related genes splice forms between oestrus and diestrus stages (Huang et al., 2023), indicating a dynamic regulation of

[1] Filtering out any chimeric transcripts, transcripts that appear only once, and transcripts present in only one sample, even if multiple times.

gene expression through alternative splicing in the ovaries. This can be a point for further investigation as to what processes are being organized by alternative splicing in *Drosophila* ovaries. These findings collectively with our new findings suggest that the ovaries have a unique alternative splicing profile compared to other tissues, with specific regulation and expression patterns that need to be further analysed to show how distinct and essential for ovarian function and reproductive processes they are. One possible explanation for this high incidence of alternative splicing is the need for a vast protein and mRNA array in early embryo before embryonic genome activation.

While testes exhibit broad gene expression (Telonis-Scott et al., 2008), this study expands on the findings of Gibilisco et al. (2016) and others, showing that ovaries use more alternative splicing than other tissues (Figure 10(B)), especially for X-linked genes (Figure 12(C)). This suggests different patterns of alternative splicing between sexes that may extend beyond the ovaries, as seen in research by Rogers et al. (2020), Gibilisco et al. (2016), Telonis-Scott et al. (2008), and Hartmann et al. (2011). Our GO analyses revealed an enrichment of meiotic and cell division GO terms specific to ovary splice forms, likely used in the maturation and development of oocytes and eggs. However, these GO terms are also found in genes from other tissues, which might be an artifact since such processes are unlikely to occur outside the ovaries. Additionally, genes with non-ovary specific splice forms also have cell division-related GO terms, reflecting the widespread occurrence of this process in various tissues.

Some GO terms appearing across different tissues may be involved in alternative splicing regulation mechanisms, such as mRNA polyadenylation (GO:0006378), positive regulation of chromatin organization (GO:1905269), and piRNA-mediated gene silencing via mRNA destabilization (GO:0140991). These mechanisms could influence splicing (Naftelberg et al., 2015; Teixeira et al., 2017; Jimeno-González et al., 2015; Luco et al., 2010; Tian et al., 2007; Movassat et al., 2016). Since ovaries seem to have many new splice forms, further investigation is necessary to understand their function in oogenesis and other related processes.

In relation to chromosome location, considering the evolutionary background of these chromosomes, i.e., that the 4th chromosome was a sex chromosome in basal Dipterans (Vicoso & Bachtrog, 2015), this finding was not unexpected. The X-chromosome exhibits a higher incidence of alternative splicing compared to other chromosomes, for example, in mammals, random differences in Xist RNA levels influence the production of spliced RNA in the two X-chromosomes, leading to the inactivation of one chromosome (Federico et al., 2012). The unique chromatin state and organization of the X-chromosome also impact alternative splicing outcomes, with histone modifications and chromatin marks corresponding to exons recruiting auxiliary factors that influence alternative splicing decisions (Luco et al., 2010). Studies have also shown that mutations affecting X-chromosome inactivation can lead to differential splicing patterns, potentially influencing disease processes in humans (Ilagan et al., 2014). Likewise, the role of the X-chromosome in sex determination and dosage compensation mechanisms in organisms like *Drosophila* also highlights its unique alternative splicing characteristics (González et al., 2008). With the 4th chromosome having been an X-chromosome in the past in Diptera (Vicoso & Bachtrog, 2015), it is expected to have undergone similar pressures, explaining its higher splice form incidence compared to chromosomes that do not have such an evolutionary history.

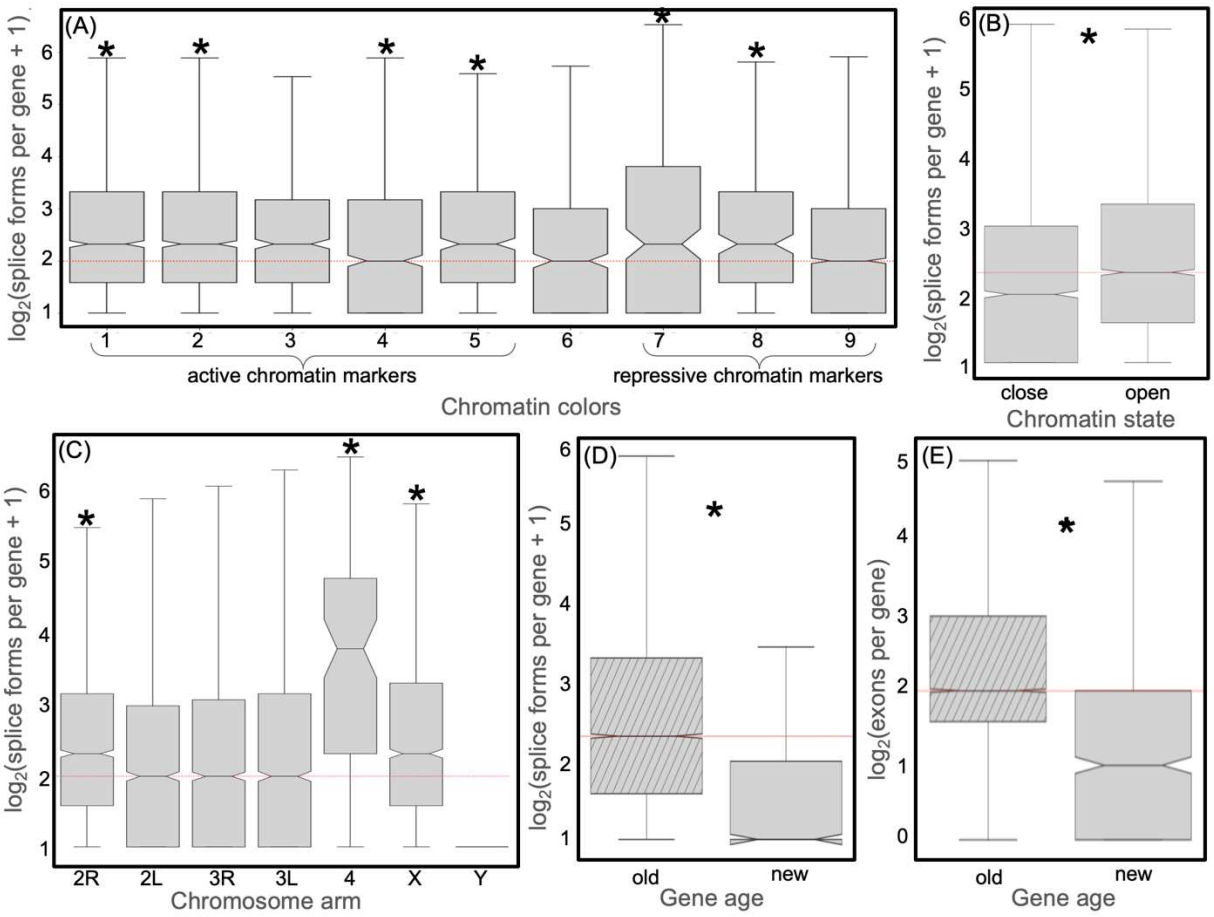

Figure 12. (A) Chromatin environment doesn't influence number of splice forms per gene beyond open or closed chromatin. *It has been shown that the chromatin environment is very important for gene expression and possibly for alternative splicing. Here we use the chromatin subdivisions from Kharchenko et al. (2011) but could not find any specific conformation associated to a different number of splice forms per gene. Most groups were significantly different from the combination of all others, but the number of genes in the group or an association to active chromatin would explain it more simply. It is noticeable that colours 4, 6, and 9 have less splice forms per gene than the other colours, but no one colour environment behaves in a unique way.* ***(B) Chromatin state influences the number of splice forms per gene.*** *Genes in open chromatin are expected to have more splice forms per gene as they are expected to be more easily accessible for expression and mutation agents. Close chromatin is expected to have less splice forms per gene. Chromatin state as in Milon et al., 2014.* ***(C) Gene location impacts the number of splice forms per gene.*** *Here we see the chromosome arm (equivalent to Muller elements) impacts the number of unique splice forms per gene. We can see a significant deviation from the mean for the Y chromosome and chromosomes X and 4 (Muller elements A and F respectively). That is in line with our expectations as very few genes are present on the Y and we expect them to have less splice forms as they are often related to male fertility. The X and 4th chromosomes have more transcripts per gene on average than other chromosomes, it is expected that the X has more splice forms and as the dot chromosome has been theorized to have been an X-chromosome before in the evolution of this group, therefore it is in line that the dot chromosome also has more splice forms per gene. Gene location data from FlyBase.* ***(D) Gene age influences splice form incidence.*** *Old genes (present in Drosophila and Sophophora subgenus) tend to have more splice forms per gene, probably as a consequence of them having had more time to evolve such mechanisms. It is not expected that new genes in Drosophila show many splice forms due to a great number of such genes being generated via retroposition, which means those genes will not have introns and therefore will have fewer chances of creating different splice forms. Gene ages from Zhang et al. (2010).* ***(E) Gene age versus number of exons.*** *Old genes (present in Drosophila and Sophophora subgenus) tend to have more exons per gene. As new genes in Drosophila often arise due to retroposition it is expected that they have less exons. Asterisk (*) indicates p-value < 0.05 in a Wilcoxon test between the group it is above, and the set of all other groups combined.*

Here we show the importance of chromatin state (open or closed), as seen in Figure 12(B), and chromosomal location, as seen in Figure 12(C). It is important to notice the variations between and within chromosomes of open and close chromatin, which still need further study. In the case of chromatin state, it significantly influences alternative splicing by affecting the accessibility of the splicing machinery to different parts of the pre-mRNA, so that exons are more often spliced-in to the mature mRNA compared to introns (Keren-Shaul et al., 2013). We indeed find more splice forms without introns than with them, further supporting that finding. The interplay between chromatin structure and pre-mRNA splicing involves factors like histone acetylation, histone variants, and histone tail post-translational modifications, which can either activate or silence gene expression and impact alternative splicing (Moreno et al., 2015), and in this study we did not directly investigate histone variations leaving an open question as to how much each of those possible modifications impact the patterns of exon and intron usage we found in each tissue.

Along with gene location, exon number is also important to splice incidence (Figure 20(B) in Appendix 1). Exon structure was shown to influence splice form incidence and tissue specificity (Zhao et al., 2023; Xing & Lee, 2005; Mandadi & Scholthof, 2015), in accordance with our data, where higher exon numbers lead to more splice forms. Figure 12(D) and Figure 20(B) shows how exon number per gene and gene age are factors influencing splice incidence per gene. Older genes or those with more exons tend to have more splice forms, possibly due to evolutionary adaptations or structural variability. Gene duplication and sub- or neofunctionalization may parallel these processes (Connallon & Clark, 2011; Des Marais & Rausher, 2008; Zhang et al., 1998; Gleixner et al., 2012). As new duplicated genes can undergo subfunctionalization -with each gene being responsible for a part of the ancestral function (Huang et al., 2016; Froyd et al., 2011)- so can different splice forms become responsible for different parts of a function (Lambert et al., 2015), similar to *Sxl* in sex determination where different splice forms will lead to female or male sex differentiation. The same could occur relating to neofunctionalization -when a duplicated gene finds a new function or is expressed in a new tissue (Assis & Bachtrog, 2013; Teshima & Innan, 2008; Ding et al., 2010)- so can splice forms of a same gene be expressed in different tissues or produce different protein products. It is though important to notice older genes tend to have more exons than new ones as seen in Figure 12(E), due to the same processes that lead to a higher splice form number.

Finally, our new method for the identification of full splice forms has its limitations. Our pipeline is not currently able to identify 3’ and 5’ splice site variations, due to its current annotation format. This could be addressed through changes to the script to identify such variations. These modifications would enhance the ability of the script to analyse splice forms and make it more comparable to other studies in various organisms, and experimental conditions.

3 Recruitment of an Ancestral Gene Network for Sex Determination in Drosophila

3.1 Introduction

Sex determination is a complex and diverse process that has been extensively researched. *Drosophila melanogas*ter, a well-studied model organism in genetics, has provided significant insights into the biology and evolution of sex determination and sex chromosomes (Vicoso & Bachtrog, 2015). Diptera species have a conserved karyotype of six chromosomal arms, including five large rods and a small dot chromosome (Vicoso & Bachtrog, 2015). This chromosomal composition is believed to represent the ancestral karyotype in Diptera (Vicoso & Bachtrog, 2015). The gene content of these chromosomal elements, known as Muller elements A to F, is highly conserved across Diptera (Vicoso & Bachtrog, 2015).

Despite the conservation of chromosomal architecture and genotypic sex determination in Diptera, comparative studies have revealed variations in master switch genes for sex determination (Meisel et al., 2015; Meccariello et al., 2019; Sharma et al., 2017) and the presence and nature of sex chromosomes among different groups (Vicoso & Bachtrog, 2015). In *Drosophila melanogaster*, the gene *Sex-lethal* (*Sxl*) initiates sex determination, regulated by a set of counter genes. The end of the sex determination cascade in *Drosophila* and other Diptera is marked by the gene *doublesex* (*dsx*), considered the ancestral sex determining gene. The honeybee *Apis mellifera*, belonging to Hymenoptera - the most basal lineage of holometabolous insects, also shows sex-specific splicing of their *dsx* gene. Both fly-type and moth-type splicing forms, indicating that the use of different splicing forms in controlling sexual differentiation, was present in the common ancestor of holometabolous insects (Cho et al., 2007).

This study aims to investigate the beginning of the sex determination pathway in *D. melanogaster*, focusing on the counter genes that control the splicing of *Sxl*, and their behaviour in related species.

3.1.1 Sex determination in Diptera

Although sex determination mechanism significantly varies in insects, some commonalities exist, such as the presence of the gene *transformer* and a master gene initiating the process (Verhulst et al., 2010; Gempe & Beye, 2011; Schütt & Nöthiger, 2000). In particular, in Dipterans -a group with heteromorphic sex chromosomes- numerous transitions of sex chromosomes have been observed between species and groups (Vicoso & Bachtrog, 2015).

In Diptera, chromosome arms retain their gene content but undergo frequent rearrangements. These chromosome arms, known as Muller elements, facilitate studies across different species. In some Dipteran groups, one Muller element is linked to sex determination, while in others, a different element is involved (Vicoso & Bachtrog, 2013). Additionally, even when master sex determining genes change, the downstream genes remain mostly the same (Graham et al., 2003). This change in sex determining genes is not uncommon, with the downstream switch *doublesex* being found in all insects (as reviewed in Geuverink & Beukeboom, 2013). In many cases, changes in the master sex determination switch are also related to sex chromosome turnover, but recently it was proposed that new sex determining upstream switches need not be related to sex chromosome turnover (Meisel,

2020). Sex chromosome turnover has been described to occur multiple times in dipterans (Vicoso & Bachtrog, 2015), not necessarily involving changes to the master sex determination genes.

Sex chromosomes can harbour a sex determining gene, such as the SRY gene on the mammalian Y chromosome (Yang *et al.*, 2006), or initiate sex determination in response to the X-to-Autosome ratio, as seen in *Drosophila* (Salz & Erickson, 2010). In *Drosophila* genes on the X and autosomes serve as counter genes that determine how many X-chromosomes the cells have in relation to autosomes, a ratio of one X-chromosome for each pair of autosomes leads to male phenotype, while two X-chromosomes for each pair of autosomes leads to female phenotype.

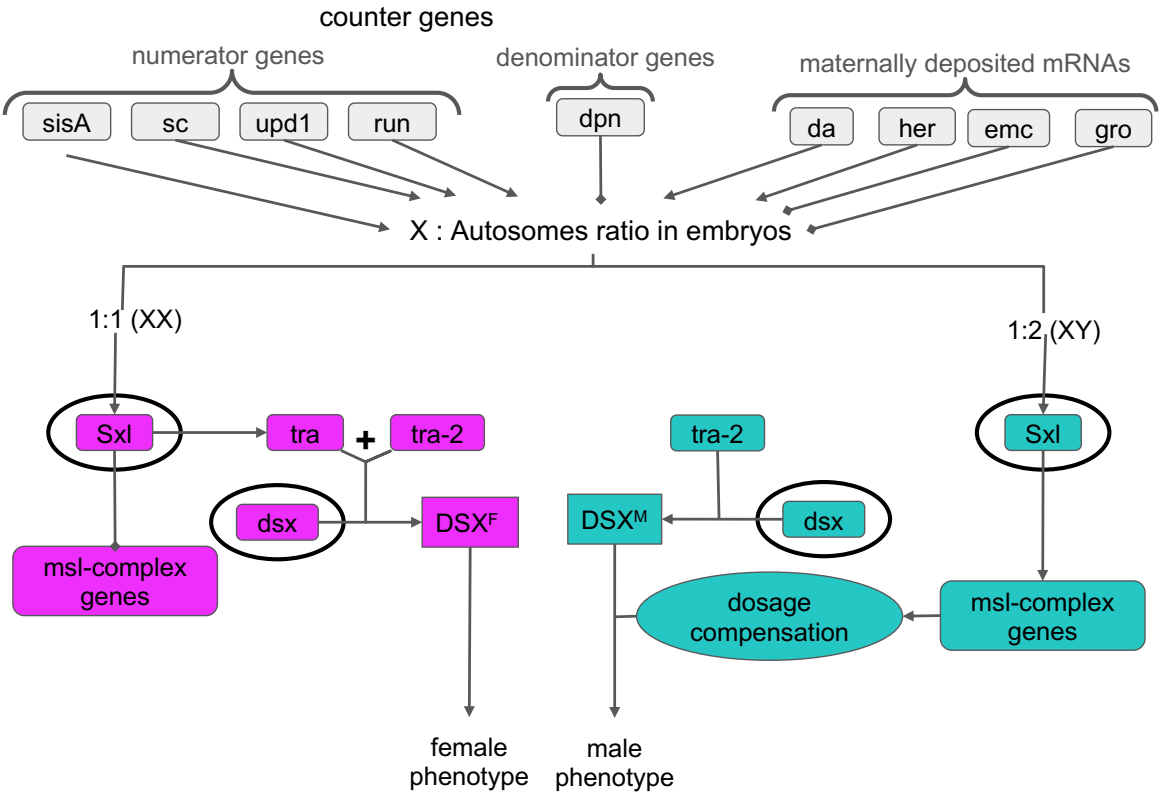

Figure 13. Alternative splicing in Drosophila sex determination. *General diagram of sex determination in Drosophila melanogaster. The male pathway is depicted on the right side (turquoise), and the female pathway is represented on the left side (pink). The top part (in grey) shows the genes involved in the determination of the autosomes to X-chromosomes proportion, known as counter genes. The transcripts with sex-specific splice forms are marked with circles. We can observe that even though many of the genes used are the same their different splice forms lead to different protein production and phenotypes. Arrowheads indicate activation and square heads indicate repression of a given product. Top part adapted from MacCarthy et al. (2003).*

Sex-lethal (*Sxl*), a well-documented *Drosophila* gene with sex-specific splice forms, is often considered the primary gene for sex determination (Graham et al., 2003). However, before *Sxl* expression begins, other genes are necessary to assess the X-to-Autosome ratio, leading to the expression of the appropriate *Sxl* form (Gabrielli, 2010; Gouw, 2009). Figure 13 shows a schematic view of the sex determination process where it is possible to see the counter

genes responsible for accessing X dosage, as well as the genes that have different splice forms between males and females (circled in black in Figure 13). Counter genes are expressed from autosomes and the X-chromosome and interact with transcription regulators leading to the expression of *Sxl*, which is the main sex determination gene in *Drosophila* (Cline, 1988; Cline, 1984; Mahadeveraju et al., 2020). Our study also examines these characteristics to determine if the genes had suitable expression patterns, levels, and locations to be co-opted for this function. Specifically, the research here investigates how genes upstream of the master sex determination gene in *Drosophila* were co-opted as counter genes.

3.1.2 Sex determination in Drosophila

In *Drosophila melanogaster*, the *dsx* gene is key for differentiating sexually dimorphic traits. It has a conserved role in sex determination in other organisms for example *Aedes aegypti*, *Anopheles gambiae*, and *Apis mellifera* (Meccariello et al., 2019; Cho et al., 2007). The less conserved part of the sex determination cascade is at the opposite end, where the master sex determining gene (*Sex-lethal*) determines sex in *D. melanogaster* through splicing into male or female forms. Counter genes recruit this process.

In *Drosophila*, *Sxl* is the primary switch gene for starting sex determination in early embryos (Graham et al., 2003). It acts through a genetic pathway leading to either male or female differentiation (Figure 13). Although *Sxl* is the main switch for sex determination in Drosophilids, its recruitment for this function is unique to this group (Mullon et al., 2012; Gabrieli et al., 2010; Thomas et al., 2010). *Sxl* activation is triggered by the ratio of X-chromosomes to autosomes in each cell, relying on X-linked signal genes such as *scute*, *sisA*, and *runt*, balanced against autosomal genes (*gro*, *da*, *her*, *dpn*, and *emc*) (Sawanth et al., 2016).

The evolution of counter genes in *D. melanogaster* and other dipterans is not well-studied. The shift from their ancestral function to their new role in sex determination may have been influenced by natural selection, genetic drift, and sexual selection. These forces can affect the regulatory regions and coding sequences of counter genes, leading to changes in expression levels, protein structure, and interaction partners (Parhad et al., 2017; Langley et al., 1982; Bettencourt & Feder, 2001).

Gene duplication events can also influence the evolutionary dynamics of genes in *D. melanogaster* (Lynch & Conery, 2000; Connallon & Clark, 2011; Taylor & Raes, 2004). Duplicated counter genes might undergo sub-functionalization or neo-functionalization, acquiring novel roles in sex determination. The diversity in gene expression and regulation contributes to the evolution of counter genes, allowing for the fine-tuning of their activity in various tissues and developmental stages (Meisel et al., 2012; Ellegren & Parsch, 2007; Fong et al., 2005). Understanding the evolution of counter genes in *D. melanogaster* is crucial for deciphering the molecular basis of sex determination and its role in the diversity of dipteran species. Comparative studies across different *Drosophila* species and other dipterans can reveal the conserved and variable aspects of counter genes and their contribution to sex determination. By exploring the evolutionary forces and mechanisms driving the evolution of counter genes, a deeper understanding of the genetic and developmental processes underlying sex determination in *D. melanogaster* and other organisms can be gained (Hediger et al., 2004; Dübendorfer et al., 2002).

In *Drosophila*, the initial signal for sex determination is the ratio between X-linked and autosomes, whereas in other Diptera species, such as *Ceratitis capitata*, it is a masculinizing factor on the Y-chromosome (Graham et al., 2003). This indicates that counter genes in *Drosophila* have been recently recruited for this function, raising questions about the mechanisms of this recruitment (Cline et al., 2010; Mullon et al., 2012; Gabrieli et al., 2010). One possibility is that the optimal chromosomal location of these genes facilitated their recruitment as counter genes.

The evolution of the sex determining system, particularly the emergence of *Sxl* in the Drosophilid lineage, has been a subject of extensive study (such as in Salz and Erickson 2010). *Sxl* arose in Drosophilids through duplication and extensive rearrangement (Traut et al., 2006; Cline et al., 2010; Meise et al., 1998; Mullon et al., 2012). The multiple transitions of sex chromosomes in Diptera have been investigated (Vicoso & Bachtrog, 2015), highlighting the need to explore the drivers of these changes. *Drosophila*, as a model organism, has been central to this research, but the diversity of sex determination mechanisms and influences from the environment in other Diptera species have also been studied (Andere et al., 2002; Agarie et al., 2023). Other insects, such as *Bombyx mori*, have been used to explore sex determination and sex chromosome dynamics (Xu et al., 2017), contributing to the understanding of the evolution of sex determination pathways. The diversity of sex determination strategies in Diptera necessitates integrative and innovative approaches for further understanding (Kratochvíl et al., 2021; Abbott et al., 2017).

A key question in this field is identifying the evolutionary forces driving the diversity of sex determination mechanisms in the groups. This study examines the role of serendipity in the recruitment of specific genes for the initial *Sxl* splice form recruitment in *D. melanogaster*. By comparing data from *D. melanogaster* with other organisms, the study aims to infer a possible ancestral state and understand the regulation of *Sxl* as the master sex determining switch and the previous relationships of the involved counter genes.

3.2 Methods

All data used in this study were previously published and accessible on the database from NCBI. The researchers searched SRA archive from NCBI for expression data and Ensembl for genome data. Table 2 lists the set of counter genes analysed, their FlyBase accession numbers, and their chromosomal location in *D. melanogaster*.

3.2.1 Genome Location

The study involved comparing the chromosomal location of counter genes from *Drosophila melanogaster* with those in *Ceratitis capitata*. For this, we used Ensembl Biomart data (Zerbino et al., 2018) to determine the chromosomal location of all *D. melanogaster* genes and FlyBase data for CDS of all *D. melanogaster* genes (Larkin et al., 2021). The chromosomal genome assembly of *C. capitata* from Ward et al. (2021) was employed to align *D. melanogaster* sex-determining genes (Table 2) using tblastn (Altschul, 1997).

Table 2. Counter genes analysed and their characteristics as per FlyBase (Larkin et al., 2021)

Gene name	*Gene symbol*	*FlyBase accession number*	*Chromosomal location*
sisterless A	*sisA*	FBgn0003411	X
runt	*run*	FBgn0003300	X
scute	*sc*	FBgn0004170	X
Sex-lethal	*Sxl*	FBgn0264270	X
daughterless	*da*	FBgn0267821	2L
hermaphrodite	*her*	FBgn0001185	2L
deadpan	*dpn*	FBgn0010109	2R
transformer	*tra*	FBgn0003741	3L
extra macrochaetae	*emc*	FBgn0000575	3L
groucho	*gro*	FBgn0001139	3R

3.2.2 Gene Expression

The recent recruitment of *Sxl* in sex determination and its non-essential nature in other groups prompted an investigation into how genes at the beginning of the pathway were recruited for this function. Besides location, the counter genes in *Drosophila* might have had an ideal expression during early embryogenesis, aiding their recruitment to the sex determination pathway.

Table 3 Correlations between each B. jarvisi embryo expression replicates and D. melanogaster cell cycle. *MSL complex and dosage compensation is expected to start at late cycle 14, while sex determination starts during late cell cycle 14 (Lott et al., 2011; Yuan et al., 2016; Salz & Erickson, 2010). Here those stages have a darker background, and the best correlation to each Bactrocera replicate is in bold lettering. rep. stands for replicate; c.c. stands for cell cycle.*

		D. melanogaster *embryonic cell cycle*								
		c.c. 10	c.c. 11	c.c. 12	c.c. 14a	c.c. 14b, rep. 1	c.c. 14b, rep. 2	c.c. 14c, rep. 1	c.c. 14c, rep. 2	c.c. 14d
B. jarvisi *embryonic expression*	Early rep. 1	0.3748	0.35352	**0.37722**	0.37312	0.34165	0.36387	0.34972	0.36036	0.36837
	Early rep. 2	0.45781	0.43329	0.45756	**0.46562**	0.42347	0.44472	0.43083	0.44116	0.45701
	Late rep. 1	0.53625	0.50762	0.53863	**0.55163**	0.501106	0.53071	0.51567	0.52957	0.54769
	Late rep. 2	0.53587	0.50939	0.53903	0.5512	0.50407	0.53203	0.52158	0.53395	**0.55278**

To assess this hypothesis, the expression of such genes in *Drosophila melanogaster* embryos (Lott et al., 2011) was compared with that of *Bactrocera jarvisi* embryos (Morrow et al., 2014). *B. jarvisi*, a Tephritid fly, does not have *Sxl*-based sex determination and served as a proxy for the ancestral state. This comparison tested if the expression of the counter genes is similar in the ancestral lineage of *Drosophila*.

We downloaded RNA-seq reads from the SRA database from NCBI and used Trimmomatic (Bolger et al., 2014) version 0.36 to trim reads for quality. *B. jarvisi* data was assembled with SOAPde novo (Luo et al., 2012), and transcripts were mapped to the longest CDS for each gene from *D. melanogaster* (using BLAT with a translated query and dataset). We kept only the best *D. melanogaster* hit for each transcript. Transcripts not mapped to *D. melanogaster* were excluded from downstream analyses. Expression values for *B. jarvisi* were obtained using Kallisto (Bray et al., 2016). The databases for *B. jarvisi* are listed, as well as for *D. melanogaster*.

A total of 147,498 reads from *B. jarvisi* were mapped to *D. melanogaster* genes, encompassing 9,720 genes. Seven of these belong to genes in the sex determination genetic initial pathway. The researchers also used BLAST (Altschul et al., 1990) to supplement the gene set with more sex determination related genes.

We used four female samples for *B. jarvisi* and various samples for different cell cycles of *D. melanogaster*. The correlation between each possible pair of samples from the two species was plotted (Table 3), and the pair with the most significant correlation was selected for subsequent analyses. In Figure 14 we see four scatter plots for the best correlations of each *B. jarvisi* embryonic stage to a cell cycle of *D. melanogaster*. Considering that these were the highest correlation between both species they were the selected pairings between embryonic phases in the two species, i.e., *B. jarvisi* early replicate 1 was paired to cell cycle 12 of *D. melanogaster*, early replicate 2 and late replicate 1 were paired to cell cycle 14a, and late replicate 2 to cell cycle 14d.

We also analysed expression data from *Bactrocera oleae* (Bioproject PRJNA196340). As there was only one sample for *B. oleae* (SRP021044), as with *B. jarvisi* we found the best correlating *Drosophila* cell cycle for comparison and normalized expressions using the Loess normalization process. In Appendix 2, it is possible to see scatter plots of the best correlation between *B. oleae* embryonic expression replicate and cell cycles of *D. melanogaster*, which is cell cycle 14a from *D. melanogaster*.

We used Kallisto (Bray et al., 2016) to quantify reads and provide TPM data. To make these comparisons we used the TPM measure, and normalized the samples using Loess' normalization function in R. Zero values were adjusted by adding a small amount to all gene expressions.

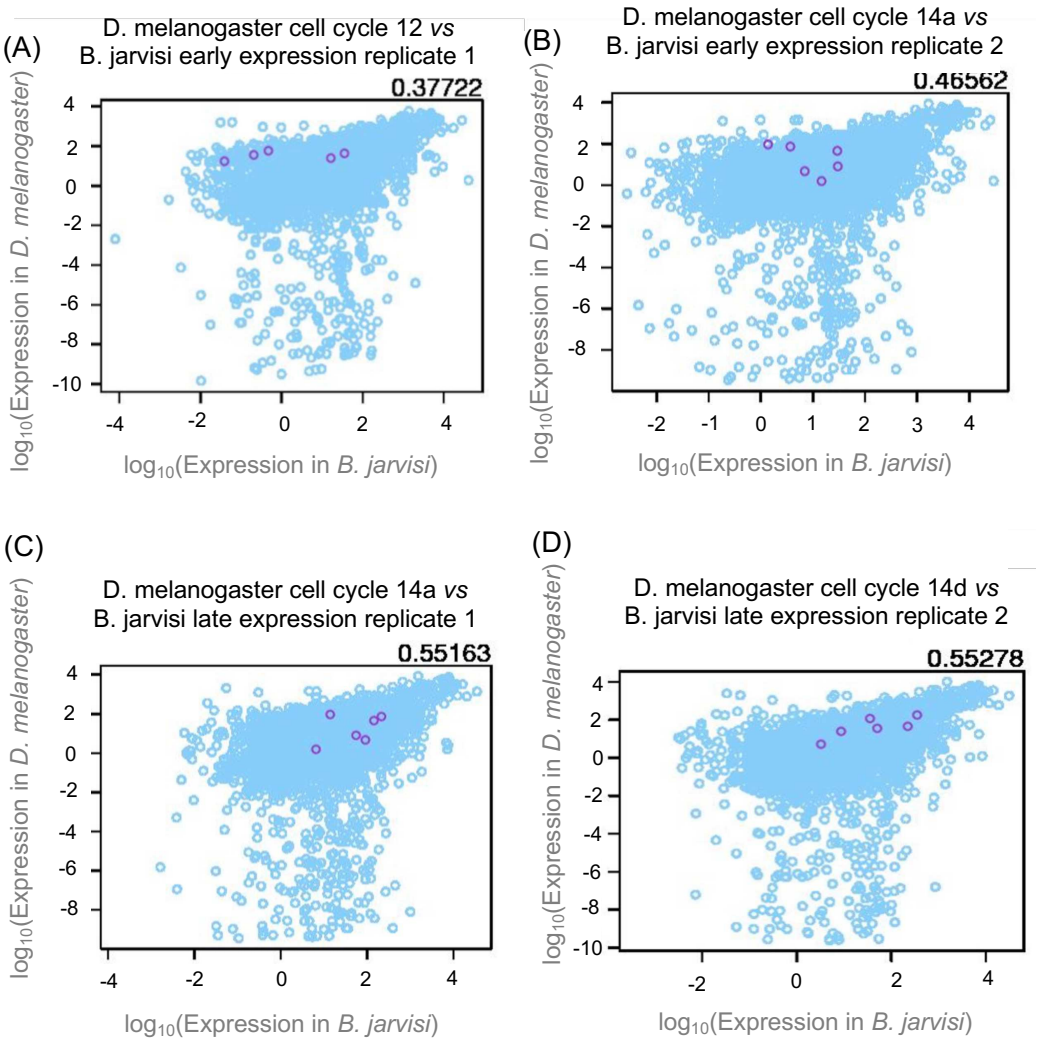

Figure 14. Expression in different Bactrocera jarvisi and Drosophila melanogaster embryonic stages. *It is possible to see a high correlation of the expression between the species, and using the highest correlation between each B. jarvisi stage and D. melanogaster cell cycle the most appropriate pairing between embryonic phases in the two species was selected. Each plot refers to the highest correlated expression in a different sample of embryonic stages in each species.*

3.2.3 Gene Interactions

We hypothesize that multiple counter genes were simultaneously recruited to the sex determination network due to their ancestral interactions, which facilitated their integration into a new network. Alternatively, each gene might have been independently recruited to the sex determination network. To explore this hypothesis, we used data on gene orthologs and their gene interactions from FlyBase (Larkin et al., 2014) and StringDB (Szklarczyk et al., 2019).

FlyBase v.FB2017_04 provided the orthologs for the selected genes for *Caenorhabditis elegans*, *Mus musculus*, *Homo sapiens*, and *Saccharomyces cerevisiae*. FlyBase employs the DIOPT (DRSC Integrative Ortholog Prediction Tool) v.6 to identify possible orthologs of queried *Drosophila melanogaster* genes. We noticed that while some genes lacked orthologs in certain species, others had multiple orthologs for the same gene.

We downloades the complete protein interactions and protein alias data from the "Downloads" section of String-DB (Szklarczyk et al., 2019). We used a batch script to segregate this data according to each species, resulting in separate files for gene interactions in each species and a file listing all gene aliases with their corresponding StringDB identification numbers. Interactions were acknowledged only if there was experimental evidence, such as Yeast2Hybrid experiments or protein-protein interactions, identified between the genes or their orthologs in any organism.

We analysed the number of interactions each ortholog had and which of those were with other orthologs from the *D. melanogaster* counter genes. The researchers then ascertained the number of interactions each ortholog had with other orthologs.

While indirect interactions are also crucial in the interactions of these genes, examining how orthologs of each gene interact might reveal if these genes participate in genetic networks in other organisms. The DIOPT orthologs are not highly conservative, so we conducted the same analysis using OrthoDB orthologs for humans, mice, *C. elegans*, and *S. cerevisiae* (Kuznetsov et al., 2023). In OrthoDB, yeast did not have orthologs to any of the *D. melanogaster* sex determination network genes. The worm had four orthologs (*Sex-lethal*, *groucho*, *deadpan*, and *hermaphrodite*), but no experimental interactions were recorded in StringDB for them. For mice, only two of the thirteen orthologs (to four *D. melanogaster* genes) had experimentally verified interactions in StringDB.

Table 4. Genomic location of counter genes from D. melanogaster in C. capitata. *Locations in D. melanogaster are given as Muller elements, where A corresponds to the X-chromosome, B to 2L, C to 2R, D to 3L, E to 3R, and F to chromosome 4. In C. capitata locations are given as Muller elements, and in parenthesis the scaffold to which they were aligned is shown. Scaffolds as in Ward et al., 2021.*

Gene	Location in *D. melanogaster*	Location in *C. capitata*
sisterless A	A	B (CAJHJT010000012.1)
runt	A	A (CAJHJT010000045.1)
scute	A	A (CAJHJT010000045.1)
Sex-lethal	A	A (CAJHJT010000045.1)
daughterless	B	B (CAJHJT010000012.1)
hermaphrodite	B	B (CAJHJT010000012.1)
deadpan	C	C (CAJHJT010000056.1)
transformer	D	D (CAJHJT010000034.1)
extra macrochaetae	D	D (CAJHJT010000034.1)
groucho	E	E (CAJHJT010000001.1)

3.3 Results

3.3.1 Conserved Genome Location of Counter Genes

The genomic location of counter genes is critical for their function, so there are two hypotheses for their recruitment: the first hypothesis is that they were initially located on appropriate chromosomes, and the second hypothesis is that their original chromosomal configuration was different, but they moved or were duplicated onto the correct chromosomal configuration with the evolution of the new *Sxl*-based sex determining system. To test this, we inferred their location in *C. capitata*, which carries the ancestral X-chromosome and does not use *Sxl* for sex determination. Table 4 displays the genomic location of counter genes in *C. capitata* and *D. melanogaster*. Most counter genes in *C. capitata* are located in the same Muller elements as in *D. melanogaster*, suggesting they were already in the ideal chromosomal location for counter genes.

3.3.2 Counter Genes Ancestrally Expressed in Early Embryogenesis

We next examined the expression patterns of counter genes during early embryogenesis. Figure 24 in Appendix 2 compares the expression levels of orthologues in *D. melanogaster* and *B. jarvisi*. Genes with expression ratios below the 2.5th percentile and above the 97.5% percentile are marked in dark blue, while those within the 95% expression ratio between species are in light blue, and genes from the sex determination network are shown in pink (X chromosome genes) and purple (autosomal genes). Most counter genes fall within the 95% expression ratio, with exceptions noted for *scute* and *Sxl*. The inclusion of two extra genes, *upd1* and *sisA*, recovered using BLAST, did not significantly alter these results (Figure 25 in Appendix 2).

We then compared the expression levels in embryogenesis of sex determination-related genes and genes unrelated to sex determination (Figure 15). For early expression pairs of samples, there was a significant difference in the expression ratio for the dataset with seven sex determination-related genes. However, the addition of X-linked genes *upd1* and *sisA* removed this significant difference (Figure 26 in Appendix 2), suggesting similar relative expression levels between *B. jarvisi* and *D. melanogaster* or a possibly too early stage for significant expression changes to be identified. For a full comparison of embryonic stages between *D. melanogaster* and *B. jarvisi*, see Table 3.

Another comparison used *Bactrocera oleae* and *D. melanogaster*. The correlation analyses matched the *B. oleae* sample to cycle 14A in *D. melanogaster* (Table 5, Appendix 2). In Figure 27(A) in Appendix 2, all counter genes from *B. oleae* fit within the 95% intermediary ratio of expressions between *D. melanogaster* and B*. oleae*, supporting the hypothesis that genes were already expressed in early embryos in the ancestral lineage. The relative expression analyses between species for genes involved and not involved in sex determination (Figure 27(B) in Appendix 2) also showed no difference, further supporting this hypothesis.

3.3.3 Gene Interactions from Counter Genes Conserved in Distant Clades

The preceding sections suggest counter genes were recruited due to their correct chromosomal configuration and expression during relevant embryogenesis stages. The researchers then investigated if they were already part of the same gene network, potentially facilitating their joint recruitment for sex determination. Figure 16(A) shows direct

interactions of original *Drosophila* genes according to StringDB data (Szklarczyk et al., 2019). Then, Figure 16(B), (C), and (D) display the interactions between *Drosophila* orthologs in other species (worm, mouse, and human), with each species represented in its heatmap. We identified interactions based on experimental evidence. Notably, a single *Drosophila* gene can have multiple orthologs in these species, allowing for numerous interactions.

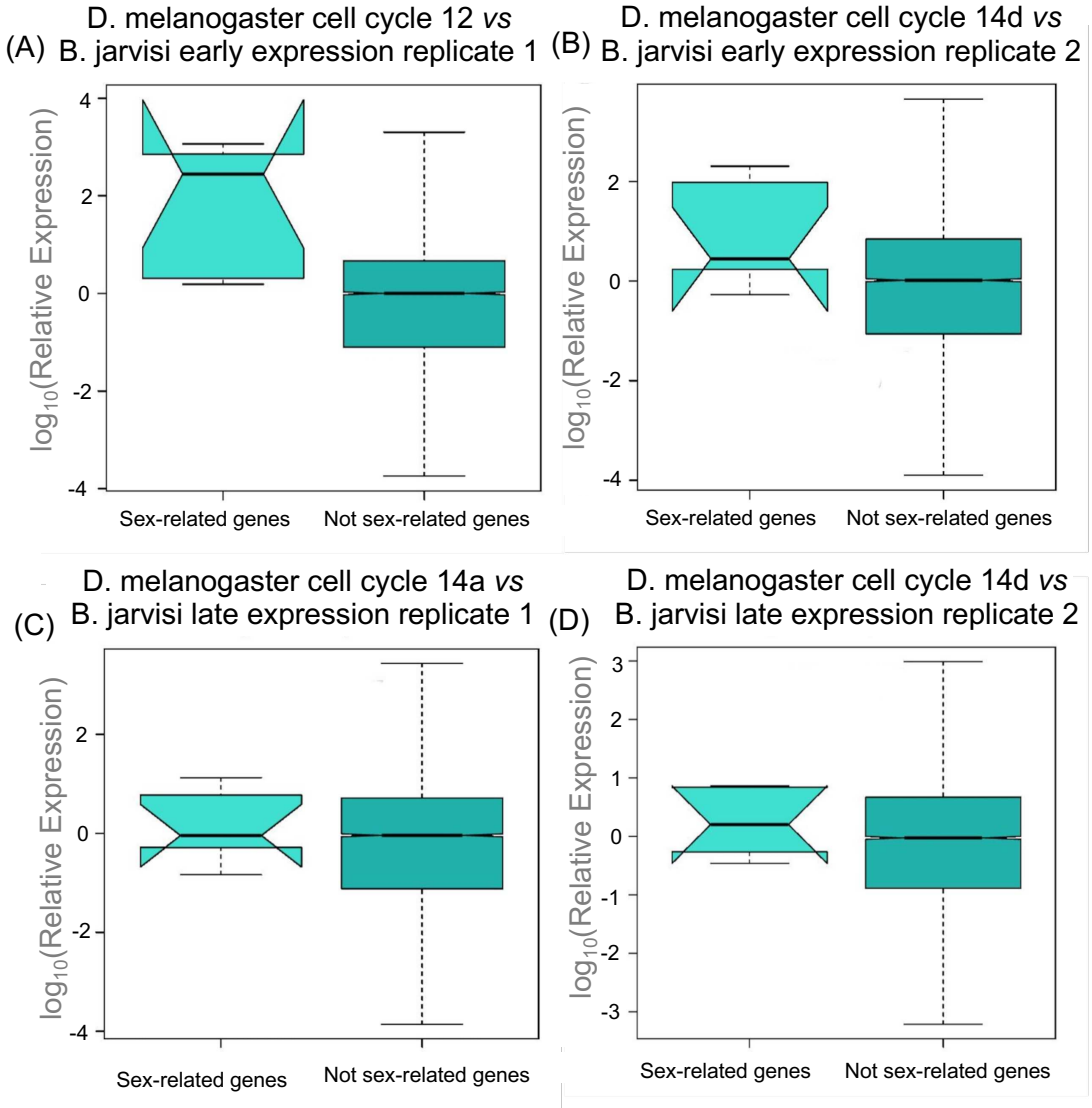

Figure 15. Relative log expression in the pairs of expression stages in D. melanogaster and B. jarvisi, with outliers removed from the plot*. For the early expression pairs (A and B) the difference between sex determination related genes relative expression and the relative expression of genes not related to sex determination is significant, and higher for sex determination related genes. This indicates that these genes have a higher expression in D. melanogaster than in B. jarvisi, supporting the idea that these genes got to their optimal expression levels after being recruited. But, in the late phases, we cannot see such a pattern (C and D). This lack of difference may indicate that for the late phases, these genes had an optimal expression in the ancestral lineage. Still, this can also be an artefact due to a less-than-ideal match between the embryonic phases in D. melanogaster and B. jarvisi. Asterisk (*) indicates p-value < 0.05 in a Wilcoxon test.*

In mice and humans, we observed interactions between *daughterless* and *extra macrochaetae* and between *deadpan* and *scute*, also found in *D. melanogaster*. The recurrent interaction between *runt* and *groucho* orthologs in mice and humans, as well as in *Drosophila*, even in conservative analyses, indicates a possibly conserved network. These last analyses, even with more conservative orthologs, found interactions consistent with those in *D. melanogaster* in distant animals, suggesting a potentially conserved genetic network.

3.4 Discussion

Understanding how new functions evolve is a central theme in evolutionary biology. Sex-determining pathways, which are often subject to change, serve as excellent models to explore how genes and networks are co-opted to control fundamental biological processes. It is known that an ancestral chromatin regulatory network was repurposed for X-chromosome dosage compensation in *Drosophila melanogaster* (Smith et al., 2005). More broadly, gene networks often evolve in a coordinated manner (Ciliberti et al., 2007; Amoutzias et al., 2004), suggesting that the interaction of genes in a network across different organisms might indicate they were already part of the same network, aiding their incorporation into a new one. This research focused on whether similar recruitment of an ancestral pathway could explain the simultaneous inclusion of several genes in the sex-determining network in *Drosophila* and the factors that might have contributed to their integration.

In the current chapter, we showed how the recruitment of counter genes in *Drosophila* was probably due to those genes already presenting ideal locations, expression patterns, and gene networks prior to recruitment. The findings from this study endorse the notion that at least a portion of the network predated the evolution of new sex-determination mechanisms in *Drosophila*. These findings highlight a potential role for serendipity in the recruitment of counter genes for sex determination: these genes were ancestrally expressed during critical embryonic stages for sex determination (mitotic divisions 13 and 14 (Lott et al., 2011)). The presence of several genes from the network on Muller element A, which evolved into an X-chromosome, along with others on different chromosomes, likely eased their recruitment. While these results are compelling, a comprehensive analysis of sex determination in closely related outgroups is necessary to illuminate the timing of the recruitment of different genes fully and to determine whether this happened concurrently with the emergence of Muller element A as the X-chromosome or after it. Similarly, conducting analogous analyses in other groups that have experienced shifts in sex determination will provide insights into the commonality of recruiting pre-existing networks and whether embryonic expression and appropriate chromosomal location are prerequisites or can evolve as part of the shift in sex determination.

Our study concludes that the genomic locations and expression patterns of counter genes played a crucial role in their recruitment into the sex-determination network in *Drosophila*. Specifically, most counter genes were already situated in the optimal chromosomal configurations necessary for their function before the evolution of the *Sxl*-based sex determination system. We evidenced this by the conserved location of these genes in *C. capitata*, a Tephritid fly more basal in the phylogeny than *Drosophila*. Additionally, the counter genes were apparently ancestrally expressed during critical stages of early embryogenesis, suggesting that their expression patterns were also preadapted for their roles

in sex determination. This ancestral expression, especially during key embryonic stages, likely facilitated their integration into the sex-determination pathway.

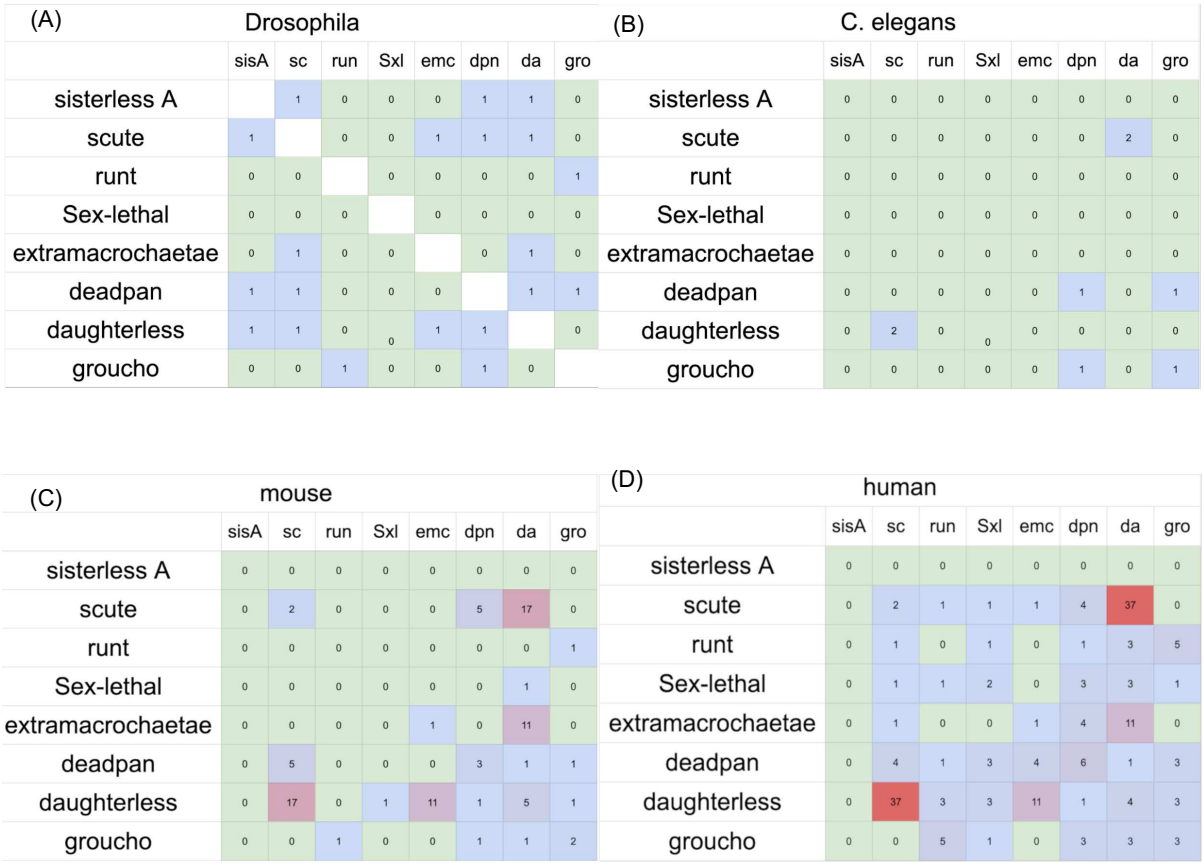

Figure 16. Heatmaps of the number of gene interactions in (A) D. melanogaster, (B) C. elegans, (C) Mus musculus, and (D) humans. *Gene interactions of genes involved in the early sex-determining pathway in D. melanogaster. Interaction data taken from StringDB. Due to the presence of orthologs in species other than D. melanogaster, orthologs of the same gene can interact with each other, and more than one ortholog can interact with the ortholog(s) of another gene.*

Our conclusions underscore the importance of preexisting gene configurations and expression patterns in the evolution of new biological functions and pathways. The conservation of gene interactions across diverse species, including worms, mice, and humans, indicates that these counter genes were part of a conserved genetic network even before their recruitment for sex determination in *Drosophila*. This suggests a potential role of serendipity in evolutionary processes, where genes with favorable attributes are co-opted for new functions. Our findings motivate further research into the recruitment mechanisms of gene networks, emphasizing the need to examine sex determination in closely related outgroups and other species with shifts in sex determination systems. Such studies could reveal whether the recruitment of preexisting networks and embryonic expression patterns are common strategies in the evolution of sex determination pathways, since gene networks often evolve in a synchronized way (Ciliberti et al., 2007; De Smet & Van de Peer, 2012; Ramos & Barolo, 2012; Amoutzias et al., 2004),

4 Conclusions

The study of sex determination and sexual dimorphism, particularly in the fruit fly *Drosophila melanogaster*, has been a focal point for researchers seeking to understand complex biological processes. This thesis examined the recruitment of ancestral gene networks for sex determination in *D. melanogaster*, focusing on counter genes that initiate the sex determination cascade and the use of alternative splice forms by males and females of the species, which can aid in understanding the mechanisms of sexual dimorphism.

The exploration of alternative splicing and its variances between males and females stands at the forefront of biomedical and evolutionary research, presenting a pivotal avenue to unravel the complexities underpinning sex differences. Differences in alternative splicing between males and females have been mostly explored through short reads sequencing, which can be less effective than long reads in identifying splice forms (see Figure 2). Here, we use long-read sequencing with a new pipeline to identify specific splice forms in different tissues of *D. melanogaster*. Alternative splicing is essential, for instance, for sex determination in *D. melanogaster*. Among the myriad mechanisms governing sex determination, the counter genes employed by *Drosophila* for selecting the male or female splice form of *Sex-lethal* to be expressed represent a relatively uncharted territory. Despite the extensive study of the sex-determination system of Drosophila, these counter genes have yet to be thoroughly investigated, suggesting a significant gap in our understanding of sex-determination evolution.

In Chapter 3, on sex determination evolution, the research highlighted the significance of comprehending sex determination in *Drosophila*. Despite the conservation of chromosomal architecture and genotypic sex determination in Dipterans, variations in master switch genes for sex determination were observed. In *D. melanogaster*, the *Sex-lethal* (*Sxl*) gene, regulated by a set of counter genes, initiates sex determination. The study explored potential factors influencing the evolution of these counter genes, including natural selection, genetic drift, sexual selection, gene duplication, and gene expression regulation. The analyses indicated that genomic location, expression patterns during early embryogenesis, and gene interactions of counter genes were crucial for their recruitment. In Figure 15 and Figure 27(B) (Appendix 2), it can be seen that the expression of counter genes appears to be conserved between *Bactrocera* and *Drosophila*, and Figure 16 and Table 2 show the conservation of gene interactions and genomic location between counter genes in different clades. The findings suggest that these counter genes in *D. melanogaster* may have had conserved genomic locations, expression patterns, and interactions across different species, hinting at an optimal expression and interaction network in their ancestral lineage that facilitated their recruitment for sex determination.

The investigation into sex determination in *Drosophila melanogaster* offers insights into the evolutionary dynamics of gene networks and their influence on the diversity of sex-determination systems in Dipteran species. The recruitment of counter genes, a vital aspect of the sex determination cascade, appears to be influenced by genomic location, gene interactions, and conserved gene expression patterns. These findings underscore the need to consider multiple factors in understanding the molecular basis and evolution of sex determination. Comparative studies across diverse Dipteran species contribute to a broader understanding of sex determination mechanisms and their genetic and developmental

underpinnings. The adaptability of sex determination networks across Dipteran species highlights the complexity of this fundamental biological process.

In Chapter 2, on alternative splicing in males and females, we employed long-read sequencing technologies to investigate the splice form landscape within *Drosophila melanogaster* males and females. The study identified more splice forms than those catalogued in FlyBase (which can be seen in Figure 8(B) and (C)), suggesting a potential underestimation of splice form diversity in current knowledge. The analyses revealed a particularly high incidence of splice-form expression in ovaries, indicating the significance of alternative splicing in shaping the ovarian transcriptome. The study also explored factors influencing splice form incidence per gene, including gene age (Figure 12(D)), gene structure (Figure 20(B)), chromosomal location (Figure 12(C)), and chromatin state (Figure 12(B)). These insights into gene regulation and expression variation across tissues and sexes enhance understanding of gene expression regulation in complex organisms.

Both chapters shed light on the multifaceted nature of gene usage and evolution, whether in different splice forms or entire gene networks in each tissue and how serendipitous factors can influence gene evolution and usage. The findings highlight the intricate interplay between gene regulation and alternative splicing, with expression site and genomic location playing key roles in orchestrating gene expression diversity.

Further investigation using long-read sequencing in different tissues and species is necessary to identify potential artefacts and common splice forms. Expanding analyses to more tissues and additional species of *Drosophila* and other organisms is crucial. Additionally, more extensive analyses of counter genes, including embryonic expression in more species and experimental validation of gene product interactions, are needed to corroborate the general interactions found in the StringDB database and understand the evolution of these genes over time.

Our study falls along with others such as Gibilisco et al. (2016), Rogers et al. (2022), and Telonis-Scott et al. (2008) in showing the prevalence of alternative splicing in males and females along with their differences. We expand on those studies and add a new method to evaluate the differences in full-length reads. Furthermore, we show how different splice forms in early development can be recruited from a set of genes with a serendipitous expression pattern, location, and network, as those found in *Drosophila* counter genes.

5 References

Abbott, J. K., Norden, A. K., & Hansson, B. (2017). Sex chromosome evolution: Historical insights and future perspectives. Proc Biol Sci, 284(1854): 20162806.

Adamopoulos, P., Theodoropoulou, M., & Scorilas, A. (2018). Alternative splicing detection tool—a novel perl algorithm for sensitive detection of splicing events, based on next-generation sequencing data analysis. Ann Transl Med, 6(12): 244.

Agarie, A., Miyaguni, Y., Sugio, K., Tsuji, K., & Kobayashi, K. (2023). Males influence the primary sex ratio bias in Ryukyu dry wood termites. Front Ecol Evol, 10: 1081043.

Agirre, E., Oldfield, A., Bellora, N., Segelle, A., & Luco, R. F. (2021). Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. Nat Commun, 12(1): 682.

Almqvist, C., Worm, M., & Leynaert, B. (2007). Impact of gender on asthma in childhood and adolescence: a ga2len review. Allergy, 63(1): 47-57.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. J Mol Biol, 215(3): 403-410.

Amoutzias, G. D., Robertson, D. L., Oliver, S. G., & Bornberg-Bauer, E. (2004). Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. EMBO Rep, 5(3): 274-279.

Andere, A., Pimsler, M., Tarone, A., & Picard, C. (2020). The genomes of a monogenic fly: views of primitive sex chromosomes. Sci Rep, 10(1): 15728.

Anderson, S. T., & FitzGerald, G. A. (2020). Sexual dimorphism in body clocks. Science, 369(6508): 1164-1165.

Arbuthnott, D., Fedina, T. Y., Pletcher, S. D., & Promislow, D. E. L. (2017). Mate choice in fruit flies is rational and adaptive. Nat Commun, 8: 13953.

Assis, R. & Bachtrog, D. (2013). Neofunctionalization of young duplicate genes in drosophila. PNAS, 110(43): 17409-17414.

Bachtrog, D., Mank, J., Peichel, C., Kirkpatrick, M., Otto, S., Ashman, T.-L., Hahn, M., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J., & The Tree of Sex Consortium. (2014). Sex determination: Why so many ways of doing it? PLoS Biol, 12(7): e1001899.

Bain, S. A., Marshall, H., Filia, A. G. d. l., Laetsch, D. R., Husník, F., & Ross, L. (2021). Sex-specific expression and DNA methylation in a species with extreme sexual dimorphism and paternal genome elimination. Mol Ecol, 30(22): 5687-5703.

Baralle, F. E., & Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol, 18(7): 437-451.

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. PNAS, 74(8): 3171-3175.

Bettencourt, B. R., & Feder, M. E. (2001). Hsp70 Duplication in the Drosophila melanogaster Species Group: How and When Did Two Become Five? Mol Biol Evol, 18(7): 1272-1282.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15): 2114-2120.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol, 34(5): 525-527.

Breitbart, R. E., Andreadis, A., & Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Ann Rev Biochem, 56: 467-495.

Bridges, C. B. (1925). Sex in relation to chromosomes and genes. Am Nat, 59(661): 127-137.

Bridges, C. B (1921) Triploid Intersexes in Drosophila melanogaster. Science, 54(1394): 252-254.

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., Wan, K. H., Yu, C., Zhang, D., Carlson, J. W., Cherbas, L., Eads, B. D., Miller, D., Mockaitis, K., Roberts, J., ... Celniker, S. E. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature, 512(7515): 393-399.

Brown, S., Stoilov, P., & Xing, Y. (2012). Chromatin and epigenetic regulation of pre-mRNA processing. Hum Mol Genet, 21(R1): R90-R96.

Burtis, K. C. (1993). The regulation of sex determination and sexually dimorphic differentiation in Drosophila. Curr Opin Cell Biol, 5(6): 1006-1014.

Camara, N., Whitworth, C., & Doren, M. V. (2008). The creation of sexual dimorphism in the drosophila soma. Curr Top Dev Biol, 83: 65-107.

Carvalho, A. (2002). Origin and evolution of the Drosophila Y chromosome. Curr Opin Genet Dev, 12(6): 664-668.

Carvalho, A. B., Dobo, B. A., Vibranovski, M. D., & Clark, A. G. (2001). Identification of five new genes on the Y chromosome of Drosophila melanogaster. PNAS, 98(23): 13225-13230.

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.

Chen, W., Zhou, W., Li, Q., & Mao, X. (2023). Sex differences in gene expression and alternative splicing in the chinese horseshoe bat. PeerJ, 11: e15231.

Cho, S., Huang, Z. Y., & Zhang, J. (2007). Sex-Specific Splicing of the Honeybee doublesex Gene Reveals 300 Million Years of Evolution at the Bottom of the Insect Sex-Determination Pathway. Genetics, 177(3): 1733-1741.

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5’ ends of adenovirus 2 messenger RNA. Cell, 12(1): 1-8.

Ciliberti, S., Martin, O. C., & Wagner, A. (2007). Innovation and robustness in complex regulatory gene networks. PNAS, 104(34): 13591-13596.

Cline, T. W. (1984). Autoregulatory Functioning of a Drosophila Gene Product That Establishes and Maintains the Sexually Determined State. Genetics, 107(2): 231-77.

Cline, T. W. (1988). Evidence that sisterless-a and sisterless-b are two of several discrete "numerator elements" of the X/A sex determination signal in Drosophila that switch Sxl between two alternative stable expression states. Genetics, 119(4): 829-862.

Cline, T. W., Dorsett, M., Sun, S., Harrison, M. M., Dines, J., Sefton, L., & Megna, L. (2010). Evolution of the drosophila feminizing switch gene Sex-lethal. Genetics, 186(4): 1321-1336.

Coelho, S., Gueno, J., Lipinska, A., Cock, J., & Umen, J. (2018). UV chromosomes and haploid sexual systems. Trends Plant Sci, 23(9): 794-807.

Coelho, S., Mignerot, L., & Cock, J. (2019). Origin and evolution of sex-determination systems in the brown algae. New Phytol, 222(4): 1751-1756.

Connallon, T., & Clark, A. G. (2011). The Resolution of Sexual Antagonism by Gene Duplication. Genetics, 187(3): 919-937.

Cornelius, V., Fulton, J., & Margariti, A. (2021). Alternative splicing: a key mediator of diabetic vasculopathy. Genes, 12(9): 1332.

Cui, Z., Hui, M., Liu, Y., Song, C., Li, X., Li, Y., ... & Chu, K. (2015). High-density linkage mapping aided by transcriptomics documents zw sex determination system in the chinese mitten crab eriocheir sinensis. Heredity, 115(3): 206-215.

Czaja, W., Miller, K., Skinner, M., & Miller, B. (2014). Structural and functional conservation of fungal mata and human sry sex-determining proteins. Nat Commun, 5: 5434.

Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., ... & Chen, R. (2010). The drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res, 21(2): 315-324.

De Smet, R., & Van de Peer, Y. (2012). Redundancy and rewiring of genetic networks following genome- wide duplication events. Curr Opin Plant Biol, 15(2): 168-176.

Demir, E., & Dickson, B. J. (2005). Fruitless Splicing Specifies Male Courtship Behavior in Drosophila. Cell, 121(5): 785-794.

Des Marais, D. L., & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature, 454(7205): 762-765.

Ding, Y., Zhao, L., Yang, S., Jiang, Y., Chen, Y., Zhao, R., ... & Wang, W. (2010). A young drosophila duplicate gene plays essential roles in spermatogenesis by regulating several y-linked male fertility genes. PLoS Genet, 6(12): e1001255.

Dübendorfer, A., Hediger, M., Burghardt, G., & Bopp, D. (2002). Musca domestica, a window on the evolution of sex-determining mechanisms in insects. Int J Dev Biol, 46(1): 75-79.

Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet, 8(9): 689-698.

English, A. C., Patel, K. S., & Loraine, A. E. (2010). Prevalence of alternative splicing choices in Arabidopsis thaliana. BMC Plant Biol, 10: 102.

Ezaz, T., Moritz, B., Waters, P., Graves, J., Georges, A., & Sarre, S. (2009). The zw sex microchromosomes of an australian dragon lizard share no homology with those of other reptiles or birds. Chromosome Res, 17(8): 965-973.

Federico, A., Cirillo, D., Bolognesi, B., & Tartaglia, G. G. (2012). X-inactivation: quantitative predictions of protein interactions in the xist network. Nucleic Acids Res, 41(1): e31.

Ferguson-Smith, M. (2006). The evolution of sex chromosomes and sex determination in vertebrates and the key role of dmrt1. Sex Dev, 1(1): 2-11.

Fiszbein, A., Krick, K. S., Begg, B. E., & Burge, C. B. (2019). Exon-Mediated Activation of Transcription Starts. Cell, 179(7): 1551-1565.

Fong, S. S., Joyce, A. R., & Palsson, B. Ø. (2005). Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Res, 15(10): 1365-1372.

Fraser, J., Diezmann, S., Subaran, R., Allen, A., Lengeler, K., Dietrich, F., ... & Heitman, J. (2004). Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol, 2(12): e384.

Froyd, C. A. & Rusché, L. N. (2011). The duplicated deacetylases sir2 and hst1 subfunctionalized by acquiring complementary inactivating mutations. Mol Cell Biol, 31(16): 3351-3365.

Fu, X. and Ares, M. (2014). Context-dependent control of alternative splicing by rna-binding proteins. Nat Rev Genet, 15(10): 689-701.

Gabrieli, P., Falaguerra, A., Siciliano, P., Gomulski, L. M., Scolari, F., Zacharopoulou, A., Franz, G., Malacrida, A. R., & Gasperi, G. (2010). Sex and the single embryo: Early deveiopment in the Mediterranean fruit fly, Ceratitis capitata. BMC Dev Biol, 10: 12.

Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K., & Chen, X. (2010). Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res, 20(7): 763-783.

Gempe, T., & Beye, M. (2011). Function and evolution of sex determination mechanisms, genes and pathways in insects. BioEssays, 33(1): 52-60.

Geuverink, E. & Beukeboom, L. W. (2013). Phylogenetic distribution and evolutionary dynamics of the sex determination genes doublesex and transformer in insects. Sex Dev, 8(1-3): 38-49.

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.

Gibson, J. R., Chippindale, A. K., & Rice, W. R. (2002). The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc Biol Sci, 269(1490): 499-505.

Gnan, S., Matelot, M., Weiman, M., Arnaiz, O., Guerin, F., Sperling, L., Betermier, M., Thermes, C., Chen, C.-L., & Duharcourt, S. (2022). GC content but not nucleosome positioning directly contributes to intron- splicing efficiency in Paramecium. Genome Res, 32(4): 699-709.

González, A., Lü, H., & Erickson, J. (2008). A shared enhancer controls a temporal switch between promoters during drosophila primary sex determination. PNAS, 105(47): 18436-18441.

Gouw, J., Pinkse, M., Vos, H., Moshkin, Y., Verrijzer, C., Heck, A., … & Krijgsveld, J. (2009). In vivo stable isotope labelling of fruit flies reveals post-transcriptional regulation in the maternal-to-zygotic transition. Mol Cell Proteomics, 8(7): 1566-1578.

Graham, P., Penn, J. K. M., & Schedl, P. (2003). Masters change, slaves remain. BioEssays, 25(1): 1-4.

Grange, P. d. l., Gratadou, L., Delord, M., Dutertre, M., & Auboeuf, D. (2010). Splicing factor and exon profiling across human tissues. Nucleic Acids Res, 38(9): 2825-2838.

Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., van Baren, M. J., Boley, N., Booth, B. W., Brown, J. B., Cherbas, L., Davis, C. A., Dobin, A., Li, R., Lin, W., Malone, J. H., Mattiuzzo, N. R., Miller, D., ... Celniker, S. E. (2011). The developmental transcriptome of Drosophila melanogaster. Nature, 471(7339): 473-479.

Graves, J. (1996). Mammals that break the rules: genetics of marsupials and monotremes. Annu Rev Genet, 30: 233-260.

Graves, J. (2016). Did sex chromosome turnover promote divergence of the major mammal groups?. Bioessays, 38(8): 734-743.

Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, F., … & Shi, Y. (2014). Bs69/zmynd11 reads and connects histone h3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mrna processing. Mol Cell, 56(2): 298-310.

Hafezi, Y., Sruba, S. R., Tarrash, S. R., Wolfner, M. F., & Clark, A. G. (2020). Dissecting Fertility Functions of Drosophil a Y Chromosome Genes with CRISPR. Genetics, 214(4): 977-990.

Hansson, Å. & Olsson, M. (2018). Incubation temperature and parental identity determine sex in the australian agamid lizard ctenophorus pictus. Ecol Evol, 8(19): 9827-9833.

Hartmann, B., Castelo, R., Minana, B., Peden, E., Blanchette, M., Rio, D. C., Singh, R., & Valcarcel, J. (2011). Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. RNA, 17(3): 453-468.

Hediger, M., Burghardt, G., Siegenthaler, C., Buser, N., Hilfiker-Kleiner, D., Dbendorfer, A., & Bopp, D. (2004). Sex determination in Drosophila melanogaster and Musca domestica converges at the level of the terminal regulator doublesex. Dev Genes Evol, 214(1): 29-42.

Hedrick, A. V., & Temeles, E. J. (1989). The Evolution of Sexual Dimorphism in Animals: Hypotheses and Tests. Trends Ecol Evol, 4(5): 136-138.

Hill, P., Shams, F., Burridge, C., Wapstra, E., & Ezaz, T. (2021). Differences in homomorphic sex chromosomes are associated with population divergence in sex determination in carinascincus ocellatus (scincidae: lygosominae). Cells, 10(2): 291.

Hornett, E., Kageyama, D., & Hurst, G. (2022). Sex determination systems as the interface between male-killing bacteria and their hosts. Proc Biol Sci, 289(1972): 20212781.

Huang, B., Chen, Y. W., Huang, C. C., Gao, J., & Liao, P. C. (2016). Imbalanced positive selection maintains the functional divergence of duplicated dihydrokaempferol 4-reductase genes. Sci Rep, 6: 39031.

Huang, S., Dou, J., Li, Z., Hu, L., Yu, Y., & Wang, Y. (2022). Analysis of genomic alternative splicing patterns in rat under heat stress based on RNA-seq data. Genes, 13(2): 358.

Hudry, B., de Goeij, E., Mineo, A., Gaspar, P., Hadjieconomou, D., Studd, C., Mokochinski, J.B., Kramer, H.B., Plaçais, P.Y., Preat, T., Miguel-Aliaga, I. (2019). Sex Differences in Intestinal Carbohydrate Metabolism Promote Food Intake and Sperm Maturation. Cell, 178(4): 901-918.

Ilagan, J., Ramakrishnan, A., Hayes, B., Murphy, M., Zebari, A., Bradley, P., ... & Bradley, R. (2014). u2af1 mutations alter splice site recognition in hematological malignancies. Genome Res, 25(1): 14-26.

Ilik, I. A., Quinn, J. J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R. C., Luscombe, N., Backofen, R., Chang, H. Y., & Akhtar, A. (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. Mol Cell, 51(2): 156-173.

Jabre, I., Chaudhary, S., Guo, W., Kalyna, M., Reddy, A. S. N., Chen, W., Zhang, R., Wilson, C., & Syed, N. H. (2021). Differential nucleosome occupancy modulates alternative splicing in Arabidopsis thaliana. New Phytol, 229(4): 1937-1945.

Jimeno-González, S., Payán-Bravo, L., Muñoz-Cabello, A., Guijo, M., Gutiérrez, G., Prado, F., … & Reyes, J. (2015). Defective histone supply causes changes in RNA polymerase ii elongation rate and cotranscriptional pre-mRNA splicing. PNAS, 112(48): 14840-14845.

Kageyama, D., Narita, S., & Watanabe, M. (2012). Insect sex determination manipulated by their endosymbionts: incidences, mechanisms and implications. Insects, 3(1): 161-199.

Karlebach, G., Veiga, D. F. T., Mays, A. D., Chatzipantsiou, C., Barja, P. P., Chatzou, M., Kesarwani, A. K., Danis, D., Kararigas, G., Zhang, X. A., George, J., Steinhaus, R., Hansen, P., Seelow, D., McMurry, J.A., Haendel, M. A., Yang, J., Oprea, T., Anczukow, O., ... Robinson, P. N. (2020). The impact of biological sex on alternative splicing. bioRxiv. DOI: 10.1101/490904.

Karp, N. A., Mason, J., Beaudet, A. L., Benjamini, Y., Bower, L., Braun, R. E., Brown, S. D. M., Chesler, E. J., Dickinson, M. E., Flenniken, A. M., Fuchs, H., Angelis, M. H. de, Gao, X., Guo, S., Greenaway, S., Heller, R., Herault, Y., Justice, M. J., Kurbatova, N., ... White, J. K. (2017). Prevalence of sexual dimorphism in mammalian phenotypic traits. Nat Commun, 8: 15475.

Kato, Y., Kobayashi, K., Watanabe, H., & Iguchi, T. (2011). Environmental sex determination in the branchiopod crustacean daphnia magna: deep conservation of a doublesex gene in the sex-determining pathway. PLoS Genet, 7(3): e1001345.

Keren, H., Lev-Maor, G., & Ast, G. (2010). Alternative splicing and evolution: Diversification, exon definition and function. Nat Rev Genet, 11(5): 345-355.

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M. Y., Luquette, L. J., Xi, R., Jung, Y. L., Park, R. W., Bishop, E. P., ... Park, P. J. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature, 471(7339): 480-485.

Kimura, K.-I., Ote, M., Tazawa, T., & Yamamoto, D. (2005). Fruitless specifies sexually dimorphic neural circuitry in the Drosophila brain. Nature, 438(7065): 229-233.

Kirkpatrick, M., & Guerrero, R. F. (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. Genetics, 197(2): 531-541.

Koerich, L. B., Wang, X., Clark, A. G., & Carvalho, A. B. (2008). Low conservation of gene content in the Drosophila Y chromosome. Nature, 456(7224): 949-951.

Koterniak, B., Pilaka, P. P., Gracida, X., Schneider, L. M., Pritišanac, I., Zhang, Y., & Calarco, J. A. (2020). Global regulatory features of alternative splicing across tissues and within the nervous system of C. elegans. Genome Res, 31(12): 1766-1780.

Kraaij, R., Verhoef-Post, M., Grootegoed, J. A., & Themmen, A. P. N. (1998). Alternative splicing of follicle-stimulating hormone receptor pre-mRNA: cloning and characterization of two alternatively spliced mRNA transcripts. J Endocrinol, 158(1): 127-136.

Kratochvíl, L., Stöck, M., Rovatsos, M., Bullejos, M., Herpin, A., Jeffries, D. L., Peichel, C. L., Perrin, N., Valenzuela, N., & Pokorná, M. J. (2021). Expanding the classical paradigm: What we have learnt from vertebrates about sex chromosome evolution. Philos Trans R Soc Lond B Biol Sci, 376(1833): 20200097.

Kufel, J., Diachenko, N., & Golisz, A. (2022). Alternative splicing as a key player in the fine-tuning of the immunity response in Arabidopsis. Mol Plant Pathol, 23(8): 1226-1238.

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., & Zdobnov, E. M. (2023). OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Research, 51(D1): D445-D451.

Lambert, M. J., Cochran, W. O., Wilde, B. M., Olsen, K. G., & Cooper, C. D. (2015). Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. Genome Res, 25(5): 624-632.

Lang, A. S., Austin, S. H., Harris, R. M., Calisi, R. M., & MacManes, M. D. (2020). Stress-mediated convergence of splicing landscapes in male and female rock doves. BMC Genomics, 21(1): 251.

Langley, C. H., Montgomery, E., & Quattlebaum, W. F. (1982). Restriction map variation in the Adh region of Drosophila. PNAS, 79(18): 5631-5635.

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., Thurmond, J., FlyBase Consortium, Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, M., Dos Santos, G., ... Lovato, T. (2021). FlyBase: Updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res, 49(D1): D899-D907.

Leader, D. P., Krause, S. A., Pandit, A., Davies, S. A., & Dow, J. A. T. (2018). FlyAtlas 2: A new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Res, 46(D1): D809-D815.

Lee, Y. and Rio, D. (2015). Mechanisms and regulation of alternative pre-mrna splicing. Annu Rev Biochem, 84: 291-323.

Lengeler, K., Fox, D., Fraser, J., Allen, A., Forrester, K., Dietrich, F., ... & Heitman, J. (2002). Mating-type locus of cryptococcus neoformans: a step in the evolution of sex chromosomes. Eukaryot Cell, 1(5): 704-718.

Leung, C., Douglass, S., Morselli, M., Obusan, M., Pavlyukov, M., Pellegrini, M., … & Johnson, T. (2019). H3k36 methylation and the chromodomain protein eaf3 are required for proper cotranscriptional spliceosome assembly. Cell Rep, 27(13): 3760-3769.

Li, Y., Dai, C., Hu, C., Liu, Z., & Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. Plant J, 90(1): 164-176.

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 Biol, 9(2): e1000590.

Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage Compensation in Drosophila. Cold Spring Harb Perspect Biol, 7(5): a019398.

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science, 327(5968): 996-1000.

Ludwig, S. A., Smith, R. E., & Ibrahim, N. (2023). Sexual dimorphism in dinosaurs. eLife, 12: e89158.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. GigaScience, 1(1): 18.

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. Science, 290(5494): 1151-1155.

Ma, W., Vavre, F., & Beukeboom, L. (2013). Manipulation of arthropod sex determination by endosymbionts: diversity and molecular mechanisms. Sex Dev, 8(1-3): 59-73.

MacCarthy, T., Seymour, R., & Pomiankowski, A. (2003). The evolutionary potential of the Drosophila sex determination gene network. J Theor Biol, 225(4): 461-468.

MacDougall, C., Harbison, D., & Bownes, M. (1995). The developmental consequences of alternate splicing in sex determination and differentiation in Drosophila. Dev Biol, 172(2): 353-376.

Mahadeveraju, S., Jung, Y.-H., & Erickson, J. W. (2020). Evidence That Runt Acts as a Counter-Repressor of Groucho During Drosophila melanogaster Primary Sex Determination. G3, 10(7): 2487-2496.

Mandadi, K. K. & Scholthof, K. G. (2015). Genomic architecture and functional relationships of intronless, constitutively- and alternatively-spliced genes in brachypodium distachyon. Plant Signal Behav, 10(8): e1042640.

Mane, S. P., Evans, C., Cooper, K. L., Crasta, O. R., Folkerts, O., Hutchison, S. K., Harkins, T. T., Thierry-Mieg, D., Thierry-Mieg, J., & Jensen, R. V. (2009). Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing. BMC Genomics, 10: 264.

Mank, J. E. (2009). Sex Chromosomes and the Evolution of Sexual Dimorphism: Lessons from the Genome. Am Nat, 173(2): 141-150.

Martín, G., Márquez, Y., Mantica, F., Duque, P., & Irimia, M. (2021). Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. Genome Biol, 22(1): 35.

Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. Nature Reviews Molecular Cell Biology, 15(2): 108-121.

McIntyre, L. M., Bono, L. M., Genissel, A., Westerman, R., Junk, D., Telonis-Scott, M., Harshman, L., Wayne, M. L., Kopp, A., & Nuzhdin, S. V. (2006). Sex-specific expression of alternative transcripts in Drosophila. Genome Biol, 7(9): R79.

Meccariello, A., Salvemini, M., Primo, P., Hall, B., Koskinioti, P., Dalíková, M., Gravina, A., Gucciardino, M. A., Forlenza, F., Gregoriou, M.-E., Ippolito, D., Monti, S. M., Petrella, V., Perrotta, M. M., Schmeing, S., Ruggiero, A., Scolari, F., Giordano, E., Tsoumani, K. T., ... Saccone, G. (2019). Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests. Science, 365(6460): 1457-1460.

Meise, M., Hilfiker-Kleiner, D., Dübendorfer, A., Brunner, C., Nöthiger, R., & Bopp, D. (1998). Sex-lethal, the master sex-determining gene in Drosophila, is not sex-specifically regulated in Musca domestica. Development, 125(8): 1487-1494.

Meisel R.P. (2020). Evolution of Sex Determination and Sex Chromosomes: A Novel Alternative Paradigm. Bioessays, 42(9): e1900212.

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Faster-X Evolution of Gene Expression in Drosophila. PLoS Genet, 8(10): e1003013.

Meisel, R. P., Scott, J. G., & Clark, A. G. (2015). Transcriptome Differences between Alternative Sex Determining Genotypes in the House Fly, Musca domestica. Genome Biol Evol, 7(7): 2051-2061.

Menkis, A., Jacobson, D., Gustafsson, T., & Johannesson, H. (2008). The mating-type chromosome in the filamentous ascomycete neurospora tetrasperma represents a model for early evolution of sex chromosomes. PLoS Genet, 4(3): e1000030.

Merkin, J. J., Chen, P., Alexis, M. S., Hautaniemi, S. K., & Burge, C. B. (2015). Origins and Impacts of New Mammalian Exons. Cell Rep, 10(12): 1992-2005.

Millington, J. W., & Rideout, E. J. (2018). Sex differences in Drosophila development and physiology. Curr Opin Psysiol, 6: 46-56.

Milon, B., Sun, Y., Chang, W., Creasy, T., Mahurkar, A., Shetty, A., Nurminsky, D., & Nurminskaya, M. (2014). Map of open and closed chromatin domains in Drosophila genome. BMC Genomics, 15(1): 988.

Modrek, B. and Lee, C. (2002). A genomic view of alternative splicing. Nat Genet, 30(1): 13-19.

Mohr, C., Kleiner, S., Blanchette, M., Pyrowolakis, G., & Hartmann, B. (2017). Sex-Specific Transcript Diversity in the Fly Head Is Established during Pupal Stages and Adulthood and Is Largely Independent of the Mating Process and the Germline. Sex Dev, 11(2): 94-108.

Movassat, M., Crabb, T., Busch, A., Yao, C., Reynolds, D., Shi, Y., … & Hertel, K. (2016). Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. RNA Biol, 13(7): 646-655.

Moreno, N. N., Giono, L. E., Botto, A. E. C., Muñoz, M. J., & Kornblihtt, A. R. (2015). Chromatin, dna structure and alternative splicing. FEBS Lett, 589(22): 3370-3378.

Morgan, T. H. (1910). Sex limited inheritance in drosophila. Science, 32(812): 120-122.

Morrow, J. L., Riegler, M., Gilchrist, A., Shearman, D. C., & Frommer, M. (2014). Comprehensive transcriptome analysis of early male and female Bactrocera jarvisi embryos. BMC Genet, 15(S2): S7.

Mullon, C., Pomiankowski, A., & Reuter, M. (2012). Molecular evolution of Drosophila Sex-lethal and related sex determining genes. BMC Evol Biol, 12: 5.

Naftelberg, S., Schor, I. E., Ast, G., & Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem, 84: 165-198.

Nakano, Y., Kelly, M. C., Rehman, A. U., Boger, E. T., Morell, R. J., Kelley, M. W., Friedman, T. B., & Bánfi, B. (2018). Defects in the Alternative Splicing-Dependent Regulation of REST Cause Deafness. Cell, 174(3): 536-548.

Nanda, I., Schlegelmilch, K., Haaf, T., Schartl, M., & Schmid, M. (2008). Synteny conservation of the z chromosome in 14 avian species (11 families) supports a role for z dosage in avian sex determination. Cytogenet Genome Res, 122(2): 150-156.

Nissen, K., Homer, C., Ryan, C., Shales, M., Krogan, N., Patrick, K., … & Guthrie, C. (2017). The histone variant h2a.z promotes splicing of weak introns. Genes Dev, 31(7): 688-701.

Oltean, S. and Bates, D. (2013). Hallmarks of alternative splicing in cancer. Oncogene, 33(46): 5311-5318.

Olthof, A. M., White, A. K., Mieruszynski, S., Doggett, K., Lee, M. F., Chakroun, A., Abdel Aleem, A. K., Rousseau, J., Magnani, C., Roifman, C. M., Campeau, P. M., Heath, J. K., & Kanadia, R. N. (2021). Disruption of exon-bridging interactions between the minor and major spliceosomes results in alternative splicing around minor introns. Nucleic Acids Res, 49(6): 3524-3545.

Painter, T.S. (1922) The sex chromosomes of the monkey. Science, 56(1445): 286-287.

Painter, T.S. (1923) Further observations on the sex chromosomes of mammals. Science, 58(1500): 247-248.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet, 40(12): 1413-1315.

Parhad, S. S., Tu, S., Weng, Z., & Theurkauf, W. E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev Cell, 43(1): 60-70.e5.

Parisi, M., Nuttall, R., Edwards, P., Minor, J., Naiman, D., Lü, J., Doctolero, M., Vainer, M., Chan, C., Malley, J., Eastman, S., & Oliver, B. (2004). A survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. Genome Biol, 5(6): R40.

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am J Hum Genet, 102(1): 11-26.

Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., & Graveley, B. R. (2004). Identification of alternative splicing regulators by RNA interference in Drosophila. PNAS, 101(45): 15974-15979.

Paronetto, M., Achsel, T., Massiello, A., Chalfant, C., & Sette, C. (2007). The rna-binding protein sam68 modulates the alternative splicing of bcl-x. J Cell Biol, 176(7): 929-939.

Peters, S. A., Muntner, P., & Woodward, M. (2019). Sex differences in the prevalence of, and trends in, cardiovascular risk factors, treatment, and control in the United States, 2001 to 2016. Circulation, 139(8): 1025-1035.

Petrova, V., Song, R., Nordström, K., Walter, J., Wong, J., Armstrong, N. J., ... & Schmitz, U. (2022). Increased chromatin accessibility facilitates intron retention in specific cell differentiation states. Nucleic Acids Res, 50(20): 11563-11579.

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: Their purpose and place. Hum Mol Genet, 27(R2): R234-R241.

Punzalan, D., & Hosken, D. J. (2010). Sexual Dimorphism: Why the Sexes Are (and Are Not) Different. Curr Biol, 20(22): R972-R973.

Ramos, A. I., & Barolo, S. (2012). Evolution of Gene Regulation: Hybrid Networks Breed Diversity. Curr Biol, 22(23): R1009-11.

Rice, W. R. (1984). Sex Chromosomes and the Evolution of Sexual Dimorphism. Evolution, 38(4): 735-742.

Rogers, T. F., Palmer, D. H., & Wright, A. E. (2021). Sex-Specific Selection Drives the Evolution of Alternative Splicing in Birds. Mol Biol Evol, 38(2): 519-530.

Roy, N., Choi, I., Um, T., Jeon, M., Kim, B., Kim, Y., … & Kim, N. (2021). Gene expression and isoform identification of pacbio full-length cdna sequences for berberine biosynthesis in berberis koreana. Plants, 10(7): 1314.

Salvemini, M., Mauro, U., Lombardo, F., Milano, A., Zazzaro, V., Arcà, B., ... Saccone, G. (2011). Genomic organization and splicing evolution of the doublesex gene, a drosophila regulator of sexual differentiation, in the dengue and yellow fever mosquito aedes aegypti. BMC Evol Biol, 11: 41.

Salz, H. K. (2011). Sex determination in insects: A binary decision based on alternative splicing. Curr Opin Genet Dev, 21(4): 395-400.

Salz, H., & Erickson, J. W. (2010). Sex determination in Drosophila: The view from the top. Fly, 4(1): 60-70.

Sawanth, S. K., Gopinath, G., Sambrani, N., & Arunkumar, K. P. (2016). The autoregulatory loop: A common mechanism of regulation of key sex determining genes in insects. J Biosci, 41(2): 283-294.

Schor, I., Llères, D., Risso, G., Pawellek, A., Ule, J., Lamond, A., … & Kornblihtt, A. (2012). Perturbation of chromatin structure globally affects localization and recruitment of splicing factors. PLoS One, 7(11): e48084.

Schütt, C., & Nöthiger, R. (2000). Structure, function and evolution of sex-determining systems in Dipteran insects. Development, 127(4): 667-677.

Seo, P. J., Park, M., & Park, C. (2013). Alternative splicing of transcription factors in plant responses to low temperature stress: mechanisms and functions. Planta, 237(6): 1415-1424.

Sharma, A., Heinze, S. D., Wu, Y., Kohlbrenner, T., Morilla, I., Brunner, C., Wimmer, E. A., Van De Zande, L., Robinson, M. D., Beukeboom, L. W., & Bopp, D. (2017). Male sex in houseflies is determined by Mdmd, a paralog of the generic splice factor gene CWC22. Science, 356(6338): 642-645.

Shi, X., Waiho, K., Li, X., Ikhwanuddin, M., Miao, G., Lin, F., ... & Ma, H. (2018). Female-specific snp markers provide insights into a wz/zz sex determination system for mud crabs scylla paramamosain, s. tranquebarica and s. serrata with a rapid method for genetic sex identification. BMC Genomics, 19(1): 981.

Shim, E., Zuccarello, G., & Kim, G. (2020). Sex-specific genes and their expression in the life history of the red alga bostrychia moritziana (ceramiales, rhodomelaceae). J Phycol, 57(2): 528-540.

Shukla, S. & Oberdoerffer, S. (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. Biochim Biophys Acta, 1819(7): 673-683.

Smith, C. A., & Sinclair, A. H. (2004). Sex determination: Insights from the chicken. BioEssays, 26(2): 120-132.

Smith, G., Chen, Y. R., Blissard, G. W., & Briscoe, A. D. (2014). Complete dosage compensation and sex-biased gene expression in the moth Manduca sexta. Genome Biol Evol, 6(3): 526-537.

Soto, E., Gandal, M., Gonatopoulos-Pournatzis, T., Heller, E., Luo, D., & Zheng, S. (2019). Mechanisms of neuronal alternative splicing and strategies for therapeutic interventions. J Neurosci, 39(42): 8193-8199.

Stegeman, R., Hall, H., Escobedo, S. E., Chang, H. C., & Weake, V. M. (2018). Proper splicing contributes to visual function in the aging Drosophila eye. Aging Cell, 17(5): e12817.

Stevens N.M. (1905). Studies in spermatogenesis. Washington, DC: Carnegie Institution of Washington.

Syed, N. H., Kalyna, M., Márquez, Y., Barta, A., & Brown, J. W. (2012). Alternative splicing in plants – coming of age. Trends Plant Sci, 17(10): 616-623.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res, 47(D1): D607-D613.

Taylor, J. S., & Raes, J. (2004). Duplication and divergence: The evolution of new genes and old ideas. Annu Rev Genet, 38: 615-643.

Teixeira, F., Okuniewska, M., Malone, C., Coux, R., Rio, D., & Lehmann, R. (2017). piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. Nature, 552(7684): 268-272.

Telonis-Scott, M., Kopp, A., Wayne, M. L., Nuzhdin, S. V., & McIntyre, L. M. (2008). Sex-Specific Splicing in Drosophila: Widespread Occurrence, Tissue Specificity and Evolutionary Conservation. Genetics, 181(2): 421-434.

Teshima, K. & Innan, H. (2008). Neofunctionalization of duplicated genes under the pressure of gene conversion. Genetics, 178(3): 1385-1398.

Thomas, G. W. C., Dohmen, E., Hughes, D. S. T., Murali, S. C., Poelchau, M., Glastad, K., Anstead, C. A., Ayoub, N. A., Batterham, P., Bellair, M., Binford, G. J., Chao, H., Chen, Y. H., Childers, C., Dinh, H., Doddapaneni, H. V., Duan, J. J., Dugan, S., Esposito, L. A., ... Richards, S. (2020). Gene content evolution in the arthropods. Genome Biol, 21(1): 15.

Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.P., Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci, 31(1): 8-22.

Tian, B., Pan, Z., & Lee, J. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res, 17(2): 156-165.

Tikhonov, M., Utkina, M., Maksimenko, O., & Georgiev, P. (2018). Conserved sequences in the Drosophila mod(mdg4) intron promote poly(A)-independent transcription termination and trans-splicing. Nucleic Acids Res, 46(20): 10608-10618.

Toyota, K., Miyakawa, H., Hiruta, C., Satō, T., Katayama, H., Ohira, T., ... & Iguchi, T. (2021). Sex determination and differentiation in decapod and cladoceran crustaceans: an overview of endocrine regulation. Genes, 12(2): 305.

Tower, J. & Arbeitman, M. N. (2009). The genetics of gender and life span. J Biol, 8(4): 38.

Traut, W., Niimi, T., Ikeo, K., & Sahara, K. (2006). Phylogeny of the sex-determining gene Sex-lethal in insects. Genome, 49(3): 254-262.

Umen, J. and Coelho, S. (2019). Algal sex determination and the evolution of anisogamy. Annu Rev Microbiol, 73(1): 267-291.

Verhulst, E. C., Van De Zande, L., & Beukeboom, L. W. (2010). Insect sex determination: It all evolves around transformer. Curr Opin Genet Dev, 20(4): 376-383.

Vicoso, B., & Bachtrog, D. (2013). Reversal of an ancient sex chromosome to an autosome in Drosophila. Nature, 499(7458): 332-335.

Vicoso, B., & Bachtrog, D. (2015). Numerous Transitions of Sex Chromosomes in Diptera. PLoS Biol, 13(4): e1002078.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature, 456(7221): 470-476.

Wang, M., Branco, A. T., & Lemos, B. (2018). The Y Chromosome Modulates Splicing and Sex-Biased Intron Retention Rates in Drosophila. Genetics, 208(3): 1057-1067.

Wang, Q., Abruzzi, K. C., Rosbash, M., & Rio, D. C. (2018). Striking circadian neuron diversity and cycling of Drosophila alternative splicing. eLife, 7: e35618.

Ward, C. M., Aumann, R. A., Whitehead, M. A., Nikolouli, K., Leveque, G., Gouvi, G., Fung, E., Reiling, S. J., Djambazian, H., Hughes, M. A., Whiteford, S., Caceres-Barrios, C., Nguyen, T. N. M., Choo, A., Crisp, P., Sim, S. B., Geib, S. M., Marec, F., Häcker, I., ... Schetelig, M. F. (2021). White pupae phenotype of tephritids is caused by parallel mutations of a MFS transporter. Nat Commun, 12(1): 491.

Watanabe, T. (2019). Evolution of the neural sex-determination system in insects: does fruitless homologue regulate neural sexual dimorphism in basal insects?. Insect Mol Biol, 28(6): 807-827.

Wilhelm, D., Palmer, S., & Koopman, P. (2007). Sex determination and gonadal development in mammals. Physiol Rev, 87(1): 1-28.

Willis, R. (2006). World Mythology: The Illustrated Guide. Oxford University Press.

Wu, B., Ma, L., Zhang, E., Du, J., Liu, S., Price, J., Li, S., & Zhao, Z. (2018). Sexual dimorphism of sleep regulated by juvenile hormone signaling in Drosophila. PLoS Genet, 14(4): e1007318.

Xing, Y. & Lee, C. J. (2005). Protein modularity of alternatively spliced exons is associated with tissue- specific regulation of alternative splicing. PLoS Genet, 1(3): e34.

Xu, J., Zhan, S., Chen, S., Zeng, B., Li, Z., James, A. A., Tan, A., & Huang, Y. (2017). Sexually dimorphic traits in the silkworm, Bombyx mori, are regulated by doublesex. Insect Biochem Mol Biol, 80: 42-51.

Xu, R., Martelossi, J., Smits, M., Iannello, M., Peruzza, L., Babbucci, M., … & Ghiselli, F. (2022). Multi-tissue rna-seq analysis and long-read-based genome assembly reveal complex sex-specific gene regulation and molecular evolution in the manila clam. Genome Biol Evol, 14(12): evac171.

Yang, X. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res, 16(8): 995-1004.

Yeo, G., Holste, D., Kreiman, G., & Burge, C. B. (2004). Variation in alternative splicing across human tissues. Genome Biol, 5(10): R74.

Yuan, K., Seller, C. A., Shermoen, A. W., & O'Farrell, P. H. (2016). Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. Trends Genet, 32(8): 496-507.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., ... Flicek, P. (2018). Ensembl 2018. Nucleic Acids Res, 46: D754-D761.

Zhang, T., Rosenberg, H. F., & Nei, M. (1998). Positive Darwinian Selection after Gene Duplication in Primate Ribonuclease Genes. PNAS, 95(7): 3708-3713.

Zhang, Y. E., Vibranovski, M. D., Krinsky, B. H., & Long, M. (2010). Age-dependent chromosomal distribution of male-biased genes in Drosophila. Genome Res, 20(11): 1526-1533.

Zhao, F., Yan, Y., Wang, Y., Liu, Y., & Yang, R. (2023). Splicing complexity as a pivotal feature of alternative exons in mammalian species. BMC Genomics, 24: 198.

Zheng, Z.-Z., Sun, X., Zhang, B., Pu, J., Jiang, Z.-Y., Li, M., Fan, Y.-J., & Xu, Y.-Z. (2019). Alternative splicing regulation of doublesex gene by RNA-binding proteins in the silkworm Bombyx mori. RNA Biol, 16(6): 809-820.

Zhou, H., Whitworth, C., Pozmanter, C., Neville, M. C., & Van Doren, M. (2021). Doublesex regulates fruitless expression to promote sexual dimorphism of the gonad stem cell niche. PLoS Genet, 17(3): e1009468.

6 Appendix 1

Various genes exhibit sex-specific splice forms, as shown in Figure 13. This includes genes involved in sex determination, such as *doublesex* in Figure 17, and somatic genes such as *fruitless* in Figure 18, which are illustrated with their male and female forms.

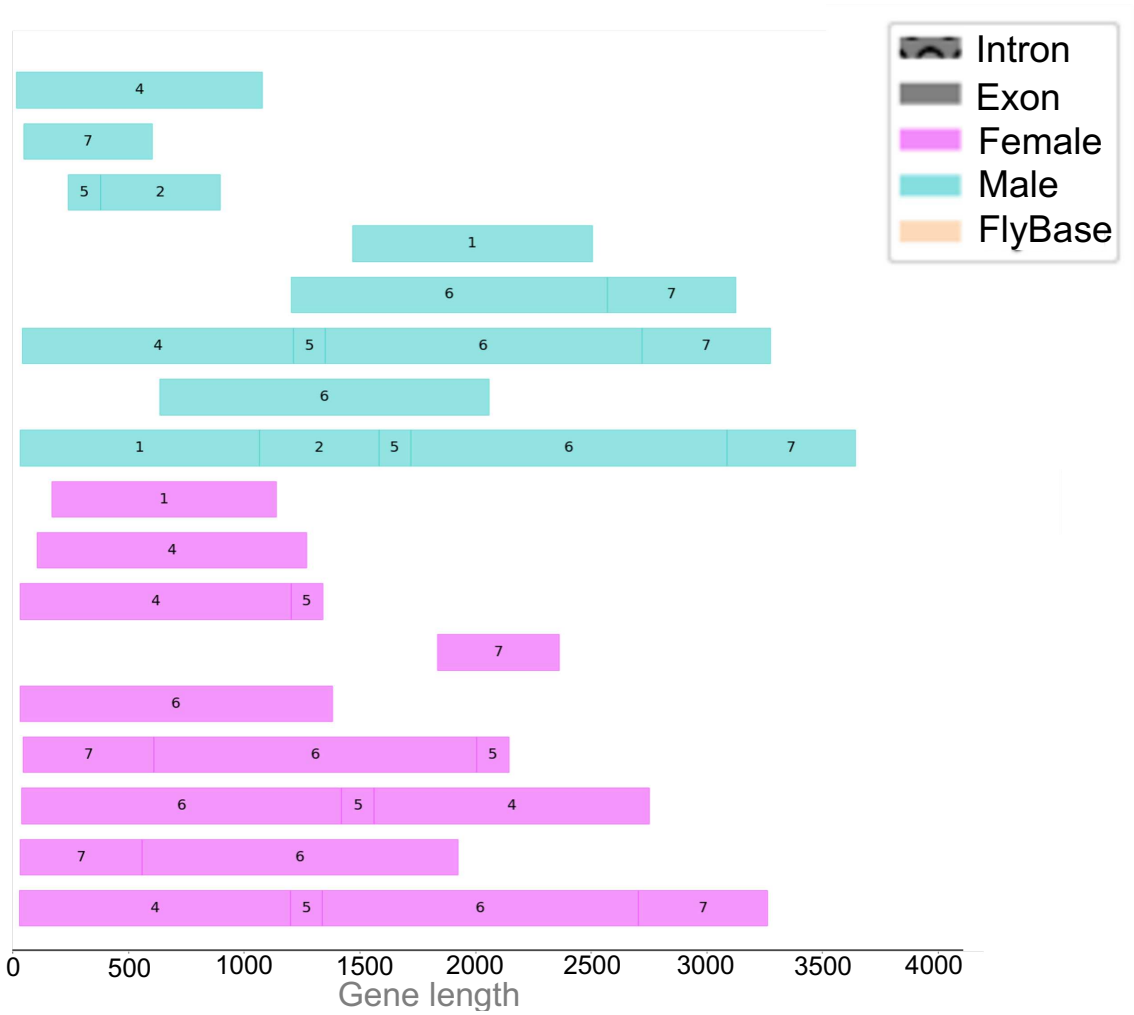

Figure 17. Splice forms of doublesex gene found in males (turquoise) and females (fuchsia).

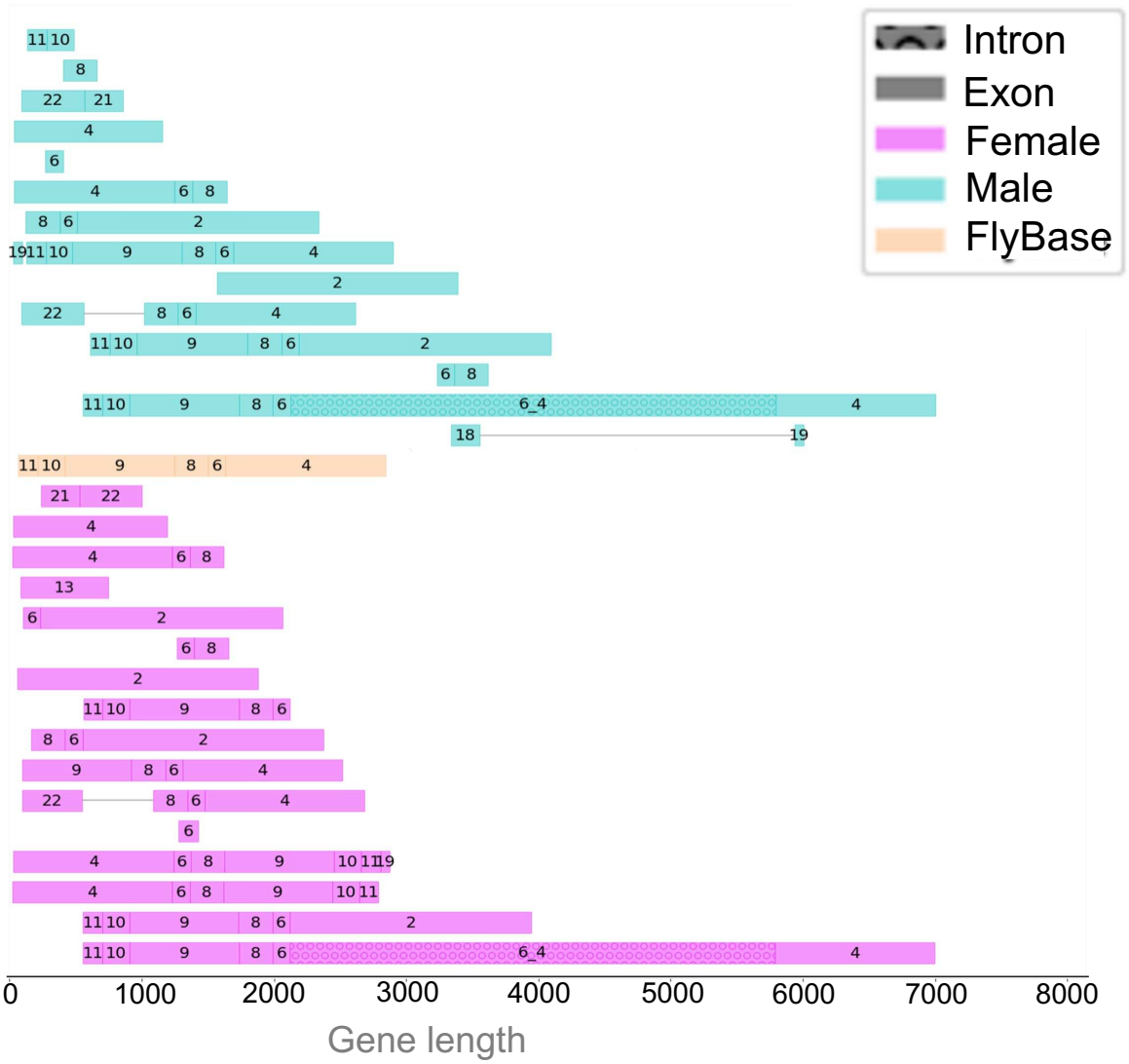

Figure 18. Splice forms of the gene fruitless in males (turquoise) and females (fuchsia).

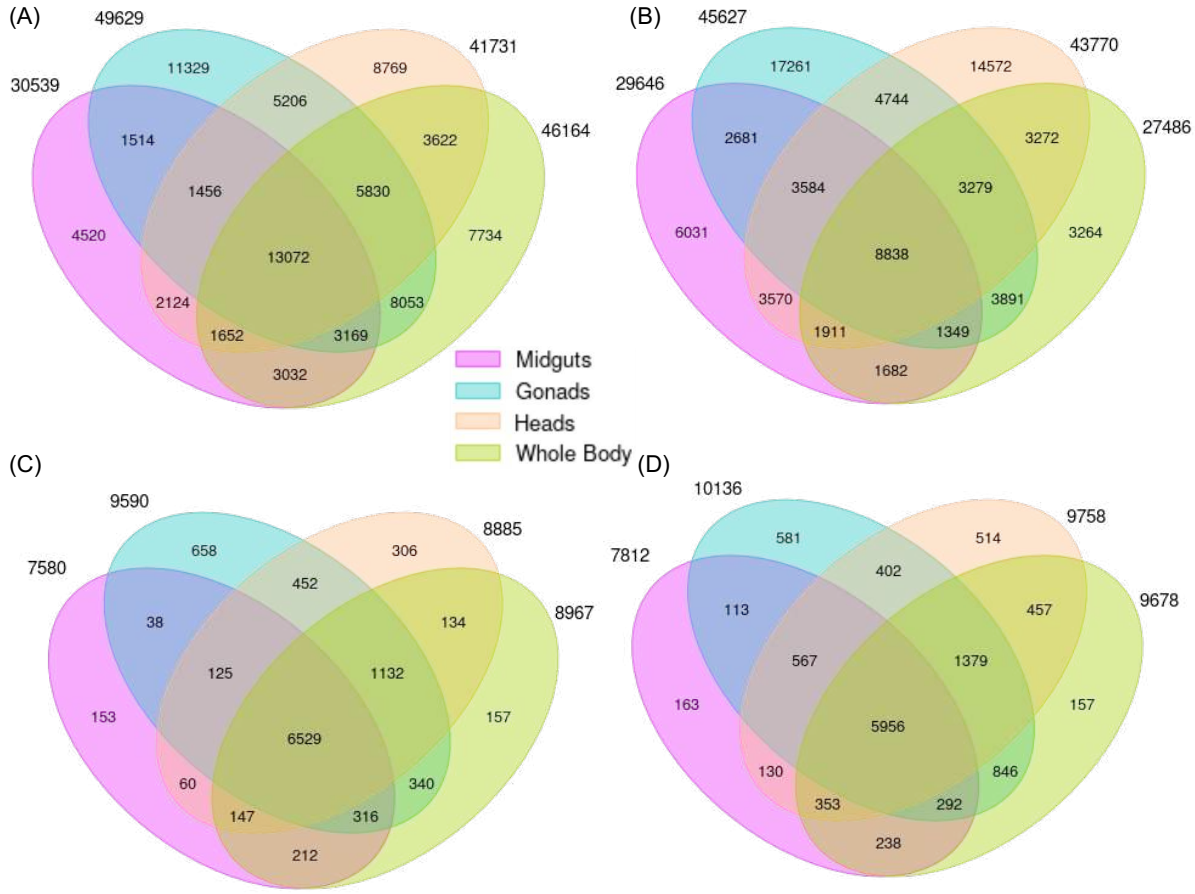

Figure 19. Venn diagrams of shared and exclusive splice forms in (a) male and (b) female tissues and the number of genes in (c) male and (d) female tissues. *It is possible to see a higher number of transcripts in females, despite the number of expressed genes being similar.*

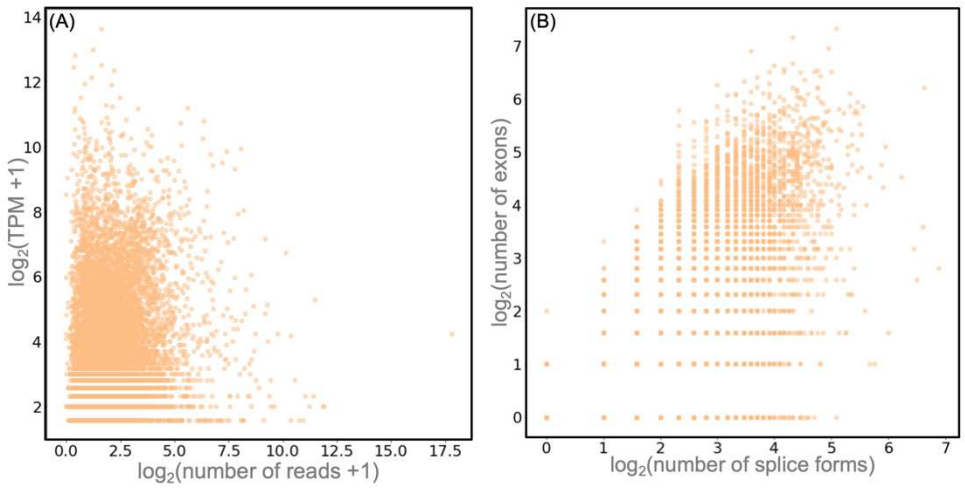

Figure 20. (A) Scatter plot of the number of reads versus expression in whole body. *Expression in TPM (transcripts per million) from FlyAtlas2 (Leader et al., 2018).* ***(B) Number of exons vs. the number of splice forms found per gene.*** *Exons present an important influence in the number of splice forms each gene presents. Genes with more exons present more splice forms than those with less exons.*

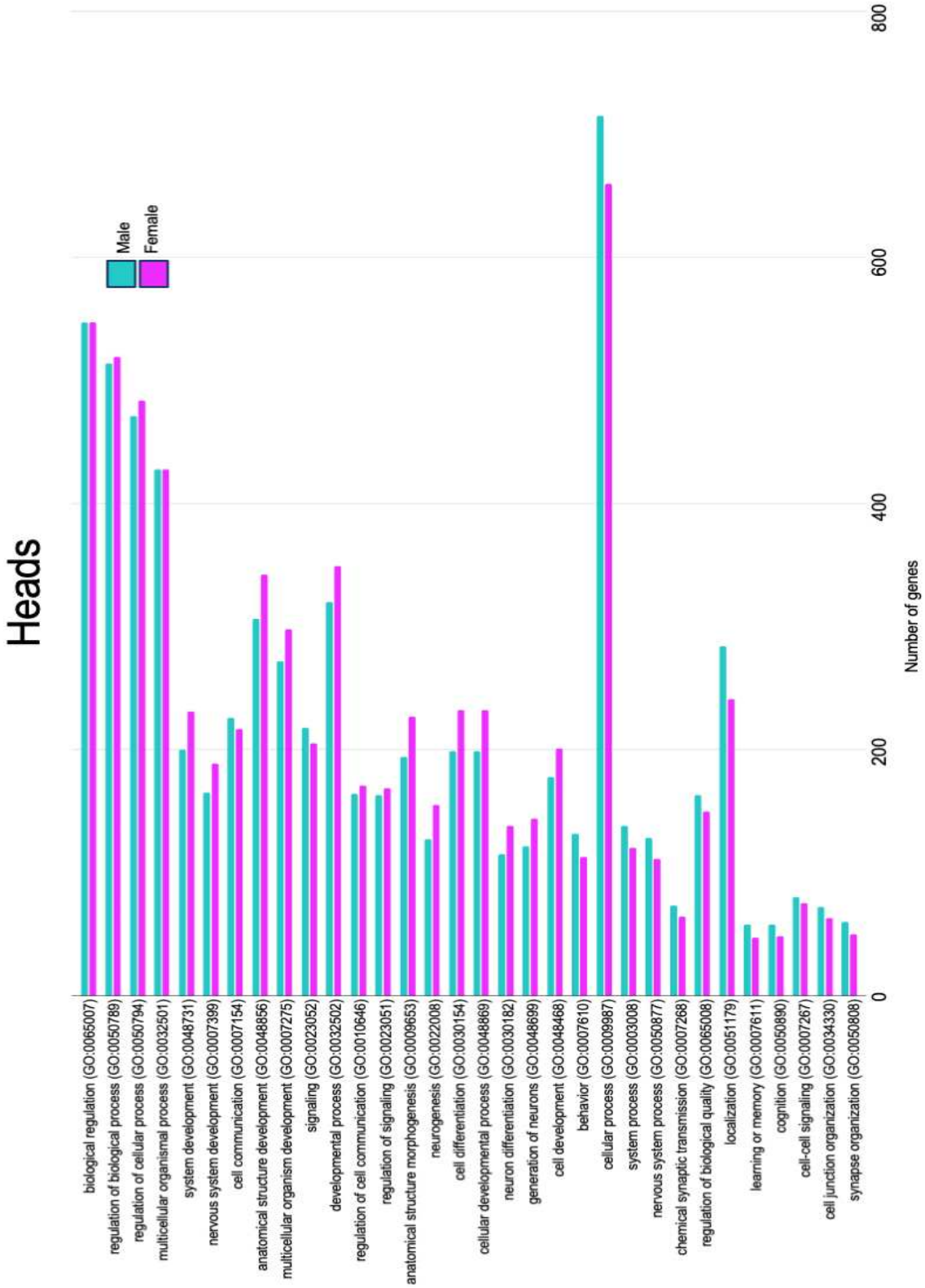

Figure 21. GO enrichment analysis for genes with head exclusive splice forms in males (turquoise) and females (pink). *Depicted are the numbers of genes for each of the twenty GO terms with lowest corrected p-values in males and females (many of the terms being repeated between both sexes in these 20 most significant category).*

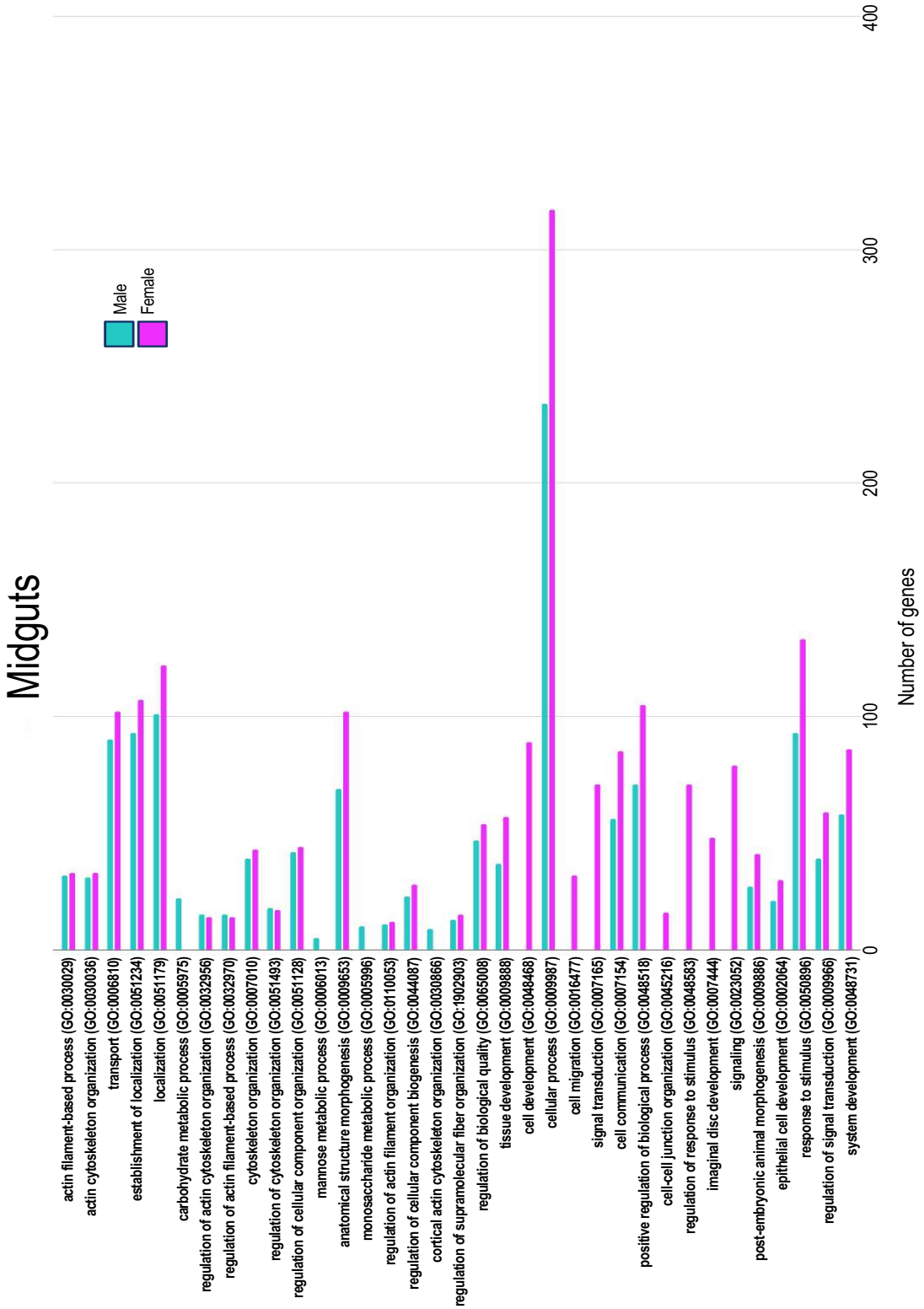

Figure 22. GO enrichment analysis for genes with midguts exclusive splice forms in males (turquoise) and females (pink). *Depicted are the numbers of genes for each of the twenty GO terms with the lowest corrected p-values in males and females (many of the terms being repeated between both sexes in this 20 most significant category). We see less terms shared between both sexes with a few only represented by genes in one sex.*

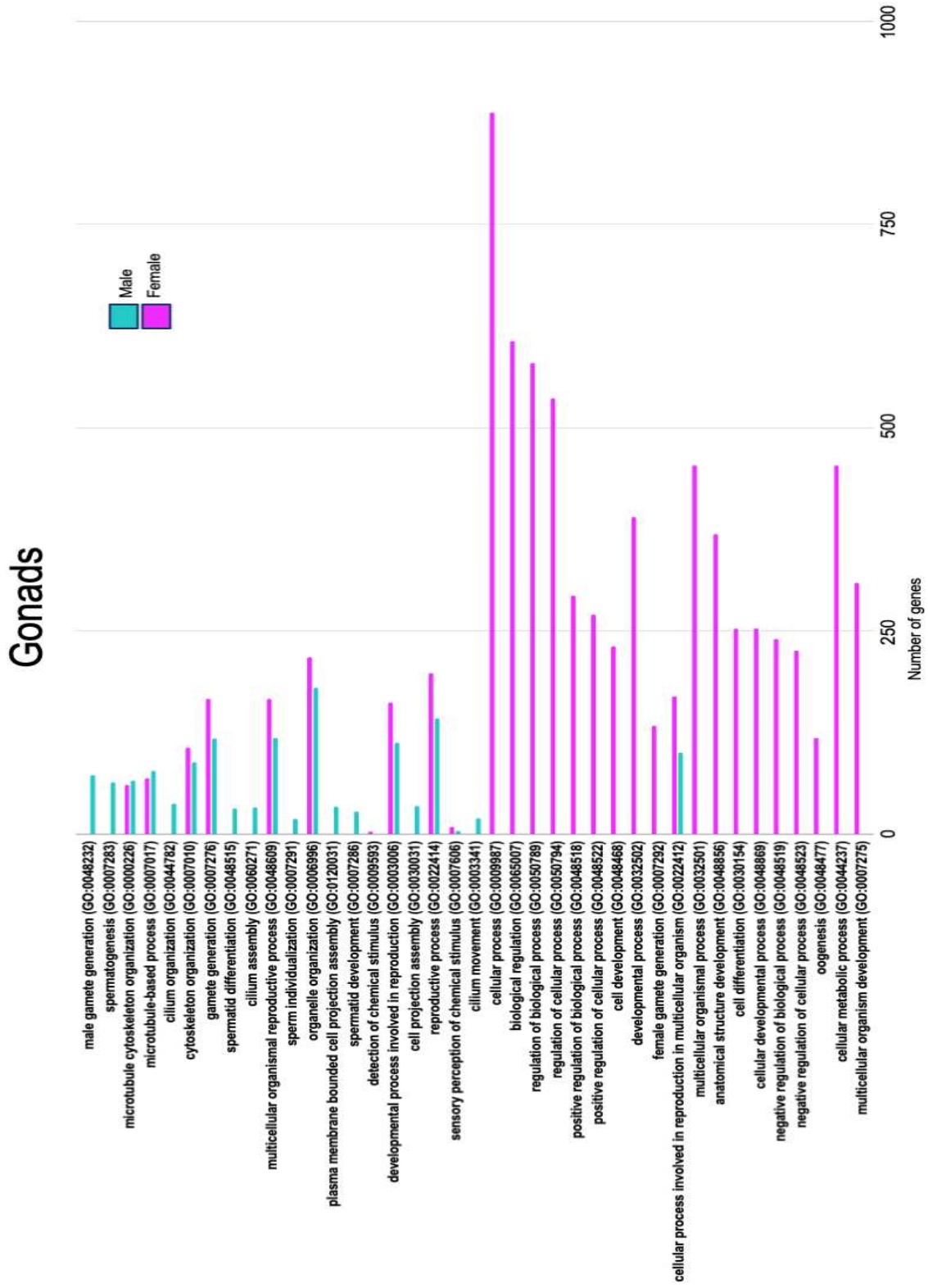

Figure 23. GO enrichment analysis for genes with gonad exclusive splice forms in males (turquoise) and females (pink). *Depicted are the numbers of genes for each of the twenty GO terms with the lowest corrected p-values in males and females (many of the terms being repeated between both sexes in this 20 most significant category). We see most terms are not shared between both sexes with most represented by genes in only one sex.*

7 Appendix 2

The embryonic stages of *D. melanogaster* and *B. jarvisi* are compared to understand the expression of counter genes in different groups. Table 3 presents all possible comparisons, highlighting the highest correlation for each *B. jarvisi* replicate used in the main text. Similar analyses is conducted for *D. melanogaster* and *B. oleae* expressions in Table 5.

Additionally, the inclusion of two genes involved in sex determination in the *B. jarvisi* expression dataset does not alter the conclusions drawn in Chapter 3. Figure 22 demonstrates this, showing the expression plots between *D. melanogaster* and *B. jarvisi* with the added genes.

Table 5. Correlation between each Drosophila melanogaster embryonic cell cycle and B. oleae embryo expression. *MSL complex and dosage compensation are expected to start at late cycle 14, sex determination starts during late cell cycle 14 (Lott et al., 2011; Yuan et al., 2016; Salz & Erickson, 2010). Here those stages have a darker background, and the best correlation to the Bactrocera expression is in bold lettering. Rep stands for replicate; c.c. stands for cell cycle.*

	D. melanogaster *embryonic cell cycle*								
	c.c. 10	c.c. 11	c.c. 12	c.c. 14a	c.c. 14b, rep. 1	c.c. 14b, rep. 2	c.c. 14c, rep. 1	c.c. 14c, rep. 2	c.c. 14d
B. oleae *embryo expression*	0.67019	0.66194	0.66361	**0.68872**	0.34165	0.6769	0.66972	0.67255	0.67764

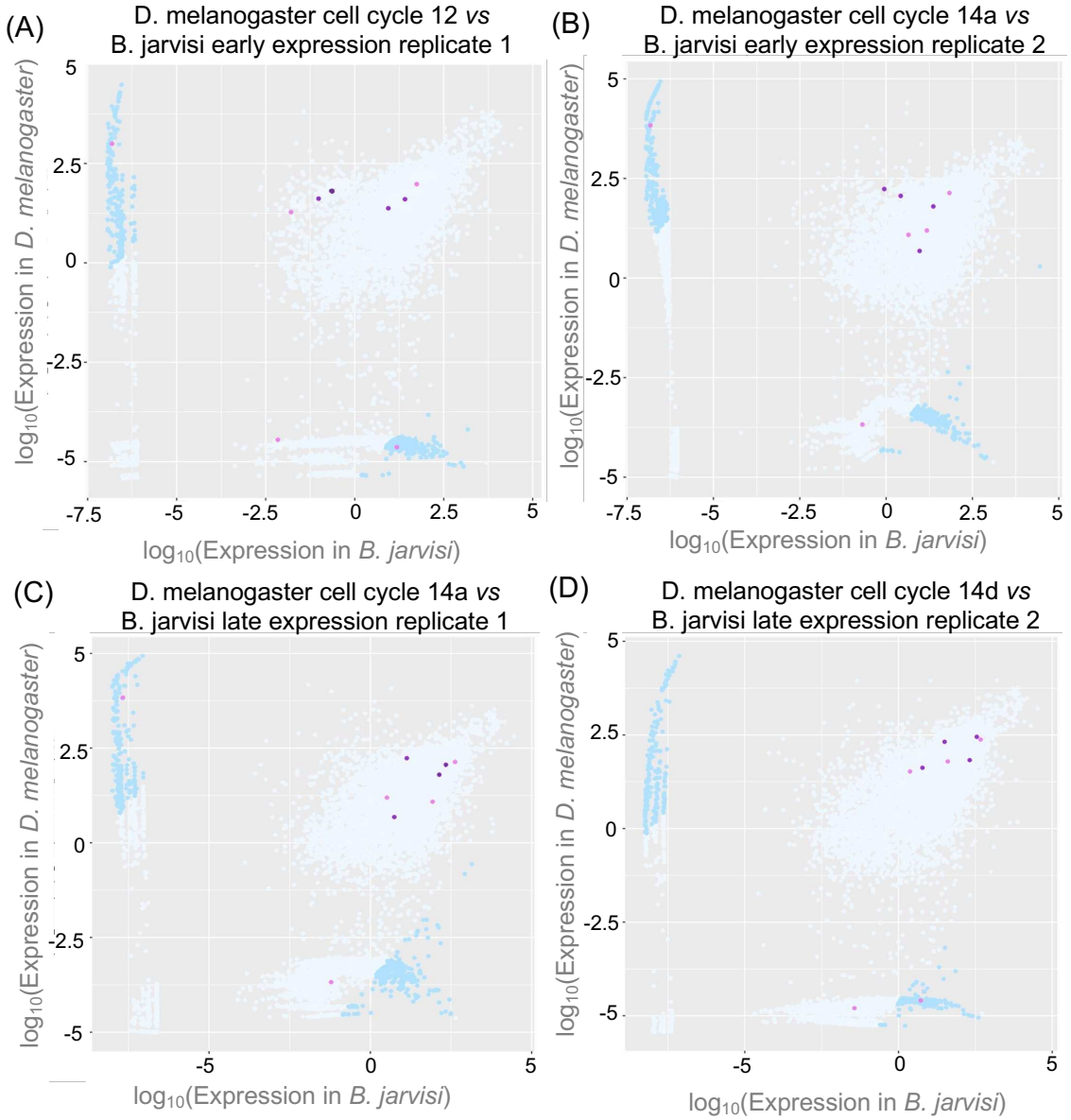

Figure 24. Comparison of expression in embryos of D. melanogaster and B. jarvisi in each corresponding cell cycle. *(A) D. melanogaster cell cycle 12 (y-axis) and B. jarvisi first replicate of early expression (x-axis). (B) D. melanogaster cell cycle 14a (y-axis) and B. jarvisi second replicate of early expression (x-axis). (C) D. melanogaster cell cycle 14a (y-axis) and B. jarvisi first replicate of late expression (x-axis). (D) D. melanogaster cell cycle 14d (y-axis) and B. jarvisi second replicate of late expression (x-axis). Darker blue indicates the 2.5% of genes higher and lower ratios of expression between D. melanogaster and B. jarvisi. The genes involved in the sex determination pathway are highlighted in purple (autosomal genes in D. melanogaster) and pink (X-linked genes in D. melanogaster). All expressions are using TPM values, normalized and in a log scale.*

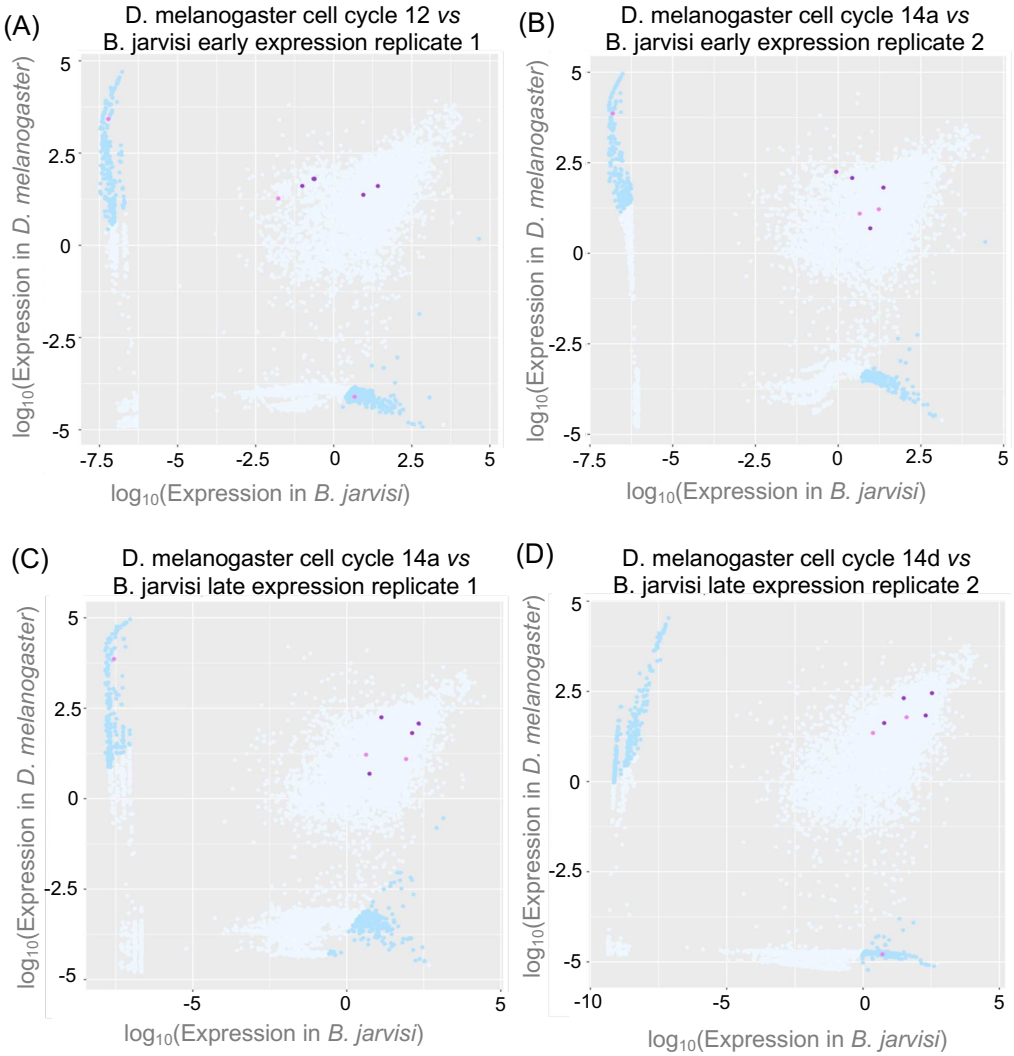

Figure 25. Comparison of expression in embryos of D. melanogaster and B. jarvisi in each corresponding cell cycle with two extra genes (sisA and upd1). *Here using the expanded set of 9 genes related to sex determination. (A) D. melanogaster cell cycle 12 (y-axis) and B. jarvisi first replicate of early expression (x-axis). (B) D. melanogaster cell cycle 14a (y-axis) and B. jarvisi second replicate of early expression (x-axis). (C) D. melanogaster cell cycle 14a (y-axis) and B. jarvisi first replicate of late expression (x-axis). (D) D. melanogaster cell cycle 14d (y-axis) and B. jarvisi second replicate of late expression (x-axis). Blue indicates the 2.5% of genes highest and lowest ratios of expression between D. melanogaster and B. jarvisi. The genes involved in the sex determination pathway are highlighted in purple (autosomal genes in D. melanogaster) and pink (X-linked genes in D. melanogaster). All expressions are using TPM values, normalized and in a log scale. We can see that except for scute (sc) and -in two cases- Sex-lethal (Sxl) all genes are in the light blue region, indicating they are in the 95% middle expression group between genes in both species.*

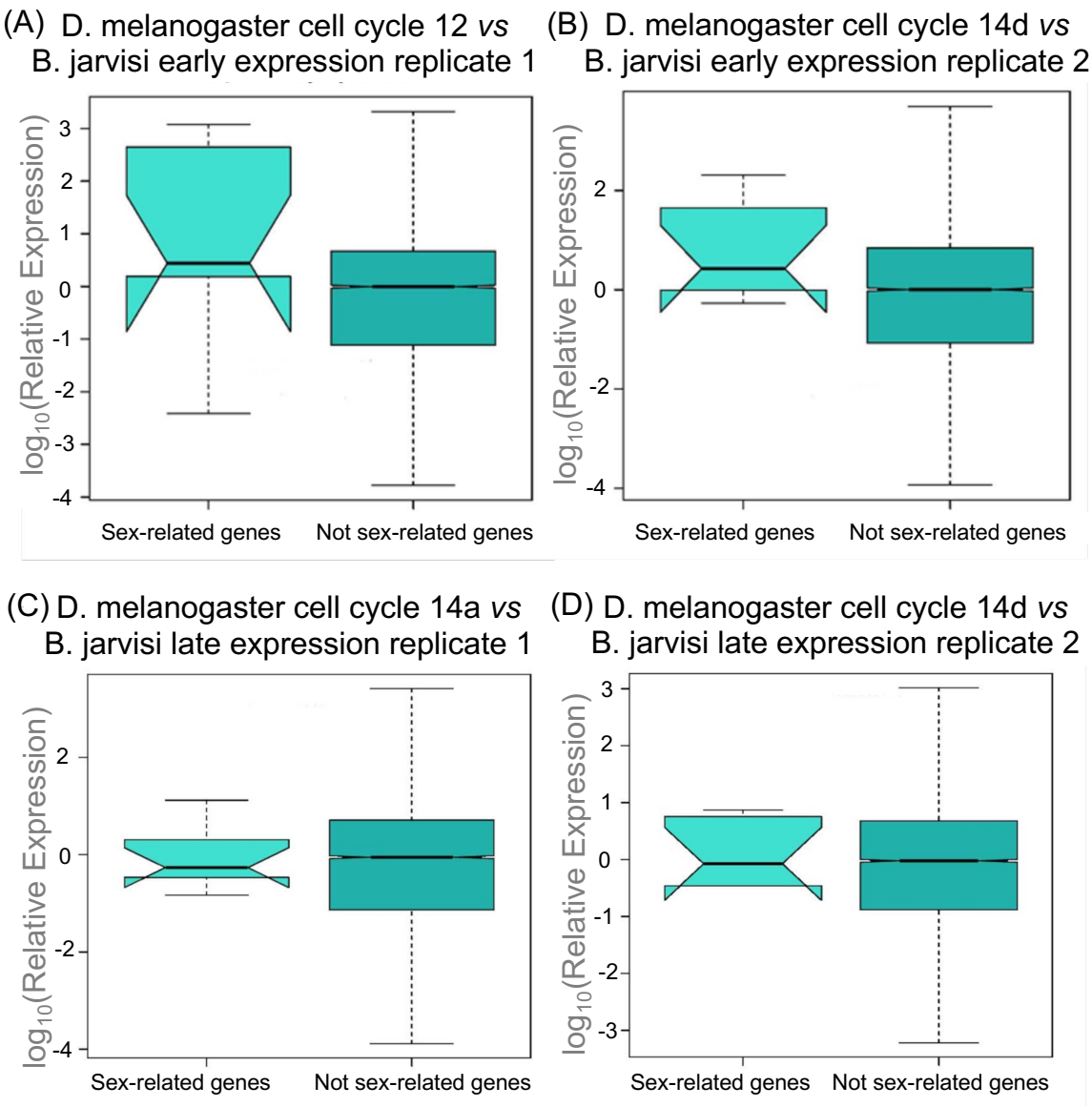

Figure 26. Relative log expression in the pairs of expression stages in D. melanogaster and B. jarvisi using the two extra genes (sisA and upd1), with outliers removed from the plot. *With the addition of those two genes, we can no longer see a significative difference between sex determination related genes relative expression and the relative expression of genes not related to sex determination in the early embryonic phases (A and B). And, in the late phases we still cannot see such pattern (C and D). This lack of difference may indicate that these genes had an optimal expression in the ancestral lineage. Still, this can also be an artefact due to a less than ideal match between the embryo phases in D. melanogaster and B. jarvisi. Asterisk (*) indicates p-value < 0.05 in a Wilcoxon test.*

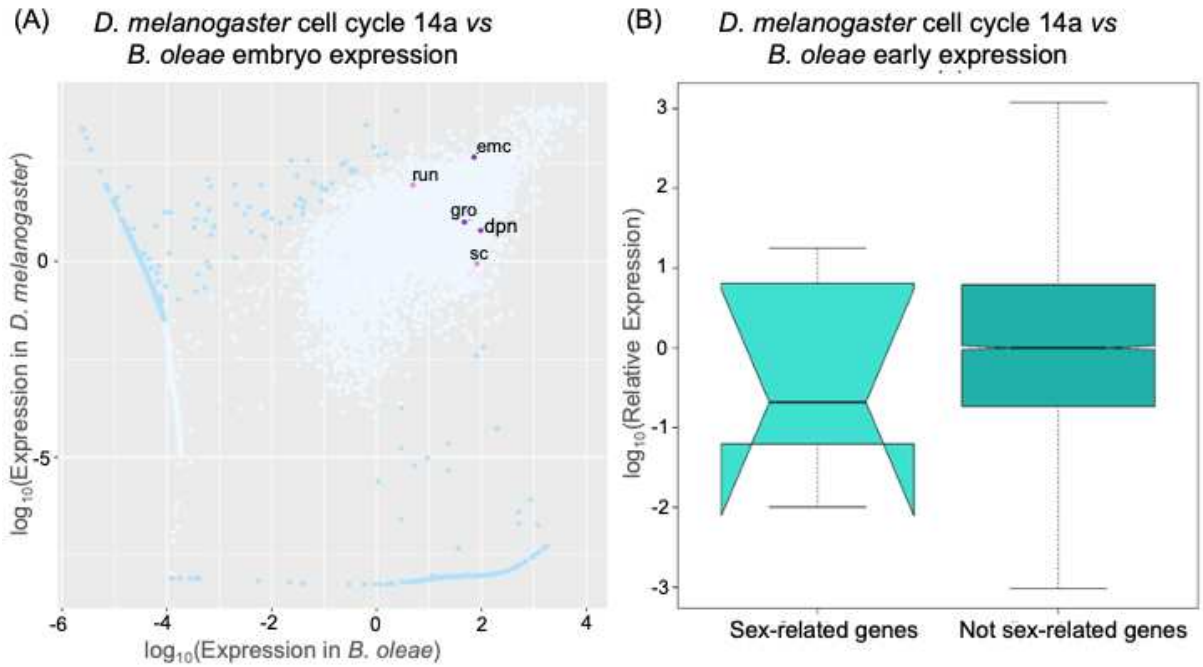

Figure 27. (A) Comparison of expression in D. melanogaster cell cycle 14a (y-axis) and B. oleae (x-axis). *Darker blue indicates the 5% genes with higher and lower ratios of expression between D. melanogaster and B. oleae. The genes involved in the sex determination pathway are highlighted in purple (autosomal genes in D. melanogaster) and pink (X-linked genes in D. melanogaster). All expressions are using TPM values, normalized, and in a log scale. We can see that all genes involved in sex determination analysed fall into the 95% intermediate ratios. This suggests that there has been no dramatic change in the expression of such genes from the ancestral lineage. And such results indicate that it is most likely that these genes already had an ideal expression level in the ancestral lineage, instead of getting to these levels after being recruited for sex determination pathway.* ***(B) Relative log expression in the pairs of expression stages in D. melanogaster and B. oleae, with outliers removed from the plot.*** *We can see no significative difference between the ratio of expression between genes involved and not involved in sex determination. This lack of difference may indicate that for these genes had an optimal expression in the ancestral lineage. Together with the same results in B. jarvisi this hypothesis seems more likely than the alternative hypothesis of the genes getting to this ideal expression levels after being recruited for sex determination pathways. Asterisk (*) indicates p-value < 0.05 in a Wilcoxon test.*