EVOLUTION OF SEX-LINKED GENE EXPRESSION IN THREESPINE STICKLEBACK FISH

by

DANIEL E. SHAW

(Under the Direction of Michael A. White)

ABSTRACT

Sex chromosomes have evolved independently many times across the tree of life. Ancient sex chromosomes have been a powerful model to understand both the evolution of sex chromosomes, as well as improve our understanding of fundamental genetics. But the sequencing of recently evolved sex chromosomes provides the opportunity to identify how selection operates on sex chromosomes in earlier stages of sex chromosome evolution and how testis genes accumulate on Y chromosomes. My dissertation focused on two projects; the first was to identify how *cis*-regulatory elements have evolved on the X and Y chromosome and the second was to identify how gene content and gene expression has evolved for testis genes across different stages of spermatogenesis. I utilized a combination of transcriptomics and epigenomics, overlayed with molecular evolution between species and within populations to identify patterns of regulatory evolution on the X and Y chromosomes. I found that *cis*-regulatory elements shared between the X and Y chromosome harbor molecular signatures of rapid evolution consistent with adaptive loss of expression on the Y chromosome. These findings help provide empirical evidence for theoretical predictions about Y chromosome evolution. I also showed that threespine stickleback sex chromosomes have evolved post-meiotic sex chromosome repression,

perhaps in response to intragenomic conflicts evolving on between the X and Y chromosome. Together, my dissertation provides insight into the evolution of sex chromosomes in stickleback fish, and advances our understanding of sex chromosome evolutionary theory.

INDEX WORDS: Sex chromosomes; Gene regulation; Molecular evolution; Spermatogenesis; Meiosis; Threespine stickleback fish

EVOLUTION OF SEX-LINKED GENE EXPRESSION IN THREESPINE STICKLEBACK FISH

by

DANIEL E. SHAW

B.S., Clemson University, 2016

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

EVOLUTION OF SEX-LINKED GENE EXPRESSION IN THREESPINE STICKLEBACK FISH

by

DANIEL E. SHAW

Major Professor:	Michael White
Committee:	Robert Schmitz
	Mary Goll
	Douglas Menke
	Casey Bergman

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

DEDICATION

This dissertation is dedicated to my dad, Fred Shaw (1958-2017).

My dad dropped out of high school and never went to a single college class. Instead, he went to work. He worked a job that he didn't care for, until the day he died. He did that so that my sister and I could grow up and live comfortably enough to pursue our own dreams.

Despite a lack of education, my dad was one of the smartest people I've ever known. While he may not have known it, we shared a similar fondness of knowledge. Many scientists I meet grew up chasing bugs with ecologists, going to space camp, or winning the science fair. I did none of those. While growing up, we watched a lot of TV, perhaps more than the recommended amount. However, some of the programming I remember the most were the science shows, documentaries, and of course, science fiction movies. The funny thing to me now, is that to my dad, the scientists in the movies were only fiction. But for me, I've been able to take that interest and turn it into a career.

I'm extremely fortunate to get to wake up every day and think about interesting questions that excite me. Few people get to do that every day, and some live their whole lives and never get to. For the lucky folks that do, I think it's important for us to remember the sacrifices of others, that paved the way so that we could follow our dreams.

My dad passed away unexpectedly a few months before I began graduate school. While I know he was proud of me, he did not get to see all of the work I have done, and he won't be here to see the first PhD in our family be conferred. But he was present the whole time, because there wasn't a day of grad school that I didn't stop and think about him and how thankful I am to have him as a dad.

Throughout college and grad school, I always felt that I was not the smartest student, because I didn't come from the best background, and I often felt like academia wasn't a good fit for me. It was in grad school, that things clicked for me that I wasn't going to be able to change any of those factors. The only thing that I could dictate was how hard I was willing to work. And the only strength I had compared to my peers was how excited I was to be here. I love science because hard work and excitement goes a long way. And for both of those features, I inherited them from my dad.

ACKNOWLEDGEMENTS

I have many people to thank for supporting me while I completed the work of this dissertation.

I would like to thank my family for supporting me on my life and academic journey. Notably my mom: Linda, sister and brother-in law: Amanda and Jim, my grandparents: James and Louise Mulligan, and of course, my nephew and niece: Jack and Cora. I appreciate their patience and understanding while I left home to pursue my degree. My mom and sister were my very first teachers. When I was younger, I was fortunate to observe perseverance, first hand, when my mom decided to complete her GED, and then associates degree, all while working full time and being a mom. Compared to that, the path of a PhD is considerably less remarkable. In addition to being a role model to me, my mom is my biggest fan. And will undoubtedly make an attempt to read this dissertation in its entirety. We'll see how far she gets.

A PhD cannot be completed without a strong support system at home. My partner, Maddie, has been exhaustively supportive of me on my journey. It is a powerful weapon to have someone who believes in you and sees things in you that others don't. Maddie has been an integral part of my progress through graduate school. I would also like to thank my dog Indy, who has worked tirelessly to keep me in tip-top shape throughout graduate school. Together, Maddie and Indy have helped make our time in Athens memorable and enjoyable.

I owe a large amount of gratitude to the White lab, including my advisor, Mike White. The White lab is a great place to learn how to do science. A supportive and inclusive environment allowed me to develop my research plans. Mike encouraged and supported my ideas and helped me grow into a scientist.

The Department of Genetics has been a great place to complete PhD work. I appreciated the support of my committee, other faculty and staff in the Department of Genetics. I also became friends with many graduate students who were an essential part of my social enrichment during graduate school. I hope to remain close with the friends and colleagues I have made here.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

IINTRODUCTION

Evolution of heteromorphic sex chromosomes

Heteromorphic sex chromosomes (i.e., X and Y) have evolved independently across multiple species (Bachtrog, 2013; Bachtrog et al., 2014). The formation of sex chromosomes is driven by the suppression of recombination, followed by the subsequent accumulation of mutations on the Y chromosome, rapidly increasing divergence between the X and Y (Bachtrog, 2013; Charlesworth & Charlesworth, 2000). Once genetic modifiers of recombination evolve (e.g. inversions), selection likely maintains suppression of recombination either to link sexual antagonistic mutations (Charlesworth, 2021) or to link Y-linked regulatory mutations that alter gene expression balance (Lenormand & Roze, 2022). Suppression of recombination reduces the efficacy of selection through processes like Muller's ratchet, genetic hitchhiking, and background selection (Charlesworth & Charlesworth, 2000), leading to the accumulation of deleterious mutations. The build-up of deleterious mutations likely leads to the degeneration of sequence.

Despite the loss of most sequence on ancient Y chromosomes, evidence of purifying selection maintaining Y gametologs (shared genes on the X and Y) is clear (Bellott et al., 2014; Bellott et al., 2017; Crowson et al., 2017; Singh et al., 2014; White et al., 2015). The conserved gametologs on Y chromosomes have been found to be enriched for dosage sensitive genes and are broadly expressed across many tissues (Bellott et al., 2014; Bellott & Page, 2021; Bellott et

al., 2017). However, despite the maintenance of ancestral genes, gene expression appears to be lower on Y alleles compared to X alleles across many species (Beaudry et al., 2017; Bellott et al., 2014; White et al., 2015). Loss of gene expression could be caused by reduced purifying selection leading to the accumulation of *cis*-regulatory variation. Or positive selection could favor the downregulation of Y alleles, in favor of increased X expression (Lenormand et al., 2020; Lenormand & Roze, 2022). In chapter 2, I will review the theoretical expectations and experimental outcomes related to the evolution of gene regulation on sex-limited chromosomes across multiple species and establish hypotheses about the evolution of *cis*-regulatory evolution on the X and Y chromosome.

In addition to the ancestral Y gametologs that are maintained, Y chromosomes often contain genes that have translocated from autosomes (Bachtrog et al., 2019; Chang et al., 2022; Peichel et al., 2020; Soh et al., 2014; Tobler et al., 2017). In contrast to the broad expression of ancestral genes, translocated genes tend to be expressed in a testis-biased manner (Bachtrog et al., 2019; Chang et al., 2022; Mueller et al., 2008; Peichel et al., 2020). These genes may have functioned, ancestrally, as male-biased, and our favored to accumulate on male-limited chromosomes. Many translocated genes have increased in copy number on the Y chromosome, that have created large ampliconic arrays of gene families (Bachtrog et al., 2019; Chang et al., 2022; Soh et al., 2014).

X chromosomes also undergo distinct patterns of molecular evolution compared to autosomes. Genes that degenerate on the Y chromosome, remain on the X chromosome in single copy (hemizygous). Selection can act on X-linked mutations in males that would otherwise be recessive on autosomal loci. This leads to a ‘faster X effect’, where the X chromosome often evolves more rapidly compared to autosomes (Charlesworth et al., 2018). X-linked genes in fruit

flies and mice have been shown to have greater rates of coding evolution compared to autosomal loci (Kopania. et al., 2022; E. L. Larson et al., 2016; Meisel & Connallon, 2013; Richard P. Meisel et al., 2012). This is especially true for genes with male specific functions, like testis biased genes. In addition to the ancestral gametologs, the X chromosome in some species have accumulated gene duplications of translocated genes that have important functions in testis (Cocquet et al., 2012; Cocquet et al., 2009; Mueller et al., 2008; Mueller et al., 2013). Additionally, extensive traffic of genes is observed on and off of X chromosomes, especially for genes involved in spermatogenesis (Bradley et al., 2004; Emerson et al., 2004; Koslowski et al., 2006).

Distinct stages of spermatogenesis exhibit diverse patterns of molecular evolution.

Testes are made up of a diverse population of somatic and germ cells. Germ cells will eventually undergo meiosis and divide into haploid cells that will give rise to sperm. The first stages of spermatogenesis include the maintenance of undifferentiated spermatogonia, and their subsequent proliferation into differentiated spermatogonia. Earlier stages of spermatogenesis are highly constrained (Bonilla & Xu, 2008) to ensure proper cell divisions and avoid infertility (Burgoyne et al., 2009). This also applies to the sex chromosomes. However, the X chromosome is predicted to accumulate male-biased genes, or genes that are expressed in male specific tissues (Rice, 1984). Indeed, multiple X chromosomes have been found to have accumulated an enrichment of pre-meiotic genes in mammals (Khil et al., 2004; Koslowski et al., 2006; Wang et al., 2001; Wang et al., 2005). However, many pre-meiotic genes may not be specialized, and are important for both sexes (Michael Koslowski et al., 2006). While most ancestral genes on the Y chromosome do not have testis-biased gene expression, a few Y-linked mammalian genes have been identified to play a role in spermatogonia (Subrini & Turner, 2021). Characterization of Y

genes is limited, but XO mice with *sry* transgenic expression are phenotypically male but do not produce differentiated spermatogonia (Mazeyrat et al., 2001), suggesting other Y-linked genes are important in this stage.

In some ways, genes involved in meiosis are highly conserved across species. Many of the core processes of meiotic cell division, such as chromosome pairing, and segregation are conserved across sexually reproducing species (Ramesh et al., 2005; Villeneuve & Hillers, 2001). Even copy number of meiotic genes seems strongly conserved, as species with whole genome duplication, have been found to preferentially lose meiotic gene duplicates (Lloyd et al., 2014). However, some genes involved in meiosis and recombination show evidence of rapid evolution and positive selection (Chowdhury et al., 2009; Dapper & Payseur, 2019; Turner et al., 2008; Wright et al., 2014). This may suggest that meiotic structural components are highly conserved, but genes involved may evolve rapidly in parallel (Bomblies et al., 2015). Our understanding of meiotic genes evolving on the sex chromosomes has been largely biased to the mammalian X and Y chromosome, which undergo complete transcriptional silencing during male meiosis, and therefore are under-enriched for male meiotic genes (See sex-linked gene expression across meiosis).

In contrast to early stages of spermatogenesis, later stages are under pressure to evolve rapidly. Sperm account for one of the most morphologically diverse cell types in vertebrates and are often under strong sexual selection (Good & Nachman, 2005; Kopania. et al., 2022; Larson et al., 2016; Swanson et al., 2003). Changes in coding sequence and expression may be due to sperm competition either between or within individuals has been associated with morphically evolution of sperm (Lüpold & Pitnick, 2018). Additionally, interactions between sperm and egg can lead to cryptic female choice (Alonzo et al., 2016), driving rapid sperm evolution.

The X and Y chromosome often play key roles in spermatogenesis. Y chromosomes are often a hotspot for accumulating male-specific genes (Bachtrog et al., 2019; Peichel et al., 2020; Soh et al., 2014) likely because Y chromosomes are sex limited, male-specific genes can evolve freely without creating deleterious effects in female. But testis genes can also evolve rapidly on the X chromosome (Kopania. et al., 2022; Mueller et al., 2008; Mueller et al., 2013) due to the faster X effect. Genes expressed in testis have been found to be enriched on both X and Y chromosomes (Bachtrog et al., 2019; Mueller et al., 2008; Soh et al., 2014), and rapidly evolving (Good & Nachman, 2005; Kopania. et al., 2022; Meisel et al., 2012; Meisel et al., 2012).

Later stages of spermatogenesis are prone to genetic conflicts between sex chromosomes. The division of cells into haploid gametes requires heterogametic sexes to divide their genome unequally, creating X- bearing and Y-bearing spermatids. This sets the stage for meiotic conflict if X or Y bearing sperm have an opportunity to increase their changes of being passed on to offspring (transmission distortion). In many species, cytoplasmic bridges connect spermatids (Greenbaum et al., 2011) allowing transfer of transcripts produced within each haploid cell. Genes on the X or Y chromosome can take advantage of the transcript transfer to increase their own transmission, if the transcripts from one chromosome can disrupt or kill gametes with the opposite allele (i.e. sperm killing). Sperm killing can result in genetic arms races, where suppressors of drive evolve to maintain sex ratio.

The most studied sex-linked meiotic drive systems are copy number variants on sex chromosomes. Sex chromosomes have been found to accumulate massive copy numbers of translocated genes (Bachtrog et al., 2019; Hughes et al., 2020; Mueller et al., 2008; Soh et al., 2014). Ampliconic sequence is hypothesized to play a role in sex-ratio distortion when they are expressed in late spermatogenesis and co-amplified on the X and Y chromosome. One example

of this is the co-amplification of *slx* and *sly* in house mice (Cocquet et al., 2012; Cocquet et al., 2009; Mueller et al., 2013; Soh et al., 2014). Co-amplification is hypothesized to be driven by the interplay between meiotic drivers and suppressors, where increases in copy number of *slx* reduces transmission of Y bearing sperm and increases in copy number of *sly* act to restore sex ratios (Cocquet et al., 2012; Cocquet et al., 2009; Kruger et al., 2019). While, expressed in post-meiotic haploid cells, *sly* and *slx* transcripts can be identified in both X-bearing and Y-bearing spermatids, as they likely are exchanged via cytoplasmic bridges. SLY and SLX proteins compete for binding domains to regulate similar genes and may alter the expression of genes in opposing directions (Moretti et al., 2020). Depletion of SLY, either through experimental knockdowns (Cocquet et al., 2009; Moretti et al., 2017), or hybrids with unbalanced *sly*/*slx* copy numbers (Larson et al., 2016), leads to infertility, and differential expression of both X-linked and autosomal genes.

Sex-linked gene expression across meiosis

Gene expression of sex chromosomes during male meiosis is often tightly regulated. The most well characterized example of this is meiotic sex chromosome inactivation (MSCI). In mammals, gene expression is completely shut down on the X and Y chromosomes during the pachytene stage of meiosis (McKee & Handel, 1993; Turner et al., 2005; Turner, 2015). No genes remain expressed during this crucial timepoint, as programmed double stranded breaks are being repaired across the genome. The reason for this drastic reduction in transcription is not known. There are several interesting ideas. MSCI could prevent transcription of regions that have DNA double stranded breaks, as these breaks last longer in mammalian meiosis compared to autosomal DSBs that are repaired with crossovers (Turner, 2007, 2015). Alternatively, MSCI may be important to physically sequester the sex chromosomes in one place, to prevent

accidental recombination with non-homologous sequence, or to ensure proper segregation in later stages (Turner, 2007). Additionally, the fact that disruption of MSCI leads to meiotic arrest, has led to the hypothesis that MSCI operates as a checkpoint to eliminate cells with meiotic defects related to synapsis (Turner, 2007, 2015).

The molecular mechanisms underlying meiotic silencing have been well studied in mammals. In mammals, MSCI has been linked to the massive accumulation of γH2AX, a phosphorylated histone variant, that forms an extensive accumulation around the asynapsed X and Y chromosome (Hirota et al., 2018; Turner et al., 2005; Turner, 2007, 2015). Canonically, γH2AX is a marker of double stranded breaks (DSBs). γH2AX foci accumulate first around DSBs across all chromosomes. As DSBs are repaired, γH2AX signal is lost across autosomes, but continue to spread along the entirety of the unsynaped regions of the X and Y chromosome during pachytene. γH2AX may form a complex with heterochromatin machinery to create the transcriptionally repressive landscape associated with complete silencing (Hirota et al., 2018). After meiosis ends, sex chromosomes remain largely repressed (Namekawa et al., 2006; Turner et al., 2006), although many genes have been discovered to reactivate and have important roles in the developing spermatid such as the amplified copies of *sly* and *slx* (Cocquet et al., 2012; Cocquet et al., 2009; Mueller et al., 2008). The mechanism responsible for establishing sex-linked heterochromatin in meiosis, may extend post-meiotically.

Alternatively, meiotic silencing and post-meiotic repression could be regulated independently. Profiling histone modifications across distinct stages of mice spermatogenesis revealed that drastic changes between meiotic and post- meiotic stages (Moretti et al., 2016). The sex chromosomes accumulate both repressive and active histone modifications in post-meiotic cells, that differ from meiotic stages with a particular enrichment of a repressive modification,

H3K9me3 (Moretti et al., 2016). Additionally, similar patterns of histone modification enrichment were observed around ampliconic sequence on an autosome, α-*takusan*. Interestingly, this gene family has been found to interact with the previously mentioned amplified genes on the sex chromosomes (*sly/slx*) in late spermatogenesis (Larson et al., 2016; Moretti et al., 2020; Moretti et al., 2016). This finding may highlight that post-meiotic chromatin changes are essential to regulate ampliconic gene duplicates that are involved in late meiosis and suggest that post-meiotic sex chromosome regulation evolved in response to intragenomic conflict and independent to meiotic sex chromosome inactivation (Moretti et al., 2016). In mammals, where both meiotic silencing and post-meiotic sex chromosome repression occur, it may be difficult to untangle the function of both processes in mice alone.

Beyond mammals, it remains unclear how universal meiotic or post-meiotic silencing is. It has been proposed to evolve from an ancestral mechanism in *C. elegans,* where regions surrounding unsynapsed regions of chromosomes undergo transcriptional silencing (Maine et al., 2005; Turner, 2015). Some evidence of gene silencing or lower sex-linked expression has been identified in other species (Guioli et al., 2012; Vibranovski, 2014; Viera et al., 2021). Additionally, meiotic silencing has only been studied in the context of ancient sex chromosomes that are heavily degenerated. Recently evolved sex chromosomes offer a unique context to study the evolution of meiotic gene expression. Identifying similar silencing in other species with recently evolved sex chromosomes would help explain the generality of these mechanisms across species. Alternatively, the existence of sex chromosomes evolving without meiotic silencing may help untangle some steps of how complex programs evolve. For example, the mammalian X has experienced extensive gene traffic of meiotic genes translocating onto other autosomes to avoid

silencing (Bradley et al., 2004; Emerson et al., 2004). Studying younger sex chromosomes could explain if these genes translocate before or after silencing evolves.

Using threespine stickleback fish as a model to understand regulatory evolution on sex chromosomes

Fish have evolved a wide range of mechanisms to determine biological sex (Bachtrog et al., 2014). In stickleback fish, sex chromosomes have evolved independently multiple times (Jeffries et al., 2022; Ross et al., 2009)**.** The threespine stickleback fish (*Gasterosteus aculeatus*) has evolved a Y chromosome within the last 26 million years (Peichel et al., 2020; Ross et al., 2009). The sex chromosomes likely evolved after recombination was suppressed through the fixation of large structural inversions. These large inversions have resulted in the formation of three distinguishable strata (Peichel et al., 2020; White et al., 2015). Each stratum has a distinct level of synonymous sequence divergence suggesting that recombination was suppressed in each region at a different time (Peichel et al., 2020). The recombination suppression and subsequent Y differentiation has occurred on a more recent time scale than Y chromosome evolution in mammals and many *Drosophila* species (Bellott et al., 2014; Bellott & Page, 2021; Bellott et al., 2017; Chang & Larracuente, 2019; Hughes et al., 2020; Soh et al., 2014).

The threespine stickleback fish Y chromosome shows early signs of degeneration, where many genes have been lost, and remaining genes contain many substitutions compared to the X gametolog (Peichel et al., 2020). Despite degeneration, at least 505 ancestral gametologs can still be identified on the Y chromosome and many of these gametologs still show expression, suggesting they are functional (Peichel et al., 2020; White et al., 2015). In its current state of differentiation, there are many more genes retained on the Y chromosome, compared to previously assembled Y chromosomes in other vertebrates (Bellott et al., 2014; Hughes et al.,

2012; Hughes et al., 2010; Hughes et al., 2020; Soh et al., 2014). In this early stage of degeneration, we can compare the relative rates of coding and regulatory variants that have accumulated on Y-linked genes. Additionally, the recent evolution of the Y chromosome allows for powerful comparative approaches between species of stickleback fish. For example, the chromosome that gave rise to the X and Y chromosome in threespine fish, has remained an autosome in multiple closely related stickleback fishes, like ninespine stickleback fish (*Pungitius pungitius*) (Jeffries et al., 2022; Ross et al., 2009). The threespine stickleback fish has a high-quality genome assembly that includes the Y chromosome (Nath et al., 2021; Peichel et al., 2020). Comparative approaches have previously identified that genes on the Y chromosome are under purifying selection and enriched with dosage sensitive regulators (Peichel et al., 2020; White et al., 2015). Interestingly, despite degeneration, there is no evidence of chromosome wide dosage compensation (White et al., 2015). Together, this makes stickleback a powerful species clade to study the evolution of allele specific gene expression and test models of rapid regulatory evolution on of ancestral gametologs on the X and Y chromosomes. In addition to ancestral genes, there is evidence of genes that have translocated on to the Y chromosome, and have testis biased expression (Peichel et al., 2020). These duplications have smaller copy number than genes observed on ancient Y chromosomes (Chang et al., 2022; Hughes et al., 2020; Mueller et al., 2008; Soh et al., 2014), and allow us to study the evolution of Y specific genes in the early stage of evolution.

Research Aims

The major goal of my thesis was to identify molecular signatures of gene expression evolution on the X and Y chromosome. First, I wanted to identify how selection operates on allele specific expression in the context of sex chromosomes. I investigated the hypothesis of

adaptive downregulation across the Y chromosome using multiple forms of genomic sequencing. Second, I investigated meiotic gene expression of sex chromosomes using single-cell RNA sequencing to test if chromosome-wide silencing occurs in meiotic or post-meiotic cells, and whether spermatogenesis genes show distinct patterns of molecular evolution. Together, these findings highlight the use of threespine stickleback fish as a model to investigate drivers of regulatory evolution on sex chromosomes.

In Chapter 2, I provide a literature review of the current state of research investigating the regulatory evolution on sex chromosomes across plants and animals. I describe the theoretical expectations of non-recombing sex chromosomes in the context of how gene regulation can evolve through multiple mechanisms. I synthesize empirical findings across a wide range of species in the context of theoretical models for sex chromosome evolution. And I discuss how to use recently evolved sex chromosomes to tests models of positive selection acting on *cis*-regulatory elements.

In Chapter 3, I identify putative *cis*-regulatory elements on the X and Y chromosome and compare rates of molecular evolution between *cis*-regulatory regions, coding regions, and intergenic sequence. I calculate rates of sequence divergence across the Y chromosome between species, and calculated nucleotide diversity within species. I compare rates of regulatory evolution to changes in gene expression to test hypotheses of adaptive downregulation of Y-linked gametologs.

In Chapter 4, I identify spermatogenesis cell types of threespine stickleback fish using single-cell RNA sequencing. I compare gene expression across spermatogenesis for sex-linked genes to test if transcriptional silencing occurs on the stickleback sex chromosomes. I identify

patterns of molecular evolution across each developmental timepoint of spermatogenesis. I identify cell-type specific gene expression profiles of sex-linked duplicates.

In chapter 5, I describe future directions of studying regulatory and molecular evolution on the threespine stickleback fish sex chromosomes. I focus on the power of utilizing additional stickleback species, such as ninespine stickleback fish, to further our understanding of regulatory evolution on sex chromosomes. I highlight the challenges of using single-cell gene expression in stickleback fish and suggest alternative approaches to study the evolution of spermatogenesis. I propose several functional genetics and genomics approaches that could be used to follow up on hypotheses related to the evolution of *cis*-regulatory evolution and stickleback spermatogenesis.

References

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113-124. https://doi.org/10.1038/nrg3366

Bachtrog, D., Mahajan, S., & Bracewell, R. (2019). Massive gene amplification on a recently formed Drosophila Y chromosome. *Nature Ecology & Evolution*, *3*(11), 1587-1597. https://doi.org/10.1038/s41559-019-1009-9

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C. (2014). Sex Determination: Why So Many Ways of Doing It? *PLoS Biology*, *12*(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899

Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2017). Genomic Loss and Silencing on the Y Chromosomes of Rumex. *Genome Biology and Evolution*, *9*(12), 3345-3355. https://doi.org/10.1093/gbe/evx254

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., Kremitzki, C., Fulton, R. S., Dugan, S., Ding, Y., Morton, D., Khan, Z., Lewis, L., Buhay, C., Wang, Q., Watt, J., Holder, M., Lee, S., Nazareth, L., Alföldi, J., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*, *508*(7497), 494-499. https://doi.org/10.1038/nature13206

Bellott, D. W., & Page, D. C. (2021). Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Research*, *31*(2), 198-210. https://doi.org/10.1101/gr.268516.120

Bellott, D. W., Skaletsky, H., Cho, T.-J., Brown, L., Locke, D., Chen, N., Galkina, S., Pyntikova, T., Koutseva, N., Graves, T., Kremitzki, C., Warren, W. C., Clark, A. G., Gaginskaya, E., Wilson, R. K., & Page, D. C. (2017). Avian W and mammalian Y chromosomes

convergently retained dosage-sensitive regulators. *Nature Genetics*, *49*(3), 387-394. https://doi.org/10.1038/ng.3778

Bomblies, K., Higgins, J. D., & Yant, L. (2015). Meiosis evolves: adaptation to external and internal environments. *New Phytologist*, *208*(2), 306-323. https://doi.org/10.1111/nph.13499

Bonilla, E., & Xu, E. Y. (2008). Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human. *Molecular Human Reproduction*, *14*(3), 137-142. https://doi.org/10.1093/molehr/gan002

Bradley, J., Baltus, A., Skaletsky, H., Royce-Tolland, M., Dewar, K., & Page, D. C. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice. *Nature Genetics*, *36*(8), 872-876. https://doi.org/10.1038/ng1390

Burgoyne, P. S., Mahadevaiah, S. K., & Turner, J. M. A. (2009). The consequences of asynapsis for mammalian meiosis. *Nature Reviews Genetics*, *10*(3), 207-216. https://doi.org/10.1038/nrg2505

Chang, C.-H., Gregory, L. E., Gordon, K. E., Meiklejohn, C. D., & Larracuente, A. M. (2022). Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. *eLife*, *11*. https://doi.org/10.7554/elife.75795

Chang, C.-H., & Larracuente, A. M. (2019). Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. *Genetics*, *211*(1), 333-348. https://doi.org/10.1534/genetics.118.301765

Charlesworth, B., Campos, J. L., & Jackson, B. C. (2018). Faster-X evolution: Theory and evidence fromDrosophila. *Molecular Ecology*, *27*(19), 3753-3771. https://doi.org/10.1111/mec.14534

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1403), 1563-1572. https://doi.org/10.1098/rstb.2000.0717

Charlesworth, D. (2021). When and how do sex-linked regions become sex chromosomes? *Evolution*, *75*(3), 569-581. https://doi.org/10.1111/evo.14196

Chowdhury, R., Bois, P. R. J., Feingold, E., Sherman, S. L., & Cheung, V. G. (2009). Genetic Analysis of Variation in Human Meiotic Recombination. *PLoS Genetics*, *5*(9), e1000648. https://doi.org/10.1371/journal.pgen.1000648

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genetics*, *8*(9), e1002900. https://doi.org/10.1371/journal.pgen.1002900

Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS Biology*, *7*(11), e1000244. https://doi.org/10.1371/journal.pbio.1000244

Crowson, D., Barrett, S. C. H., & Wright, S. I. (2017). Purifying and Positive Selection Influence Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome System. *Molecular Biology and Evolution*, *34*(5), 1140-1154. https://doi.org/10.1093/molbev/msx064

Dapper, A. L., & Payseur, B. A. (2019). Molecular evolution of the meiotic recombination pathway in mammals. *Evolution*, *73*(12), 2368-2389. https://doi.org/10.1111/evo.13850

Emerson, J. J., Kaessmann, H., Betrán, E., & Long, M. (2004). Extensive Gene Traffic on the Mammalian X Chromosome. *Science*, *303*(5657), 537-540. https://doi.org/doi:10.1126/science.1090042

Good, J. M., & Nachman, M. W. (2005). Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Molecular Biology and Evolution*, *22*(4), 1044-1052. https://doi.org/10.1093/molbev/msi087

Greenbaum, M. P., Iwamori, T., Buchold, G. M., & Matzuk, M. M. (2011). Germ Cell Intercellular Bridges. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a005850-a005850. https://doi.org/10.1101/cshperspect.a005850

Guioli, S., Lovell-Badge, R., & Turner, J. M. A. (2012). Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. *PLoS Genetics*, *8*(3), e1002560. https://doi.org/10.1371/journal.pgen.1002560

Hirota, T., Blakeley, P., Sangrithi, M. N., Mahadevaiah, S. K., Encheva, V., Snijders, A. P., Elinati, E., Ojarikre, O. A., De Rooij, D. G., Niakan, K. K., & Turner, J. M. A. (2018). SETDB1 Links the Meiotic DNA Damage Response to Sex Chromosome Silencing in Mice. *Developmental Cell*, *47*(5), 645-659.e646. https://doi.org/10.1016/j.devcel.2018.10.004

Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Graves, T., Fulton, R. S., Dugan, S., Ding, Y., Buhay, C. J., Kremitzki, C., Wang, Q., Shen, H., Holder, M., Villasana, D., Nazareth, L. V., Cree, A., Courtney, L., Veizer, J., Kotkiewicz, H., Cho, T.-J., Koutseva, N., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C. (2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*, *483*(7387), 82-86. https://doi.org/10.1038/nature10843

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., Van Daalen, S. K. M., Minx, P. J., Fulton, R. S., McGrath, S. D., Locke, D. P., Friedman, C., Trask, B. J., Mardis, E. R., Warren, W. C., Repping, S., Rozen, S., Wilson, R. K., & Page, D. C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, *463*(7280), 536-539. https://doi.org/10.1038/nature08700

Hughes, J. F., Skaletsky, H., Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., Bellott, D. W., Cho, T.-J., Dugan-Rocha, S., Khan, Z., Kremitzki, C., Fronick, C., Graves-Lindsay, T. A., Fulton, L., Warren, W. C., Wilson, R. K., Owens, E., Womack, J. E., Murphy, W. J., Muzny, D. M., Worley, K. C., Chowdhary, B. P., Gibbs, R. A., & Page, D. C. (2020). Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Research*, *30*(12), 1716-1726. https://doi.org/10.1101/gr.269902.120

Jeffries, D. L., Mee, J. A., & Peichel, C. L. (2022). Identification of a candidate sex determination gene in Culaea inconstans suggests convergent recruitment of an Amh duplicate in two lineages of stickleback. *Journal of Evolutionary Biology*. https://doi.org/10.1111/jeb.14034

Khil, P. P., Smirnova, N. A., Romanienko, P. J., & Camerini-Otero, R. D. (2004). The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature Genetics*, *36*(6), 642-646. https://doi.org/10.1038/ng1368

Kopania, E. E. K., Larson, E. L., Callahan, C., Keeble, S., & Good, J. M. (2022). Molecular Evolution across Mouse Spermatogenesis. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msac023

Koslowski, M., Sahin, U., Huber, C., & Türeci, O. (2006). The human X chromosome is enriched for germline genes expressed in premeiotic germ cells of both sexes. *Human Molecular Genetics*, *15*(15), 2392-2399. https://doi.org/10.1093/hmg/ddl163

Koslowski, M., Sahin, U., Huber, C., & Türeci, Ö. (2006). The human X chromosome is enriched for germline genes expressed in premeiotic germ cells of both sexes. *Human Molecular Genetics*, *15*(15), 2392-2399. https://doi.org/10.1093/hmg/ddl163

Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., De Rooij, D. G., Hu, Y.-C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, *29*(21), 3699-3706.e3695. https://doi.org/10.1016/j.cub.2019.08.057

Larson, E. L., Keeble, S., Vanderpool, D., Dean, M. D., & Good, J. M. (2016). The Composite Regulatory Basis of the Large X-Effect in Mouse Speciation. *Molecular Biology and Evolution*, *34*(2), 282-295. https://doi.org/10.1093/molbev/msw243

Larson, E. L., Vanderpool, D., Keeble, S., Zhou, M., Sarver, B. A., Smith, A. D., Dean, M. D., & Good, J. M. (2016). Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. *Genetics*, *203*(4), 1841-1857. https://doi.org/10.1534/genetics.116.186825

Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex Chromosome Degeneration by Regulatory Evolution. *Current Biology*, *30*(15), 3001-3006.e3005. https://doi.org/10.1016/j.cub.2020.05.052

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663-666. https://doi.org/doi:10.1126/science.abj1813

Lloyd, A. H., Ranoux, M., Vautrin, S., Glover, N., Fourment, J., Charif, D., Choulet, F., Lassalle, G., Marande, W., Tran, J., Granier, F., Pingault, L., Remay, A., Marquis, C., Belcram, H., Chalhoub, B., Feuillet, C., Bergès, H., Sourdille, P., & Jenczewski, E. (2014). Meiotic Gene Evolution: Can You Teach a New Dog New Tricks? *Molecular Biology and Evolution*, *31*(7), 1724-1727. https://doi.org/10.1093/molbev/msu119

Maine, E. M., Hauth, J., Ratliff, T., Vought, V. E., She, X., & Kelly, W. G. (2005). EGO-1, a Putative RNA-Dependent RNA Polymerase, Is Required for Heterochromatin Assembly on Unpaired DNA during C. elegans Meiosis. *Current Biology*, *15*(21), 1972-1978. https://doi.org/10.1016/j.cub.2005.09.049

Mazeyrat, S., Saut, N., Grigoriev, V., Mahadevaiah, S. K., Ojarikre, O. A., Rattigan, Á., Bishop, C., Eicher, E. M., Mitchell, M. J., & Burgoyne, P. S. (2001). A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nature Genetics*, *29*(1), 49-53. https://doi.org/10.1038/ng717

McKee, B. D., & Handel, M. A. (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma*, *102*(2), 71-80. https://doi.org/10.1007/BF00356023

Meisel, R. P., & Connallon, T. (2013). The faster-X effect: integrating theory and data. *Trends in Genetics*, *29*(9), 537-544.

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research*, *22*(7), 1255-1265. https://doi.org/10.1101/gr.132100.111

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Faster-X Evolution of Gene Expression in Drosophila. *PLoS Genetics*, *8*(10), e1003013. https://doi.org/10.1371/journal.pgen.1003013

Moretti, C., Blanco, M., Ialy-Radio, C., Serrentino, M. E., Gobé, C., Friedman, R., Battail, C., Leduc, M., Ward, M. A., Vaiman, D., Tores, F., & Cocquet, J. (2020). Battle of the Sex Chromosomes: Competition between X and Y Chromosome-Encoded Proteins for Partner Interaction and Chromatin Occupancy Drives Multicopy Gene Expression and Evolution in Muroid Rodents. *Molecular Biology and Evolution*, *37*(12), 3453-3468. https://doi.org/10.1093/molbev/msaa175

Moretti, C., Serrentino, M.-E., Ialy-Radio, C., Delessard, M., Soboleva, T. A., Tores, F., Leduc, M., Nitschké, P., Drevet, J. R., Tremethick, D. J., Vaiman, D., Kocer, A., & Cocquet, J. (2017). SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death & Differentiation*, *24*(6), 1029-1044. https://doi.org/10.1038/cdd.2017.32

Moretti, C., Vaiman, D., Tores, F., & Cocquet, J. (2016). Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics & Chromatin*, *9*(1). https://doi.org/10.1186/s13072-016-0099-8

Mueller, J. L., Mahadevaiah, S. K., Park, P. J., Warburton, P. E., Page, D. C., & Turner, J. M. A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics*, *40*(6), 794-799. https://doi.org/10.1038/ng.126

Mueller, J. L., Skaletsky, H., Brown, L. G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W. C., Wilson, R. K., & Page, D. C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics*, *45*(9), 1083-1087. https://doi.org/10.1038/ng.2705

Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D., & Lee, J. T. (2006). Postmeiotic Sex Chromatin in the Male Germline of Mice. *Current Biology*, *16*(7), 660-667. https://doi.org/10.1016/j.cub.2006.01.066

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkab007

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02097-x

Ramesh, M. A., Malik, S.-B., & Logsdon, J. M. (2005). A Phylogenomic Inventory of Meiotic Genes. *Current Biology*, *15*(2), 185-191. https://doi.org/10.1016/j.cub.2005.01.003

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, 735-742.

Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., & Peichel, C. L. (2009). Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). *PLoS Genetics*, *5*(2), e1000391. https://doi.org/10.1371/journal.pgen.1000391

Singh, N. D., Koerich, L. B., Carvalho, A. B., & Clark, A. G. (2014). Positive and Purifying Selection on the Drosophila Y Chromosome. *Molecular Biology and Evolution*, *31*(10), 2612-2623. https://doi.org/10.1093/molbev/msu203

Soh, Y. Q., Alföldi, J., Pyntikova, T., Laura, Graves, T., Patrick, Robert, Kremitzki, C., Koutseva, N., Jacob, Rozen, S., Jennifer, Owens, E., James, William, Cao, Q., Pieter, Wesley, Richard, Skaletsky, H., & David. (2014). Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*, *159*(4), 800-813. https://doi.org/10.1016/j.cell.2014.09.052

Subrini, J., & Turner, J. (2021). Y chromosome functions in mammalian spermatogenesis. *eLife*, *10*. https://doi.org/10.7554/elife.67345

Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. *Molecular Biology and Evolution*, *20*(1), 18-20. https://doi.org/10.1093/oxfordjournals.molbev.a004233

Tobler, R., Nolte, V., & Schlötterer, C. (2017). High rate of translocation-based gene birth on the Drosophila Y chromosome. *Proceedings of the National Academy of Sciences*, *114*(44), 11721-11726. https://doi.org/10.1073/pnas.1706502114

Turner, J. M., Mahadevaiah, S. K., Fernandez-Capetillo, O., Nussenzweig, A., Xu, X., Deng, C. X., & Burgoyne, P. S. (2005). Silencing of unsynapsed meiotic chromosomes in the mouse. *Nature Genetics*, *37*(1), 41-47. https://doi.org/10.1038/ng1484

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, *134*(10), 1823-1831. https://doi.org/10.1242/dev.000018

Turner, J. M. A. (2015). Meiotic Silencing in Mammals. *Annual Review of Genetics*, *49*(1), 395-412. https://doi.org/10.1146/annurev-genet-112414-055145

Turner, J. M. A., Mahadevaiah, S. K., Ellis, P. J. I., Mitchell, M. J., & Burgoyne, P. S. (2006). Pachytene Asynapsis Drives Meiotic Sex Chromosome Inactivation and Leads to Substantial Postmeiotic Repression in Spermatids. *Developmental Cell*, *10*(4), 521-529. https://doi.org/10.1016/j.devcel.2006.02.009

Turner, T. L., Levine, M. T., Eckert, M. L., & Begun, D. J. (2008). Genomic analysis of adaptive differentiation in Drosophila melanogaster. *Genetics*, *179*(1), 455-473.

Vibranovski, M. D. (2014). Meiotic sex chromosome inactivation in Drosophila. *J Genomics*, *2*, 104-117. https://doi.org/10.7150/jgen.8178

Viera, A., Parra, M. T., Arévalo, S., García De La Vega, C., Santos, J. L., & Page, J. (2021). X Chromosome Inactivation during Grasshopper Spermatogenesis. *Genes*, *12*(12), 1844. https://doi.org/10.3390/genes12121844

Villeneuve, A. M., & Hillers, K. J. (2001). Whence meiosis? *Cell*, *106*(6), 647-650.

Wang, P. J., McCarrey, J. R., Yang, F., & Page, D. C. (2001). An abundance of X-linked genes expressed in spermatogonia. *Nature Genetics*, *27*(4), 422-426.

Wang, P. J., Page, D. C., & McCarrey, J. R. (2005). Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Human Molecular Genetics*, *14*(19), 2911-2918.

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Molecular Biology and Evolution*, *32*(8), 1981-1995. https://doi.org/10.1093/molbev/msv078

Wright, K. M., Arnold, B., Xue, K., Šurinová, M., O'Connell, J., & Bomblies, K. (2014). Selection on Meiosis Genes in Diploid and Tetraploid Arabidopsis arenosa. *Molecular Biology and Evolution*, *32*(4), 944-955. https://doi.org/10.1093/molbev/msu398

CHAPTER 2

LITERATURE REVIEW: THE EVOLUTION OF GENE REGULATION ON SEX CHROMOSOMES[1]

[1] Shaw, D.E. and White, M.A. 2022. *Trends in Genetics*. Volume 38. Issue 8. P844-855. Reprinted here with permission.

Evolution of heteromorphic sex chromosomes

Sex chromosomes evolve from homologous autosome pairs. The classic model of sex chromosome evolution predicts that linkage between sexually antagonistic mutations and the master sex determination locus selects for the suppression of recombination between the X and Y (Bachtrog, 2013; Charlesworth, 1978). Since the classic model of sex chromosome evolution was proposed, empirical support for the role of sexually antagonistic mutations has been scarce (Bergero et al., 2019; Charlesworth, 2017; Ironside, 2010). This has led to the synthesis of several other models. Alternative explanations for recombination suppression have included meiotic drive (Úbeda et al., 2015), heterozygote advantage (Charlesworth & Wall, 1999), and genetic drift (reviewed in (Ponnikas et al., 2018)). Recently, a new model (Degeneration by Regulatory Evolution; DRE) has been developed where inversions can be favored to be retained on sex chromosomes after rapid divergence of regulatory regions between the X and Y (Lenormand & Roze, 2022). This model has been supported through simulations and does not require the accumulation of sexually antagonistic mutations to suppress recombination.

For all models, once recombination is suppressed, the Y chromosome rapidly undergoes sequence degeneration through the accumulation of deleterious mutations. This process can occur through selective interference (Bachtrog, 2008; Charlesworth & Charlesworth, 2000; Orr & Kim, 1998). Sequence degeneration can also occur in the absence of selective interference. In this scenario, regulatory and coding sequence substitutions accumulate on the Y chromosome in a gene-by-gene manner, leading to silencing of Y-linked gametologs (Lenormand et al., 2020; Lenormand & Roze, 2022). In many species, recombination suppression has continued in a stepwise manner, extending the non-cross over region into multiple, independent evolutionary strata (Bellott et al., 2014; Lahn & Page, 1999; Papadopulos et al., 2015; Peichel et al., 2020).

While extensive sequence degeneration does occur, it is not the inevitable outcome in all species. Many species possess small sex determination regions with a majority of the remaining sex chromosomes freely recombining (Grossen et al., 2012; Ma et al., 2018).

At the sequence level, the accumulation of deleterious mutations has been extensively demonstrated within coding regions of genes that are ancestrally shared between the X and Y chromosomes (gametologs). Over time, the accumulation of missense, nonsense, and frameshift mutations render most coding regions nonfunctional on the Y chromosome. On some of the most highly degenerated Y chromosomes, only a handful of ancestral Y-linked alleles remain (Bellott et al., 2014; Soh et al., 2014). Degeneration of regulatory regions has been indirectly shown by identifying gametolog-specific expression changes. This has been most thoroughly studied using RNA-seq technologies to interrogate the entire sex-linked transcriptome (Bellott et al., 2017; Martin et al., 2019; Muyle et al., 2012; Veltsos et al., 2019; White et al., 2015). However, the overall lack of high-quality Y assemblies, combined with limited annotations of sex-linked regulatory elements, has prevented directly quantifying the number and location of mutations important for regulatory evolution on sex chromosomes. To date, only a handful of chromosome-level scaffolds have been produced, restricted to mammals (Bellott et al., 2014; Hughes et al., 2020; Rhie et al., 2021; Soh et al., 2014; Xiao et al., 2021), fish (Li et al., 2021; Peichel et al., 2020; Shao et al., 2020), birds (Bellott et al., 2017; Rhie et al., 2021), and *Drosophila* (Chang et al., 2022; Mahajan et al., 2018). With new sequencing technologies, the completion of additional sex chromosome assemblies (see Box 1), will provide a foundation to characterize factors that influence gene regulation like *cis*-regulatory evolution, DNA methylation, and chromatin changes. Here, we will review the recent theoretical and empirical advances that have expanded our understanding of regulatory evolution on heteromorphic sex chromosomes. While we focus

on Y chromosomes, these concepts also apply to W chromosomes, which evolve in a similar manner (see Box 2).

Sex-linked gametologs rapidly evolve expression differences

As the X and Y chromosomes differentiate, gametolog expression can diverge substantially, with the Y-linked allele following one of several different evolutionary fates. Due to the widespread accumulation of deleterious alleles, theory predicts that expression should be lost from most gametologs across the Y chromosome either through silencing of deleterious coding substitutions (Engelstädter, 2008; Orr & Kim, 1998), or through degeneration by regulatory evolution (Lenormand et al., 2020; Lenormand & Roze, 2022). Transcriptome sequencing has confirmed lowered Y expression widespread across Y chromosomes of animals (R. P. Meisel et al., 2012; Singh et al., 2014; Wei & Bachtrog, 2019; White et al., 2015) and plants (Beaudry et al., 2017; Martin et al., 2019; Muyle et al., 2012; Muyle et al., 2018; Rodríguez Lorenzo et al., 2018; Veltsos et al., 2019). Studies of young sex chromosomes have revealed that gametologs can rapidly evolve sex-biased expression prior to the accumulation of deleterious mutations within coding regions (Martin et al., 2019; Veltsos et al., 2019; Yoshida et al., 2014) and this may be due to the accumulation of mutations within *cis*-regulatory regions (see Models of Regulatory Evolution**)**. In many species, dosage compensation can occur through up-regulation of the X-linked allele to restore expression of genes lost from the Y chromosome (Graves, 2016). This can be essential to maintain early embryonic viability of the heterogametic sex (Bellott et al., 2014; Bellott & Page, 2021). In mammals, there is a second stage, where one X chromosome is inactivated to restore balanced gene expression in females (Jegalian & Page, 1998). This process evolved gene-by-gene and does not apply universally across gametologs (Tukiainen et al., 2017).

A second class of gametologs retain expression from the Y chromosome. These genes have signatures of strong purifying selection to maintain function on the Y chromosome (Bellott et al., 2014; Bellott & Page, 2021; Bellott et al., 2017; White et al., 2015). Genes with essential cellular functions, that are broadly expressed across multiple tissues, have been independently retained across multiple lineages of mammalian Y chromosomes (Bellott et al., 2017). Similar patterns of retention have also been documented more broadly across the independently evolved Y and W chromosomes of amniotes (Bellott & Page, 2021), as well as in fish (Peichel et al., 2020), and *Drosophila* (Kaiser et al., 2011). These retained genes are often enriched for haploinsufficiency phenotypes (Bellott et al., 2014; Bellott & Page, 2021; Bellott et al., 2017; Peichel et al., 2020), suggesting they may be dosage sensitive. Gene editing approaches that knockout the Y-specific gametolog will be essential to test whether many of these retained genes are actually dosage sensitive. Retention of essential genes has also been demonstrated on plant sex-limited chromosomes. Genes that are important for the haploid phase of the plant life cycle are retained on Y chromosomes, despite widespread sequence degeneration (Crowson et al., 2017; Margarita & Dmitry, 2011).

Although there are clear signs of purifying selection within the coding regions of broadly expressed gametologs, comparative analyses of the regulatory regions are lacking. An outstanding question is whether ancestral regulatory elements are maintained through purifying selection. Alternatively, regulatory elements could evolve rapidly, but still maintain optimal expression levels by gaining new transcription factor binding sites that compensate for the repeated loss of ancestral regulatory elements during Y chromosome degeneration. Although it remains to be demonstrated whether this process has a prevalent role in maintaining expression of broadly expressed Y gametologs, the evolution of compensatory transcription factor binding

sites have been documented elsewhere in the genomes of *Drosophila* (Arnold et al., 2014) and mammals (Vermunt et al., 2016; Villar et al., 2014) for genes that are functionally critical.

Y-linked gametologs can also gain novel expression patterns relative to the X chromosome (Godfrey et al., 2020; White et al., 2015). Many gametologs on the Y chromosome are enriched for male-specific functions in spermatogenesis. This pattern can be caused by the biased retention of genes important for spermatogenesis on the Y chromosome followed by the loss of spermatogenesis function from the X-linked gametolog through regulatory evolution (Mahajan & Bachtrog, 2017). This could also occur through the evolution of spermatogenesis-specific regulation of the Y-linked gametolog after the X and Y chromosomes diverged (Martínez-Pacheco et al., 2020). Most Y-linked gametologs have been associated with spermatogenesis, but recent evidence shows that some Y-linked gametologs can also exhibit elevated expression in non-reproductive tissues in humans (Godfrey et al., 2020), suggesting male specific functions that extend beyond testes.

In addition to evolution of gametologs on the Y chromosome, many novel genes are acquired on the Y chromosome via translocation from autosomes. These genes are often further duplicated within the Y chromosome and evolve testis-specific functions. Ampliconic gene families evolve rapidly following the suppression of recombination (Bachtrog et al., 2019; Peichel et al., 2020) and can grow into massive arrays of duplicated genes on highly degenerated Y chromosomes (Chang et al., 2022; Hughes et al., 2010; Hughes et al., 2020; Skaletsky et al., 2003; Soh et al., 2014; Tomaszkiewicz et al., 2016). While gene duplication occurs genome wide (Fernández et al., 2020; Hughes et al., 2018), it appears that Y chromosomes are especially susceptible, which may be driven by intra-chromosomal conflicts between the sex chromosomes (Cocquet et al., 2012). Comparisons between sex chromosomes have shown that gene

amplification may be accelerated on X, Y, and Z chromosomes, which all pass through the male germline, indicating selection pressure on male functions may drive the expansion (Bellott et al., 2010; Mueller et al., 2013; Soh et al., 2014) (Box 2). Manipulating copy number through gene editing will provide insight whether amplicon expansion is essential for male fertility, as observed with the *Slx* and *Slx1* amplicons in house mice (Kruger et al., 2019). Additionally, Y gene duplication may be an alternative way to maintain expression of genes important for spermatogenesis. Recombination between duplicated genes through gene conversion could aid in purging deleterious mutations that accumulate in coding or regulatory regions (Sakamoto & Innan, 2022).

Multiple mechanisms contribute to altered expression patterns of gametologs

Although gain of expression occurs on Y chromosomes, the mechanism that has garnered most empirical and theoretical attention is the loss of expression of Y-linked gametologs. Loss of gene expression can be caused by many factors, including *cis*- and *trans*- regulatory changes (Hill et al., 2021; Wittkopp & Kalay, 2012). To date, *trans*-acting factors that broadly act to downregulate Y-linked gametologs have not been identified. The gene-by-gene loss of expression independently observed across taxa is most likely due to *cis*-regulatory evolution acting locally, including changes in heterochromatin, accumulation of transposable elements, differential DNA methylation, and *cis*-regulatory DNA changes at promoters and enhancers (Figure 2.1).

Mutations within *cis*-regulatory regions (promoters and enhancers) that alter transcription factor binding sites, or disrupt chromatin accessibility, can lead to down-regulation if binding of transcriptional activators are inhibited (Boyer et al., 2006). For example, deletions in the regulatory region of the mammalian Y-linked sex determination gene, *Sry,* prevents expression

of a reporter gene in the primordial gonad (Boyer et al., 2006), indicating the presence of an essential regulatory element. Mutations within regulatory regions can also lead to deleterious up-regulation of genes if sites of repressors are lost or novel activators are gained (Wittkopp & Kalay, 2012).

The accumulation of repetitive transposable elements may contribute to allele-specific gene expression across sex chromosomes (Figure 2.1). After suppression of crossing over, transposable elements accumulate on the sex-limited chromosome. It is well established that repetitive transposable elements can be silenced through the accumulation of heterochromatin (Iglesias & Moazed, 2017) and the spreading of heterochromatin can affect expression of adjacent coding regions (Elgin & Reuter, 2013; Lee, 2015; Lee & Karpen, 2017). Accumulation can occur rapidly, as apparent by the high density of transposable elements on younger sex-limited chromosomes (Mahajan et al., 2018; Peichel et al., 2020). On the neo-Y chromosomes of *Drosophila miranda*, the accumulation of transposable elements has been associated with the onset of constitutive heterochromatin formation on the neo-Y (Zhou et al., 2013), visualized cytogenetically as well as through ChIP-seq, targeting two repressive histone modifications (H3K9me3 and H3K9me2). The accumulation of transposable elements and spreading of heterochromatin is associated with the silencing of gametologs throughout the neo-Y chromosome, raising the possibility that transposable element accumulation may directly lead to silencing of genes. However, it remains difficult to determine the temporal order of events. It is possible that transcriptionally repressed genes may instead be an opportunistic location for a transposable element insertion, leading to heterochromatinization after Y gametolog silencing.

The accumulation of transposable elements does not always lead to the down-regulation of gametologs. Transposable elements can also rewire transcriptional networks by the

introduction of *cis*-regulatory elements (Chung et al., 2007). In *D. miranda,* some transposable elements contain binding sites for *MSL* (male specific lethal). Accumulation of these elements on the X chromosome cause the up-regulation of gametologs, resulting in dosage compensation in males (Ellison & Bachtrog, 2019a; Ellison & Bachtrog, 2013). Transposable elements are also responsible for the evolution of genetic sex determination in multiple species of fish (Herpin et al., 2010; Herpin et al., 2021; Schartl et al., 2018). In each case, novel *cis*-regulatory elements introduced by the elements allowed for the neo-functionalization of existing genes to initiate male development. Given the considerable number of transposable elements that accumulate on the Y chromosome, the co-option of novel binding sites from these insertions may be a mechanism to transcriptionally rewire genes throughout Y chromosome evolution. Additional analyses of regulatory evolution on the Y chromosome, will reveal the pervasiveness of this mechanism.

DNA methylation is another gene regulatory modifier likely playing a role in sex chromosome evolution (Figure 2.1). Changes in DNA methylation can have context dependent effects on gene expression, but it is often associated with the repression of transcription, especially when found at repetitive regions and CpG islands in mammals (Moore et al., 2013). Similar mechanisms may affect expression of gametologs on sex-limited chromosomes. An analysis of a set of XY gametologs in *Silene latifolia* revealed that many Y-linked gametologs within the oldest evolutionary strata had high levels of methylation in the promoter, relative to their X chromosome counterparts and many of these genes were in close proximity to transposable elements (Rodríguez Lorenzo et al., 2018). Additional work that compares X- and Y-specific methylomes will clarify the relationship between DNA methylation and expression on sex chromosomes.

Models of regulatory evolution

Mutations that randomly occur within regulatory elements can accumulate through selective interference (Beaudry et al., 2017) or through adaptive evolution. In the case of adaptive evolution, mutations that downregulate gene expression would be beneficial to prevent expression of coding regions that have accumulated deleterious mutations, rendering the protein sub- or non-functional. Selection could also favor downregulation of gametologs that produce toxic proteins, although this pattern has not been observed empirically, as these types of mutations may be rare and quickly lost from populations. Recent models suggest regulatory evolution could occur first before the degeneration of coding regions that leads to adaptative silencing (Lenormand et al., 2020; Lenormand & Roze, 2022). In the DRE model, *cis*-regulatory mutations evolve first on the Y chromosome, lowering expression. Deleterious mutations that accumulate within the coding regions of gametologs with lowered expression are more recessive, which selects for the accumulation of additional *cis*-regulatory mutations to further reduce expression. Importantly, this model does not require selective interference for deleterious mutations to initially accumulate. As silencing continues, stabilizing selection to maintain dosage balance leads to up-regulation of the X-linked allele through the evolution of *trans* acting transcription factors (Lenormand et al., 2020)**.**

The DRE model aligns with many patterns observed across sex chromosomes. First, Y chromosome gametologs can be silenced before amino acid substitutions have accumulated, as observed across two species of *Rumex (*Beaudry et al., 2017) and *D. miranda* (Bachtrog, 2006). This is consistent with mutations first accumulating in *cis*-regulatory elements in the DRE model. Second, if there is an absence of stabilizing selection for dosage compensation, Y gametolog silencing and degeneration still occurs (Lenormand et al. 2020). This helps explain

the widespread loss of Y chromosome alleles in species that do not exhibit dosage compensation for every gene (Nozawa et al., 2018; White et al., 2015; Zhou et al., 2021). In these species, many Y gametologs are lost without coordinated up-regulation of the X gametolog to compensate for dosage. Third, if stabilizing selection to maintain expression from both sex chromosomes is strong, but *trans* factors do not exist to permit gene-by-gene dosage compensation, Y chromosome degeneration does not occur (Lenormand et al., 2020). This finding may help explain the retention of similar genes across multiple sex chromosomes (Bellott et al., 2014; Bellott & Page, 2021; Bellott et al., 2017; Peichel et al., 2020). Many gametologs that are retained on Y chromosomes through strong purifying selection are enriched for expression across multiple tissues (Bellott et al., 2014). A single *trans* factor may not easily evolve to universally upregulate the X gametolog to compensate for dosage across multiple developmental stages and tissues.

The later stages of adaptive silencing through the DRE model have not yet been clearly observed on sex chromosomes. As amino acid substitutions accumulate, selection should favor substitutions within *cis*-regulatory elements on the sex-limited chromosome to further downregulate the gametolog. Adaptive silencing would be evident if there was a negative correlation between amino acid substitutions and Y gametolog expression level. However, loss of Y-linked expression was not associated with high levels of divergence in coding regions on the recently evolved sex chromosome in *Rumex rothschildianus* (Beaudry et al., 2017), failing to find support for widespread adaptive silencing. In addition, this pattern has not been observed across the young neo-Y chromosomes of *Drosophila*. Expression of Y-linked gametologs was not correlated with amino acid divergence on the young neo-Y of *D. albomicans* (Wei & Bachtrog, 2019; Zhou & Bachtrog, 2012), which evolved only ~120,000 years ago, or on the

slightly older neo-Y of *D. miranda* (~1 million years old) (Bachtrog, 2006; Bachtrog et al., 2008). A correlation may only be observed during a narrow window of degeneration after amino acid substitutions begin to accumulate, but before widespread degeneration and complete silencing has occurred. Surveys of additional sex chromosomes at varied states of degeneration will help test this hypothesis.

The DRE model provides a theoretical framework for how degeneration occurs first at the regulatory region, followed by the accumulation of deleterious mutations within the coding region. To date, empirical studies have not broadly surveyed the molecular evolution of regulatory regions relative to coding sequences on sex chromosomes (Figure 2.2). Some signatures of adaptive evolution have been noted within non-coding sequences on the *D. melanogaster* X chromosome (Andolfatto, 2005), but there has not been a widespread focus on the Y chromosome. Future work will also need to focus on how regulatory evolution proceeds when genes with multiple types of functional constraints evolve on the same linked Y chromosome (Figure 2.3). One clear finding from comparative analyses of Y chromosome assemblies is that not all gametologs have the same fate of pseudogenization and dosage compensation. Additionally, the DRE model does not integrate the evolution of beneficial mutations on the Y chromosome, where Y chromosome gametologs evolve neo- or sub-functionalization through the gain of novel regulatory elements. Signatures of positive selection are prevalent across Y gametologs (Chang et al., 2022; Crowson et al., 2017; Singh et al., 2014), and expansion of Y gametolog expression domains may be more common than previously assumed (Godfrey et al., 2020).

A greater challenge will be assessing the overall importance of the different types of mechanisms that alter gametolog expression. Current models use a single mutation rate that

affects all regulatory and coding regions equally. These mutations generally have small deleterious effects and act in an additive fashion, progressively eroding Y gametolog function. In actuality, regulatory mutations will have a mix of effect sizes and will occur at different rates. Transposable element insertions, for instance, may immediately trigger complete Y gametolog repression through the spreading of repressive chromatin or DNA methylation. Understanding which types of regulatory change are most prevalent early in the evolution of Y chromosomes will be aided by producing additional high-quality reference sequences that focus on annotating functional regulatory regions for comparative analyses.

Concluding Remarks

One of the major obstacles limiting our understanding of regulatory evolution on sex chromosomes is the lack of high-quality assemblies. The high repeat content of Y chromosomes makes assemblies and mapping reads a challenge. However, recent sequencing approaches that combine long-read technologies with chromatin-interaction maps or optical mapping have made the production of reference chromosomes feasible in many cases (Chang & Larracuente, 2019; Mahajan et al., 2018; Peichel et al., 2020) (Box 1).

As high-quality assemblies are completed for the sex chromosomes across diverse taxa, detailed transcriptome and epigenome annotations (Box 3) will be required in order to empirically validate theoretical models of regulatory evolution. By utilizing approaches targeted towards regions of chromatin accessibility, DNA methylation, and histone modifications, we can gain an understanding of how gene regulatory modifications are evolving on young and old sex chromosomes. An ongoing challenge will be the identification of long-distance enhancer elements important for gene regulation on the Y chromosome (Box 3). An additional challenge will be discovering regulatory mechanisms that are unique to the sex-limited chromosome. For

example, some genes on the Y chromosome in *D. melanogaster* contain gigantic introns composed of simple satellite repeats (Fingerhut et al., 2019). These introns are transcribed and recruit specialized RNA binding proteins that are required for successful transcription and processing.

An important question remaining will be untangling the timing and importance of different types of regulatory evolution over the course of Y chromosome evolution (Figure 3). Y chromosome degeneration occurs rapidly after the suppression of recombination (Bachtrog, 2013; Lenormand et al., 2020). One empirical approach is to study overall rates of regulatory evolution in species with more recently evolved sex chromosomes. Fish (Kirkpatrick et al., 2021; Li et al., 2021; Ross et al., 2009), plants (Beaudry et al., 2017; Harkess et al., 2020; Rifkin et al., 2021), and insects (Bracewell et al., 2017; Gu et al., 2019) have proven useful taxa to study the convergent evolution of sex chromosomes over relatively short timescales.

Box 1. Improved technology has vastly accelerated the assembly of sex chromosomes

The repetitive nature of Y chromosomes makes sequencing and assembly particularly challenging. To date, most reference chromosome assemblies have avoided the heterogametic sex entirely. To overcome the highly repetitive nature of sex-limited chromosomes, long-read sequencing approaches are necessary. The earliest assemblies of Y chromosomes were completed for mammalian genomes using Single Haplotype Iterative Mapping and Sequencing (SHIMS) of bacterial artificial chromosome inserts (Bellott et al., 2014; Bellott et al., 2017; Hughes et al., 2012; Hughes et al., 2010; Hughes et al., 2020; Mueller et al., 2013; Skaletsky et al., 2003; Soh et al., 2014). These inserts contained ~150 kb fragments, which were able to span the large repeat units of the Y chromosome. The SHIMS method has been used to create higher quality assemblies than those assembled using enrichment based (Kichigin et al., 2016),

reference guided approaches (Bidon et al., 2015), and de novo short read assemblies (Jevit et al., 2021).

In addition to the SHIMS method, long-read technologies (PacBio sequencing and Oxford Nanopore Technology) have created an increasingly efficient way to obtain highly accurate chromosome-scale assemblies. These techniques can sequence long single molecule-reads up to 100 kb (Wenger et al., 2019), enabling the assembly of highly repetitive regions. These assemblies can be further error-corrected using short-read technologies to increase the accuracy of base pair calling (Zhang et al., 2020). Chromosome Conformation Sequencing (Hi-C) techniques can also be used to link contigs into full chromosome-scale scaffolds. The first two non-mammalian high quality Y chromosomes were sequenced using a combination of PacBio sequencing, Illumina short-read sequencing and Hi-C scaffolding in *Drosophila miranda* (Mahajan et al., 2018) and threespine stickleback fish (Peichel et al., 2020). Since then, multiple sex chromosomes have been assembled using similar approaches in insects (Chang et al., 2022), fish (Li et al., 2021; Shao et al., 2020), birds (Rhie et al., 2021), and mammals (Miga et al., 2020; Rhie et al., 2021; Xiao et al., 2021). These approaches are even robust enough to assemble dense heterochromatic Y regions (Chang & Larracuente, 2019). In some situations, the SHIMS method may be superior to long-read sequencing approaches. A *de novo* assembly of the human Y chromosome was found to be less contiguous than the SHIMS-assembled Y chromosome (Kuderna et al., 2019). However, the speed of the long-read approach may be an important tradeoff to rapidly increasing the number of representative Y chromosome assemblies across the tree of life.

Box 2. Similarities and differences of Y and W chromosome evolution

Heteromorphic sex chromosomes can occur in males (X/Y) and females (Z/W), both of which have evolved independently across taxa (Bachtrog et al., 2014). The formation of W and Y chromosomes are predicted to occur through similar processes and are driven by the suppression of recombination (Wright et al., 2016). Following recombination suppression, sequence degeneration, loss of gene expression from the sex-limited chromosome, and retention of dosage sensitive genes through purifying selection have all been documented on ancient W chromosomes (Ayers et al., 2013; Bellott et al., 2017). Despite the similarities, there are differences in sex-specific gene content on Y and W chromosomes. Interestingly, the chicken W chromosome has not accumulated genes with functions restricted to female tissues (Bellott et al., 2017). This is in stark contrast to the overrepresentation of testis-biased genes on Y chromosomes (Mahajan & Bachtrog, 2017; Soh et al., 2014). In addition, while some multi-copy gene families have been found on avian W chromosomes (Backström et al., 2005; Komissarov et al., 2018; Rogers et al., 2021), gene family amplification appears in higher copy numbers on Y chromosomes (Chang et al., 2022; Hughes et al., 2020; Soh et al., 2014). It remains unclear why fewer female specific genes have been acquired and amplified on W chromosomes. The difference may depend on differences in gametogenesis. For example, spermatogenesis has a longer developmental window that is prone to meiotic conflict. During oogenesis, competition between gametes must be resolved before the end of Meiosis I, before gametes are split into polar bodies. During spermatogenesis, transcripts are shared through cytoplasmic bridges (Greenbaum et al., 2011), resulting in potential conflicts throughout meiosis and in developing spermatids. Additionally, testis tissue has been found to be highly transcribed, compared to other

tissues including ovaries (Parisi et al., 2004). Elevated rates of transcription in testis may provoke the formation of new genes (Witt et al., 2019).

Box 3. Challenges remain in annotating regulatory regions across the Y chromosome

Our ability to understand how regulatory regions evolve on sex chromosomes is dependent on broadly annotating enhancer and promoter regions. Highly contiguous assemblies enable the identification of previously unannotated regions of the genome, such as repeat-dense heterochromatic regions (Chang et al., 2019; Chang & Larracuente, 2019; Nguyen & Bachtrog, 2021) including centromeres and pericentromeric regions (Chang et al., 2019; Jain et al., 2018; Logsdon et al., 2021; Nath et al., 2021; Peichel et al., 2020). Annotation of these regions will require using chromatin immunoprecipitation (ChIP-seq) in many different species to identify repressive histone modifications and variants (e.g., H3K9me3, HeK9me2, CENPA).

Regulatory regions are challenging to annotate throughout the genome and on sex chromosomes. Regulatory regions can be located far from the genes they regulate, may interact with multiple genes, and do not always have distinct sequence signatures. Putative enhancer regions can be identified using next generation sequencing targeted towards accessible chromatin regions (ATAC-seq). Additionally, ChIP-seq can be used to locate functional regulatory regions. H3K4me3 modifications are associated with promoters and H3K4me1 and H3K27ac modifications are associated with enhancers (Gates et al., 2017). Hi-C contact maps can be used to find chromatin interactions between enhancer elements and promoter regions of genes (Golov et al., 2020; Ron et al., 2017). It is important to note that genomic approaches for regulatory annotation utilize short-read sequencing, which may be challenging to accurately align to highly repetitive, degenerate sex chromosomes. In addition, sequence identity between recently evolved

sex chromosomes may be high, limiting the number of variants available to map X- and Y-specific reads.

In order to empirically test models of regulatory evolution on Y chromosomes, it will be necessary to detect signatures of selection in annotated regions. Detection of recent positive selection may be challenging on non-recombining sex chromosomes because the entire linked region will uniformly have reduced genetic variation within populations (Larracuente & Clark, 2013; Wilson Sayres, 2018; Wilson Sayres et al., 2014). Scans of selection will need to be coupled with estimates of divergence between the X and Y over longer evolutionary time scales to understand whether variants are accumulating at a higher rate within *cis*-regulatory elements compared to unconstrained regions.

Once candidate regulatory elements are identified, there are a growing number of approaches that can be used to test function. Reporter constructs, like luciferase expression assays, can be used to validate whether an element of DNA has regulatory function *in vitro* (Chabot et al., 2007; Cherry et al., 2020). CRISPR-Cas9 editing has also proven to be a promising approach to test regulatory function throughout the genome by inducing mutations within promoter regions (Liu et al., 2021) or by targeting regulatory regions based on chromatin accessibility (Borys & Younger, 2020; Mochizuki et al., 2018). To date, gene editing has only been used to target limited sites on the ancient Y chromosomes of *D. melanogaster* (Buchman & Akbari, 2019*)* and the house mouse (Imaimatsu et al., 2018). Gene editing may have a higher efficiency on younger Y chromosomes, which are generally composed of fewer repeats and less heterochromatin.

Figures

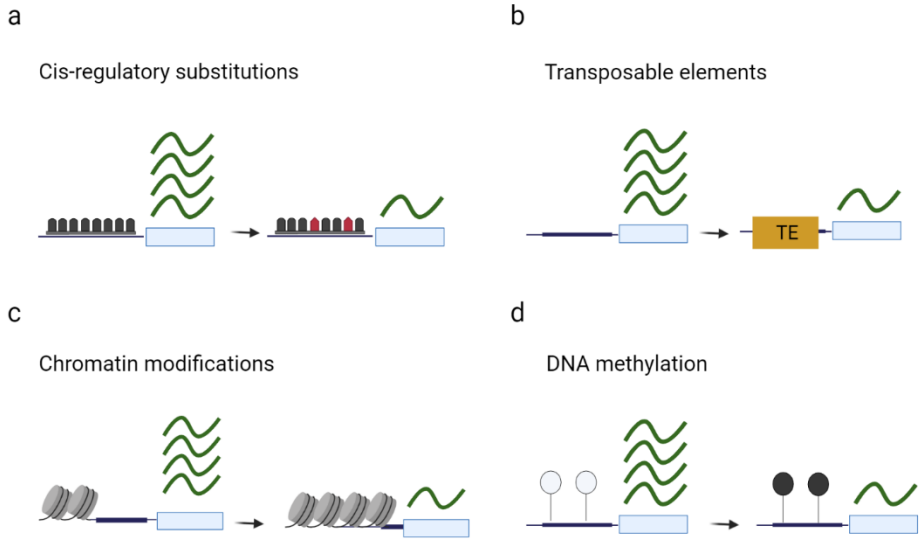


Figure 2.1 Gene expression differences on sex-limited chromosomes evolve through multiple mechanisms. (A) *Cis*-regulatory substitutions, (B) transposable elements, (C) nucleosome occupancy due to chromatin changes, and (D) DNA methylation all likely contribute to expression evolution of gametologs on sex chromosomes. Each of these mechanisms evolve rapidly and can lead to both the gain and the loss of gene expression. Created with Biorender.com.

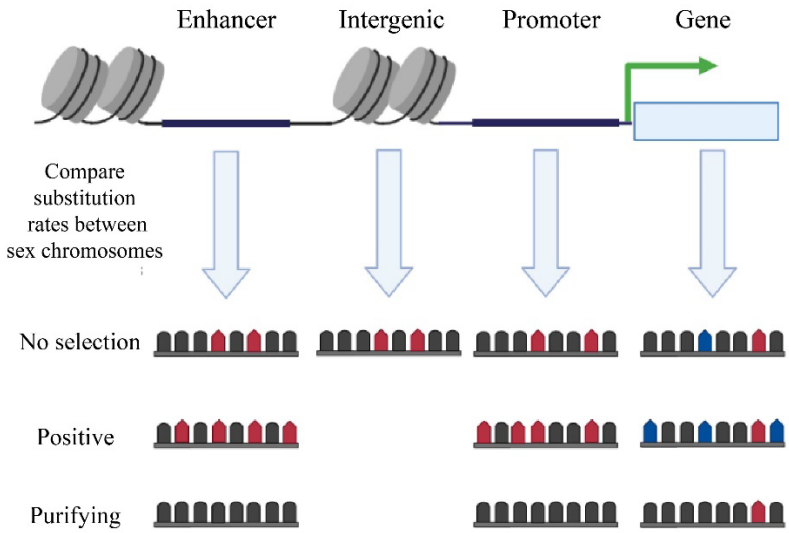


Figure 2.2 Testing for models of selection on sex-linked regulatory regions. Positive and purifying selection in functional regions throughout the genome can be identified by comparisons with neutrally evolving intergenic regions. If loss of gene expression is not driven by selection, regulatory regions would be predicted to accumulate mutations (red) at a similar rate as intergenic regions. A higher substitution rate (red mutations) within enhancers or promoters compared to intergenic regions would indicate positive selection. A lower substitution rate (red mutations) within enhancers or promoters compared to intergenic regions would indicate negative selection. Within coding regions, positive selection is indicated by more non-synonymous mutations (blue), relative to synonymous mutations (red), whereas purifying selection is indicated by fewer non-synonymous mutations, relative to synonymous mutations. Positive selection on coding regions does not have to be associated with regulatory elements under positive selection. Created with Biorender.com.

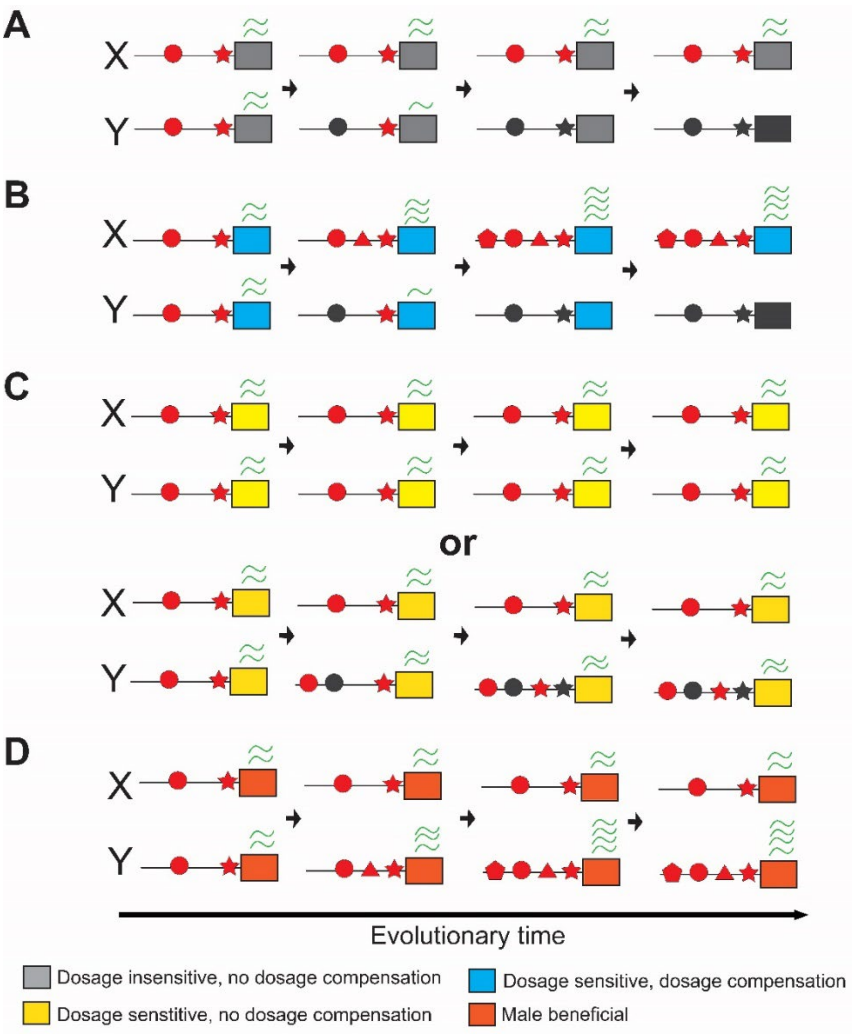


Figure 2.3. Evolution of regulatory elements on sex chromosomes. (A) Gametologs that are not dosage sensitive can accumulate mutations within *cis*-regulatory elements, down-regulating Y-linked genes. These gametologs may degenerate without X dosage compensation. Mutation accumulation can also lead to the up-regulation of Y-linked genes (not shown). (B) Stabilizing selection selects for the loss of Y expression and gain of X expression for gametologs that are dosage sensitive (DRE model). (C) Gametologs that are dosage sensitive may maintain regulatory function on the Y chromosome through purifying selection, circumventing the need for dosage compensation. These genes could also evolve new *cis*-regulatory regions to compensate for the loss of ancestral regulatory elements to maintain expression levels. (D) Gametologs can evolve new functions important for male fitness. These genes will accumulate regulatory mutations that lead to novel expression patterns (e.g., testis-specific expression). The coding regions are represented by rectangles. *Cis*-regulatory elements are represented by the shapes upstream of the coding region. Regions that have degenerated are shown in black.

References

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in Drosophila. *Nature*, *437*(7062), 1149-1152. https://doi.org/10.1038/nature04107

Arnold, C. D., Gerlach, D., Spies, D., Matts, J. A., Sytnikova, Y. A., Pagani, M., Lau, N. C., & Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nature Genetics*, *46*(7), 685-692. https://doi.org/10.1038/ng.3009

Ayers, K. L., Davidson, N. M., Demiyah, D., Roeszler, K. N., Grützner, F., Sinclair, A. H., Oshlack, A., & Smith, C. A. (2013). RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biology*, *14*(3), R26. https://doi.org/10.1186/gb-2013-14-3-r26

Bachtrog, D. (2006). Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. *Current Biology*, *16*(17), 1694-1699. https://doi.org/10.1016/j.cub.2006.07.053

Bachtrog, D. (2008). The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration. *Genetics*, *179*(3), 1513-1525. https://doi.org/10.1534/genetics.107.084012

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113-124. https://doi.org/10.1038/nrg3366

Bachtrog, D., Hom, E., Wong, K. M., Maside, X., & De Jong, P. (2008). Genomic degradation of a young Y chromosome in Drosophila miranda. *Genome Biology*, *9*(2), R30. https://doi.org/10.1186/gb-2008-9-2-r30

Bachtrog, D., Mahajan, S., & Bracewell, R. (2019). Massive gene amplification on a recently formed Drosophila Y chromosome. *Nature Ecology & Evolution*, *3*(11), 1587-1597. https://doi.org/10.1038/s41559-019-1009-9

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C. (2014). Sex Determination: Why So Many Ways of Doing It? *PLoS Biology*, *12*(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899

Backström, N., Ceplitis, H., Berlin, S., & Ellegren, H. (2005). Gene Conversion Drives the Evolution of HINTW, an Ampliconic Gene on the Female-Specific Avian W Chromosome. *Molecular Biology and Evolution*, *22*(10), 1992-1999. https://doi.org/10.1093/molbev/msi198

Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2017). Genomic Loss and Silencing on the Y Chromosomes of Rumex. *Genome Biology and Evolution*, *9*(12), 3345-3355. https://doi.org/10.1093/gbe/evx254

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., Kremitzki, C., Fulton, R. S., Dugan, S., Ding, Y., Morton, D., Khan, Z., Lewis, L., Buhay, C., Wang, Q., Watt, J., Holder, M., Lee, S., Nazareth, L., Alföldi, J., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*, *508*(7497), 494-499. https://doi.org/10.1038/nature13206

Bellott, D. W., & Page, D. C. (2021). Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Research*, *31*(2), 198-210. https://doi.org/10.1101/gr.268516.120

Bellott, D. W., Skaletsky, H., Cho, T.-J., Brown, L., Locke, D., Chen, N., Galkina, S., Pyntikova, T., Koutseva, N., Graves, T., Kremitzki, C., Warren, W. C., Clark, A. G., Gaginskaya, E., Wilson, R. K., & Page, D. C. (2017). Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature Genetics*, *49*(3), 387-394. https://doi.org/10.1038/ng.3778

Bellott, D. W., Skaletsky, H., Pyntikova, T., Mardis, E. R., Graves, T., Kremitzki, C., Brown, L. G., Rozen, S., Warren, W. C., Wilson, R. K., & Page, D. C. (2010). Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature*, *466*(7306), 612-616. https://doi.org/10.1038/nature09172

Bergero, R., Gardner, J., Bader, B., Yong, L., & Charlesworth, D. (2019). Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *Proceedings of the National Academy of Sciences*, *116*(14), 6924-6931. https://doi.org/10.1073/pnas.1818486116

Bidon, T., Schreck, N., Hailer, F., Nilsson, M. A., & Janke, A. (2015). Genome-Wide Search Identifies 1.9 Mb from the Polar Bear Y Chromosome for Evolutionary Analyses. *Genome Biology and Evolution*, *7*(7), 2010-2022. https://doi.org/10.1093/gbe/evv103

Borys, S. M., & Younger, S. T. (2020). Identification of functional regulatory elements in the human genome using pooled CRISPR screens. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-020-6497-0

Boyer, A., Pilon, N., Raiwet, D. L., Lussier, J. G., & Silversides, D. W. (2006). Human and pigSRY 5′ flanking sequences can direct reporter transgene expression to the genital ridge and to migrating neural crest cells. *Developmental Dynamics*, *235*(3), 623-632. https://doi.org/10.1002/dvdy.20670

Bracewell, R. R., Bentz, B. J., Sullivan, B. T., & Good, J. M. (2017). Rapid neo-sex chromosome evolution and incipient speciation in a major forest pest. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01761-4

Buchman, A., & Akbari, O. S. (2019). Site-specific transgenesis of the Drosophila melanogaster Y-chromosome using CRISPR/Cas9. *Insect Molecular Biology*, *28*(1), 65-73. https://doi.org/10.1111/imb.12528

Chabot, A., Shrit, R. A., Blekhman, R., & Gilad, Y. (2007). Using Reporter Gene Assays to Identify cis Regulatory Differences Between Humans and Chimpanzees. *Genetics*, *176*(4), 2069-2076. https://doi.org/10.1534/genetics.107.073429

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B. J., Wu, C.-T., Larracuente, A. M., & Mellone, B. G. (2019). Islands of retroelements are major components of Drosophila centromeres. *PLoS Biology*, *17*(5), e3000241. https://doi.org/10.1371/journal.pbio.3000241

Chang, C.-H., Gregory, L. E., Gordon, K. E., Meiklejohn, C. D., & Larracuente, A. M. (2022). Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. *eLife*, *11*. https://doi.org/10.7554/elife.75795

Chang, C.-H., & Larracuente, A. M. (2019). Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. *Genetics*, *211*(1), 333-348. https://doi.org/10.1534/genetics.118.301765

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences*, *75*(11), 5618-5622. https://doi.org/10.1073/pnas.75.11.5618

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1403), 1563-1572. https://doi.org/10.1098/rstb.2000.0717

Charlesworth, B., & Wall, J. D. (1999). Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1414), 51-56. https://doi.org/10.1098/rspb.1999.0603

Charlesworth, D. (2017). Evolution of recombination rates between sex chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160456. https://doi.org/10.1098/rstb.2016.0456

Cherry, T. J., Yang, M. G., Harmin, D. A., Tao, P., Timms, A. E., Bauwens, M., Allikmets, R., Jones, E. M., Chen, R., De Baere, E., & Greenberg, M. E. (2020). Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proceedings of the National Academy of Sciences*, *117*(16), 9001-9012. https://doi.org/10.1073/pnas.1922501117

Chung, H., Bogwitz, M. R., McCart, C., Andrianopoulos, A., Ffrench-Constant, R. H., Batterham, P., & Daborn, P. J. (2007). Cis-Regulatory Elements in the Accord Retrotransposon Result in Tissue-Specific Expression of the Drosophila melanogaster Insecticide Resistance Gene Cyp6g1. *Genetics*, *175*(3), 1071-1077. https://doi.org/10.1534/genetics.106.066597

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genetics*, *8*(9), e1002900. https://doi.org/10.1371/journal.pgen.1002900

Crowson, D., Barrett, S. C. H., & Wright, S. I. (2017). Purifying and Positive Selection Influence Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome System. *Molecular Biology and Evolution*, *34*(5), 1140-1154. https://doi.org/10.1093/molbev/msx064

Elgin, S. C. R., & Reuter, G. (2013). Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology*, *5*(8), a017780-a017780. https://doi.org/10.1101/cshperspect.a017780

Ellison, C., & Bachtrog, D. (2019). Contingency in the convergent evolution of a regulatory network: Dosage compensation in Drosophila. *PLoS Biology*, *17*(2), e3000094. https://doi.org/10.1371/journal.pbio.3000094

Ellison, C. E., & Bachtrog, D. (2013). Dosage Compensation via Transposable Element Mediated Rewiring of a Regulatory Network. *Science*, *342*(6160), 846-850. https://doi.org/10.1126/science.1239552

Engelstädter, J. (2008). Muller's Ratchet and the Degeneration of Y Chromosomes: A Simulation Study. *Genetics*, *180*(2), 957-967. https://doi.org/10.1534/genetics.108.092379

Fernández, R., Marcet-Houben, M., Legeai, F., Richard, G., Robin, S., Wucher, V., Pegueroles, C., Gabaldón, T., & Tagu, D. (2020). Selection following Gene Duplication Shapes Recent Genome Evolution in the Pea Aphid Acyrthosiphon pisum. *Molecular Biology and Evolution*, *37*(9), 2601-2615. https://doi.org/10.1093/molbev/msaa110

Fingerhut, J. M., Moran, J. V., & Yamashita, Y. M. (2019). Satellite DNA-containing gigantic introns in a unique gene expression program during Drosophila spermatogenesis. *PLoS Genetics*, *15*(5), e1008028. https://doi.org/10.1371/journal.pgen.1008028

Gates, L. A., Foulds, C. E., & O'Malley, B. W. (2017). Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends in Biochemical Sciences*, *42*(12), 977-989. https://doi.org/10.1016/j.tibs.2017.10.004

Godfrey, A. K., Naqvi, S., Chmátal, L., Chick, J. M., Mitchell, R. N., Gygi, S. P., Skaletsky, H., & Page, D. C. (2020). Quantitative analysis of Y-Chromosome gene expression across 36 human tissues. *Genome Research*, *30*(6), 860-873. https://doi.org/10.1101/gr.261248.120

Golov, A. K., Abashkin, D. A., Kondratyev, N. V., Razin, S. V., Gavrilov, A. A., & Golimbet, V. E. (2020). A modified protocol of Capture-C allows affordable and flexible high-resolution promoter interactome analysis. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-72496-4

Graves, J. A. M. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. *Nature Reviews Genetics*, *17*(1), 33-46. https://doi.org/10.1038/nrg.2015.2

Greenbaum, M. P., Iwamori, T., Buchold, G. M., & Matzuk, M. M. (2011). Germ Cell Intercellular Bridges. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a005850-a005850. https://doi.org/10.1101/cshperspect.a005850

Grossen, C., Neuenschwander, S., & Perrin, N. (2012). The Balanced Lethal System of Crested Newts: A Ghost of Sex Chromosomes Past? *The American Naturalist*, *180*(6), E174-E183. https://doi.org/10.1086/668076

Gu, L., Reilly, P. F., Lewis, J. J., Reed, R. D., Andolfatto, P., & Walters, J. R. (2019). Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. *Current Biology*, *29*(23), 4071-4077.e4073. https://doi.org/10.1016/j.cub.2019.09.056

Harkess, A., Huang, K., Van Der Hulst, R., Tissen, B., Caplan, J. L., Koppula, A., Batish, M., Meyers, B. C., & Leebens-Mack, J. (2020). Sex Determination by Two Y-Linked Genes in Garden Asparagus. *The Plant Cell*, *32*(6), 1790-1796. https://doi.org/10.1105/tpc.19.00859

Herpin, A., Braasch, I., Kraeussling, M., Schmidt, C., Thoma, E. C., Nakamura, S., Tanaka, M., & Schartl, M. (2010). Transcriptional Rewiring of the Sex Determining dmrt1 Gene Duplicate by Transposable Elements. *PLoS Genetics*, *6*(2), e1000844. https://doi.org/10.1371/journal.pgen.1000844

Herpin, A., Schartl, M., Depincé, A., Guiguen, Y., Bobe, J., Hua-Van, A., Hayman, E. S., Octavera, A., Yoshizaki, G., Nichols, K. M., Goetz, G. W., & Luckenbach, J. A. (2021). Allelic diversification after transposable element exaptation promoted gsdf as the master sex determining gene of sablefish. *Genome Research*, *31*(8), 1366-1380. https://doi.org/10.1101/gr.274266.120

Hill, M. S., Vande Zande, P., & Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*, *22*(4), 203-215. https://doi.org/10.1038/s41576-020-00304-w

Hughes, G. M., Boston, E. S. M., Finarelli, J. A., Murphy, W. J., Higgins, D. G., & Teeling, E. C. (2018). The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. *Molecular Biology and Evolution*, *35*(6), 1390-1406. https://doi.org/10.1093/molbev/msy028

Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Graves, T., Fulton, R. S., Dugan, S., Ding, Y., Buhay, C. J., Kremitzki, C., Wang, Q., Shen, H., Holder, M., Villasana, D., Nazareth, L. V., Cree, A., Courtney, L., Veizer, J., Kotkiewicz, H., Cho, T.-J., Koutseva, N., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C.

(2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*, *483*(7387), 82-86. https://doi.org/10.1038/nature10843

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., Van Daalen, S. K. M., Minx, P. J., Fulton, R. S., McGrath, S. D., Locke, D. P., Friedman, C., Trask, B. J., Mardis, E. R., Warren, W. C., Repping, S., Rozen, S., Wilson, R. K., & Page, D. C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, *463*(7280), 536-539. https://doi.org/10.1038/nature08700

Hughes, J. F., Skaletsky, H., Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., Bellott, D. W., Cho, T.-J., Dugan-Rocha, S., Khan, Z., Kremitzki, C., Fronick, C., Graves-Lindsay, T. A., Fulton, L., Warren, W. C., Wilson, R. K., Owens, E., Womack, J. E., Murphy, W. J., Muzny, D. M., Worley, K. C., Chowdhary, B. P., Gibbs, R. A., & Page, D. C. (2020). Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Research*, *30*(12), 1716-1726. https://doi.org/10.1101/gr.269902.120

Iglesias, N., & Moazed, D. (2017). Silencing repetitive DNA. *eLife*, *6*. https://doi.org/10.7554/elife.29503

Imaimatsu, K., Fujii, W., Hiramatsu, R., Miura, K., Kurohmaru, M., & Kanai, Y. (2018). CRISPR/Cas9-mediated knock-in of the murine Y chromosomal Sry gene. *Journal of Reproduction and Development*, *64*(3), 283-287. https://doi.org/10.1262/jrd.2017-161

Ironside, J. E. (2010). No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *Bioessays*, *32*(8), 718-726. https://doi.org/10.1002/bies.200900124

Jain, M., Olsen, H. E., Turner, D. J., Stoddart, D., Bulazel, K. V., Paten, B., Haussler, D., Willard, H. F., Akeson, M., & Miga, K. H. (2018). Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology*, *36*(4), 321-323. https://doi.org/10.1038/nbt.4109

Jegalian, K., & Page, D. C. (1998). A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature*, *394*(6695), 776-780. https://doi.org/10.1038/29522

Jevit, M. J., Davis, B. W., Castaneda, C., Hillhouse, A., Juras, R., Trifonov, V. A., Tibary, A., Pereira, J. C., Ferguson-Smith, M. A., & Raudsepp, T. (2021). An 8.22 Mb Assembly and Annotation of the Alpaca (Vicugna pacos) Y Chromosome. *Genes*, *12*(1), 105. https://doi.org/10.3390/genes12010105

Kaiser, V. B., Zhou, Q., & Bachtrog, D. (2011). Nonrandom Gene Loss from the Drosophila miranda Neo-Y Chromosome. *Genome Biology and Evolution*, *3*(0), 1329-1337. https://doi.org/10.1093/gbe/evr103

Kichigin, I. G., Giovannotti, M., Makunin, A. I., Ng, B. L., Kabilov, M. R., Tupikin, A. E., Barucchi, V. C., Splendiani, A., Ruggeri, P., Rens, W., O'Brien, P. C. M., Ferguson-Smith, M. A., Graphodatsky, A. S., & Trifonov, V. A. (2016). Evolutionary dynamics of Anolis sex chromosomes revealed by sequencing of flow sorting-derived microchromosome-specific DNA. *Molecular Genetics and Genomics*, *291*(5), 1955-1966. https://doi.org/10.1007/s00438-016-1230-z

Kirkpatrick, M., Sardell, J. M., Pinto, B. J., Dixon, G., Peichel, C. L., & Schartl, M. (2021). Evolution of the canonical sex chromosomes of the guppy and its relatives. *G3 Genes|Genomes|Genetics*. https://doi.org/10.1093/g3journal/jkab435

Komissarov, A. S., Galkina, S. A., Koshel, E. I., Kulak, M. M., Dyomin, A. G., O'Brien, S. J., Gaginskaya, E. R., & Saifitdinova, A. F. (2018). New high copy tandem repeat in the content of the chicken W chromosome. *Chromosoma*, *127*(1), 73-83. https://doi.org/10.1007/s00412-017-0646-5

Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., De Rooij, D. G., Hu, Y.-C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, *29*(21), 3699-3706.e3695. https://doi.org/10.1016/j.cub.2019.08.057

Kuderna, L. F. K., Lizano, E., Julià, E., Gomez-Garrido, J., Serres-Armero, A., Kuhlwilm, M., Alandes, R. A., Alvarez-Estape, M., Juan, D., Simon, H., Alioto, T., Gut, M., Gut, I., Schierup, M. H., Fornas, O., & Marques-Bonet, T. (2019). Selective single molecule sequencing and assembly of a human Y chromosome of African origin. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-07885-5

Lahn, B. T., & Page, D. C. (1999). Four Evolutionary Strata on the Human X Chromosome. *Science*, *286*(5441), 964-967. https://doi.org/doi:10.1126/science.286.5441.964

Larracuente, A. M., & Clark, A. G. (2013). Surprising Differences in the Variability of Y Chromosomes in African and Cosmopolitan Populations of Drosophila melanogaster. *Genetics*, *193*(1), 201-214. https://doi.org/10.1534/genetics.112.146167

Lee, Y. C. G. (2015). The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. *PLoS Genetics*, *11*(6), e1005269. https://doi.org/10.1371/journal.pgen.1005269

Lee, Y. C. G., & Karpen, G. H. (2017). Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife*, *6*. https://doi.org/10.7554/elife.25762

Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex Chromosome Degeneration by Regulatory Evolution. *Current Biology*, *30*(15), 3001-3006.e3005. https://doi.org/10.1016/j.cub.2020.05.052

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663-666. https://doi.org/doi:10.1126/science.abj1813

Li, M., Zhang, R., Fan, G., Xu, W., Zhou, Q., Wang, L., Li, W., Pang, Z., Yu, M., Liu, Q., Liu, X., Schartl, M., & Chen, S. (2021). Reconstruction of the Origin of a Neo-Y Sex Chromosome and Its Evolution in the Spotted Knifejaw, Oplegnathus punctatus. *Molecular Biology and Evolution*, *38*(6), 2615-2626. https://doi.org/10.1093/molbev/msab056

Liu, L., Gallagher, J., Arevalo, E. D., Chen, R., Skopelitis, T., Wu, Q., Bartlett, M., & Jackson, D. (2021). Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nature Plants*, *7*(3), 287-294. https://doi.org/10.1038/s41477-021-00858-5

Logsdon, G. A., Vollger, M. R., Hsieh, P., Mao, Y., Liskovykh, M. A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P. C., Rhie, A., De Lima, L. G., Dvorkina, T., Porubsky, D., Harvey, W. T., Mikheenko, A., Bzikadze, A. V., Kremitzki, M., Graves-Lindsay, T. A., Jain, C., Hoekzema, K., Murali, S. C., Munson, K. M., Baker, C., Sorensen, M., Lewis, A. M., Surti, U., Gerton, J. L., Larionov, V., Ventura, M., Miga, K. H., Phillippy, A. M., & Eichler, E. E. (2021). The structure, function and evolution of a complete human chromosome 8. *Nature*, *593*(7857), 101-107. https://doi.org/10.1038/s41586-021-03420-7

Ma, W.-J., Veltsos, P., Sermier, R., Parker, D. J., & Perrin, N. (2018). Evolutionary and developmental dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1548-4

Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of Y chromosome gene content in flies. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00653-x

Mahajan, S., Wei, K. H. C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biology*, *16*(7), e2006348. https://doi.org/10.1371/journal.pbio.2006348

Margarita, & Dmitry. (2011). Plant Y Chromosome Degeneration Is Retarded by Haploid Purifying Selection. *Current Biology*, *21*(17), 1475-1479. https://doi.org/10.1016/j.cub.2011.07.045

Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., Marais, G. A. B., & Touzet, P. (2019). Evolution of Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination Systems. *Genome Biology and Evolution*, *11*(2), 350-361. https://doi.org/10.1093/gbe/evz001

Martínez-Pacheco, M., Tenorio, M., Almonte, L., Fajardo, V., Godínez, A., Fernández, D., Cornejo-Páramo, P., Díaz-Barba, K., Halbert, J., Liechti, A., Székely, T., Urrutia, A. O., & Cortez, D. (2020). Expression Evolution of Ancestral XY Gametologs across All Major Groups of Placental Mammals. *Genome Biology and Evolution*, *12*(11), 2015-2028. https://doi.org/10.1093/gbe/evaa173

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research*, *22*(7), 1255-1265. https://doi.org/10.1101/gr.132100.111

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G. G., Chang, A. M., Hansen, N. F., Wilfert, A. B., Thibaud-Nissen, F., Schmitt, A. D., Belton, J.-M., Selvaraj, S., Dennis, M. Y., Soto, D. C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N. J., Holmes, N., Loose, M., Surti, U., Risques, R. A., Graves Lindsay, T. A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J. C., Pevzner, P. A., Gerton, J. L., Sullivan, B. A., Eichler, E. E., & Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, *585*(7823), 79-84. https://doi.org/10.1038/s41586-020-2547-7

Mochizuki, Y., Chiba, T., Kataoka, K., Yamashita, S., Sato, T., Kato, T., Takahashi, K., Miyamoto, T., Kitazawa, M., Hatta, T., Natsume, T., Takai, S., & Asahara, H. (2018). Combinatorial CRISPR/Cas9 Approach to Elucidate a Far-Upstream Enhancer Complex for Tissue-Specific Sox9 Expression. *Developmental Cell*, *46*(6), 794-806.e796. https://doi.org/10.1016/j.devcel.2018.07.024

Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, *38*(1), 23-38. https://doi.org/10.1038/npp.2012.112

Mueller, J. L., Skaletsky, H., Brown, L. G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W. C., Wilson, R. K., & Page, D. C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics*, *45*(9), 1083-1087. https://doi.org/10.1038/ng.2705

Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. B. (2012). Rapid De Novo Evolution of X Chromosome Dosage Compensation in Silene latifolia, a Plant with Young Sex Chromosomes. *PLoS Biology*, *10*(4), e1001308. https://doi.org/10.1371/journal.pbio.1001308

Muyle, A., Zemp, N., Fruchard, C., Cegan, R., Vrana, J., Deschamps, C., Tavares, R., Hobza, R., Picard, F., Widmer, A., & Marais, G. A. B. (2018). Genomic imprinting mediates dosage compensation in a young plant XY system. *Nature Plants*, *4*(9), 677-680. https://doi.org/10.1038/s41477-018-0221-y

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkab007

Nguyen, A. H., & Bachtrog, D. (2021). Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. *PLoS Genetics*, *17*(4), e1009438. https://doi.org/10.1371/journal.pgen.1009438

Nozawa, M., Ikeo, K., & Gojobori, T. (2018). Gene-by-gene or localized dosage compensation on the neo-X chromosome in Drosophila miranda. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evy148

Orr, H. A., & Kim, Y. (1998). An Adaptive Hypothesis for the Evolution of the Y Chromosome. *Genetics*, *150*(4), 1693-1698. https://doi.org/10.1093/genetics/150.4.1693

Papadopulos, A. S. T., Chester, M., Ridout, K., & Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences*, *112*(42), 13021-13026. https://doi.org/10.1073/pnas.1508454112

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). *Genome Biology*, *5*(6), R40. https://doi.org/10.1186/gb-2004-5-6-r40

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02097-x

Ponnikas, S., Sigeman, H., Abbott, J. K., & Hansson, B. (2018). Why Do Sex Chromosomes Stop Recombining? *Trends in Genetics*, *34*(7), 492-503. https://doi.org/10.1016/j.tig.2018.04.001

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S. C., Lama, T. M., Grutzner, F., Warren, W. C., Balakrishnan, C. N., Burt, D., George, J. M., Biegler, M. T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A. F., Franchini, P., Detrich, H. W., Svardal, H., Wagner, M., Naylor, G. J. P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B. T., Pesout, T., Houck, M., Misuraca, A., Kingan, S. B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R. E., Putnam, N. H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R. E., Guan, D., London, S. E., Clayton, D. F., Mello, C. V., Friedrich, S. R., Lovell, P. V., Osipova, E., Al-Ajli, F. O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R. S., Makova, K. D., Medvedev, P., Hoffman, J., Masterson, P., Clark,

K., Martin, F., Howe, K., Flicek, P., Walenz, B. P., Kwak, W., Clawson, H., Diekhans, M., Nassar, L., Paten, B., Kraus, R. H. S., Crawford, A. J., Gilbert, M. T. P., Zhang, G., Venkatesh, B., Murphy, R. W., Koepfli, K.-P., Shapiro, B., Johnson, W. E., Di Palma, F., Marques-Bonet, T., Teeling, E. C., Warnow, T., Graves, J. M., Ryder, O. A., Haussler, D., O'Brien, S. J., Korlach, J., Lewin, H. A., Howe, K., Myers, E. W., Durbin, R., Phillippy, A. M., & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737-746. https://doi.org/10.1038/s41586-021-03451-0

Rifkin, J. L., Beaudry, F. E. G., Humphries, Z., Choudhury, B. I., Barrett, S. C. H., & Wright, S. I. (2021). Widespread Recombination Suppression Facilitates Plant Sex Chromosome Evolution. *Molecular Biology and Evolution*, *38*(3), 1018-1030. https://doi.org/10.1093/molbev/msaa271

Rodríguez Lorenzo, J. L., Hobza, R., & Vyskot, B. (2018). DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant Silene latifolia. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4936-y

Rogers, T. F., Pizzari, T., & Wright, A. E. (2021). Multi-Copy Gene Family Evolution on the Avian W Chromosome. *Journal of Heredity*, *112*(3), 250-259. https://doi.org/10.1093/jhered/esab016

Ron, G., Globerson, Y., Moran, D., & Kaplan, T. (2017). Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02386-3

Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., & Peichel, C. L. (2009). Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). *PLoS Genetics*, *5*(2), e1000391. https://doi.org/10.1371/journal.pgen.1000391

Sakamoto, T., & Innan, H. (2022). Muller's ratchet of the Y chromosome with gene conversion. *Genetics*, *220*(1). https://doi.org/10.1093/genetics/iyab204

Schartl, M., Schories, S., Wakamatsu, Y., Nagao, Y., Hashimoto, H., Bertin, C., Mourot, B., Schmidt, C., Wilhelm, D., Centanin, L., Guiguen, Y., & Herpin, A. (2018). Sox5 is involved in germ-cell regulation and sex determination in medaka following co-option of nested transposable elements. *BMC Biology*, *16*(1). https://doi.org/10.1186/s12915-018-0485-8

Shao, F., Ludwig, A., Mao, Y., Liu, N., & Peng, Z. (2020). Chromosome-level genome assembly of the female western mosquitofish (Gambusia affinis). *GigaScience*, *9*(8). https://doi.org/10.1093/gigascience/giaa092

Singh, N. D., Koerich, L. B., Carvalho, A. B., & Clark, A. G. (2014). Positive and Purifying Selection on the Drosophila Y Chromosome. *Molecular Biology and Evolution*, *31*(10), 2612-2623. https://doi.org/10.1093/molbev/msu203

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R. H., Wilson, R. K., Rozen, S., & Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, *423*(6942), 825-837. https://doi.org/10.1038/nature01722

Soh, Y. Q., Alföldi, J., Pyntikova, T., Laura, Graves, T., Patrick, Robert, Kremitzki, C., Koutseva, N., Jacob, Rozen, S., Jennifer, Owens, E., James, William, Cao, Q., Pieter, Wesley, Richard, Skaletsky, H., & David. (2014). Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*, *159*(4), 800-813. https://doi.org/10.1016/j.cell.2014.09.052

Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Sanchez, R. C., Fescemyer, H. W., Harris, R., Ye, D., O'Brien, P. C. M., Chikhi, R., Ryder, O. A., Ferguson-Smith, M. A., Medvedev, P., & Makova, K. D. (2016). A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. *Genome Research*, *26*(4), 530-540. https://doi.org/10.1101/gr.199448.115

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B. B., Castel, S. E., Karczewski, K. J., Aguet, F., Byrnes, A., Lappalainen, T., Regev, A., Ardlie, K. G., Hacohen, N., & Macarthur, D. G. (2017). Landscape of X chromosome inactivation across human tissues. *Nature*, *550*(7675), 244-248. https://doi.org/10.1038/nature24265

Úbeda, F., Patten, M. M., & Wild, G. (2015). On the origin of sex chromosomes from meiotic drive. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798), 20141932. https://doi.org/10.1098/rspb.2014.1932

Veltsos, P., Ridout, K. E., Toups, M. A., González-Martínez, S. C., Muyle, A., Emery, O., Rastas, P., Hudzieczek, V., Hobza, R., Vyskot, B., Marais, G. A. B., Filatov, D. A., & Pannell, J. R. (2019). Early Sex-Chromosome Evolution in the Diploid Dioecious Plant Mercurialis annua. *Genetics*, *212*(3), 815-835. https://doi.org/10.1534/genetics.119.302045

Vermunt, M. W., Tan, S. C., Castelijns, B., Geeven, G., Reinink, P., De Bruijn, E., Kondova, I., Persengiev, S., Bontrop, R., Cuppen, E., De Laat, W., & Creyghton, M. P. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature Neuroscience*, *19*(3), 494-503. https://doi.org/10.1038/nn.4229

Villar, D., Flicek, P., & Odom, D. T. (2014). Evolution of transcription factor binding in metazoans — mechanisms and functional implications. *Nature Reviews Genetics*, *15*(4), 221-233. https://doi.org/10.1038/nrg3481

Wei, K. H. C., & Bachtrog, D. (2019). Ancestral male recombination in Drosophila albomicans produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. *PLoS Genetics*, *15*(11), e1008502. https://doi.org/10.1371/journal.pgen.1008502

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., Depristo, M. A., Ruan, J., Marschall, T., Sedlazeck, F. J., Zook, J. M., Li, H., Koren, S., Carroll, A., Rank, D. R., & Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, *37*(10), 1155-1162. https://doi.org/10.1038/s41587-019-0217-9

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Molecular Biology and Evolution*, *32*(8), 1981-1995. https://doi.org/10.1093/molbev/msv078

Wilson Sayres, M. A. (2018). Genetic Diversity on the Sex Chromosomes. *Genome Biology and Evolution*, *10*(4), 1064-1078. https://doi.org/10.1093/gbe/evy039

Wilson Sayres, M. A., Lohmueller, K. E., & Nielsen, R. (2014). Natural Selection Reduced Diversity on Human Y Chromosomes. *PLoS Genetics*, *10*(1), e1004064. https://doi.org/10.1371/journal.pgen.1004064

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. *eLife*, *8*. https://doi.org/10.7554/elife.47138

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59-69. https://doi.org/10.1038/nrg3095

Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications*, *7*(1), 12087. https://doi.org/10.1038/ncomms12087

Xiao, C., Li, J., Xie, T., Chen, J., Zhang, S., Elaksher, S. H., Jiang, F., Jiang, Y., Zhang, L., Zhang, W., Xiang, Y., Wu, Z., Zhao, S., & Du, X. (2021). The assembly of caprine Y chromosome sequence reveals a unique paternal phylogenetic pattern and improves our understanding of the origin of domestic goat. *Ecology and Evolution*, *11*(12), 7779-7795. https://doi.org/10.1002/ece3.7611

Yoshida, K., Makino, T., Yamaguchi, K., Shigenobu, S., Hasebe, M., Kawata, M., Kume, M., Mori, S., Peichel, C. L., Toyoda, A., Fujiyama, A., & Kitano, J. (2014). Sex Chromosome Turnover Contributes to Genomic Divergence between Incipient Stickleback Species. *PLoS Genetics*, *10*(3), e1004223. https://doi.org/10.1371/journal.pgen.1004223

Zhang, H., Jain, C., & Aluru, S. (2020). A comprehensive evaluation of long read error correction methods. *BMC Genomics*, *21*(S6). https://doi.org/10.1186/s12864-020-07227-0

Zhou, Q., & Bachtrog, D. (2012). Chromosome-Wide Gene Silencing Initiates Y Degeneration in Drosophila. *Current Biology*, *22*(6), 522-525. https://doi.org/10.1016/j.cub.2012.01.057

Zhou, Q., Ellison, C. E., Kaiser, V. B., Alekseyenko, A. A., Gorchakov, A. A., & Bachtrog, D. (2013). The Epigenome of Evolving Drosophila Neo-Sex Chromosomes: Dosage Compensation and Heterochromatin Formation. *PLoS Biology*, *11*(11), e1001711. https://doi.org/10.1371/journal.pbio.1001711

Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J. C., Peel, E., Cheng, Y., Pajpach, F., Bradley, N., Suzuki, H., Nikaido, M., Damas, J., Daish, T., Perry, T., Zhu, Z., Geng, Y., Rhie, A., Sims, Y., Wood, J., Haase, B., Mountcastle, J., Fedrigo, O., Li, Q., Yang, H., Wang, J., Johnston, S. D., Phillippy, A. M., Howe, K., Jarvis, E. D., Ryder, O. A., Kaessmann, H., Donnelly, P., Korlach, J., Lewin, H. A., Graves, J., Belov, K., Renfree, M. B., Grutzner, F., Zhou, Q., & Zhang, G. (2021). Platypus and echidna genomes reveal mammalian biology and evolution. *Nature*, *592*(7856), 756-762. https://doi.org/10.1038/s41586-020-03039-0

CHAPTER 3

POSITIVE SELECTION DRIVES *CIS*-REGULATORY EVOLUTION ACROSS THE THREESPINE STICKLEBACK Y CHROMOSOME [2]

[2] Shaw, D.E. , Naftaly, A.S. and White, M.A. Submitted to *Molecular Biology and Evolution*.

Abstract

Allele-specific gene expression evolves rapidly across gametologs on heteromorphic sex chromosomes (X and Y). Current models of sex chromosome evolution suggest this occurs through the rapid accumulation of mutations within *cis*-regulatory regions. However, these patterns have not been demonstrated empirically due to the limited number of Y chromosome assemblies available to survey sequence evolution outside of coding regions. The threespine stickleback (*Gasterosteus aculeatus*) Y chromosome is an ideal model to test hypotheses of rapid *cis*-regulatory evolution due to its intermediate state of divergence from the X chromosome. A large number of Y-linked gametologs still exist across three differently aged evolutionary strata to test these hypotheses. We found that putative enhancer regions (accessible chromatin regions defined through ATAC-seq) on the Y chromosome exhibited elevated substitution rates when compared to intergenic regions and synonymous sites within coding regions. This strongly suggests that many *cis*-regulatory regions are under positive selection on the Y chromosome. This divergence was correlated with X-biased gametolog expression, indicating the loss of Y-linked gametolog expression may be favored by selection. Our findings provide evidence that Y-linked regulatory regions exhibit signs of positive selection quickly after the suppression of recombination, supporting recent theoretical models that show the rapid divergence of regulatory regions may be favored to mask deleterious mutations on the Y chromosome.

Introduction

The evolution of heteromorphic sex chromosomes has occurred many times across species (Bachtrog, 2013; Bachtrog et al., 2014). Heteromorphic sex chromosomes evolve once recombination is suppressed between the X and Y (or Z and W) (Charlesworth, 1978; Charlesworth & Charlesworth, 2000; Muller, 1918). After recombination is suppressed, the Y

chromosome rapidly accumulates mutations, leading to sequence degeneration (Charlesworth & Charlesworth, 2000). Empirical evidence of this process has focused on the sequence evolution of coding regions on sex chromosomes. Broad comparative work has revealed that coding sequence evolution can follow one of several different evolutionary trajectories. Although many ancestral Y-linked genes are lost because of sequence degeneration through the accumulation of deleterious mutations, growing evidence indicates not all genes can be lost. Some genes appear to be dosage sensitive and are under strong purifying selection to be retained on the Y chromosome (Bellott et al., 2014; Bellott et al., 2017; Peichel et al., 2020; White et al., 2015). Novel genes with sex-specific functions can also accumulate on sex-limited chromosomes via translocation and subsequent gene duplication (Chang et al., 2022; Ellison & Bachtrog, 2019b; Hughes et al., 2020; Mahajan & Bachtrog, 2017; Peichel et al., 2020; Soh et al., 2014).

The accumulation of mutations is not restricted to coding sequences and can also occur in regulatory regions, leading to changes in expression from the sex-limited chromosome. Allele-specific gene expression changes has been shown to evolve rapidly on degenerating sex chromosomes often resulting in the loss of gene expression on the sex-limited chromosome (Y and W) (Ayers et al., 2013; Beaudry et al., 2017; Martin et al., 2019; R. P. Meisel et al., 2012; Muyle et al., 2012; Muyle et al., 2018; Rodríguez Lorenzo et al., 2018; Shaw & White, 2022; Singh et al., 2014; Veltsos et al., 2019; Wei & Bachtrog, 2019; White et al., 2015). Although many deleterious regulatory mutations may accumulate through selective interference (Bachtrog, 2008; Charlesworth & Charlesworth, 2000), selection may also favor mutations within regulatory regions to downregulate coding regions with deleterious mutations (Bachtrog, 2006; Orr & Kim, 1998)**.** Recent theory supports the role of positive selection driving the rapid

accumulation of mutations to downregulate the Y-linked allele and upregulate the X-linked allele to maintain ancestral dosage balance (Lenormand et al., 2020; Lenormand & Roze, 2022).

One challenge to empirically testing theoretical models that focus on regulatory evolution is complete Y-chromosome assemblies are only available for a limited number of taxa. Reference assemblies are needed to identify the substitutions that are accumulating within intergenic regulatory regions on the sex-limited chromosome. Additionally, many of the available Y-chromosome assemblies are from model organisms that have sex chromosomes that are highly degenerated (Bellott et al., 2014; Hughes et al., 2010; Hughes et al., 2020; Mahajan et al., 2018; Soh et al., 2014; Tomaszkiewicz et al., 2016). These species only have a few remaining ancestral gametologs on the Y chromosome, thus limiting the number of genes available to study how regulatory evolution potentially lead to expression differences between the X and Y. Species with recently derived sex chromosomes that still harbor many ancestral Y-linked gametologs at varied stages of degeneration are needed to understand how substitutions within regulatory regions lead to the evolution of expression differences between the X and Y.

In addition to having a chromosome-scale assembly of the Y chromosome, an additional challenge is annotating regulatory regions (Reviewed in Shaw & White, 2022). Recent approaches have focused on identifying regions of the genome with accessible chromatin that may contain transcription factor binding sites to regulate gene expression (Ricci et al., 2019). Chromatin accessibility profiling techniques, like (Assay for Transposase Accessible Chromatin) ATAC-seq (Buenrostro et al., 2013), utilize short-read sequencing to profile accessible chromatin regions (ACRs) at a fine-scale across the genome. Across autosomes, ACRs show signatures of purifying selection relative to other intergenic regions, consistent with these regions containing important functional elements for gene regulation (Connelly et al., 2014; Horvath et

al., 2021; Lu et al., 2019). Profiling ACRs on sex chromosomes would therefore provide a means to study the molecular evolution of *cis*-regulatory regions in the context of Y degeneration.

The threespine stickleback fish (*Gasterosteus aculeatus*) is an excellent model to study the evolution of regulatory regions on sex chromosomes. Threespine stickleback fish have a high-quality reference Y chromosome that is more recently derived compared to many other chromosome-scale assemblies (Bellott et al., 2014; Bellott et al., 2017; Hughes et al., 2012; Hughes et al., 2010; Hughes et al., 2020; Peichel et al., 2020). The Y chromosome contains multiple evolutionary strata at different stages of degeneration (Peichel et al., 2020; White et al., 2015), allowing for comparisons of sequence evolution at different temporal scales. While some of the Y chromosome has degenerated, over half of X-linked gametologs remain on the Y chromosome (Peichel et al., 2020). Many of the coding regions of the shared gametologs show signatures of purifying selection and are enriched for dosage sensitive functions (Peichel et al., 2020; White et al., 2015). Interestingly, chromosome-wide dosage compensation has not evolved to counter the loss of expression for genes where the coding sequence has degenerated on the Y chromosome (White et al., 2015). The lack of chromosome-wide dosage compensation enables a gene-by-gene comparison of *cis*-regulatory changes to understand how expression of X-linked alleles evolves in response to degenerating coding regions on the Y.

Here, we leveraged ATAC-seq from two different tissues to characterize ACRs shared between the X and Y chromosomes. To understand how selection may be operating within these regions, we estimated nucleotide divergence of ACRs between the X and Y chromosomes of the threespine stickleback fish (*Gasterosteus aculeatus*) as well as within-population nucleotide diversity of ACRs. We found signatures of positive selection across sex-linked ACRs that indicate a rapid accumulation of nucleotide substitutions on the Y chromosome, relative to other

intergenic regions. Our findings complement existing sex chromosome evolution theory that suggests the accumulation of mutations within regulatory regions may be beneficial to silence expression from a degenerating Y chromosome.

Methods

Molecular evolution of accessible chromatin regions and coding regions

We used accessible chromatin regions (ACRs) from two different tissues. Liver ACRs were previously identified from two replicates (Naftaly et al., 2021) (PRJNA667175). We also collected testes from two juvenile males (~4.4 cm in standard length) of laboratory-reared threespine stickleback fish, originally isolated from Lake Washington (Seattle, Washington, USA) (PRJNA686097). The testis cells were immediately dissociated through homogenization in 1X PBS containing proteinase inhibitor cocktail (PIC, cOmplete tablets Roche) (PBS+PIC). The cells were fixed with 16% formaldehyde and washed twice with PBS+PIC, followed by lysis in 1M Tris-HCl, pH=8, 0.5M EDTA, 10% NP-40, 50% glycerol/molecular grade H_2O, and 1X PIC. Nuclei were stained with DAPI and counted with a hemocytometer. We diluted samples to 60,000 – 80,000 nuclei. ATAC-seq library preparation was conducted using previously established protocols (Lu et al., 2017; Naftaly et al., 2021). The libraries were sequenced on an Illumina NextSeq (2 x 150 bp; Georgia Genomics and Bioinformatics Core). Reads were trimmed with Trimmomatic (v. 0.36) (Bolger et al., 2014), using a sliding window of four bases, trimming the remainder of the read when the average quality within a window dropped below 20. Residual adapter sequences were removed using Trimmomatic ILLUMINACLIP. Reads were filtered for a minimum length using MINLEN:30. We aligned the trimmed reads using Bowtie2 (v. 2.4.1) with default parameters (Langmead & Salzberg, 2012). We filtered for alignments with

a mapping quality greater than 20 using SAMtools (v1.14). We also removed PCR duplicates by using the MarkDuplicates function in Picard (https://github.com/broadinstitute/picard).

We used a tn5 control sample to normalize ATAC-seq reads to remove the effect of tn5 bias. For the control, genomic DNA was extracted from a caudal fin clip of one male fish using a standard phenol-chloroform extraction. The ATAC-seq library preparation, whole genome sequencing library preparation, and sequencing (Illumina HiSeq 2 x 150 bp) of the control sample were completed by GENEWIZ (New Jersey, USA). Whole genome sequencing was conducted to show the DNA sample was of sufficient quality to construct the Tn5 bias control. We trimmed residual adapters and low-quality sequences using Trimmomatic (Bolger et al., 2014) as previously described. Trimmed reads were aligned to the threespine stickleback genome using Bowtie2 (Langmead & Salzberg, 2012). The read coverage per base pair was calculated using BEDTools (v2.29, -d) (Quinlan & Hall, 2010). The whole genome sequencing sample had an average read depth of 90x where only ~10% of the genome was supported with less than ten reads. Within these regions, only 6% had zero reads per base pair, indicating 94% of the genome could be queried for biased integration of Tn5. The Tn5 bias control produced over 177 million reads. The reads were trimmed from these reads using Trimmomatic with the same parameters. The trimmed reads were aligned to the threespine stickleback genome (v. 5) using Bowtie2 (Langmead & Salzberg, 2012). Reads mapping to the mitochondria and unscaffolded regions were removed. PCR duplicates were removed using MarkDuplicates from Picard (https://github.com/broadinstitute/picard).

We explored overall concordance between replicates by assaying whether ATAC-seq reads were enriched around transcription start sites in the two testis and two liver samples. Iso-Seq long read sequencing was previously conducted in threespine stickleback fish to curate

accurate transcription start site annotations across multiple tissues (Naftaly et al., 2021). We used deepTools (v3.5.1) computeMatrix reference-point (Ramírez et al., 2016) on the ATAC-seq alignments to assay read depth 3kb around the complete set of annotated transcription start sites. We plotted enrichment of ATAC-seq reads around transcription start sites (TSS) and compared to RNA-seq expression using deepTools plotHeatmap with default settings, except that we used the -sortUsingsamples to sort the regions by expression. We found replicates across both tissues were highly concordant in which transcription start sites were accessible (Supplemental Figure 3.1**)**. We also found accessible transcription start sites were a good predictor of gene expression in each tissue (Supplemental Figure 3.1). Because both replicates in each tissue were highly concordant, we pooled the aligned reads for each tissue to maximize read depth for all the analyses. ACRs were called for each tissue sample using MACS2 (v2.2.7.1) callpeak with the –keep--dup all parameter, and read depth was normalized with Tn5 control sequencing with the -c parameter (Y. Zhang et al., 2008).

We mapped the nucleotide sequence of each ACR on the X chromosome to homologous regions on the Y chromosome using previously generated alignments (Peichel et al., 2020). We identified orthologous regions between the X chromosome, Y chromosome, and orthologous autosome (chromosome 19) from the ninespine stickleback fish *(Pungitius pungitius*; (Varadharajan et al., 2019), 2019) using BLAST+ (blastn v. 2.11.0) with default blastn parameters, with -perc_identity set to 75. (Camacho et al., 2009). To ensure high quality alignments, we filtered for uniquely mapping alignments that also had a bit score greater than 100. We used BLAST+ alignments to extract sequence to create multiple sequence alignments for downstream analysis. We aligned the threespine stickleback X and Y ACR sequences to the ninespine stickleback autosome sequence using MUSCLE (v 3.8.1551) (Edgar, 2004) with

default parameters. We used MUSCLE to create a multiple sequence alignment between the X sequence, Y sequence, and the sequence from the orthologous autosome to identify X- and Y-specific variants. We called single nucleotide variants using a custom python script modified from msa2snps.py (Junli Zhang, https://github.com/pinbo/msa2snp). We calculated the reported divergence rate by dividing the number of variants by the number of the aligned sites.

We compared ACR divergence on the sex chromosomes to ACR divergence on autosome 18, a chromosome similar in length to the X chromosome. We identified orthologous regions between the threespine stickleback autosome 18 and autosome 18 from the ninespine stickleback fish using BLAST+ (blastn v. 2.11.0) (Camacho et al., 2009) as previously described. We called variants between autosomal regions in a similar manner as the sex chromosomes, by creating a pairwise alignment using MUSCLE (v 3.8.1551) with default parameters and called single nucleotide variants using the same python script as above. We calculated a divergence rate by dividing the number of single nucleotide variants by the size of the aligned region.

Divergence within ACRs was compared with the divergence of coding regions of neighboring genes. We identified the closest gene to each ACR by running annotatepeaks.pl in homer (v 4.11) (Heinz et al., 2010) using homer default settings. Per default settings, ACRs were assigned to the closest gene within 50kb of the TSS. For coding divergence comparisons, we used previously reported estimates of synonymous (d_S) and non-synonymous (d_N) divergence of gametologs between the X and Y chromosomes (Peichel et al., 2020). Coding regions with d_N/d_S ratios of 99 were omitted.

Estimating a neutral substitution rate

We used randomly drawn intergenic regions throughout the X chromosome, Y chromosome, or chromosome 18 to estimate a neutral substitution rate. Regulatory regions tend

to be GC-rich, which can be prone to higher mutation rates through mechanisms like GC-biased gene conversion or spontaneous deamination of methylated cytosines (Nesta et al., 2021). We therefore GC-matched the randomly drawn intergenic regions with the GC content of ACRs. We calculated the GC percentage of ACRs using a modified Perl script (countbp.pl, Nicholas Navin, http://www.navinlab.com/bioperl/bioperl/gc_content.html). For each ACR, a random intergenic region was drawn from the X chromosome equal in length and with a GC content within 2.0%. Intergenic regions were defined as any region that that falls within annotated functional regions from a combination of Ensembl annotations (Cunningham et al., 2021), Isoseq transcripts (Naftaly et al., 2021), Y chromosome annotations (Peichel et al., 2020), repetitive elements (Nath et al., 2021; Peichel et al., 2020) and ACRs from this study. This was performed for each stratum individually using bedtools shuffle –I ACRs.bed -incl XYalignment.bed -excl annotations.bed. We extracted 10,000 sets of GC-matched intergenic regions from each of the three evolutionary strata on the threespine stickleback sex chromosomes (set size same as number of Liver ACRs: stratum one: 442, stratum two: 228, stratum three: 199; Testis ACRs: stratum one: 30, stratum two: 37, stratum three: 51). Due to limited number of intergenic sites shared between the sex chromosomes, we sampled with replacement to align a sufficient number of intergenic regions. We identified the orthologous regions to the ninespine genome assembly using BLAST+ (blastn v2.11.0) (Camacho et al., 2009). We filtered for alignments that mapped uniquely to chromosome 19 in the ninespine stickleback fish, had an alignment length equal to 75% of the threespine query sequence, and had a bit-score greater than 100. We generated multiple sequence alignments using MUSCLE (v 3.8.1551), for each intergenic permutation and calculated X and Y specific substitution rate as previously described.

Estimating within population nucleotide diversity

We used whole-genome short-read sequencing of 12 males from the Lake Washington population (Washington, USA; SRP137809, NCBI Short Read Archive) to estimate nucleotide diversity across the sex chromosomes. The raw reads were trimmed with Trimmomatic (v. 0.39) (Bolger et al., 2014), using a sliding window of four bases, trimming the remainder of the read when the average quality within a window dropped below 15. The leading and trailing base pairs below quality three of every read were removed along with any residual adapter sequence. After trimming, any reads below a minimum length of 36 were discarded. We aligned the trimmed reads using Bowtie2 (v. 2.4.5) with default parameters (Langmead & Salzberg, 2012). We also marked PCR duplicates by using the MarkDuplicates function in Picard (v. 2.26.10) (https://github.com/broadinstitute/picard). SNP genotyping was conducted following the GATK (v. 4.2.5.0) best practices for germline short variant discovery. Variants were called per sample using HaplotypeCaller in GVCF mode. We then joint-called variants using GenotypeGVCFs using a genomics database created by GenomicsDBImport.

We only considered biallelic SNPs in our estimates of nucleotide diversity and we filtered for high quality genotypes using several different methods. On the sex chromosomes, sites should only have hemizygous genotypes in males in the non-crossover region outside of the pseudoautosomal region. Any sites that are heterozygous would be caused by errors in read alignment. We therefore did not consider sites that were heterozygous in the non-crossover region on the X or Y chromosomes. We also filtered out sites that exhibited too low or high of read depth, which would be indicative of alignment errors. To do this, we did not consider sites that were less than one-half or greater than double the median read depth of each chromosome (X and Y chromosomes: positions were retained if the read depth was between 3.5 and 14;

chromosome 18: positions were retained if the read depth was between 6.5 and 26). Read filtering was conducted using custom Perl scripts.

Nucleotide diversity (pi) was estimated on a per-site basis (--site-pi) across each chromosome using vcftools (v. 0.1.16) (Danecek et al., 2011)). Nucleotide diversity for ACRs was computed as the average diversity across all sites within ACRs, including non-variant (zero pi) sites that were not removed by read depth filters. We compared nucleotide diversity within ACRs with nucleotide diversity in randomly drawn intergenic sites. Intergenic sites were defined as previously described. Intergenic sites were drawn with replacement equal to the total number of sites within ACRs on each chromosome (10,000 total permutations).

Comparison with allele-specific expression patterns

We measured allele-specific expression of transcripts on the X and Y chromosome using three replicate liver samples (PRJNA591630) and three replicate testis (PRJNA591630) samples. We trimmed the RNA-seq reads with Trimmomatic (v. 0.36) (Bolger et al., 2014) using a sliding window of 4 bases, trimming the remainder of the read when the average quality within a window dropped below 20 (SLIDINGWINDOW:4:20). Residual sequencing adapters were also removed using ILLUMINACLIP. We aligned RNA-seq reads using Tophat2 (v2.2.1) with default parameters (Kim et al., 2013). We filtered for alignments with a map quality greater than 25 using SAMtools (v1.14) (Li et al., 2009) to identify reads that map uniquely on the X and Y allele. Similar map quality filters have been used to distinguish short reads between sex chromosomes in *Drosophila* and stickleback fish (Ellison & Bachtrog, 2019b; Peichel et al., 2020) and between sexes with heteromorphic sex chromosomes in guppies (Kirkpatrick et al., 2021). Read counts were obtained using htseq-count (v 0.9.1) (Anders et al., 2015). Default parameters were used with the addition of --stranded=no and --nonunique all to maximize our

ability to count reads on the Y chromosome. We used a custom GTF file for htseq-count that included previously determined start and end sites of all Ensembl predicted transcripts that are shared between the X and Y chromosomes (Peichel et al., 2020). Reads were counted on the X and Y across the entire transcript as a feature. Transcripts were removed from the analysis if they had an RNA expression count of zero in all samples. For each tissue, we calculated the average read count for each gametolog across the three tissue replicates. We calculated the allele-specific expression as $\log_2$(X transcript average / Y transcript average).

Data availability

The data for this manuscript is available within the short reads archive under accession numbers: PRJNA667175, PRJNA591630, PRJNA686097, SRP137809.

Scripts and bioinformatic pipelines are available at: https://github.com/daniel-shaw1/Regulatory_divergence_paper

Results

Nucleotide divergence between the X and Y is higher in accessible chromatin regions

The completion of the threespine stickleback Y chromosome assembly (Peichel et al., 2020) allowed us to thoroughly examine regulatory evolution among sex linked gametologs shared between the sex chromosomes. Accessible chromatin regions (ACRs) are sites of the genome that putatively serve as domains for transcription factor binding. We first sought to compare divergence of ACRs between the X and Y chromosomes to previously characterized synonymous divergence throughout coding regions. Crossing over was suppressed across the threespine stickleback sex chromosomes in at least three separate events, forming strata with distinct levels of divergence (stratum one: oldest; stratum two: middle; stratum three: youngest) (Peichel et al., 2020). Based on theories of regulatory evolution on sex chromosomes

(Lenormand et al., 2020; Lenormand & Roze, 2022), we predicted ACRs may be evolving more rapidly than neutrally evolving synonymous sites within coding regions. To survey ACRs, we utilized ATAC-seq from liver tissue (Naftaly et al. 2021) and testis tissue. In the liver, we found a total of 1,279 X-linked ACRs. We were able to align 948 (74.1%) of these ACRs to orthologous regions on the Y chromosome and the ancestral ninespine stickleback autosome 19. In testis tissue, we identified 896 X-linked ACRs. Compared to the liver, we found far fewer orthologous regions on the Y chromosome and autosome 19 in the ninespine stickleback (118 total aligned; 13.2%), suggesting regulatory regions within testis ACRs may be more rapidly evolving. We estimated divergence between the X and Y chromosomes across ACRs in each tissue. Matching these predictions, among the ACRs we were able to align, testis ACRs had higher divergence between the X and Y chromosomes relative to both synonymous coding region divergence and divergence within liver ACRs (Figure 3.1). This pattern was consistent across the three evolutionary strata. Although liver divergence within ACRs was not as high as testis ACRs, it was still significantly higher than synonymous coding region divergence in the two oldest strata (one and two) (Figure 3.1), indicating liver ACRs are also evolving rapidly. In the youngest stratum (three), the liver ACRs were not significantly different than synonymous substitutions. Our results highlight that ACRs are likely under positive selection and this has the strongest effect in ACRs that harbor regulatory elements that are functional in testes.

Mutations within accessible chromatin regions are mostly accumulating on the Y chromosome

We used the orthologous sequence from the ninespine stickleback fish as to infer whether substitutions were accumulating at a higher rate on the X or Y chromosome. ACRs functional in the liver exhibited higher substitution rates on the Y chromosome than on the X chromosome. Y-

linked ACRs had significantly higher substitution rates compared to random intergenic regions for the oldest two evolutionary strata (Figure 3.2, $P < 0.001$; stratum one and two, 10,000 permutations). In contrast to the Y-linked ACRs, we found the X-linked liver ACR sequence divergence was significantly lower than intergenic permutations, consistent with purifying selection operating across all three strata (Figure 3.2; Stratum one: $P = 0.016$; Stratum two and three: $P < 0.001$; 10,000 permutations). This suggests that ACRs may be functionally constrained more on the X chromosome compared to Y-linked ACRs. We observed a similar pattern on the Y chromosome for ACRs function in testis tissue. Y-linked ACRs had significantly higher substitution rates in all three evolutionary strata, compared to random intergenic regions (Figure 3.2; $P < 0.001$, 10,000 permutations). However, we observed X-linked ACRs from testis tissue, also exhibited elevated nucleotide substitutions in the oldest two evolutionary strata (Figure 3.2; $P < 0.001$, 10,000 permutations), while the youngest stratum was significantly lower than intergenic regions (Figure 3.2; $P < 0.001$, 10,000 permutations). This indicates some ACRs may also be under positive selection on the X chromosome and this occurs in a tissue-specific fashion.

Autosomal ACRs have been shown to be under purifying selection in other species (Horvath et al., 2021). We therefore tested whether threespine stickleback autosomes ACRs also exhibited purifying selection, similar to the pattern we observed on the X chromosome in liver. Consistent with patterns of purifying selection, we found autosomal ACRs exhibited lower sequence divergence compared to random intergenic regions in both liver (Figure 3.2; $P = 0.002$; 10,000 permutations) and testis tissue (Figure 3.2; $P < 0.001$; 10,000 permutations).

Accessible chromatin regions have low nucleotide diversity

We searched for signatures of selection in ACRs from 12 sequenced males from a freshwater population of stickleback fish. Low nucleotide diversity within ACRs relative to neutrally evolving intergenic regions would be indicative of either purifying or positive selection. On autosomes, ACRs have been shown to exhibit lower nucleotide diversity than intergenic regions (Horvath et al., 2021), suggesting conserved function. We first tested whether we observed a similar pattern among testis and liver tissue ACRs on autosome 18. We found that liver ACRs on chromosome 18 had lower nucleotide diversity compared to intergenic regions (Figure 3.3; $P < 0.001$, 10,000 permutations), similar to what has been observed in other species. However, we did not detect a significant difference for autosomal testis ACRs (Figure 3.3; $P = 0.251$, 10,000 permutations). We also did not detect differences in nucleotide diversity between ACRs and intergenic regions on the X or Y chromosomes in liver and testis tissue (Figure 3.3; all comparisons $P > 0.05$, 10,000 permutations). Combined with low divergence between species, the low nucleotide diversity within ACRs identified from liver tissue indicates some ACRs are under purifying selection, likely due to functional constraints. However, this effect is not consistent across chromosomes or tissues.

Among neutrally evolving intergenic sites, we detected major differences among the autosomes and sex chromosomes. Due to a lower effective population size, the Y chromosome is expected to be one-quarter of the nucleotide diversity observed on autosomes, given an equal sex ratio in the population. We found Y chromosome nucleotide diversity within intergenic regions was much lower than expected, compared to autosomes (Y chromosome pi: 0.00013; autosome pi: 0.0050; Figure 3.3). This suggests that selection acting on linked deleterious or beneficial mutations has lowered nucleotide diversity beyond neutral expectations.

Y-linked substitutions are correlated with biased expression from the X chromosome

Selection should favor the loss of expression from the Y chromosome to silence coding regions that have accumulated deleterious mutations (Bachtrog, 2006; Lenormand et al., 2020; Orr & Kim, 1998). To test if Y-linked regulatory divergence was associated with changes in expression, we compared ACR sequence divergence to changes in allelic expression on the X and Y chromosomes from liver and testis RNA-seq transcriptomes (Peichel et al., 2020). To maximize the total number of genes for this analysis, we pooled genes across all three evolutionary strata. We found that sequence divergence in liver ACRs, that were proximal to genes, were a predictor of allele-specific expression. We observed higher expression of the X-linked gametolog, relative to the Y-linked gametolog, when ACRs on the Y chromosome had more substitutions (Figure 3.4A; N = 51*;* liver expression, P = 0.004, Spearman's rank correlation). There was a weaker correlation between expression and substitutions within ACRs on the X chromosome (Figure 3.4C: N = 51, P = 0.033, Spearman's rank correlation). Combined, these results suggest substitutions on the Y chromosome have a greater effect on increasing X-biased gametolog expression. Although we do not have gene expression from the ninespine stickleback to determine which sex chromosome is deviating from the ancestral expression level, the Y-linked substitutions indicate it is likely due to downregulating the gametolog on the Y chromosome. A similar trend was observed in testis tissue, but the correlations were not significant, likely due to the small number of genes available for the analysis (Figure 3.4B N = 31, P = 0.111, Spearman's rank correlation).

Regulatory evolution is not correlated with coding sequence evolution

The downregulation of Y alleles could be adaptive if loss of expression follows the accumulation of deleterious coding substitutions (Orr & Kim, 1998). We searched for evidence of adaptive silencing on the Y chromosome by comparing the ACR divergence to the ratio of non-synonymous to synonymous substitutions (d_N / d_S) within the coding region of each gene. An adaptive silencing model would be supported if Y-linked gametologs with elevated non-synonymous substitutions relative to synonymous substitutions (d_N / d_S) have elevated regulatory divergence. Among genes shared between the X and Y chromosomes, we found no correlation between ACR nucleotide divergence and d_N / d_S (Supplemental Figure 3.2; each stratum on X and Y: $P > 0.05$; Spearman's rank correlation), indicating coding regions with elevated ACR divergence are not more likely to have elevated substitutions in the amino acid sequence. We also examined whether ACR divergence was higher within genes that contained frameshift mutations, nonsense mutations, or in-frame deletions. These types of mutations are more likely to produce a non-functional or sub-optimal peptide. We found no significant difference in the number of ACR substitutions from liver and testis tissue for coding regions that contained frameshift, nonsense mutations, or deletions, compared to coding regions that did not have these putatively deleterious mutations (functional) (Supplemental Figure 3.3) (all pairwise comparisons $P > 0.05$; Mann-Whitney U test). Overall, we found no evidence of regulatory divergence associated with deleterious coding sequence divergence, suggesting *cis*-regulatory evolution evolves independently of Y-linked coding variation**.**

Discussion

Gene regulation evolves rapidly on degenerating Y chromosomes. Our understanding of this process has largely been informed by characterizing gametolog-specific gene expression from degenerating sex chromosomes (Beaudry et al., 2017; Martin et al., 2019; Muyle et al., 2012; Muyle et al., 2018; Veltsos et al., 2019; White et al., 2015). Over time, gametolog expression is generally lost across most genes on the Y chromosome. With the growing number of Y chromosome reference assembles and sequencing methods to quickly profile functional regulatory regions across the genome, it has become feasible to explore the molecular evolution of these regions that lead to altered expression patterns. Here, we observed widespread elevated nucleotide substitutions within regulatory elements across the threespine stickleback Y chromosome, consistent with positive selection driving this process. We found that increased rates of *cis*-regulatory divergence were associated with X-biased expression of XY gametologs, suggesting loss of expression from the Y chromosome. Without liver and testis gene expression from the ninespine stickleback as an ancestral comparison, we are unable to confirm that X-biased gametolog expression is a result of downregulation of the Y chromosome or upregulation of the X chromosome. However, previous sequencing of a brain transcriptome showed expression patterns consistent with loss of Y expression rather than gain of X expression (White et al., 2015). It is therefore likely that the X-biased expression we observed in liver and testis tissue is also through loss of Y chromosome expression, similar to what has been observed in other species with degenerating Y chromosomes (Beaudry et al., 2017; Muyle et al., 2012; Wei & Bachtrog, 2019).

Simulations indicate rapid expression loss from the Y chromosome is driven by the divergence of *cis*-regulatory elements on both the X and Y chromosomes (Lenormand et al.,

2020; Lenormand & Roze, 2022). Mutations initially accumulate in regulatory elements within the non-recombining region of the Y chromosome, which can lead to some expression loss. As additional deleterious mutations arise in the coding sequence of genes, positive selection will favor the rapid accumulation of additional regulatory substitutions to silence expression of the Y-linked gametolog. Non-synonymous mutations are accumulating at a higher rate across the threespine stickleback Y chromosome (Peichel et al., 2020; White et al., 2015). However, we found that the number of nucleotide substitutions within regulatory regions were not correlated with deleterious mutations within neighboring coding regions. Our results suggest that adaptive regulatory evolution may happen without large changes in amino acid sequence. Our results complement previous findings that have compared coding sequence evolution to gametolog expression. The level of expression from the Y chromosome is often not correlated with the overall number of deleterious mutations within coding regions (i.e., gametologs with many deleterious mutations are not necessarily expressed at lower levels from the Y chromosome) (Bachtrog, 2006; Bachtrog et al., 2008; Beaudry et al., 2017). It is also possible that we did not capture the complete set of regulatory elements important for Y expression in our ATAC-seq datasets. A correlation between nucleotide substitutions in regulatory regions and deleterious mutations in coding regions may be observed if ACRs are profiled from additional tissues. In addition, we assigned ACRs to coding sequences based on overall proximity to a gene. Although this method of annotating ACRs is commonly used to assign regulatory regions to genes (Alexandre et al., 2018; Connelly et al., 2014; Ricci et al., 2019), it is possible that some ACRs interact with other genes through long range interactions based on chromatin configuration (Mifsud et al., 2015; Schoenfelder & Fraser, 2019), reducing our ability to detect an association between deleterious mutations in coding regions and nucleotide substitutions in ACRs. Finally,

the number of substitutions within an ACR may not be a fully accurate predictor of the overall ability to silence a given gametolog on the Y chromosome. If only a small number of substitutions are needed to ablate transcription factor binding within an ACR, we would not expect to see a strong correlation between the number of deleterious mutations within coding sequences and the number of substitutions within ACRs. Additional functional work will be necessary to explore these alternatives.

Degeneration of *cis*-regulatory elements on the Y chromosome and loss of expression should select for upregulation of the gametolog on the X chromosome to compensate for dosage loss (Lenormand et al., 2020). In this scenario, nucleotide substitutions should also rapidly accumulate within ACRs on the X chromosome. In contrast to this hypothesis, we found limited evidence for widespread positive selection on the X-linked regulatory regions. We found elevated rates of X-linked ACR divergence in testis but not liver. In many cases, ACRs on the X chromosome exhibited signatures of purifying selection, with substitution rates lower than random intergenic regions. This pattern was also observed on the autosomes. Chromosome-wide dosage compensation does not occur in the threespine stickleback fish. Genes that have completely lost the Y-linked gametolog exhibit half expression in males, relative to females (White et al., 2015). The lack of positive selection among X-linked ACRs is consistent with most genes on the X and Y chromosomes not being strongly dosage sensitive. In this case, most gametologs on the Y chromosome could be lost without consequence. Even in the absence of strong stabilizing selection to maintain dosage balance *cis*-regulatory degeneration is still predicted to occur rapidly (Lenormand et al., 2020). The Y-specific positive selection in ACRs we detected support this model.

Low nucleotide diversity within a population is a signature of positive or purifying selection (Nielsen, 2005). We explored whether there was an effect of positive selection in ACRs on nucleotide diversity within a freshwater population of threespine stickleback fish. However, we did not detect reduced nucleotide diversity within Y-linked ACRs relative to intergenic regions. However, Y chromosomes often have much lower diversity overall, relative to neutral expectations (Lawson Handley et al., 2006; Wilson Sayres et al., 2014). In such case, it is difficult to identify the target of any selective sweeps. Similar to humans (Wilson Sayres et al., 2014), our findings revealed that nucleotide diversity on the threespine stickleback Y chromosome is much lower than neutral expectations. Low Y-linked diversity has been attributed to purifying selection removing deleterious mutations and affecting linked neutral sites throughout the entire non-recombining region (Wilson Sayres, 2018; Wilson Sayres et al., 2014). We found ACRs exhibited high divergence between species, but exhibited low nucleotide diversity within populations, similar to intergenic regions throughout the chromosome. Our results suggest positive selection within ACRs could also be an important driver of reducing nucleotide diversity on the Y chromosome within populations. One important consideration is whether sex-biased demography may be affecting Y-linked diversity. If males are more variable in reproductive success than females, this could lower the expected effective population size of the Y chromosome, reducing nucleotide diversity. Little is known about the operating sex-ratios of stickleback fish. Some populations exhibit sex ratios that are female biased (Rollins et al., 2017). This could contribute to the low nucleotide diversity we observed. However, simulations in human populations revealed that even drastic shifts in sex ratio could not entirely explain the low diversity observed on the Y chromosome (Wilson Sayres et al., 2014).

Although expression is lost from most genes across the Y chromosome, the coding sequence of some genes are maintained through purifying selection. These genes tend to be enriched for haploinsufficiency phenotypes and presumably cannot be lost from the Y chromosome (Bellott et al., 2014; Bellott & Page, 2021; Bellott et al., 2017; Peichel et al., 2020). For this category of genes, if *cis*-regulatory elements are accumulating deleterious mutations, ancestral expression patterns could be maintained by gaining new binding sites that maintain expression. This would also result in signatures of increased substitution rates relative to other regions. This type of compensatory evolution has been proposed for functionally critical genes in mammals (Chaix et al., 2008; Vermunt et al., 2016), and *Drosophila* (Arnold et al., 2014). Functional analysis of what transcription binding sites are affected by Y-linked mutations could help identify which mutations are leading to down-regulation of gametologs compared to those that restore ancestral expression.

Our results highlight the importance of tissue specific evolution of *cis*-regulatory evolution. We observed distinct differences between liver and testis ACRs. Some of these differences may have been driven by reduced sample size of sex-linked ACRs in testis. Although the reduction in sample size was likely a result of rapid evolution at testis ACRs between the X and Y chromosome. Interestingly, regulatory regions did appear to evolve rapidly on the X chromosome in both older strata for testis ACRs. Future work focused on the evolution regulatory evolution on sex chromosomes should account for tissue specific regulation of genes. X-linked testis specific regulatory regions may evolve more rapidly than ACRs that also regulate more broadly expressed genes.

Conclusion

Together our results provide evidence of positive selection driving accelerated rates of nucleotide substitutions in *cis*-regulatory elements. Consistent signatures across all three evolutionary strata indicate that *cis*-regulatory can proceed rapidly following the suppression of recombination, leading to reduced gene expression from the Y chromosome. Improvements in functional annotations of regulatory regions as well as an ever-growing collection of high-quality Y and W assemblies will allow continued empirical testing of new regulatory models of sex chromosome evolution.

Figures

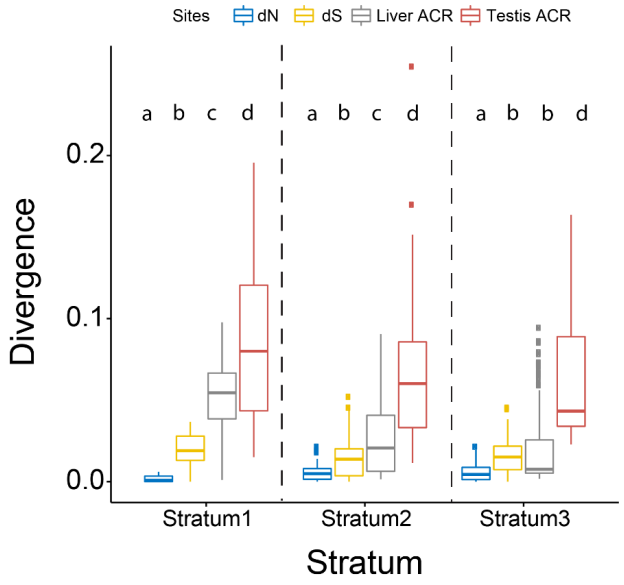


Figure 3.1. XY divergence across shared coding and non-coding elements. Regulatory regions show elevated substitution rates compared to coding sites. Protein-coding divergence was calculated between X and Y chromosomes to estimate synonymous site divergence (d_S) and nonsynonymous site divergence (d_N) and these values were averaged for each gene. Accessible Chromatin Regions (ACR) were identified on the X chromosome and mapped to the Y chromosome and ancestral autosome ACR's are defined as ATAC-seq peaks within 50kb of TSS of gametologs. X- and Y- specific mutations were identified and the sum was divided by the total length of the region. Stratum one: 12 genes, 442 liver ACRs, 30 testis ACRs; Stratum two: 71 genes, 228 liver ACRs, 37 testis ACRs; Stratum three: 46 genes, 199 liver ACRs 51 testis ACRs). Rates for each category were compared using a Mann-Whitney U test, adjusting for multiple comparisons. The letters above each plot indicate the significant differences between sites within each stratum and are considered different if $P < 0.01$.

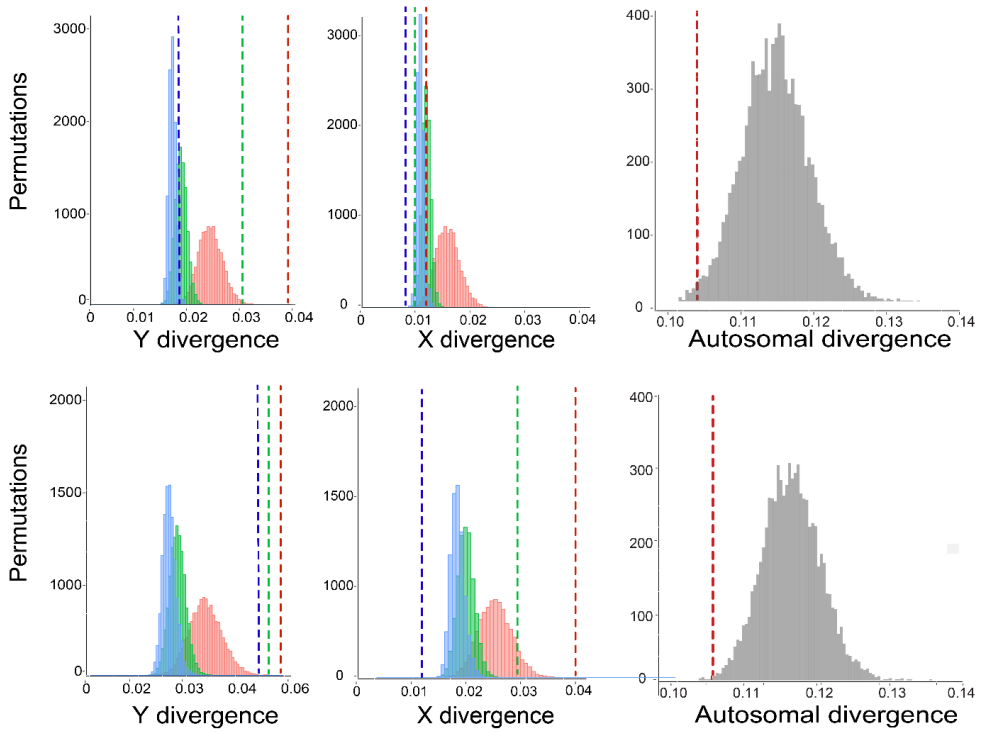


Figure 3.2. Allele-specific regulatory divergence compared to intergenic permutations. Y-linked regulatory divergence is elevated over neutrally evolving intergenic regions. X- and Y-specific average substitution rates were identified for 10,000 sets of intergenic regions per strata GC- matched to sets of regulatory regions. Averages of each set are plotted for Stratum 1 (pink), Stratum 2 (green), and Stratum 3 (blue). Accessible Chromatin Regions (ACR) were identified on the X chromosome and mapped to the Y chromosome and ninespine stickleback *(Pungititus pungititus*) autosome (chromosome 19) in liver and testis. ACRs are defined as ATAC-seq peaks assigned to the closest gametolog within 50kb of TSS. X- and Y- specific mutations were identified and the sum was divided by the total length of the region. Sets of intergenic regions match the number of ACRs for each stratum; (stratum one: liver ACRs: 442, testis ACRs: 30; stratum two: liver ACRs:228, testis ACRs 37; stratum three: liver ACRs: 129, testis ACRs: 46). Average ACR substitution rates for each stratum are shown by the dotted lines. Chromosome 18 was selected as a representative autosome with a similar number of genes as the X chromosome. ACRs assigned to genes within 50kb on chr. 18 were mapped to the ninespine stickleback (*Pungititus pungititus*) to obtain species divergence estimates. Autosomal ACR average divergence is significantly lower than permutations in liver ($P = 0.002$) and testis ($P < 0.001$) (Autosomal liver ACRs: 1056, testis ACRs: 305)

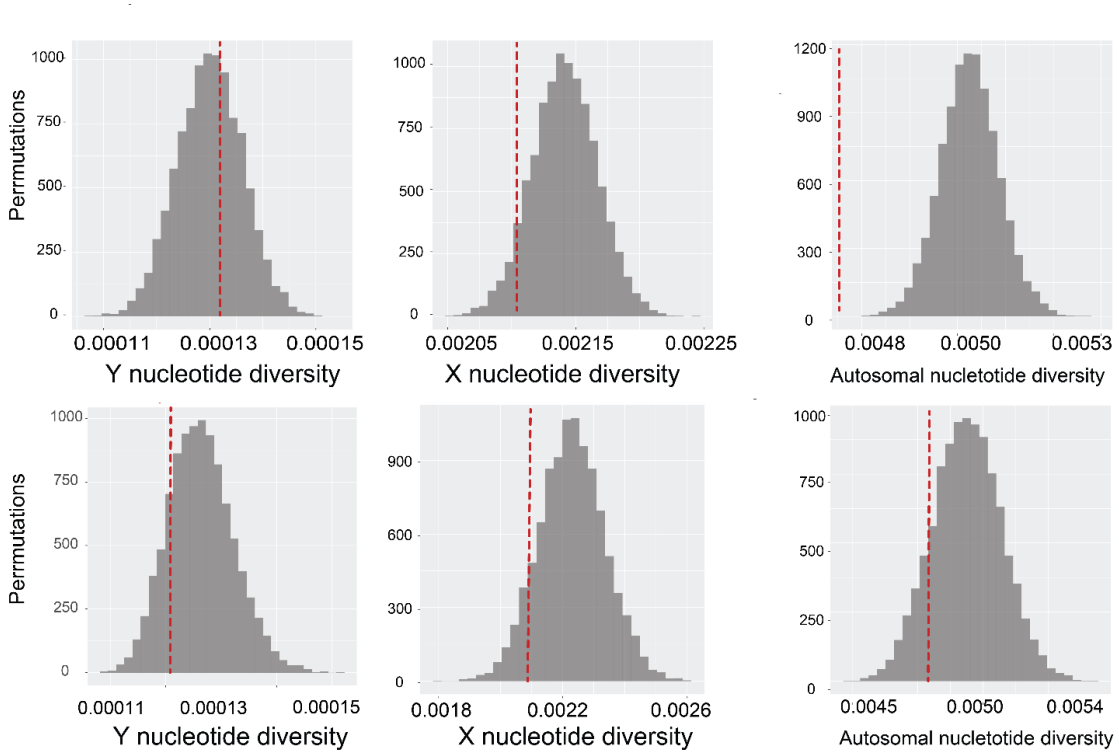


Figure 3.3. Regulatory nucleotide diversity compared to intergenic permutations. Rates of sex chromosome regulatory diversity are not different than intergenic sequence. X- and Y-specific diversity rates were identified for 10,000 sets of intergenic regions. Averages of each intergenic set are plotted. Accessible Chromatin Regions (ACR) were identified on the X chromosome and mapped to the Y chromosome and ancestral autosome. ACRs are defined as ATAC-seq peaks within 50kb of gametolog TSS. X- and Y- specific mutations were identified and the sum was divided by the total length of the region. Average ACR substitution rates for each chromosome are shown by the dotted lines. Y chromosome ACRs are not different than intergenic permutations in liver ($P = 0.623$) or testis ($P = 0.374$). X chromosome ACRs are not different than intergenic regions in liver ($P = 0.1985$), or testis ($P = 0.147$). Autosomal ACRs exhibited significantly lower diversity than intergenic permutations in liver (Liver: $P < 0.001$, Testis $P = 0.252$). Sets of intergenic regions match the number of ACRs for each stratum; (stratum one: liver ACRs: 442, testis ACRs: 30; stratum two: liver ACRs:228, testis ACRs 37; stratum three: liver ACRs: 129, testis ACRs: 46; autosomal liver ACRs: 1056, testis ACRs: 305).

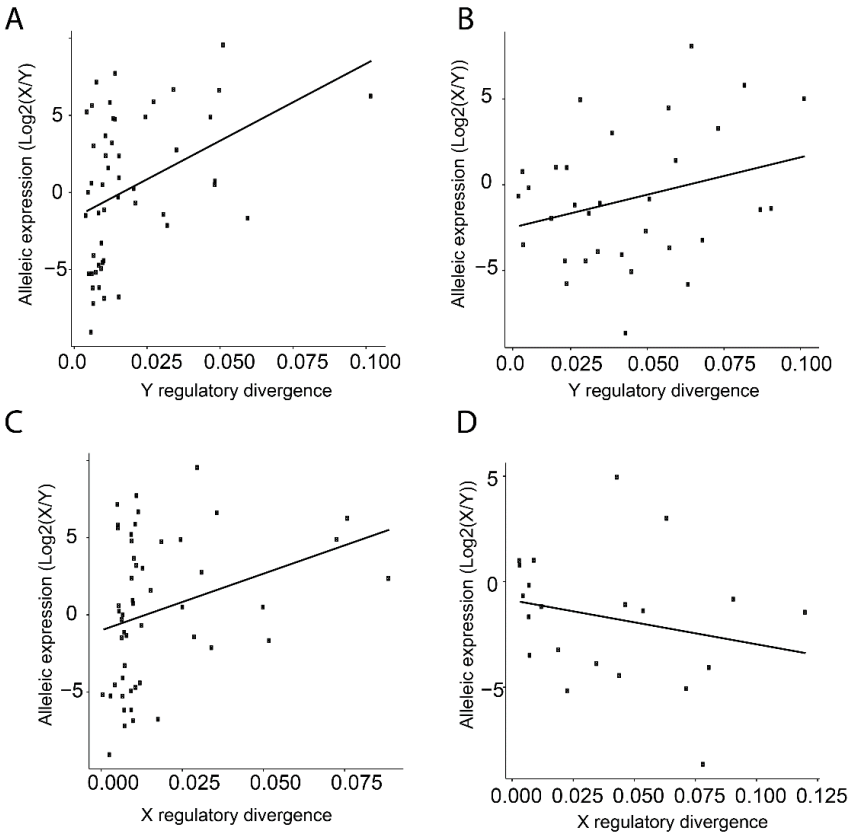
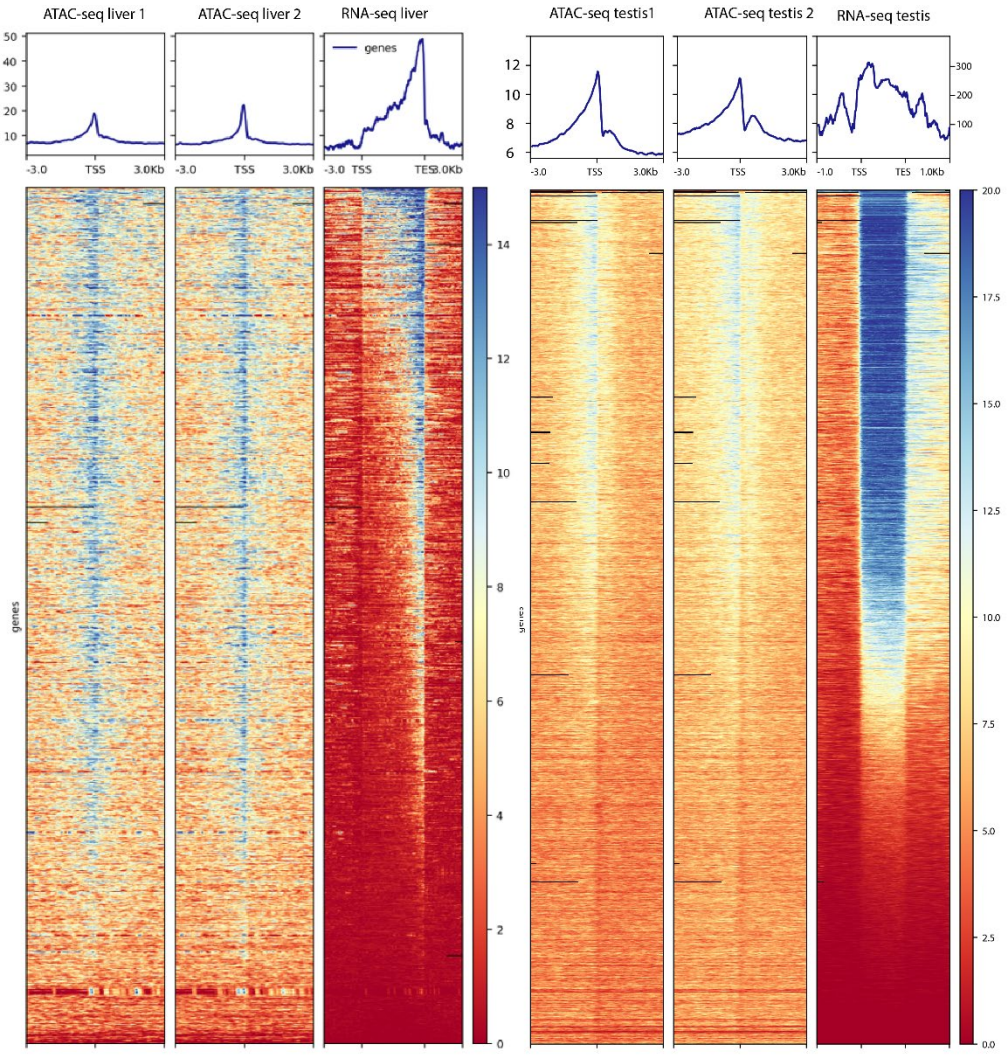
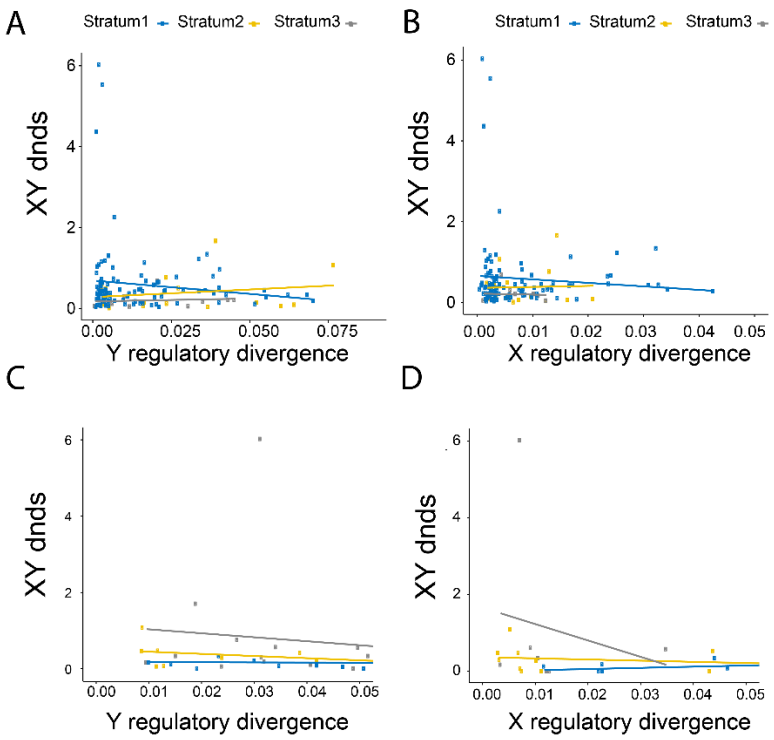


Figure 3.4. **Allele-specific regulatory evolution**
X- and Y- specific regulatory divergence is associated with X-biased gene expression across liver and testis samples. RNA-seq transcript counts uniquely mapped to the X and Y chromosome were identified to quantify allele-specific gene expression for all gametologs. Gene expression from gametologs was compared to average divergence of Y- specific (A and B) and X-specific (C and D) mutations for ACRs within 50 kb of expressed genes. (A). Y liver: R= 0.4, P = 0.0041, N = 65; (B.) Y Testis: R= 0.34, P = 0.008, N = 65; (C). X Liver: R = 0.3, P = 0.033, N = 77; (D). X Testis: R = 0.25, 0.061, N = 77.

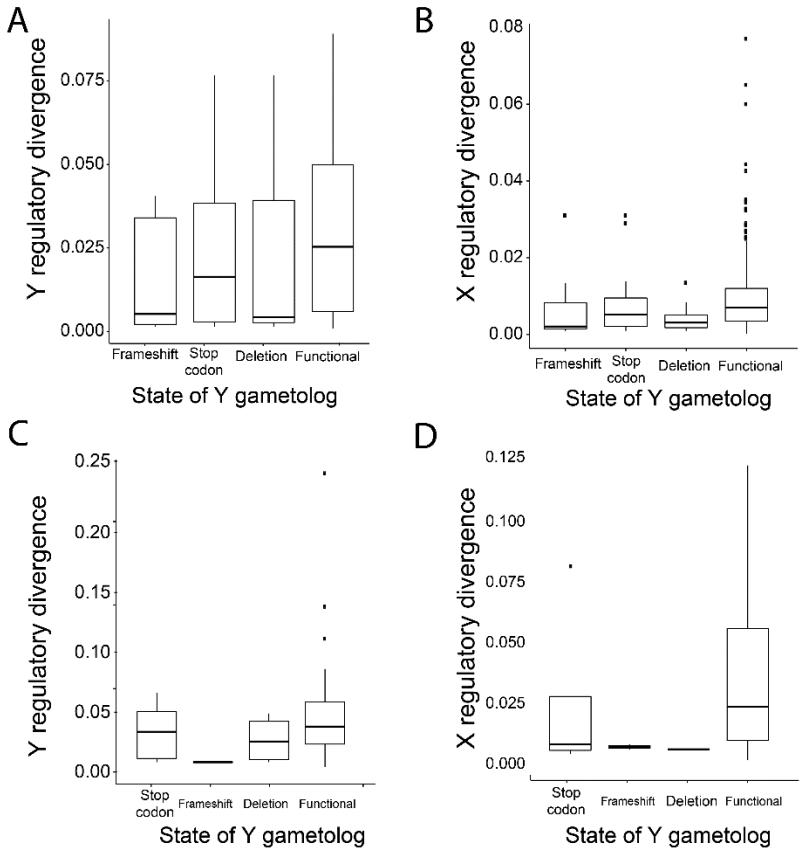
Supplemental Figures



Supplemental Figure 3.1. **High concordance of chromatin accessibility around transcription start site of expressed genes**. TSS for all genes were ordered by RNA-seq read depth. ATAC-seq read depth was assayed around within 3kb of transcription start site (TSS). Similar levels of enrichment of ATAC-seq depth were observed in replicates with high concordance of genes. ATAC-seq depth across replicates is a strong predictor of expression from RNA-seq from same tissue type.



Supplemental Figure 3.2. Regulatory divergence compared to XY coding divergence. There is no correlation between dN / dS and ACR divergence using a Spearman's Rank Correlation using ACRs from liver (A and B) or testis (C and D) tissue. Liver: (A) Y divergence: (Stratum one (Blue): R=-0.12 P = 0.230, N = 30, Stratum two (Yellow): R = 0.2, P = 0.430, N=61; Stratum three (Grey): R = 0.11, P = 0.770 N=126. (B) X divergence: (Stratum one (Blue): R = -0.080, P = 0.420, N=30; Stratum two (Yellow): R = 0.032, P = 0.9, N = 61; Stratum three (Grey): R = -0.054, P = 0.890, N = 126. Similar results in testis: (C) Y divergence: (Stratum one (Blue): R = -0.020, P = 0.650, N=14; Stratum two (Yellow): R = -0.430, P = 0.160, N = 12; Stratum three (Grey): R = -0.140, P = 0.480, N = 17. (D) X divergence: (Stratum one (Blue): R = 0.520, P = 0.120, N=14; Stratum two (Yellow): R = -0.360, P = 0.210, N = 12; Stratum three (Grey): R = -0.020, P = 0.670, N = 17.



Supplemental Figure 3.3. Regulatory divergence compared to XY coding divergence in testis.
Regulatory divergence is independent of XY coding evolution. Y- and X- linked regulatory divergence is not statistically different between gametologs with different categories of Y-linked structural variation. Y specific structural variation was identified by comparing coding sequence between the X, Y, and outgroup coding sequence. Y gametologs were determined to have premature stop codons, frameshift mutations, or in-frame deletions. All other coding sequence mapped between the X and Y is labeled functional. ACRs were assigned to the closest gametolog within 50kb. $P > 0.05$ for all pairwise comparisons for both liver (A and C) and testis (B and D) for regulatory divergence on the X and Y.

References

Alexandre, C. M., Urton, J. R., Jean-Baptiste, K., Huddleston, J., Dorrity, M. W., Cuperus, J. T., Sullivan, A. M., Bemm, F., Jolic, D., Arsovski, A. A., Thompson, A., Nemhauser, J. L., Fields, S., Weigel, D., Bubb, K. L., & Queitsch, C. (2018). Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in

Arabidopsis thaliana. *Molecular Biology and Evolution*, *35*(4), 837-854. https://doi.org/10.1093/molbev/msx326

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169. https://doi.org/10.1093/bioinformatics/btu638

Arnold, C. D., Gerlach, D., Spies, D., Matts, J. A., Sytnikova, Y. A., Pagani, M., Lau, N. C., & Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nature Genetics*, *46*(7), 685-692. https://doi.org/10.1038/ng.3009

Ayers, K. L., Davidson, N. M., Demiyah, D., Roeszler, K. N., Grützner, F., Sinclair, A. H., Oshlack, A., & Smith, C. A. (2013). RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biology*, *14*(3), R26. https://doi.org/10.1186/gb-2013-14-3-r26

Bachtrog, D. (2006). Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. *Current Biology*, *16*(17), 1694-1699. https://doi.org/10.1016/j.cub.2006.07.053

Bachtrog, D. (2008). The Temporal Dynamics of Processes Underlying Y Chromosome Degeneration. *Genetics*, *179*(3), 1513-1525. https://doi.org/10.1534/genetics.107.084012

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113-124. https://doi.org/10.1038/nrg3366

Bachtrog, D., Hom, E., Wong, K. M., Maside, X., & De Jong, P. (2008). Genomic degradation of a young Y chromosome in Drosophila miranda. *Genome Biology*, *9*(2), R30. https://doi.org/10.1186/gb-2008-9-2-r30

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C. (2014). Sex Determination: Why So Many Ways of Doing It? *PLoS Biology*, *12*(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899

Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2017). Genomic Loss and Silencing on the Y Chromosomes of Rumex. *Genome Biology and Evolution*, *9*(12), 3345-3355. https://doi.org/10.1093/gbe/evx254

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., Kremitzki, C., Fulton, R. S., Dugan, S., Ding, Y., Morton, D., Khan, Z., Lewis, L., Buhay, C., Wang, Q., Watt, J., Holder, M., Lee, S., Nazareth, L., Alföldi, J., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*, *508*(7497), 494-499. https://doi.org/10.1038/nature13206

Bellott, D. W., & Page, D. C. (2021). Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Research*, *31*(2), 198-210. https://doi.org/10.1101/gr.268516.120

Bellott, D. W., Skaletsky, H., Cho, T.-J., Brown, L., Locke, D., Chen, N., Galkina, S., Pyntikova, T., Koutseva, N., Graves, T., Kremitzki, C., Warren, W. C., Clark, A. G., Gaginskaya, E., Wilson, R. K., & Page, D. C. (2017). Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature Genetics*, *49*(3), 387-394. https://doi.org/10.1038/ng.3778

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213-1218. https://doi.org/10.1038/nmeth.2688

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Chaix, R., Somel, M., Kreil, D. P., Khaitovich, P., & Lunter, G. A. (2008). Evolution of Primate Gene Expression: Drift and Corrective Sweeps? *Genetics*, *180*(3), 1379-1389. https://doi.org/10.1534/genetics.108.089623

Chang, C.-H., Gregory, L. E., Gordon, K. E., Meiklejohn, C. D., & Larracuente, A. M. (2022). Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. *eLife*, *11*. https://doi.org/10.7554/elife.75795

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences*, *75*(11), 5618-5622. https://doi.org/10.1073/pnas.75.11.5618

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1403), 1563-1572. https://doi.org/10.1098/rstb.2000.0717

Connelly, C. F., Wakefield, J., & Akey, J. M. (2014). Evolution and Genetic Architecture of Chromatin Accessibility and Function in Yeast. *PLoS Genetics*, *10*(7), e1004427. https://doi.org/10.1371/journal.pgen.1004427

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M R., Armean, Irina M., Austine-Orimoloye, O., Azov, Andrey G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C. G., Genez, T., Martinez, Jose G., Guijarro-Clarke, C., Gymer, A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Marugán, J. C., Mohanan, S., Mushtaq, A., Naven, M., Ogeh, Denye N., Parker, A., Parton, A., Perry, M., Piližota, I., Prosovetskaia, I., Sakthivel, Manoj P., Salam, Ahamed Imran A., Schmitt, Bianca M., Schuilenburg, H., Sheppard, D., Pérez-Silva, José G., Stark, W., Steed, E., Sutinen, K., Sukumaran, R., Sumathipala, D., Suner, M.-M., Szpak, M., Thormann, A., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, Thomas A., Walts, B., Willhoft, N., Winterbottom, A., Wass, E., Chakiachvili, M., Flint, B., Frankish, A., Giorgetti, S., Haggerty, L., Hunt, Sarah E., IIsley, Garth R., Loveland, Jane E., Martin, Fergal J., Moore, B., Mudge, Jonathan M., Muffato, M., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, Stephen J., Dyer, S., Harrison, Peter W., Howe, Kevin L., Yates, Andrew D., Zerbino, Daniel R., & Flicek, P. (2021). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988-D995. https://doi.org/10.1093/nar/gkab1049

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant

call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Ellison, C., & Bachtrog, D. (2019). Recurrent gene co-amplification on Drosophila X and Y chromosomes. *PLoS Genetics*, *15*(7), e1008251. https://doi.org/10.1371/journal.pgen.1008251

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576-589. https://doi.org/10.1016/j.molcel.2010.05.004

Horvath, R., Josephs, E. B., Pesquet, E., Stinchcombe, J. R., Wright, S. I., Scofield, D., & Slotte, T. (2021). Selection on Accessible Chromatin Regions in Capsella grandiflora. *Molecular Biology and Evolution*, *38*(12), 5563-5575. https://doi.org/10.1093/molbev/msab270

Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Graves, T., Fulton, R. S., Dugan, S., Ding, Y., Buhay, C. J., Kremitzki, C., Wang, Q., Shen, H., Holder, M., Villasana, D., Nazareth, L. V., Cree, A., Courtney, L., Veizer, J., Kotkiewicz, H., Cho, T.-J., Koutseva, N., Rozen, S., Muzny, D. M., Warren, W. C., Gibbs, R. A., Wilson, R. K., & Page, D. C. (2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*, *483*(7387), 82-86. https://doi.org/10.1038/nature10843

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., Van Daalen, S. K. M., Minx, P. J., Fulton, R. S., McGrath, S. D., Locke, D. P., Friedman, C., Trask, B. J., Mardis, E. R., Warren, W. C., Repping, S., Rozen, S., Wilson, R. K., & Page, D. C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, *463*(7280), 536-539. https://doi.org/10.1038/nature08700

Hughes, J. F., Skaletsky, H., Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., Bellott, D. W., Cho, T.-J., Dugan-Rocha, S., Khan, Z., Kremitzki, C., Fronick, C., Graves-Lindsay, T. A., Fulton, L., Warren, W. C., Wilson, R. K., Owens, E., Womack, J. E., Murphy, W. J., Muzny, D. M., Worley, K. C., Chowdhary, B. P., Gibbs, R. A., & Page, D. C. (2020). Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Research*, *30*(12), 1716-1726. https://doi.org/10.1101/gr.269902.120

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kirkpatrick, M., Sardell, J. M., Pinto, B. J., Dixon, G., Peichel, C. L., & Schartl, M. (2021). Evolution of the canonical sex chromosomes of the guppy and its relatives. *G3 Genes|Genomes|Genetics*. https://doi.org/10.1093/g3journal/jkab435

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357-359. https://doi.org/10.1038/nmeth.1923

Lawson Handley, L. J., Berset-BräNdli, L., & Perrin, N. (2006). Disentangling Reasons for Low Y Chromosome Variation in the Greater White-Toothed Shrew (Crocidura russula). *Genetics*, *173*(2), 935-942. https://doi.org/10.1534/genetics.105.050203

Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex Chromosome Degeneration by Regulatory Evolution. *Current Biology*, *30*(15), 3001-3006.e3005. https://doi.org/10.1016/j.cub.2020.05.052

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663-666. https://doi.org/doi:10.1126/science.abj1813

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M., & Schmitz, R. J. (2017). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res*, *45*(6), e41. https://doi.org/10.1093/nar/gkw1179

Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., & Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants*, *5*(12), 1250-1259. https://doi.org/10.1038/s41477-019-0548-z

Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of Y chromosome gene content in flies. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00653-x

Mahajan, S., Wei, K. H. C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biology*, *16*(7), e2006348. https://doi.org/10.1371/journal.pbio.2006348

Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., Marais, G. A. B., & Touzet, P. (2019). Evolution of Young Sex Chromosomes in Two Dioecious Sister Plant Species with Distinct Sex Determination Systems. *Genome Biology and Evolution*, *11*(2), 350-361. https://doi.org/10.1093/gbe/evz001

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research*, *22*(7), 1255-1265. https://doi.org/10.1101/gr.132100.111

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., Leproust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*(6), 598-606. https://doi.org/10.1038/ng.3286

Muller, H. J. (1918). Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal Factors. *Genetics*, *3*(5), 422-499. https://doi.org/10.1093/genetics/3.5.422

Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. B. (2012). Rapid De Novo Evolution of X Chromosome Dosage Compensation in Silene latifolia, a Plant with Young Sex Chromosomes. *PLoS Biology*, *10*(4), e1001308. https://doi.org/10.1371/journal.pbio.1001308

Muyle, A., Zemp, N., Fruchard, C., Cegan, R., Vrana, J., Deschamps, C., Tavares, R., Hobza, R., Picard, F., Widmer, A., & Marais, G. A. B. (2018). Genomic imprinting mediates dosage compensation in a young plant XY system. *Nature Plants*, *4*(9), 677-680. https://doi.org/10.1038/s41477-018-0221-y

Naftaly, A. S., Pau, S., & White, M. A. (2021). Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Research*, gr.274282.274120. https://doi.org/10.1101/gr.274282.120

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkab007

Nesta, A. V., Tafur, D., & Beck, C. R. (2021). Hotspots of Human Mutation. *Trends in Genetics*, *37*(8), 717-729. https://doi.org/10.1016/j.tig.2020.10.003

Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics*, *39*(1), 197-218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Orr, H. A., & Kim, Y. (1998). An Adaptive Hypothesis for the Evolution of the Y Chromosome. *Genetics*, *150*(4), 1693-1698. https://doi.org/10.1093/genetics/150.4.1693

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02097-x

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841-842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160-W165. https://doi.org/10.1093/nar/gkw257

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-Guerra, M. K., Colomé-Tatché, M., Johannes, F., Rowley, M. J., Corces, V. G., Zhai, J., Scanlon, M. J., Buckler, E. S., Gallavotti, A., Springer, N. M., Schmitz, R. J., & Zhang, X. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nature Plants*, *5*(12), 1237-1249. https://doi.org/10.1038/s41477-019-0547-0

Rodríguez Lorenzo, J. L., Hobza, R., & Vyskot, B. (2018). DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant Silene latifolia. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4936-y

Rollins, J. L., Chiang, P., Waite, J. N., Von Hippel, F. A., & Bell, M. A. (2017). Jacks and jills: Alternative life-history phenotypes and skewed sex ratio in anadromous threespine stickleback (Gasterosteus aculeatus). *Evolutionary Ecology Research*, *18*(4), 363-382.

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, *20*(8), 437-455. https://doi.org/10.1038/s41576-019-0128-0

Shaw, D. E., & White, M. A. (2022). The evolution of gene regulation on sex chromosomes. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2022.04.004

Singh, N. D., Koerich, L. B., Carvalho, A. B., & Clark, A. G. (2014). Positive and Purifying Selection on the Drosophila Y Chromosome. *Molecular Biology and Evolution*, *31*(10), 2612-2623. https://doi.org/10.1093/molbev/msu203

Soh, Y. Q., Alföldi, J., Pyntikova, T., Laura, Graves, T., Patrick, Robert, Kremitzki, C., Koutseva, N., Jacob, Rozen, S., Jennifer, Owens, E., James, William, Cao, Q., Pieter, Wesley, Richard, Skaletsky, H., & David. (2014). Sequencing the Mouse Y Chromosome

Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*, *159*(4), 800-813. https://doi.org/10.1016/j.cell.2014.09.052

Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Sanchez, R. C., Fescemyer, H. W., Harris, R., Ye, D., O'Brien, P. C. M., Chikhi, R., Ryder, O. A., Ferguson-Smith, M. A., Medvedev, P., & Makova, K. D. (2016). A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. *Genome Research*, *26*(4), 530-540. https://doi.org/10.1101/gr.199448.115

Varadharajan, S., Rastas, P., Löytynoja, A., Matschiner, M., Calboli, F. C. F., Guo, B., Nederbragt, A. J., Jakobsen, K. S., & Merilä, J. (2019). A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evz240

Veltsos, P., Ridout, K. E., Toups, M. A., González-Martínez, S. C., Muyle, A., Emery, O., Rastas, P., Hudzieczek, V., Hobza, R., Vyskot, B., Marais, G. A. B., Filatov, D. A., & Pannell, J. R. (2019). Early Sex-Chromosome Evolution in the Diploid Dioecious Plant Mercurialis annua. *Genetics*, *212*(3), 815-835. https://doi.org/10.1534/genetics.119.302045

Vermunt, M. W., Tan, S. C., Castelijns, B., Geeven, G., Reinink, P., De Bruijn, E., Kondova, I., Persengiev, S., Bontrop, R., Cuppen, E., De Laat, W., & Creyghton, M. P. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nature Neuroscience*, *19*(3), 494-503. https://doi.org/10.1038/nn.4229

Wei, K. H. C., & Bachtrog, D. (2019). Ancestral male recombination in Drosophila albomicans produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. *PLoS Genetics*, *15*(11), e1008502. https://doi.org/10.1371/journal.pgen.1008502

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Molecular Biology and Evolution*, *32*(8), 1981-1995. https://doi.org/10.1093/molbev/msv078

Wilson Sayres, M. A. (2018). Genetic Diversity on the Sex Chromosomes. *Genome Biology and Evolution*, *10*(4), 1064-1078. https://doi.org/10.1093/gbe/evy039

Wilson Sayres, M. A., Lohmueller, K. E., & Nielsen, R. (2014). Natural Selection Reduced Diversity on Human Y Chromosomes. *PLoS Genetics*, *10*(1), e1004064. https://doi.org/10.1371/journal.pgen.1004064

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

CHAPTER 4

SINGLE-CELL GENE EXPRESSION OF SPERMATOGENESIS REVEALS EVOLUTION OF POST-MEIOITIC REPRESSION AND Y-LINKED GENE DUPLICATE ACTIVATION IN THREESPINE STICKLEBACK FISH [3]

[3] Shaw, D.E. and White, M.A. To be submitted to *Genome Biology and Evolution*.

Abstract

Heteromorphic sex chromosomes tend to accumulate testis-biased genes that exhibit rapid coding and expression evolution. Additionally, some ancient sex chromosomes have evolved dynamic gene regulation in spermatogenesis related to gene silencing. Recently evolved sex chromosomes provide an opportunity to identify into how fast testis-biased genes accumulate on Y chromosomes, and how chromosome-wide transcriptional regulation evolves. Here, we performed single-cell RNA sequencing to isolate gene expression profiles across all stages of threespine stickleback fish spermatogenesis. We found that, unlike mammals, the threespine stickleback fish sex chromosomes maintain expression throughout meiosis. However, we did find signatures of post-meiotic sex chromosome repression (PSCR), providing additional evidence that post-meiotic sex chromosome expression can be down-regulated independent of meiotic silencing. We also found that transcripts derived from Y-linked duplicates are enriched in later stages of spermatogenesis. Together our findings support the convergent evolution of post-meiotic silencing, and reactivation of important genes on sex chromosomes.

Introduction

The testis is a morphologically diverse and heterogeneous organ essential for male sexual reproduction across many species. The testis is comprised of multiple cell types with diverse functions and selective pressures. Many genes involved in early cell differentiation (spermatogonia) and meiosis (spermatocytes) are highly conserved across species (Bonilla & Xu, 2008; Dapper & Payseur, 2019; Malik et al., 2008; Ramesh et al., 2005; Sun et al., 2008; Tvedte et al., 2017), however, many genes expressed in meiotic and post-meiotic cells (spermatids) are often highly specialized and under positive selection (Dapper & Payseur, 2019; Good &

Nachman, 2005; E. L. Larson et al., 2016; Swanson et al., 2003). Sex chromosomes (i.e., X and Y) are hotspots for accumulating genes important for testis function (Mahajan & Bachtrog, 2017; Martínez-Pacheco et al., 2020; Mueller et al., 2008; Soh et al., 2014). Sex-linked genes can gain male-specific functions through novel mutation. This is evident through widespread signatures of positive selection in coding regions compared to autosomes, along with the evolution of sex-biased gene expression (Kopania. et al., 2022; Larson et al., 2016; Meisel et al., 2012). Many genes important for spermatogenesis have translocated on to the sex chromosomes, and have duplicated to form large ampliconic regions (Bachtrog et al., 2019; Hughes et al., 2020; Soh et al., 2014). Ancient sex chromosomes have revealed much about how sex-linked testis genes evolve over long periods of time, but we know less about the early stages of sex chromosome evolution, including how fast gene expression changes on recently formed sex chromosomes.

The ancient sex chromosomes, shared across mammals, undergo complete transcriptional silencing through a process known as meiotic sex chromosome inactivation (MSCI) (Turner, 2007, 2015). This process is caused by the accumulation of a phosphorylated histone variant, γH2AX, at regions of un-synapsed chromosomes (Turner et al., 2005). The sex chromosomes remain largely repressed throughout post-meiotic stages (Post-meiotic sex chromosome repression) (Namekawa et al., 2006; Turner et al., 2006), although some genes escape silencing in later stages of spermatogenesis (Cocquet et al., 2012; Cocquet et al., 2009; Moretti et al., 2016; Mueller et al., 2008). The evolutionary history of meiotic silencing and post-meiotic repression is unclear. Similar mechanisms are known to operate in *C. elegans*, where un-synapsed chromosomes across the genome are silenced (Maine et al., 2005). However, it remains unclear if many heteromorphic sex chromosomes outside of mammals exhibit similar forms of transcriptional silencing. Heteromorphic sex chromosomes in *drosophila* (Mikhaylova &

Nurminsky, 2011; Vibranovski, 2014) and chicken (Guioli et al., 2012) do not have canonical MSCI, but recent evidence in grasshopper suggests that MSCI can evolve independently (Viera et al., 2021). Model organisms with recently derived sex chromosomes present an opportunity to study the evolution of broad transcriptionally silencing across sex chromosomes in spermatogenesis.

Threespine stickleback fish have an XY sex determination system that evolved less than 26 million years ago (Peichel et al., 2020). Despite early signatures of degeneration, over half of the ancestral genes are still present on the Y chromosome. In addition to ancestral genes, small families of gene duplicates are amplifying on the Y chromosome, also with testis-biased expression (Peichel et al., 2020).

In this study, we used single-cell RNA sequencing to define cell types across threespine stickleback spermatogenesis. We identified markers for distinct cell clusters associated with stages of spermatogenesis. We found that X and Y linked genes are expressed across meiosis, suggesting that MSCI has not evolved on the stickleback sex chromosomes, despite extensive genetic differentiation. However, we did find evidence of post-meiotic gene repression on the sex chromosomes, suggesting silencing can evolve post-meiotically, independent of meiotic silencing. Our study provides insight into the early stages of sex-linked evolution of meiotic gene expression and gene content.

Methods

Dissociation of testes for single-cell RNA sequencing

Testes were dissected from two meiotically active male threespine stickleback fish. Both males were collected from separate clutches of lab strains derived from wild caught fish from Port Gardner Bay, Washington. Dissociation was performed separately for each individual.

Individuals were selected based on age (~7 months), testis size, and testis coloration, all of which are indicators that male fish are becoming reproductive and spermatogenesis is proceeding through meiosis. Spermatogenesis is synchronized in threespine stickleback fish (Sokołowska & Kulczykowska, 2006). The testes are enriched for cells at the same stages of spermatogenesis until all mature sperm are produced for the breeding season. We therefore selected a male fish that was enriched for meiotic cells and another enriched for post-meiotic cells. The first male had slightly smaller and lighter colored testes, consistent with earlier meiotic stages, while the second had larger, darker testes consistent with later stages of testis maturation. Both testes from each male were placed in ice cold PBS + 0.04% BSA. We followed a previously published single-cell dissociation method from mice testis (Lukassen et al., 2018). The tissue was minced with McPherson-Vannas scissors in 200 µL of Digestion medium (1mg/mL DNASI, 1mg/mL collagenase/dispace, 1mg/ML hyaluronidase in solution of DMEM/F12) until the tissue was broken down into small pieces. The cell solution was incubated at 37º C for 10 minutes. The cells were filtered through a 40 µm cell strainer to obtain a single-cell suspension. Cells were pelleted by centrifugation at 4º C, 400g for 10 minutes. Cells were resuspended in 1mL of cold PBS and BSA. Cell viability was assessed using trypan blue. We approximated cells to have 90% viability. The cells were immediately processed for sequencing on the 10x Genomics single-cell sequencing platform performed at the Georgia Genome and Bioinformatics Core.

Processing of Single-cell RNA-sequencing data

Reads were demultiplexed with CellRanger (v 4.0.0) mkfastq with bcl2fastq (Illumina). We ran CellRanger mkref on the threespine stickleback genome assembly (v5.0) (Nath et al., 2021; Peichel et al., 2020). We aligned fastq reads from both samples separately using CellRanger count with default parameters, plus the parameter --chemistry=SC3Pv3. To filter

cells, we used Seurat (v 4.0) (Hao et al., 2021) to remove cells with > 3000 and <200 unique features (i.e., expressed transcripts) and cells with > 5% mitochondrial counts. Cells with too many or too few identified features likely represent doublets or dead cells, respectively. Expression across each transcript were normalized using Seurat scaledata. We performed filtering on each sample first separate and then later on the merged dataset. We merged the datasets using two distinct methods; 1.) We combined cell matrices using the CellRanger aggr command with default parameters which normalizes read depth between the two samples, 2.) We filtered each individual using Seurat (v4.0), and then combined Seurat objects using Seurat merge with default parameters. Both methods lead to similar clustering (Supplemental figure 4.1) and overlap.

Clustering and UMAP

We followed the Seurat (v4.0) workflow (Hao et al., 2021) to perform cell clustering. We determined dimensionality of the dataset by performing a jackstraw permutation in Seurat. Based on the elbow plot and jackstraw plot (Supplemental Figure 4.2), we found that 12/20 principal components had P-values < 0.001 for sample 1, while 15/15 of principle components had P-values < 0.001. To keep clustering consistent, we set the number of dimensions to 12 for both samples. We ran Seurat findneighbors (Dims:1:12) and Seurat findclusters (Resolution =0.8). We ran the Seurat UMAP function set to 12 dimensions (Dims=1:12). The resolution factor is used for is recommended to be between 0.4-1.2 (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html). We experimented with a range of values from 0.4-1.2 and found that the number of clusters from 0.8 was the most consistent with marker identification and expectations based on histology (Borg, 1981; Sokołowska & Kulczykowska, 2006). We verified that most clusters had a similar number of reads per cell, and

expressed features per cell, to avoid identifying clusters based on read depth alone (Supplemental Figure 4.3).

Identification of cell clusters

Marker genes were identified using the FindAllMarkers function in Seurat. A Mann-Whitney U test was used to compared gene expression of each gene across all clusters to identify significantly different gene expression profiles between clusters. Genes were considered differentially expressed between clusters if there was a p-value < 0.001. We identified genes in the stickleback genome that had known functions in spermatogenesis and compared expression of these genes across clusters. We compared upregulated markers of clusters to characterized spermatogenesis cell-type-specific genes from mammals (Green et al., 2018; Hermann et al., 2018; S. Lukassen et al., 2018; Lukassen et al., 2018b; Shami et al., 2020; Witt et al., 2019) and fish (Qian et al., 2022; Wu et al., 2021). We also identified genes expressed during meiosis that are conserved at the coding sequence across mammals (Dapper & Payseur, 2019; Ramesh et al., 2005). These genes are likely under purifying selection to maintain function and are likely to also be expressed during meiosis in fish.

Gene ontology

We used the ShinyGO v0.741 web-based tool to identify gene ontology of upregulated genes for each cluster. We used all stickleback genes expressed in our single-cell RNA dataset as the background, to identify enrichment of biological processes involved in the significantly up-regulated genes of each cluster. For each set of genes, fold enrichment of percentage of genes within a pathway along with a false discovery rate (FDR). We filtered for GO terms with enrichment FDR < 0.01.

Gene expression across the sex chromosomes

In order to analyze chromosome-wide expression patterns across cells within a cluster, we extracted alignments from each cluster output using pysam (v 0.16.0.1) and a custom python script (https://github.com/daniel-shaw1/Single-Cell-RNA-sequencing/blob/main/Split_bam_by_cluster.py). Splitting of alignment files was ran separately for each sample based on CellRanger clustering, as BAM files were direct outputs of the CellRanger count pipeline. For each cluster alignment, we counted total transcripts for each cell using htseq (v 0.13.5) with a GTF comprised of the Y gene annotations (Peichel et al., 2020). We also counted transcripts for autosomal and X-linked genes using ensembl build 107 predicted annotations (Cunningham et al., 2021). In order to maximize the number of transcripts counted on the sex chromosomes, we ran htseq with the settings --mode union --nonunique all --minaqual 1 –stranded=no. For shared gametologs on the X and Y, we aligned the Ensembl coding sequences from the X chromosome to the Y chromosome. For X and Y genes, reads were counted across the entire transcript as one feature. We normalized htseq counts across clusters by dividing by the total library size of each cluster.

Duplicate gene expression

Y-linked gene duplications were previously annotated for the Y chromosome assembly (Peichel et al., 2020). To analyze overall expression level of each paralog family, we collapsed counts of each paralog into a single count for each family. We normalized counts for each duplicate family per cluster by dividing by the library size of each cluster.

Protein coding divergence

To compare coding sequence evolution, we aligned the threespine stickleback fish CDS (Ensembl release 107) (Cunningham et al., 2021) to the ninespine stickleback genome

(Varadharajan et al., 2019) using Exonerate (v2.4) (exonerate --model est2genome -M 2000 --bestn 15 --percent 25). We ran codeml from the PAML (phylogenetic analysis by maximum likelihood) (runmode =-2) software package to calculate d_N, d_S, and d_N/d_S for all mappable coding sequence between threespine stickleback fish and ninespine stickleback fish. Exonerate alignments were filtered to remove stop codons and gaps, and processed into codeml using a custom Perl script (https://github.com/daniel-shaw1/Single-Cell-RNA-sequencing), modified from previous SNP based comparisons (White et al., 2015), improved to compare between species with reference assemblies. d_N/d_S ratios were averaged across transcripts for, genes with multiple transcripts. Genes with a d_N/d_S value of 99 were removed.

Results

Identification of stickleback spermatogenic stages using single-cell RNA sequencing

To identify single-cell gene expression profiles for diverse cell types, we sequenced two fish at different stages of synchronized spermatogenesis (See methods). We successfully sequenced 6147 cells from a male from an earlier stage of spermatogenesis, and 2652 cells from a male at a later stage of spermatogenesis. Both individuals had a similar number of reads mapped and reads per cell (Supplemental Figure 4.2). To map the entirety of spermatogenesis, we aggregated the two samples together. Between both samples, we recovered a total of 8799 cells. Graph-based UMAP clustering revealed 12 distinct cell clusters (Figure 4.1). The distribution of UMAP clusters using both CellRanger (Supplemental Figure 4.1) and Seurat (Figure 4.1) were largely concordant. By performing a Mann-Whitney U test of each gene across all clusters, we identified 2320 differentially expressed marker genes between all clusters. Genes were considered differentially expressed within a cluster if the gene was expressed in 15% of cells within the cluster, had a $\log_2$ fold change > 0.25 for expression compared to all other

clusters, and had a significantly different level of expression evaluated with a P-value < 0.01 after correcting for multiple comparisons. We used these genes to identify markers to assign putative cell types to clusters (Table 1).

We found that the cell clustering reflected cell types at different stages of spermatogenesis (Figure 4.1; Table 4.1). In most teleost fish, a population of undifferentiated spermatogonia are maintained throughout the year (Sokołowska & Kulczykowska, 2006; Uribe et al., 2014). These cells divide through mitotic cell division and are maintained within an interstitial germinal compartment (Uribe et al., 2014). We found one cluster of cells with expression of genes known for roles in spermatogonia maintenance and homeostasis; the first had notable enrichment of *id4* (Figure 4.1; P < 0.001, Mann-whitney U test), a common marker of undifferentiated spermatogonia across species (Green et al., 2018; Helsel et al., 2017; Oatley et al., 2011). Additionally, we found nine upregulated genes in this cluster that are identified as novel on ensembl, suggesting there may be stickleback or teleost specific genes involved in this stage of spermatogenesis (Supplemental table 1).

In response to developmental signals like 11- Ketotestosterone, and retinoic acid, and growth hormone, spermatogonia differentiate, as the first step towards meiotic cell division. We found a cluster with multiple markers related to differentiated spermatogonia, identifiable by enrichment of *notch1a* , and *grna* (P < 0.001, Mann-whitney U test) (Daniel et al., 2003; Parekh et al., 2019). Additionally, early and late spermatogonia clusters were enriched for unique gene ontology terms associated with each cluster (Table 1; Supplemental table 2). The early spermatogonia was enriched with pathways related to homeostasis (FDR < 0.001), while the late spermatogonia instead had enrichment of cell differentiation pathways (FDR < 0.001). Together,

this suggest that we identified one population of undifferentiated spermatogonia being maintained for later cell divisions, and one population actively undergoing differentiation.

We found the largest population of cells were associated with specific meiotic processes. We identified four clusters as meiotic spermatocytes. We were able to identify uniquely enriched genes within each cluster well enough to characterize spermatocytes as 'early-' 'mid-', 'mid-late-' and 'late-' spermatocytes, each corresponding to distinct timepoints in meiosis (Figure 4.1; Table 4.1). In early spermatocytes, we found upregulated genes with known functions in the early stages of prophase, like genes that alter chromatin structure before DSB initiation, such as *piwil1* ($P < 0.001$, Mann-Whitney U test) (Kuramochi-Miyagawa et al., 2004) and *hells* ($P < 0.001$, Mann-whitney U test) (Spruce et al., 2020) and some genes important for forming synaptonemal complex (*smc2b, smc1al,* $P < 0.001$, Mann-whitney U test*)* (Eijpe et al., 2000; Neeb et al., 2017; Revenkova et al., 2004). The mid-spermatocytes cluster exhibited enrichment of genes involved in DNA repair and strand invasion (*rad51*, *rad54*, and *meig1*, $p < 0.001$, Mann-whitney U test) and other genes associated with the pachytene stage of prophase (Table 1) (Li et al., 2016; Mazin et al., 2010). The third meiotic cluster, we identified as mid-late spermatocytes, as they contain a mix of prophase markers involved in maintaining the synaptonemal complex (*smc3*, $P <0.001$, Mann-whitney U test) (Eijpe et al., 2000), but also genes involved in late meiosis, such as genes important for assembling spindles (*rab24*, $P < 0.001$, Mann-Whitney U test), that are required for chromosome segregation (Qiu et al., 2019). Lastly, we identified one cluster as late spermatocytes (Table 1). In this stage we found GO terms enriched for sister chromatid segregation (FDR < 0.001), and kinetochore formation (FDR = 0.002), suggesting cells in this cluster in the later stages of meiosis. (Table 1). Interestingly, we also found enrichment of genes related to changes in chromatin structure, (*cbx1*, *smarca5*, P

< 0.001, Mann-whitney U test) that may be important for transitioning into round spermatid. We additionally found expression of several meiotic genes conserved across vertebrates that were not significantly enriched in one cluster, but instead seemed broadly expressed across the 4 meiotic clusters (Supplemental Figure 4.4).

Clusters with genes involved in late spermatogenesis were largely present in the more developed testis sample. Round spermatids and elongating spermatids are in the haploid stage of cell division. Genes expressed at this stage are expected to regulate sperm formation, and have significant roles in sperm motility. We found two distinguishable clusters corresponding to early- and late- spermatids. Early spermatids were characterized with genes with known functions in spermiogenesis such as, *ak1*, *taf9*, and *ybx1* (P < 0.001, Mann-Whitney U test) (Cao et al., 2006; Klaus et al., 2016; Liang et al., 2021). GO terms enriched in this stage included axoneme assembly (FDR < 0.01), a sperm structural component important for motility (Linck et al., 2016). We identified several relevant spermatid markers including *gapdhs* (P < 0.001, Mann-Whitney U test), well-known markers for mice spermatid (Hermann et al., 2018; Margaryan et al., 2015) and *krt8* (P < 0.001, Mann-Whitney U test) identified in bull spermatozoa (Feugang et al., 2010).

Lastly, we identified a cluster as likely a mix of different testis somatic cells. In teleost fish, the somatic components of the testis are made of connective tissue, including blood vessels, myoid cells and immune cells, along with Leydig cells and Sertoli cells (Uribe et al., 2014). A large group of clusters was located the farthest from the meiotic clusters in UMAP space, suggesting the greatest difference in overall gene expression. All somatic cells likely clustered together as a group when plotted against the other closely related meiotic cells. Across the large somatic cluster, we found *gata1a* (Figure 4.1), a canonical marker of Sertoli expression (Wakabayashi et al., 2003), was significantly upregulated (P < 0.001, Mann-Whitney U test),

along with *ccng1* ($P < 0.001$, Mann-Whitney U test) a known marker of Leydig's cells (Lin et al., 2008). Additionally, we found high expression of genes known to function in connective and immune tissues like *hbae5* ($P < 0.001$, Mann-Whitney U test). Consistent with these markers, we found GO terms consistent with erythroid cell development (FDR < 0.001), myeloid cell homeostasis (FDR < 0.001), and immune development (FDR < 0.001). Additionally, we found two additional clusters with no enrichment of spermatogenesis genes, that may represent other somatic cell cluster, but we were unable to connect differentially expressed genes to known cell types in testes.

Sex-linked gene expression across spermatogenesis

In mammals, the sex chromosomes show complete depletion of transcripts compared to autosomes (meiotic sex chromosome inactivation; MSCI) (Turner et al., 2005; Turner, 2007, 2015). To determine whether meiotic silencing has evolved in threespine stickleback fish, we compared average rates of transcription between sex-linked and autosomal genes across cell clusters. When normalized for genome wide expression patterns within each cluster, we found that gene expression is not silenced on the X chromosome (Figure 4.2). While the Y chromosome does have lower expression than the X chromosome, Y-linked genes maintains similar levels of expression across all stages of meiosis (Figure 4.2). In contrast to the mammalian germline, we do not observe chromosome-wide silencing on the threespine stickleback sex chromosomes during meiosis.

In mammalian round spermatids, the sex chromosomes remain largely repressed (Namekawa et al., 2006; Turner et al., 2006), while a subset of genes are reactivated (Cocquet et al., 2012; Cocquet et al., 2009; Moretti et al., 2016; Mueller et al., 2008). This is referred to as post meiotic sex chromosome repression (PSCR) (Namekawa et al., 2006). We found that the

stickleback X and Y chromosome exhibited a significant reduction in gene expression in post-meiotic clusters (Figure 4.2). This finding suggests that the stickleback sex chromosomes are repressed post-meiotically, in a similar manner as mammals, despite absence of meiotic silencing. In mammals, PSCR is associated with the accumulation of sex-linked heterochromatin (Hirota et al., 2018; Moretti et al., 2016). Consistent with this, we found that elevated expression of heterochromatin protein coding gene *cbx1* ($P < 0.001$) enriched in late meiosis (Supplemental figure 4.5), the stage of cells prior to sex-linked reduction in expression

We next sought to profile expression across differentially diverged regions of the X and Y. The suppression of recombination on sex chromosomes often occurs in a stepwise progression, resulting in distinct evolutionary strata. In threespine stickleback fish, three evolutionary strata have evolved, all at distinct levels of sequence divergence corresponding to age (Peichel et al., 2020). We hypothesized that distinct patterns of expression may be observed across each stratum. To test this, we compared gene expression across each of the evolutionary strata compared to autosomal expression. For the X chromosome, we found that transcriptional dynamics across the younger two strata were similar (Figure 4.3), and consistent with chromosome wide patterns including post-meiotic repression. However, we found a distinct difference on the oldest region of the X chromosome, where gene expression was elevated across spermatogenesis (Figure 4.3). In the oldest region the average expression of single copy X-linked genes was elevated in multiple cell types compared to compared to autosomes. This finding suggests that with higher divergence time, X-linked gametologs have evolved higher testis expression.

Across each independently derived strata the Y -linked gametologs has similar expression dynamics across spermatogenesis (Figure 4.3). We identified downregulation across all three

strata, corresponding with the transition to post-meiotic cell types. After gene expression is drastically down regulated, average gene expression increases in late spermatids. This expression gain in the later stages of spermatogenesis may be due to genes important for sperm development. We performed a GO-term enrichment analysis for these reactivated genes and found three enriched pathways all associated with troponin ($P < 0.001$), enriched due to the expression of three unique troponin genes.

Y gene duplicate expression is enriched in late spermatogenesis

In addition to the shared genes on the X and Y, the threespine stickleback has Y- specific genes that have translocated to the Y from autosomes. Some of these genes have been duplicated and exhibit testis-biased expression (Peichel et al., 2020), and may represent the early steps of gene amplification observed on ancient Y chromosomes (Hughes et al., 2020; Soh et al., 2014). Of 168 gene duplicate families identified previously (Peichel et al., 2020), we found detected expression of over half of these in our spermatogenesis cell types. We found that most Y duplicate families were expressed in late spermatids and showed higher relative expression compared to all other cell types (Figure 4.4A). Transcripts derived from gene duplicates are found depleted in early spermatid, and then subsequently increase expression in late- spermatid to reach an average expression significantly higher than other stages across spermatogenesis (Figure 4.4B; $P < 0.01$, Mann-Whitney U test). Combined, our results suggest evolution of Y-linked amplification of post-meiotic expressed genes in threespine stickleback fish.

Relaxed selective constraint in late spermatogenesis

In other species, different stages of spermatogenesis have been shown to have opposing signatures of selection. In mice, cell types in early stages are more evolutionarily constrained, while later stages show signatures of rapid evolution (Kopania. et al., 2022). We tested if this

pattern was evident throughout threespine stickleback spermatogenesis by identifying rates of non-synonymous and synonymous substitutions within genes enriched for each of our cell-type clusters. We found the lowest d_N/d_S ratios for genes enriched in spermatogonia was within early meiotic clusters, while genes enriched in round- and late- spermatid clusters had higher ratios (Figure 4.5A). This pattern suggests there are similarities in rates of molecular evolution between fish and mammals.

The X chromosome exhibited unique patterns of molecular evolution. We found that genes upregulated in meiosis have significantly elevated d_N/d_S compared to other stages, including post-meiotic clusters (Figure 4.5B). This finding suggests that on the sex chromosomes of threespine stickleback, genes expressed during meiosis are more likely to exhibit relaxed selective constraints or perhaps positive selection. This finding may be due to transcriptional patterns observed, where meiosis genes are expressed due to absence of MSCI, while gene expression is repressed post-meiotically.

Discussion

Single-cell RNA-seq accurately identified different stages of threespine stickleback spermatogenesis

Stickleback fish spermatogenesis occurs synchronously (Borg, 1981), where sperm are not continuously replaced, but instead develop seasonally in conjunction with breeding seasons. We found that clustering of cells based on single-cell RNA transcriptomics was similar to previously described histological characterization (Borg, 1981; Borg & Veen, 1982; Sokołowska & Kulczykowska, 2006) and also supports evidence that stickleback sperm mature synchronously. However, despite largely synchronous cell division, we were able to use single-cell transcriptome to distinguish different steps in spermatogenesis. Using genes of previously

published transcriptomics spermatogenesis studies, we were able to find genetic markers to characterize many expected stages of spermatogenesis. We found a greater proportion of meiotic cells compared to other cell type, consistent with previous understanding of timing for sexually mature males in synchronous spermatogenesis, where cysts tend to develop a large proportion of cells as spermatocytes, and are not replaced as they transition through the stages of meiosis (Borg, 1981; Sokołowska & Kulczykowska, 2006). Additionally, the pachytene stage of meiosis is most commonly observed in histological sampling of fish spermatogenesis, as cells progress rapidly through the later stages of meiosis 1 and meiosis II (Uribe et al., 2014), which means experiments are less likely to capture these rarer cell types.

While we were able to distinguish multiple clusters of spermatocytes, we found some overlap in expression of previously identified meiotic genes. This could mean that all of these clusters are at a similar stage of meiosis implying that spermatocytes within cysts have similar transcriptomes as they progress synchronously. Alternatively, many core meiotic genes may be expressed in stickleback across meiosis. The latter is possible, as many meiotic genes have detected expression across spermatocytes and spermatids even in species that undergo continuous spermatogenesis (Lukassen et al., 2018). We found gene expression profiles are consistent with 'Stage 1' of spermatogenesis (Sokołowska & Kulczykowska, 2006), when the spermatogenic cysts are mostly filled with spermatocytes. Interestingly, we were able to find spermatogonia maintained in the more developed testis sample, despite finding most that most cells had advanced beyond meiosis (Supplemental figure 4.1). This finding supports histological evidence, that a small population of spermatogonia are maintained and still present in stage 5 of spermatogenesis when sperm are ready for release (Sokołowska & Kulczykowska, 2006).

Overall, we were able to recapitulate patterns of synchronous spermatogenesis using single-cell transcriptomics, and identify markers for distinct stages of spermatogenesis.

No evidence for meiotic sex chromosome inactivation in stickleback fish.

Here, we found creating a single-cell expression atlas to be useful for comparing sex-linked expression across the temporal trajectory of spermatogenesis. We found that the recently evolved sex chromosomes of threespine stickleback do not exhibit widespread silencing in gene expression across meiotic cell types. The molecular mechanism of MSCI in mammals is that silencing is initiated by the unsynapsed chromatin of the X and Y in male meiosis (Hirota et al., 2018; Turner, 2015). In mammals the lack of proper pairing along the non-PAR regions, results in a delayed DSB repair (Lu & Yu, 2015). This likely triggers the accumulation of γH2AX protein along the asynapsed region. The accumulation of γH2AX has been shown to lead to the recruitment of heterochromatin (Hirota et al., 2018). Interestingly, the stickleback sex chromosomes pair linearly during meiosis (Nath et al., 2022), despite large rearrangements between the X and Y chromosome, and accumulation of sequence divergence (Peichel et al., 2020). The pairing of the X and Y is likely mediated via non-homologous or micro-homology directed synapsis. Our results suggest that lack of homology is not sufficient to trigger meiotic silencing, at least in stickleback fish. It remains unclear if any histone modifications accumulate on the X and Y chromosome during meiotic progression in stickleback fish to help repair double stranded breaks.

Meiotic silencing in mammals may have evolved from an ancestral mechanism. Identified in *C. elegans*, meiotic silencing of unsynapsed chromosomes leads to local silencing of gene expression (Maine et al., 2005; Schimenti, 2005). But studies across species have revealed the extent of silencing is variable within mammals (Federici et al., 2015), additionally,

silencing and mechanisms are variable across other species with sex chromosomes (Guioli et al., 2012; Vibranovski, 2014; Viera et al., 2021). It remains unclear what steps would need to take place to evolve a similar silencing mechanism in threespine stickleback. For example, there may be genes essential for meiosis, present on the ancestral regions of the sex chromosomes. In mammals, these genes have duplicated on to autosomes (Bradley et al., 2004; Emerson et al., 2004). Additionally, asynapsis may lead to meiotic arrest in cells that do not undergo meiotic silencing (Turner, 2015). Stickleback fish could be an ideal model to genetically induce asynapsis to identify if silencing is required in viable asynaptic cells.

Post-meiotic repression has evolved on the recently evolved sex chromosomes before meiotic silencing.

Post-meiotic sex chromosome repression (PSCR) has multiple mechanistic hypotheses. Initially, the discovery of PSCR was proposed to be a continuation of meiotic silencing (Namekawa et al., 2006; Turner et al., 2006). In this model, meiotic silencing leads to a repressive chromatin state by inducing the accumulation of H3K9me3 (Hirota et al., 2018), and post meiotic repression is simply maintaining the chromatin state into late spermatids (Namekawa et al., 2006), with only the exception of a few genes overcoming the silencing effect (Cocquet et al., 2012; Cocquet et al., 2009; Mueller et al., 2008). However, recent evidence proposed that drastic chromatin remodeling occurs in the transition of mice spermatocytes into spermatids in (Moretti et al., 2016). This model proposed that PSCR is an independent process and may be a mechanism of silencing intragenomic conflicts between the X and Y chromosomes along with other amplified regions of the genome (Moretti et al., 2016). Our findings are consistent with this hypothesis. We found that the stickleback sex chromosomes exhibit signatures of post-meiotic repression, without first undergoing chromosome-wide silencing in

meiosis. Our findings provide additional evidence that meiotic silencing and PSCR are independent processes. While meiotic silencing has not evolved in stickleback because the chromosomes maintain synapsis, post meiotic repression has seemingly evolved first and does act on Y-linked duplicates.

Similar to mammals, we found that a burst of expression follows sex chromosome repression, particularly on the Y chromosome. In mammals, many genes 'escape' post-meiotic repression. Some of these genes are predicted to be involved intragenomic conflict because they have coamplified on the X and Y chromosome, and are expressed in late spermatogenesis (Kruger et al., 2019; Mueller et al., 2008). The stickleback Y chromosome has accumulated only a fraction of the multi-copy gene families observed on older Y chromosomes (Bachtrog et al., 2019; Hughes et al., 2020; Peichel et al., 2020; Soh et al., 2014). Despite being in the early stages of gene amplification, here we found that over half of the gene duplicates found on the stickleback Y chromosome are expressed in spermatid. This may suggest that intragenomic conflict shaping sperm morphology or sex ratio drive can explain the accumulation of Y-linked gene duplicates. Alternatively, round spermatid expressed genes may be poised to duplicate on the Y. Genes expressed in these stages be more likely to duplicate due DSB repair (Álvarez-González et al., 2022), or due to the relaxed selective constraints associated with the advanced specialization in later stages of development (Abzhanov, 2013; Kopania et al., 2022). Regardless of mechanism, the stickleback sex chromosomes provide stronger evidence of evolution of post-meiotic expressed gene duplicates on sex chromosomes. In addition to the expression of acquired gene duplicates, we found late meiotic expression of ancestral gametologs enriched with genes coding for troponin proteins. In muscle cells, troponin proteins interact with actin and myosin to bind calcium signals. (Marston & Zamora, 2020). While calcium signaling does play a role in

sperm (Correia et al., 2015), troponin does not have a previously described function in sperm. Most expressed genes on older Y chromosomes thought to have roles in intragenomic conflict have ancestral roles in testis before they translocated to the Y. Recently evolved sex chromosomes like stickleback offer an opportunity to explore the evolution of gene expression of genes that were on the Y ancestrally and may be evolving sex-biased roles.

Genes involved in late spermatogenesis are rapidly diverging.

Evolutionary divergence tends to increase in later stages of spermatogenesis (Abzhanov, 2013; Good & Nachman, 2005; Kopania. et al., 2022; Larson et al., 2016). We found that genes involved in late spermatogenesis had higher rates of protein evolution, consistent with relaxed constraint or positive selection. This finding shows that stickleback spermatogenesis genes show similar patterns of molecular evolution as *Drosophila* (Baines et al., 2008; Meiklejohn et al., 2003), pied flycatcher birds (Segami et al., 2022), and house mice (Kopania. et al., 2022; Larson et al., 2016). Sexual selection is expected to drive phenotypic evolution in sperm morphology in response to sperm competition or cryptic female choice. Sexual selection on fish spermatogenesis may be different than in other species due to external fertilization, and synchronous spermatogenesis (Fitzpatrick, 2020). Interestingly, stickleback fish exhibit paternal care (Bell et al., 2018), where males build nests and guard eggs, thus reducing the capacity for male-male competition despite having external fertilization. Despite these expected differences in selective pressures between fish and mammals, we found similar patterns of molecular evolution. This may highlight the robustness of selection acting on later stages of development in different ecological settings. The diversity of reproductive strategies in fish could offer a unique comparative approach to isolate the effect size of sexual selection.

Figures

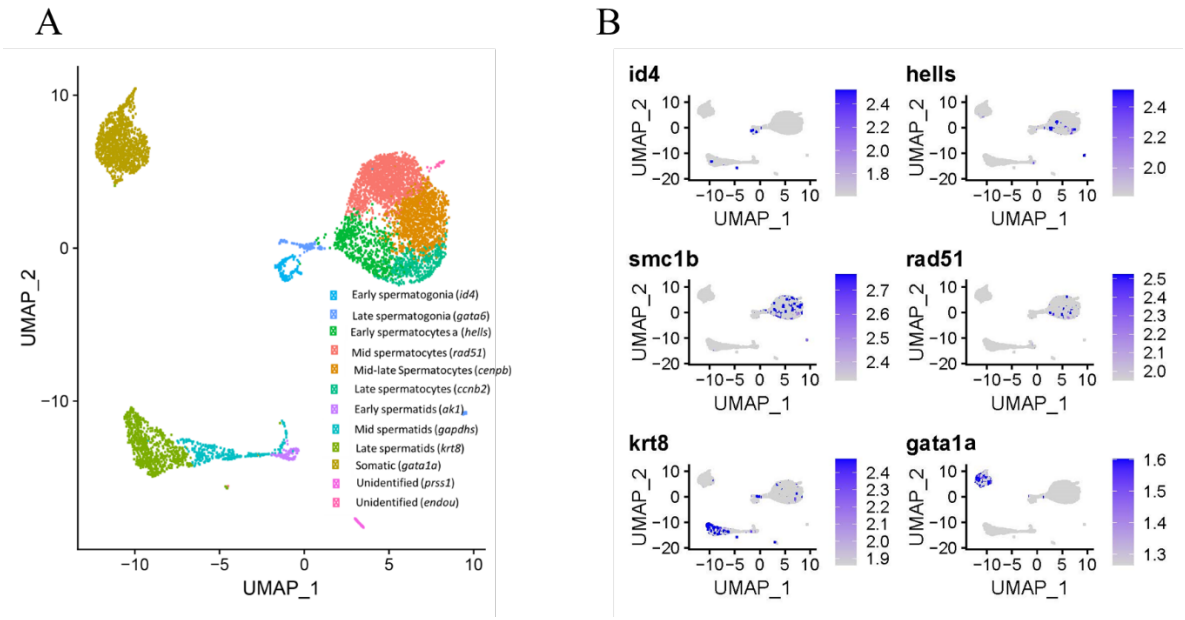


Figure 4.1. Single-cell RNA-seq of threespine stickleback spermatogenesis. Identified clusters of pre-meiotic, meiotic, post-meiotic, and somatic cell populations across two males. A. Visualization of single-cell expression profiles of two individuals in UMAP. Significantly upregulated genes were determined by a Mann-whitney U test. Significantly upregulated genes with consistent expression in other species were used to identify markers of cell populations. B. $\log_2$Fold change Enrichment of key cell type describing genes are shown.

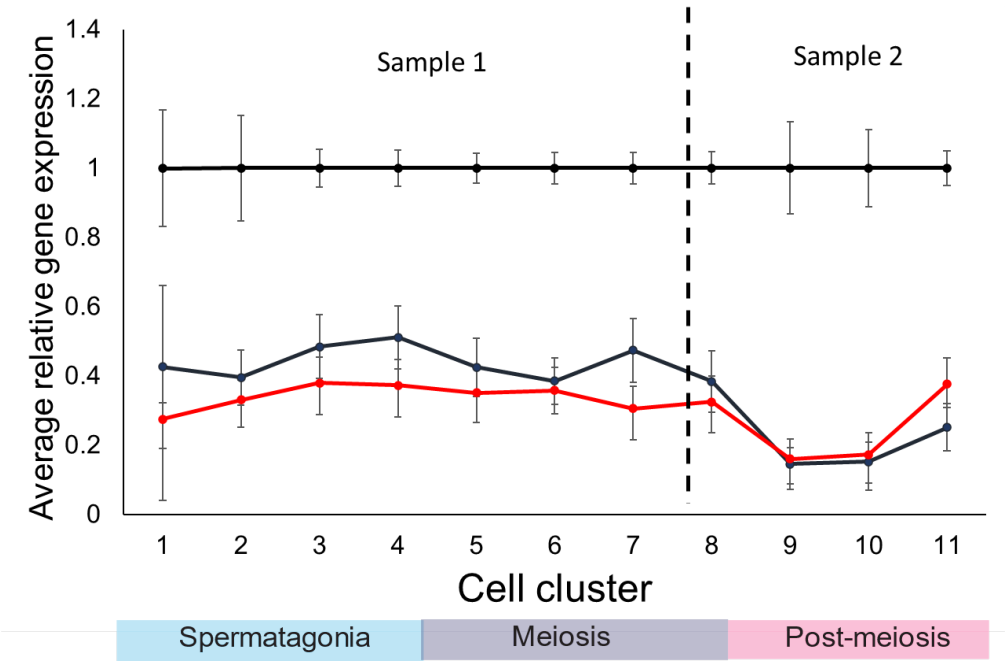


Figure 4.2. Sex-linked gene expression across stickleback spermatogenesis. X and Y chromosome maintain relative expression across spermatogonia and meiotic cell types, but experience post meiotic repression. Transcriptome alignments were extracted for each spermatogenesis related cluster. Counts were normalized for library size of each cluster alignment. Counts for each gene were divided by the average genome wide expression to calculate relative gene expression. Average relative expression for autosomes (black), X-chromosome (blue), Y chromosome (red). Error bars depict standard error of the mean.

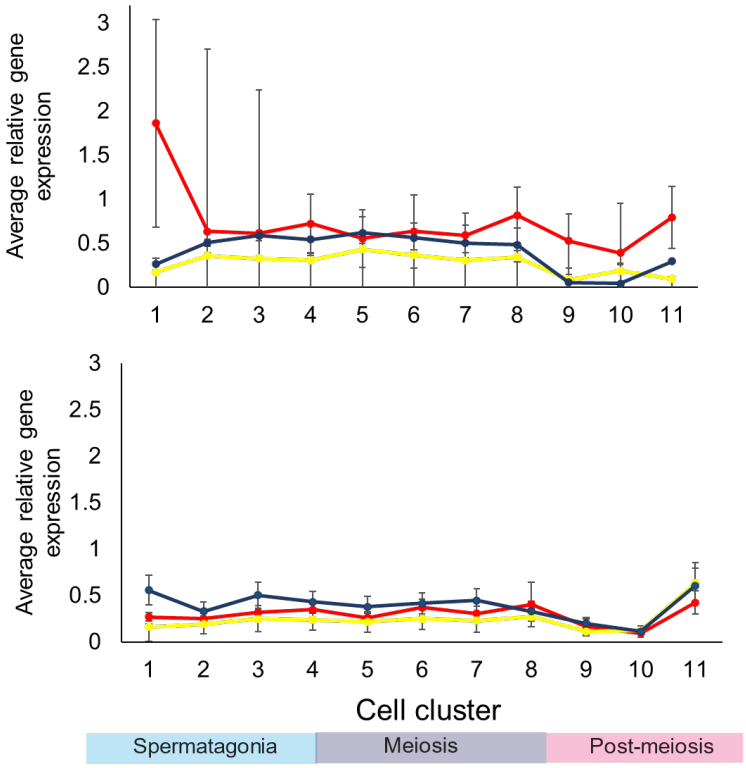


Figure 4.3. Gene expression of sex-linked gametologs across evolutionary strata. A.) The oldest region of the X chromosome shows elevated expression patterns compared to recently evolved in early and late meiosis. B.) Patterns of transcription are largely similar across spermatogenesis genes expressed between the evolutionary strata on the Y chromosome. Transcriptome alignments were extracted for each cluster. Counts were normalized for library size of each cluster alignment. Counts for each transcript were divided by the average genome wide expression to calculate relative gene expression. Genes were divided into evolutionary stratum based on their position on the X and Y chromosomes; Stratum 1 (red), Stratum 2 (yellow), Stratum 3 (blue). Error bars depict standard error of the mean.

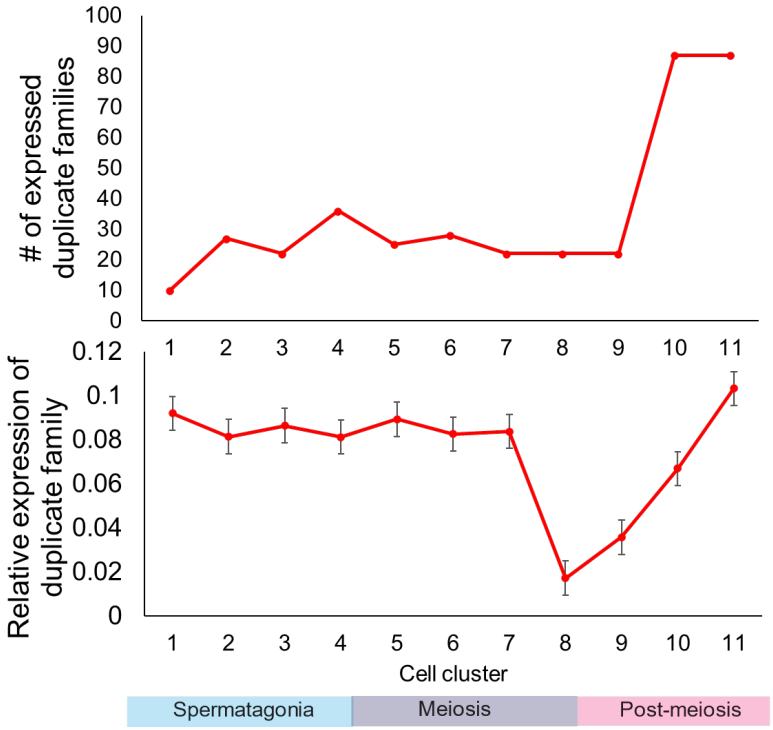


Figure 4.4. Y-specific duplicate family gene expression across spermatogenesis. Y-linked gene duplications are more likely expressed in late meiosis. 87 out of 168 total Y gene duplicates are expressed in the final two clusters of spermatogenesis. Relative rates of transcription drastically decrease at the end of meiosis and regain transcription in late spermatid. Genes with > 4 counts were counted as expressed. Counts for all paralogs were added together as a proxy for duplicate family expression. Expression was normalized by dividing number of counts by the library size of each cluster. Relative expression was calculated by dividing expression by the average genome wide expression of each cluster. Error bars represent standard error of the mean.

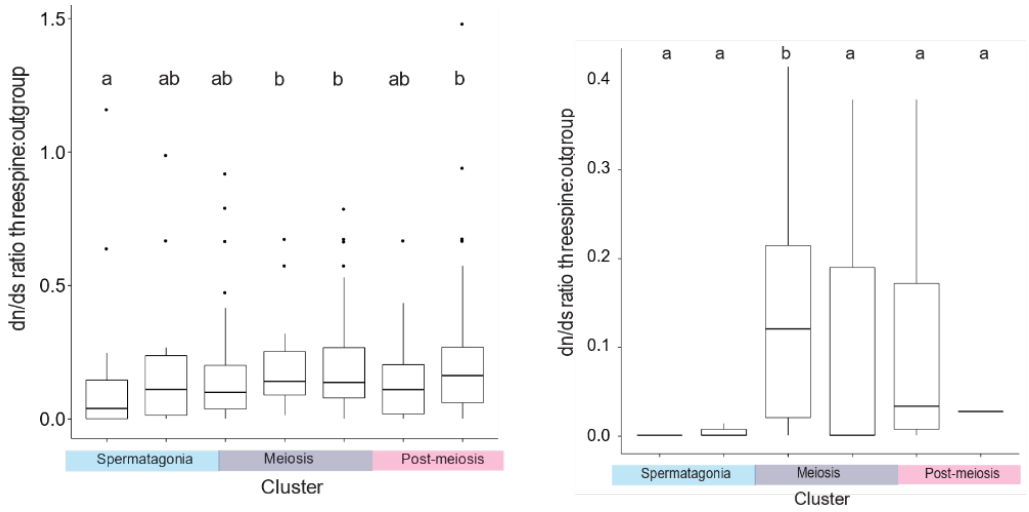


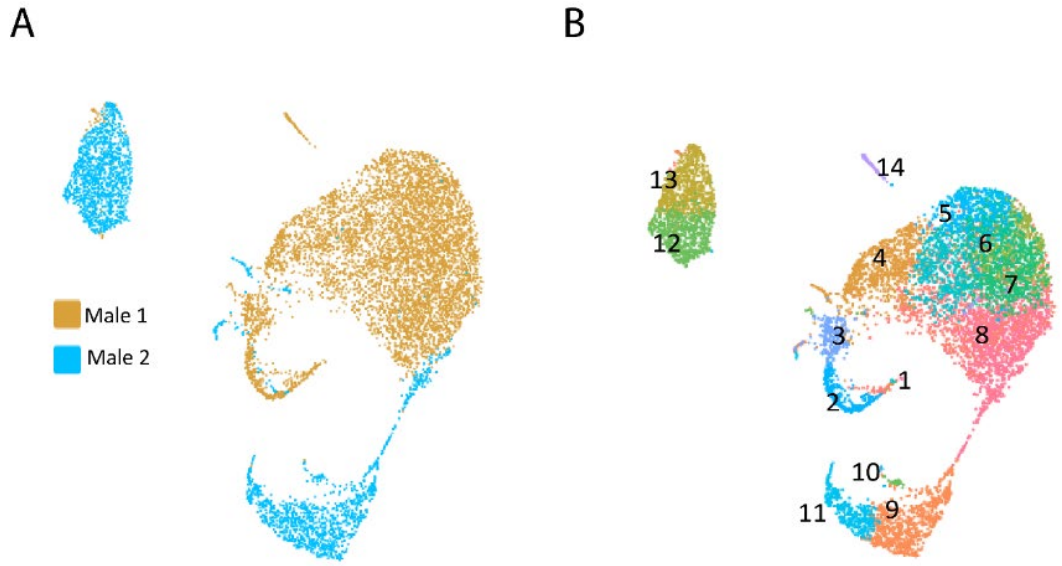
Figure 4.5. Molecular evolution across threespine stickleback spermatogenesis. d_N/d_S ratios were calculated for coding sequence that aligns between threespine stickleback fish and ninespine stickleback fish. In panel A, d_N/d_S of autosomal genes is lowest in early spermatogonia, and shows elevated divergence, in mid-late spermatocytes, and late spermatids ($P < 0.01$, Mann-whitney U test). In panel B, The X chromosome shows elevated d_N/d_S ratios only in early meiosis ($P < 0.01$).

Tables

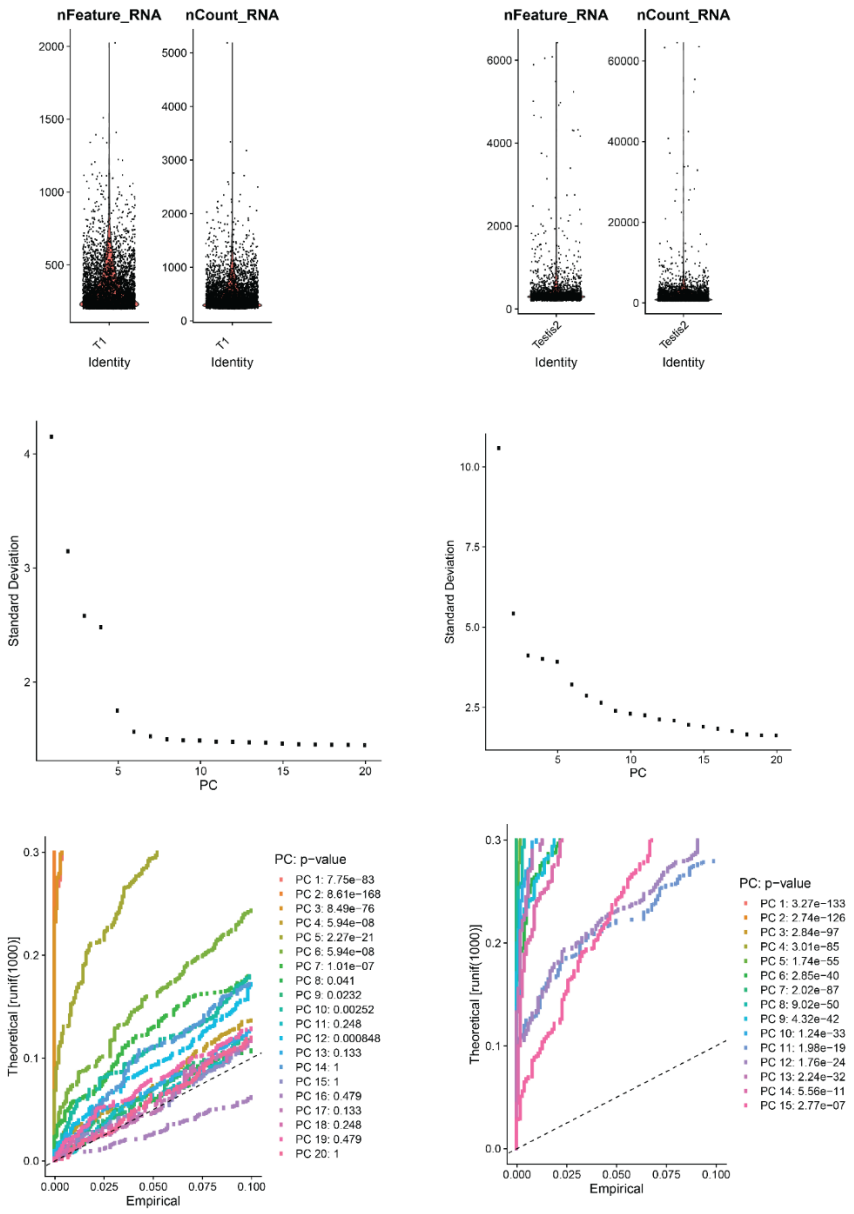
Table 4.1. Identification of clusters reveals genetic markers and gene ontology of distinct stages of spermatogenesis.

Cluster identification	Characterization	Representative spermatogenic-related markers identified in cluster (P < 0.01)	Relevant GO terms (FDR < 0.01)
Early Spermatogonia	Undifferentiation, maintenance	*id4, npsn, ptpn6, myh9a, ctss2.1 hcls1, ponzr1, coro1a, itgb2, scinlb*	cell homeostasis, cell number homeostasis, NADP metabolic process
Late Spermatogonia	Differentiating, cell proliferation, hormones	*gata6, id1, ar, grnb, ptn, cyc1, prdx3, ctsa, h2ax*	Differentiation, chromosome, cell cycle
Early Spermatocytes	Early prophase, DSB formation	*smc2b, smc1al,hells, rab5c,rab35b,cbx3a, NASP,tex11*	core histone, centromere kinetochore, tubulin, ribosomal
Mid spermatocytes	Late prophase, DSB processing and strand invasion	*rad51, rad54, meig1, ccna1, map2k6,akirin2, cdk9, mns1, dyrk2, ccnb1*	embryo development, cell cycle, regulation of cell cycle, recombination
Mid-Late spermatocytes	Late meiosis, spindle assembly	*smc3, rab24, cenpv, cetn2,dnal1,cbx8b,rab8 a,spag1,mospd1,cdc37*	chromosome organization, condensed chromosome, centromere, chromosome segregation, spindle
Late spermatocytes	Late meiosis, segregation and cell division	*cbx1, caprin1a, ccnb2, cfap57, SMARCA5, cdca8,rad23b, smc4,ncapd2,spata18,nu sap1, fzr1a*	Sister chromatid segregation, cell cycle, cilium organization, nuclear division, organelle fission, kinetochore
Early Spermatids	Spermiogenesis	*ak1, taf9, ybx1,map6a, dctn3*	Cilium, motile cilium, dynein complex, centriole
Late spermatid	Sperm maturation	*gapdh, gapdhs, krt18a.1, arpc3, myh9b*	Axoneme assembly, axoneme, ciliary plasm
Somatic cluster	Sertoli, erythrocytes, connective tissue, Leydig	*gata1a, hbae5, hbe1, ctrl, ccng1*	Myeloid cell homeostasis, erythrocyte development, immune system development

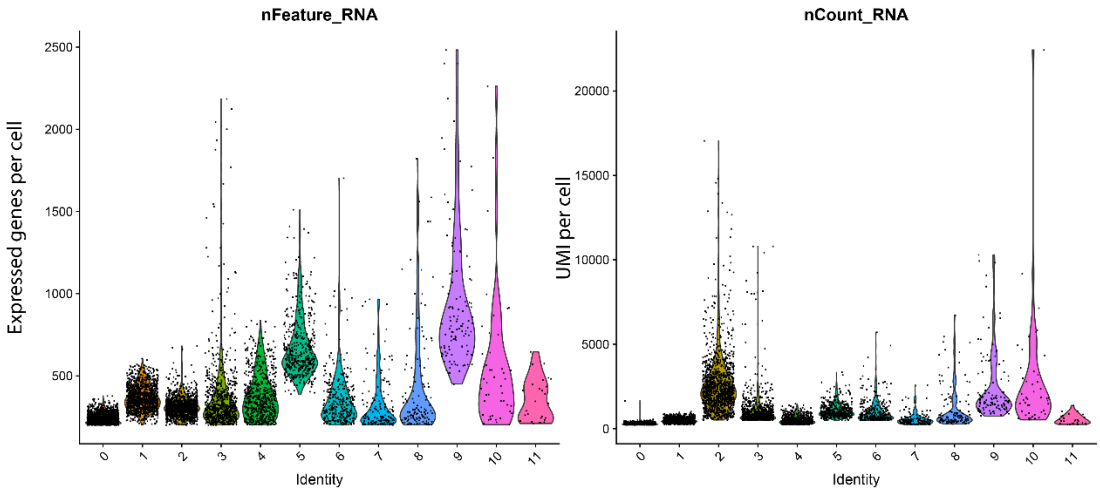
Supplemental figures



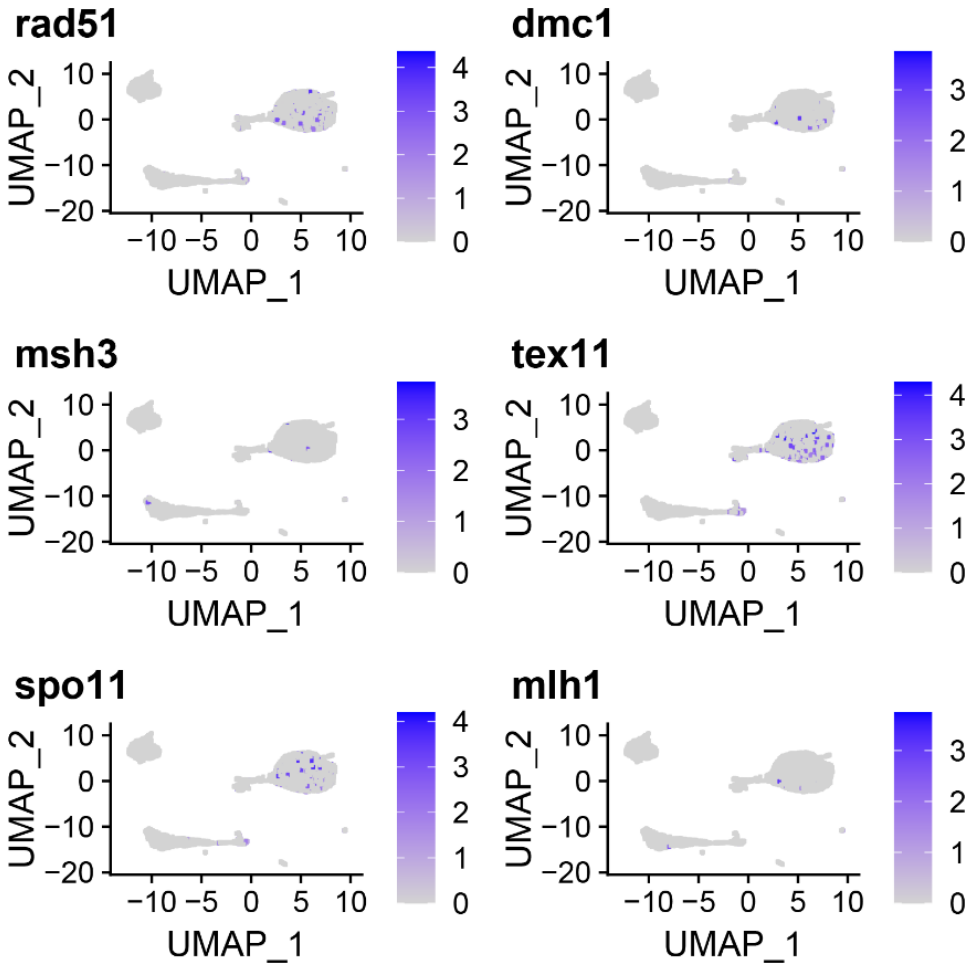
Supplemental Figure 4.1. CellRanger UMAP clustering
UMAP clustering of 2 males (A) clusters into 14 clusters using CellRanger. Both replicates show distinct stages of spermatogenesis with some overlap of cell types. Spatial clustering is similar to results from Seurat.



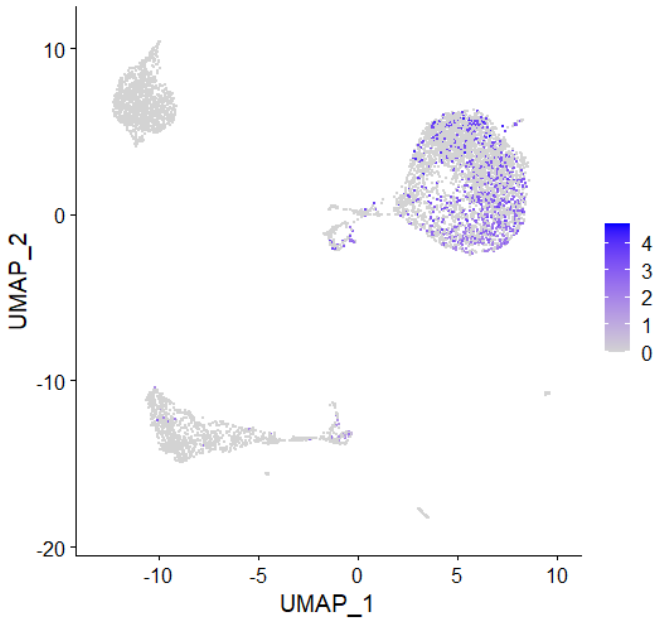
Supplemental figure 4.2. Single-cell gene expression quality control and filtering. Sample 1 (left) and sample 2 (right) showed similar numbers of reads and significant principal components across runs. We found a similar number of mapped features and reads per sample. We used elbow plots and jackstraw permutations to identify significant principal components to be used for downstream clustering.



Supplemental figure 4.3. Feature and molecule counts per cluster.
Reads and expressed genes are expressed evenly across most clusters. Feature counts were determined by counting the number genes expressed within each cell. UMI per cell is determined by counting the number unique molecular identifiers for each cell per cluster.



Supplemental figure 4.4. Expression of core meiotic genes across multiple meiotic clusters. Expression was plotted for several key meiotic genes. Many conserved genes important for meiosis were expressed at low levels across cells within our meiotic clusters, but not all were significantly upregulated within a single sub-cluster of meiosis. Of these six genes, only *rad51* and *tex11* were significantly upregulated in a cluster ($P < 0.001$). *dmc1*, and *spo11*, have low expression across a few clusters, while *msh3*, and *mlh1* were only detected in a few cells.



Supplemental Figure 4.5. Heterochromatin protein *cbx1b* expression in late meiosis. Expression was plotted for *cbx1b* and shows elevated expression in the meiotic cluster. *Cbx1b* is significantly upregulated in late meiosis compared to all other clusters ($P < 0.001$, Mann-Whitney U test).

References

Abzhanov, A. (2013). von Baer's law for the ages: lost and found principles of developmental evolution. *Trends in Genetics*, *29*(12), 712-722. https://doi.org/https://doi.org/10.1016/j.tig.2013.09.004

Álvarez-González, L., Burden, F., Doddamani, D., Malinverni, R., Leach, E., Marín-García, C., Marín-Gual, L., Gubern, A., Vara, C., Paytuví-Gallart, A., Buschbeck, M., Ellis, P. J. I., Farré, M., & Ruiz-Herrera, A. (2022). 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30296-6

Bachtrog, D., Mahajan, S., & Bracewell, R. (2019). Massive gene amplification on a recently formed Drosophila Y chromosome. *Nature Ecology & Evolution*, *3*(11), 1587-1597. https://doi.org/10.1038/s41559-019-1009-9

Baines, J. F., Sawyer, S. A., Hartl, D. L., & Parsch, J. (2008). Effects of X-Linkage and Sex-Biased Gene Expression on the Rate of Adaptive Protein Evolution in Drosophila. *Molecular Biology and Evolution*, *25*(8), 1639-1650. https://doi.org/10.1093/molbev/msn111

Bell, A. M., Trapp, R., & Keagy, J. (2018). Parenting behaviour is highly heritable in male stickleback. *Royal Society Open Science*, *5*(1), 171029. https://doi.org/10.1098/rsos.171029

Bonilla, E., & Xu, E. Y. (2008). Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human. *Molecular Human Reproduction*, *14*(3), 137-142. https://doi.org/10.1093/molehr/gan002

Borg, B. (1981). Effects of methyltestosterone on spermatogenesis and secondary sexual characters in the three-spined stickleback (Gasterosteus aculeatus L.). *General and Comparative Endocrinology*, *44*(2), 177-180. https://doi.org/10.1016/0016-6480(81)90245-8

Borg, B., & Veen, T. V. (1982). Seasonal effects of photoperiod and temperature on the ovary of the three-spined stickleback, Gasterosteus aculeatus L. *Canadian Journal of Zoology*, *60*(12), 3387-3393. https://doi.org/10.1139/z82-428

Bradley, J., Baltus, A., Skaletsky, H., Royce-Tolland, M., Dewar, K., & Page, D. C. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice. *Nature Genetics*, *36*(8), 872-876. https://doi.org/10.1038/ng1390

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genetics*, *8*(9), e1002900. https://doi.org/10.1371/journal.pgen.1002900

Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS Biology*, *7*(11), e1000244. https://doi.org/10.1371/journal.pbio.1000244

Correia, J., Michelangeli, F., & Publicover, S. (2015). Regulation and roles of Ca2+ stores in human sperm. *Reproduction*, *150*(2), R65-R76. https://doi.org/10.1530/rep-15-0102

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M R., Armean, Irina M., Austine-Orimoloye, O., Azov, Andrey G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C. G., Genez, T., Martinez, Jose G., Guijarro-Clarke, C., Gymer, A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Marugán, J. C., Mohanan, S., Mushtaq, A., Naven, M., Ogeh, Denye N., Parker, A., Parton, A., Perry, M., Piližota, I., Prosovetskaia, I., Sakthivel, Manoj P., Salam, Ahamed Imran A., Schmitt, Bianca M., Schuilenburg, H., Sheppard, D., Pérez-Silva, José G., Stark, W., Steed, E., Sutinen, K., Sukumaran, R., Sumathipala, D., Suner, M.-M., Szpak, M., Thormann, A., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, Thomas A., Walts, B., Willhoft, N., Winterbottom, A., Wass, E., Chakiachvili, M., Flint, B., Frankish, A., Giorgetti, S., Haggerty, L., Hunt, Sarah E., IIsley, Garth R., Loveland, Jane E., Martin, Fergal J., Moore, B., Mudge, Jonathan M., Muffato, M., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, Stephen J., Dyer, S., Harrison, Peter W., Howe, Kevin L., Yates, Andrew D., Zerbino, Daniel R., & Flicek, P. (2021). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988-D995. https://doi.org/10.1093/nar/gkab1049

Dapper, A. L., & Payseur, B. A. (2019). Molecular evolution of the meiotic recombination pathway in mammals. *Evolution*, *73*(12), 2368-2389. https://doi.org/10.1111/evo.13850

Emerson, J. J., Kaessmann, H., Betrán, E., & Long, M. (2004). Extensive Gene Traffic on the Mammalian X Chromosome. *Science*, *303*(5657), 537-540. https://doi.org/doi:10.1126/science.1090042

Federici, F., Mulugeta, E., Schoenmakers, S., Wassenaar, E., Hoogerbrugge, J. W., Van Der Heijden, G. W., Van Cappellen, W. A., Slotman, J. A., Van Ijcken, W. F., Laven, J. S., Grootegoed, J. A., & Baarends, W. M. (2015). Incomplete meiotic sex chromosome inactivation in the domestic dog. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1501-9

Feugang, J. M., Rodriguez-Osorio, N., Kaya, A., Wang, H., Page, G., Ostermeier, G. C., Topper, E. K., & Memili, E. (2010). Transcriptome analysis of bull spermatozoa: implications for male fertility. *Reprod Biomed Online*, *21*(3), 312-324. https://doi.org/10.1016/j.rbmo.2010.06.022

Fitzpatrick, J. L. (2020). Sperm competition and fertilization mode in fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1813), 20200074. https://doi.org/10.1098/rstb.2020.0074

Good, J. M., & Nachman, M. W. (2005). Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Molecular Biology and Evolution*, *22*(4), 1044-1052. https://doi.org/10.1093/molbev/msi087

Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S. J., Moore, B. B., Tallquist, M. D., Li, J. Z., & Hammoud, S. S. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Developmental Cell*, *46*(5), 651-667.e610. https://doi.org/10.1016/j.devcel.2018.07.025

Guioli, S., Lovell-Badge, R., & Turner, J. M. A. (2012). Error-Prone ZW Pairing and No Evidence for Meiotic Sex Chromosome Inactivation in the Chicken Germ Line. *PLoS Genetics*, *8*(3), e1002560. https://doi.org/10.1371/journal.pgen.1002560

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath, J. M., Blish, C. A., Gottardo, R., Smibert, P., & Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e3529. https://doi.org/10.1016/j.cell.2021.04.048

Helsel, A. R., Yang, Q.-E., Oatley, M. J., Lord, T., Sablitzky, F., & Oatley, J. M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. *Development*, *144*(4), 624-634. https://doi.org/10.1242/dev.146928

Hermann, B. P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K. N., Chen, I. C., Gildersleeve, H., Lehle, J. D., Mayo, M., Westernströer, B., Law, N. C., Oatley, M. J., Velte, E. K., Niedenberger, B. A., Fritze, D., Silber, S., Geyer, C. B., Oatley, J. M., & McCarrey, J. R. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Reports*, *25*(6), 1650-1667.e1658. https://doi.org/10.1016/j.celrep.2018.10.026

Hirota, T., Blakeley, P., Sangrithi, M. N., Mahadevaiah, S. K., Encheva, V., Snijders, A. P., Elinati, E., Ojarikre, O. A., De Rooij, D. G., Niakan, K. K., & Turner, J. M. A. (2018). SETDB1 Links the Meiotic DNA Damage Response to Sex Chromosome Silencing in Mice. *Developmental Cell*, *47*(5), 645-659.e646. https://doi.org/10.1016/j.devcel.2018.10.004

Hughes, J. F., Skaletsky, H., Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., Bellott, D. W., Cho, T.-J., Dugan-Rocha, S., Khan, Z., Kremitzki, C., Fronick, C., Graves-Lindsay, T. A., Fulton, L., Warren, W. C., Wilson, R. K., Owens, E., Womack, J. E., Murphy, W. J., Muzny, D. M., Worley, K. C., Chowdhary, B. P., Gibbs, R. A., & Page,

D. C. (2020). Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Research*, *30*(12), 1716-1726. https://doi.org/10.1101/gr.269902.120

Kopania, E. E. K., Larson, E. L., Callahan, C., Keeble, S., & Good, J. M. (2022). Molecular Evolution across Mouse Spermatogenesis. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msac023

Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., De Rooij, D. G., Hu, Y.-C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, *29*(21), 3699-3706.e3695. https://doi.org/10.1016/j.cub.2019.08.057

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y., & Nakano, T. (2004). Mili, a mammalian member ofpiwifamily gene, is essential for spermatogenesis. *Development*, *131*(4), 839-849. https://doi.org/10.1242/dev.00973

Larson, E. L., Vanderpool, D., Keeble, S., Zhou, M., Sarver, B. A., Smith, A. D., Dean, M. D., & Good, J. M. (2016). Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. *Genetics*, *203*(4), 1841-1857. https://doi.org/10.1534/genetics.116.186825

Lu, L.-Y., & Yu, X. (2015). Double-strand break repair on sex chromosomes: challenges during male meiotic prophase. *Cell Cycle*, *14*(4), 516-525. https://doi.org/10.1080/15384101.2014.998070

Lukassen, S., Bosch, E., Ekici, A. B., & Winterpacht, A. (2018). Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-24725-0

Lukassen, S., Bosch, E., Ekici, A. B., & Winterpacht, A. (2018). Single-cell RNA sequencing of adult mouse testes. *Scientific Data*, *5*(1), 180192. https://doi.org/10.1038/sdata.2018.192

Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of Y chromosome gene content in flies. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00653-x

Maine, E. M., Hauth, J., Ratliff, T., Vought, V. E., She, X., & Kelly, W. G. (2005). EGO-1, a Putative RNA-Dependent RNA Polymerase, Is Required for Heterochromatin Assembly on Unpaired DNA during C. elegans Meiosis. *Current Biology*, *15*(21), 1972-1978. https://doi.org/10.1016/j.cub.2005.09.049

Malik, S.-B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M., & Logsdon, J. M. (2008). An Expanded Inventory of Conserved Meiotic Genes Provides Evidence for Sex in Trichomonas vaginalis. *PLOS ONE*, *3*(8), e2879. https://doi.org/10.1371/journal.pone.0002879

Marston, S., & Zamora, J. E. (2020). Troponin structure and function: a view of recent progress. *Journal of Muscle Research and Cell Motility*, *41*(1), 71-89. https://doi.org/10.1007/s10974-019-09513-1

Martínez-Pacheco, M., Tenorio, M., Almonte, L., Fajardo, V., Godínez, A., Fernández, D., Cornejo-Páramo, P., Díaz-Barba, K., Halbert, J., Liechti, A., Székely, T., Urrutia, A. O., & Cortez, D. (2020). Expression Evolution of Ancestral XY Gametologs across All Major Groups of Placental Mammals. *Genome Biology and Evolution*, *12*(11), 2015-2028. https://doi.org/10.1093/gbe/evaa173

Meiklejohn, C. D., Parsch, J., Ranz, J. M., & Hartl, D. L. (2003). Rapid evolution of male-biased gene expression in Drosophila. *Proceedings of the National Academy of Sciences*, *100*(17), 9894-9899.

Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Faster-X Evolution of Gene Expression in Drosophila. *PLoS Genetics*, *8*(10), e1003013. https://doi.org/10.1371/journal.pgen.1003013

Mikhaylova, L. M., & Nurminsky, D. I. (2011). Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the Drosophila X chromosome. *BMC Biology*, *9*(1), 29. https://doi.org/10.1186/1741-7007-9-29

Moretti, C., Vaiman, D., Tores, F., & Cocquet, J. (2016). Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics & Chromatin*, *9*(1). https://doi.org/10.1186/s13072-016-0099-8

Mueller, J. L., Mahadevaiah, S. K., Park, P. J., Warburton, P. E., Page, D. C., & Turner, J. M. A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics*, *40*(6), 794-799. https://doi.org/10.1038/ng.126

Namekawa, S. H., Park, P. J., Zhang, L.-F., Shima, J. E., McCarrey, J. R., Griswold, M. D., & Lee, J. T. (2006). Postmeiotic Sex Chromatin in the Male Germline of Mice. *Current Biology*, *16*(7), 660-667. https://doi.org/10.1016/j.cub.2006.01.066

Nath, S., Shaw, D. E., & White, M. A. (2021). Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkab007

Nath, S., Welch, L. A., Flanagan, M. K., & White, M. A. (2022). Meiotic pairing and double-strand break formation along the heteromorphic threespine stickleback sex chromosomes. *Chromosome Research*. https://doi.org/10.1007/s10577-022-09699-0

Oatley, M., Kaucher, A., Racicot, K., & Oatley, J. (2011). Inhibitor of DNA Binding 4 Is Expressed Selectively by Single Spermatogonia in the Male Germline and Regulates the Self-Renewal of Spermatogonial Stem Cells in Mice. *Biology of Reproduction*, *85*, 347-356. https://doi.org/10.1095/biolreprod.111.091330

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02097-x

Qian, P., Kang, J., Liu, D., & Xie, G. (2022). Single Cell Transcriptome Sequencing of Zebrafish Testis Revealed Novel Spermatogenesis Marker Genes and Stronger Leydig-Germ Cell Paracrine Interactions [Original Research]. *Frontiers in Genetics*, *13*. https://www.frontiersin.org/article/10.3389/fgene.2022.851719

Ramesh, M. A., Malik, S.-B., & Logsdon, J. M. (2005). A Phylogenomic Inventory of Meiotic Genes. *Current Biology*, *15*(2), 185-191. https://doi.org/10.1016/j.cub.2005.01.003

Schimenti, J. (2005). Synapsis or silence. *Nature Genetics*, *37*(1), 11-13. https://doi.org/10.1038/ng0105-11

Segami, J. C., Semon, M., Cunha, C., Bergin, C., Mugal, C. F., & Qvarnström, A. (2022). *Single-Cell Transcriptomics reveals relaxed evolutionary constraint of spermatogenesis in two passerine birds as compared to mammals*. Cold Spring Harbor Laboratory. https://dx.doi.org/10.1101/2022.01.22.477241

Shami, A. N., Zheng, X., Munyoki, S. K., Ma, Q., Manske, G. L., Green, C. D., Sukhwani, M., Orwig, K. E., Li, J. Z., & Hammoud, S. S. (2020). Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell*, *54*(4), 529-547.e512. https://doi.org/10.1016/j.devcel.2020.05.010

Soh, Y. Q., Alföldi, J., Pyntikova, T., Laura, Graves, T., Patrick, Robert, Kremitzki, C., Koutseva, N., Jacob, Rozen, S., Jennifer, Owens, E., James, William, Cao, Q., Pieter, Wesley, Richard, Skaletsky, H., & David. (2014). Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*, *159*(4), 800-813. https://doi.org/10.1016/j.cell.2014.09.052

Sokołowska, E., & Kulczykowska, E. (2006). Annual reproductive cycle in two free living populations of three-spined stickleback (Gasterosteus aculeatus L.): Patterns of ovarian and testicular development. *Oceanologia*, *48*.

Spruce, C., Dlamini, S., Ananda, G., Bronkema, N., Tian, H., Paigen, K., Carter, G. W., & Baker, C. L. (2020). HELLS and PRDM9 form a pioneer complex to open chromatin at meiotic recombination hot spots. *Genes & Development*, *34*(5-6), 398-412. https://doi.org/10.1101/gad.333542.119

Sun, Y., Li, H., Liu, Y., Mattson, M. P., Rao, M. S., & Zhan, M. (2008). Evolutionarily Conserved Transcriptional Co-Expression Guiding Embryonic Stem Cell Differentiation. *PLOS ONE*, *3*(10), e3406. https://doi.org/10.1371/journal.pone.0003406

Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. *Molecular Biology and Evolution*, *20*(1), 18-20. https://doi.org/10.1093/oxfordjournals.molbev.a004233

Turner, J. M., Mahadevaiah, S. K., Fernandez-Capetillo, O., Nussenzweig, A., Xu, X., Deng, C. X., & Burgoyne, P. S. (2005). Silencing of unsynapsed meiotic chromosomes in the mouse. *Nature Genetics*, *37*(1), 41-47. https://doi.org/10.1038/ng1484

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, *134*(10), 1823-1831. https://doi.org/10.1242/dev.000018

Turner, J. M. A. (2015). Meiotic Silencing in Mammals. *Annual Review of Genetics*, *49*(1), 395-412. https://doi.org/10.1146/annurev-genet-112414-055145

Turner, J. M. A., Mahadevaiah, S. K., Ellis, P. J. I., Mitchell, M. J., & Burgoyne, P. S. (2006). Pachytene Asynapsis Drives Meiotic Sex Chromosome Inactivation and Leads to Substantial Postmeiotic Repression in Spermatids. *Developmental Cell*, *10*(4), 521-529. https://doi.org/10.1016/j.devcel.2006.02.009

Tvedte, E. S., Forbes, A. A., & Logsdon, J. M., Jr. (2017). Retention of Core Meiotic Genes Across Diverse Hymenoptera. *Journal of Heredity*, *108*(7), 791-806. https://doi.org/10.1093/jhered/esx062

Uribe, M. C., Grier, H. J., & Mejía-Roa, V. (2014). Comparative testicular structure and spermatogenesis in bony fishes. *Spermatogenesis*, *4*(3), e983400. https://doi.org/10.4161/21565562.2014.983400

Varadharajan, S., Rastas, P., Löytynoja, A., Matschiner, M., Calboli, F. C. F., Guo, B., Nederbragt, A. J., Jakobsen, K. S., & Merilä, J. (2019). A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evz240

Vibranovski, M. D. (2014). Meiotic sex chromosome inactivation in Drosophila. *J Genomics*, *2*, 104-117. https://doi.org/10.7150/jgen.8178

Viera, A., Parra, M. T., Arévalo, S., García De La Vega, C., Santos, J. L., & Page, J. (2021). X Chromosome Inactivation during Grasshopper Spermatogenesis. *Genes*, *12*(12), 1844. https://doi.org/10.3390/genes12121844

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome.

Molecular Biology and Evolution, *32*(8), 1981-1995. https://doi.org/10.1093/molbev/msv078

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. *eLife*, *8*. https://doi.org/10.7554/elife.47138

Wu, X., Yang, Y., Zhong, C., Wang, T., Deng, Y., Huang, H., Lin, H., Meng, Z., & Liu, X. (2021). Single-Cell Atlas of Adult Testis in Protogynous Hermaphroditic Orange-Spotted Grouper, Epinephelus coioides. *International Journal of Molecular Sciences*, *22*(22), 12607. https://doi.org/10.3390/ijms222212607

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Identifying adaptive loss of expression

The recent evolution of X and Y chromosomes in stickleback fish has provided an excellent model for testing hypotheses of regulatory evolution. I found that ancestral accessible chromatin regions (ACRs) are rapidly evolving on the Y chromosome compared to autosomal ACRs, and to neutrally evolving sequence on the Y chromosome. I found that this divergence correlates with expression changes and in some tissues may be largely associated with downregulation. This represents evidence to support recent theory (Lenormand et al., 2020; Lenormand & Roze, 2022), that rapid regulatory evolution is an integral component of sex chromosome evolution.

When dealing with the unique evolutionary constraints of Y chromosomes, it is important to consider alternative drivers of elevated divergence rates. Elevated divergence rate in Y-linked ACRs may be driven by non-adaptive alternatives. Elevated mutation rates could cause the observed trends if this only occurred in ACRs. Male-biased mutation rates have been observed in some species due to the increased rate of cell divisions in spermatogenesis compared to oogenesis (Ellegren, 2007; Hurst & Ellegren, 1998). Elevated ACR divergence seems unlikely to be driven by male-biased mutation rates, as these patterns would also be observed throughout the remainder of the Y chromosome. Synonymous substitution rates are the same on both the X and Y chromosomes, indicating male-biased mutation is not occurring in threespine stickleback fish (White et al., 2015).

It is important to consider the role of sequence composition-based mutations rates. Promoters of many species are GC- rich (Fenouil et al., 2012; Kostyuchenko et al., 2018). GC-rich regions tend to have higher mutation rates (Ellegren et al., 2003), due to cytosine deamination (Fryxell & Zuckerkandl, 2000), or GC biased conversion (Pessia et al., 2012). However, we were able to normalize for GC content by comparing to divergence within intergenic regions with similar sequence composition, which should be subject to same mutation biases, absent of selection. Instead, we still found nucleotide divergence to be lower within intergenic regions compare to ACRs allowing us to infer as signatures of selection.

DNA double strand breaks (DSBs) may occur more often in ACRs and error prone repair of these breaks could lead to elevated mutation rates. Genome wide, DSBs are expected to accumulate at promoter sequence (Shanfelter et al., 2019) and open regions of chromatin (Marand et al., 2017; Spruce et al., 2020). Homology directed repair, like crossovers, are efficient at repairing DSBs. But alternative repair pathways include repair through non-homologous end joining (NHEJ) and micro-homology mediated end joining (MHMEJ), that typically result in errors during the repair process (Chang et al., 2017; McVey & Lee, 2008). These types of repair pathways often result in deletions (McVey & Lee, 2008), that leave distinct molecular signatures that can be identified through sequencing approaches (Chang et al., 2022). Despite suppression of crossovers (Peichel et al., 2020), DSBs still occur on the stickleback X and Y chromosome at a similar rate compared to autosomes (Nath et al., 2022). Without homologous sequence to repair double strand breaks, DSBs on the sex chromosomes in males may be repaired more often with non-homology based DSB repair, resulting in an increased mutation rate within ACRs specifically. Consequently, non-homology based DSB repair mechanisms are more likely to result in small indels at sites of repair (McVey & Lee, 2008).

However, we observed elevated rates of point substitutions at ACRs, rather than deletions. However, if DSB repair errors occur at a high rate on sex-linked ACRs, these haplotypes may be deleterious. Given the choice between alleles with large deletions of ACRs, compared to alleles with only point mutations, the latter may be favorable, thus increasing the amount of sequence divergence specifically at ACRs. In other words, mutations that only lower transcription slightly by altering a few base-pairs may be preferred to large deletions that may erase transcription factor binding sites on the Y chromosome.

One way to test this hypothesis is to map DSB accumulation across a generation of meiocytes by profiling DNA binding protein associated with DSB like *rad51* or *dmc1*. Enrichment of these DSB repair DNA binding proteins would allow us to identify potential hotspots of DNA damage. We predict that DSB accumulation would be enriched around promoter sequence and regions otherwise accessible in meiosis. These sites may be subject to repair through efficient repair programs such as gene conversion, or through error-prone processes like NHEJ or MHMEJ. The latter should be associated with newly formed deletions, that would not be observed in the parent. We could identify *de novo* deletions, by sequencing pools of sperm with long read DNA sequencing. I predict that these small deletions occur in DSB hotspots such as promoters and ACRs. However, many deletions may be deleterious in these regions and selected against, either at the haploid stage directly on the sperm, or on the population level. If we observed a high number of sperm haplotypes with *de novo* deletions, despite not observing these in the population data or lab crossed offspring, we could infer that negative selection purges these errors from populations. I predict that deletions within ACRs would be under strong negative selection and exist at low allele frequencies, compared to the number that are detectable within a pool of sperm.

Functional testing of *cis*-regulatory variation on sex chromosomes.

In chapter 2, I proposed a model for weakened purifying selection driving observed patterns of accelerated substitutions on the Y chromosome. In this model, gene expression is first reduced for Y gametologs due to weakened purifying selection. Next, genes with dosage sensitive functions, are more likely to accumulate substitutions that restore expression. Distinguishing adaptive loss of expression from the compensatory model could be done by identifying if the Y-linked substitutions are more likely associated with gain or loss of expression. In the compensatory model, mutations would accumulate neutrally first, leading to downregulation, followed by elevated rates of divergence associated with restoring gene expression. Contrasted to the adaptive loss of expression model where most mutations should lead to downregulation. Identifying the direction and effect size of observed ACR mutations could help resolve this model. *In vitro* enhancer reporter assays (Román et al., 2020), could help determine the regulatory potential of previously found ACRs, to determine if mutations lead to loss or gain of function in regulatory elements. Additionally, *in vivo* approaches could be done in stickleback fish by using CRISPR/Cas9. The integration of GFP reporters has previously been done in stickleback fish to identify tissue specific expression of *cis*-regulatory elements (Wucherpfennig et al., 2022). Additionally, CRISPR/cas9 mediated mutagenesis of X-linked ACRs, could be performed to reproduce the mutations observed in the corresponding Y-linked ACRs. Expression differences could be quantified using RNA-seq across mutagenized X gametologs compared to wildtype. For example, if we found that most observed Y-linked mutations can lead to reduction of expression, this would suggest that rapid evolution of these regions on the Y chromosome are driven by adaptive loss of expression. However, if most

mutations lead to gain of expression, this will show that provide evidence that most Y-linked mutations are evolving to compensate for previously accrued deleterious mutations.

Genetic basis of chromatin accessibility on Y chromosomes

While I focused on sequence changes underlying accessible regions on the sex chromosomes, changes in chromatin accessibility itself can also lead to expression changes. There could be differences in accessibility between homologous regions on the X and Y chromosome. Y chromosomes tend to accumulate TEs (Chang et al., 2019; Chang et al., 2022; Chang & Larracuente, 2019; Mahajan et al., 2018; Peichel et al., 2020), DNA methylation (Rodríguez Lorenzo et al., 2018; W. Zhang et al., 2008), and histone methylation (Nguyen & Bachtrog, 2021; Zhou et al., 2013), that can all lead to broad scale changes in chromatin structure (Allshire & Madhani, 2018). This broad scale change in chromatin structure could cause the loss of fine scale chromatin accessibility, especially if regions are rapidly evolving and shutting down.

Additionally, profiling chromatin accessibility in outgroup species, may enable us better identify shifts in accessibility on the Y chromosome. In the case of changes in accessibility between the X and Y chromosome, it is possible that the X chromosome is gaining new accessible chromatin regions, or the Y chromosome is losing ACRs. Additionally, we focused on conserved non-coding sequence between the X and Y. Alternative methods could focus on new ACRs that are not present in both, and lead to divergent expression patterns. Profiling ACRs in an outgroup species, like ninespine stickleback fish, could help identify the evolution of chromatin accessibility. Additionally, this approach could identify if ACRs that accumulate sequence changes are more likely to lose accessibility and help explore the genetic basis of chromatin accessibility on the Y chromosome.

Profiling Y chromatin accessibility remains a challenge. Despite having high quality genome assemblies, alignment of short read data may not be sufficient to accurately call ACRs on the Y chromosome. Divergent regions of Y chromosome accumulate repetitive sequence that can complicate assemblies and short read mapping. Additionally, highly homologous regions of the X and Y, such as those recently evolved, can complicate the alignment of X and Y reads. As technology improves, techniques that combine chromatin profiling with long read DNA sequencing, would be especially useful for studying functional genomics of Y chromosomes.

Functional genomics on Y chromosomes

To study the regulatory evolution on the Y chromosome, I only focused on ACRs. Additional consideration for Y regulatory evolution should be focused on the accumulation of changes to DNA and histone methylation. These marks are known to affect the expression of nearby genes, often leading to repressive chromatin states (Allshire & Madhani, 2018; Elgin & Reuter, 2013; Lee, 2015; Lee & Karpen, 2017). Repressive chromatin marks like 5-mc DNA methylation and H3K9me2/3 histone modifications have been found enriched on Y chromosomes (Brown et al., 2020; Nguyen & Bachtrog, 2021; Rodríguez Lorenzo et al., 2018; W. Zhang et al., 2008; Zhou et al., 2013). It would be interesting to identify the accumulation of these marks on the threespine stickleback Y chromosome and compare to changes in sequence evolution and expression. Approaches that map repressive marks along the Y chromosome could help explain the loss of expression if accumulation is proximal to downregulating genes.

Additionally, exploring heterochromatin and DNA methylation evolution between the X and Y chromosome could help untangle how repressive chromatin marks spread. I predict that more sequence differentiation and accumulation of repetitive elements will have gained changes in repressive histone markers associated with heterochromatin (H3K9me2/3). Profiling these

marks across the evolutionary strata of the Y chromosome may untangle the timing of the accumulation of different marks. Heterochromatin machinery may diffuse along Y chromosomes to silencing transposable elements (Allshire & Madhani, 2018; Lee, 2015), that spread rapidly after the evolution of Y chromosomes (Mahajan et al., 2018; Peichel et al., 2020). Additionally, chromatin landscapes may evolve to help facilitate adaptive downregulation of Y-linked genes. Genes located near newly formed heterochromatin may be subject to downregulation (Elgin & Reuter, 2013). But important genes, like post-meiotic expressed genes, may evolve ways to avoid silencing (Moretti et al., 2016). Stickleback sex chromosomes are an excellent model to elucidate the complex relationship between sequence evolution and epigenomic features.

Alternative approaches to studying the evolution of spermatogenesis

Single-cell RNA-seq is a novel and quickly evolving field. Best practices would involve validating cell type expression markers with other methods such as in situ hybridization, qPCR, or histology. In the case of stickleback spermatogenesis, these approaches are limited in scope. Timing of spermatogenesis is highly variable, because stickleback spermatogenesis proceeds synchronously (Borg, 1981), and the cascade of spermatogenesis is initiated by environmental cues. Fish hatched from the same clutch may have testes at different stages of spermatogenesis. Even individual testes within an individual can have variable amounts of germ cells. All of these factors together make replicates and validation difficult. Spatial transcriptomics (Ståhl et al., 2016) could help resolve the issue by overlaying cell morphology of somatic and germ cells, histological markers, and gene expression of the same sample.

Our approaches to study gene expression were limited to identifying differences between the X and Y, without an understanding of ancestral level of expression. By performing single-cell gene expression across ninespine stickleback fish, we could identify which genes on the

threespine stickleback sex chromosomes were important ancestrally. We could also use this ancestral comparison to investigate if the X-biased gene expression we observed throughout stratum one is due to increased X-linked gametolog expression to compensate for loss of expression from the Y chromosome.

We were able to identify the convergent evolution of post-meiotic sex chromosome repression in stickleback fish. In mammals, the mechanism of repression is linked to the accumulation of heterochromatin in post-meiosis (Moretti et al., 2016). Consistent with this model, we found up-regulation of heterochromatin proteins in cell clusters of late meiosis. However, it remains unclear what would target these proteins specifically to the sex chromosome. An interesting experiment to test this idea is to isolate meiotic and post-meiotic cells and profile chromatin marks across these stages using CUT&RUN (Akdogan-Ozdilek et al., 2022). We could identify if repressive histone modifications are responsible for the loss of expression on the sex chromosomes.

In teleost fish, a whole genome duplication has led to an increase in copy number of many genes (Glasauer & Neuhauss, 2014), including multiple paralogs of genes involved in DNA methylation and chromatin modification. In zebrafish, these paralogs have been knocked down to reveal the targets of repressive chromatin modifiers (Calvird et al., 2022). It would be possible to develop a genetic screen approach to test what modifiers of chromatin are responsible for silencing the sex chromosomes. Additional inquiry is essential to identify how the sex chromosomes are specifically targeted. Molecular targets may be able to quickly evolve along the sex chromosomes linked to the accumulation of transposable elements and rapid *cis*-regulatory evolution. One interesting candidate gene is *setdb1b*. The mammalian ortholog has

been identified as crucial for lowering gene expression in meiosis specifically on the sex chromosomes (Hirota et al., 2018).

Identifying roles of gene duplications in stickleback fish

Despite post-meiotic repression, we identified an enrichment of gene duplicates expressed in the final stages of spermatogenesis. These duplicates, along with some gametologs, seem to escape the post-meiotic repression and regain expression. Genes expressed post-meiotically are essential for developing mature sperm. These genes may be under selection to increase in copy number. Increases in copy number could be selected for if the increased dosage of these genes alter sperm morphology in a beneficial way.

Gene family evolution across stickleback Y chromosomes could help elucidate the adaptiveness of gene duplication on the Y chromosome. If the duplicates in threespine stickleback fish have large effects on fitness, the same gene families could have translocated and amplified on other Y chromosomes in stickleback fish. We could test this hypothesis by annotating duplicates along the other sex-limited chromosomes of stickleback fish. Convergent evolution of the same gene duplications could be strong evidence for an adaptive role. Alternatively, duplication may occur on Y chromosomes frequently to different types of genes. In such case, increases in copy number may happen independent of gene function, and primarily serve to avoid deleterious loss of expression, or facilitate meiotic gene conversion.

In house mice, the co-amplification of *sly* and *slx* is proposed to be driven by intragenomic conflict (Cocquet et al., 2012; Cocquet et al., 2009). One potential mechanism of conflict is sex-ratio drive, where increases in copy number between the X and Y chromosomes result in skewed sex ratios (Cocquet et al., 2012; Cocquet et al., 2009; Kruger et al., 2019). This has been demonstrated in house mice, where alterations in copy number of sex-linked amplicons

led to skewed sex ratios (Kruger et al., 2019). However, in mice, it is challenging to test small differences in sex ratio, without large sample sizes of offspring. Stickleback fish offer an exciting model to test similar hypotheses. We could test the function of copy number of Y duplicates in two different approaches. 1st) We could delete copies of gene duplicates on the Y chromosome using CRISPR/cas9, and 2nd) crossing strategies can be utilized to create offspring with mis-matched sex-chromosome pairs (E. E. K. Kopania et al., 2022). We could identify populations with copy number variation of Y chromosome duplicates and create hybrids with population mis-matched XY males. In both mutant and hybrid fish, we could screen for phenotypes associated with copy number differences of duplications. Phenotypes that we may expect would include changes in sperm head morphology, or sperm tail length. As well as significant departures from 50/50 sex ratios in offspring. Identifying the function of these duplicates would help explain why these genes are enriched in post-meiotic round spermatid and if they are under selection. Mutants that have knocked-down copies of Y-linked duplicates and phenotypes could help identify the function of the genes in stickleback. Constantly, hybrids with mismatched copy number variation could help us identify the role of copy number, and identify if interactions between the X and Y are important for developing sperm.

Conclusions

The high-quality genome assembly in stickleback fish has been a powerful approach to study the evolution of recently evolved sex chromosomes. Our analysis of the Y chromosome has revealed that gene expression evolves rapidly through *cis*-regulatory evolution and may be driven by spermatogenesis evolution. Analysis of sex chromosomes in other species may reveal how universal the role of gene expression evolution is in the evolution of sex chromosomes.

References

Akdogan-Ozdilek, B., Duval, K. L., Meng, F. W., Murphy, P. J., & Goll, M. G. (2022). Identification of chromatin states during zebrafish gastrulation using <scp>CUT</scp> & <scp>RUN</scp> and <scp>CUT</scp> &Tag. *Developmental Dynamics*, *251*(4), 729-742. https://doi.org/10.1002/dvdy.430

Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*, *19*(4), 229-244. https://doi.org/10.1038/nrm.2017.119

Borg, B. (1981). Effects of methyltestosterone on spermatogenesis and secondary sexual characters in the three-spined stickleback (Gasterosteus aculeatus L.). *General and Comparative Endocrinology*, *44*(2), 177-180. https://doi.org/10.1016/0016-6480(81)90245-8

Brown, E. J., Nguyen, A. H., & Bachtrog, D. (2020). The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide. *Molecular Biology and Evolution*, *37*(10), 2808-2824. https://doi.org/10.1093/molbev/msaa082

Calvird, A. E., Broniec, M. N., Duval, K. L., Higgs, A. N., Arora, V., Ha, L. N., Schouten, E. B., Crippen, A. R., McGrail, M., Laue, K., & Goll, M. G. (2022). Uncovering Regulators of Heterochromatin Mediated Silencing Using a Zebrafish Transgenic Reporter. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.832461

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B. J., Wu, C.-T., Larracuente, A. M., & Mellone, B. G. (2019). Islands of retroelements are major components of Drosophila centromeres. *PLoS Biology*, *17*(5), e3000241. https://doi.org/10.1371/journal.pbio.3000241

Chang, C.-H., Gregory, L. E., Gordon, K. E., Meiklejohn, C. D., & Larracuente, A. M. (2022). Unique structure and positive selection promote the rapid divergence of Drosophila Y chromosomes. *eLife*, *11*. https://doi.org/10.7554/elife.75795

Chang, C.-H., & Larracuente, A. M. (2019). Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. *Genetics*, *211*(1), 333-348. https://doi.org/10.1534/genetics.118.301765

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*, *18*(8), 495-506. https://doi.org/10.1038/nrm.2017.48

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genetics*, *8*(9), e1002900. https://doi.org/10.1371/journal.pgen.1002900

Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The Multicopy Gene Sly Represses the Sex Chromosomes in the Male Mouse Germline after Meiosis. *PLoS Biology*, *7*(11), e1000244. https://doi.org/10.1371/journal.pbio.1000244

Elgin, S. C. R., & Reuter, G. (2013). Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology*, *5*(8), a017780-a017780. https://doi.org/10.1101/cshperspect.a017780

Ellegren, H. (2007). Characteristics, causes and evolutionary consequences of male-biased mutation. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1606), 1-10. https://doi.org/10.1098/rspb.2006.3720

Glasauer, S. M. K., & Neuhauss, S. C. F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, *289*(6), 1045-1060. https://doi.org/10.1007/s00438-014-0889-2

Hirota, T., Blakeley, P., Sangrithi, M. N., Mahadevaiah, S. K., Encheva, V., Snijders, A. P., Elinati, E., Ojarikre, O. A., De Rooij, D. G., Niakan, K. K., & Turner, J. M. A. (2018). SETDB1 Links the Meiotic DNA Damage Response to Sex Chromosome Silencing in Mice. *Developmental Cell*, *47*(5), 645-659.e646. https://doi.org/10.1016/j.devcel.2018.10.004

Hurst, L. D., & Ellegren, H. (1998). Sex biases in the mutation rate. *Trends in Genetics*, *14*(11), 446-452.

Kopania, E. E. K., Watson, E. M., Rathje, C. C., Skinner, B. M., Ellis, P. J. I., Larson, E. L., & Good, J. M. (2022). The contribution of sex chromosome conflict to disrupted spermatogenesis in hybrid house mice. *Genetics*. https://doi.org/10.1093/genetics/iyac151

Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., De Rooij, D. G., Hu, Y.-C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, *29*(21), 3699-3706.e3695. https://doi.org/10.1016/j.cub.2019.08.057

Lee, Y. C. G. (2015). The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. *PLoS Genetics*, *11*(6), e1005269. https://doi.org/10.1371/journal.pgen.1005269

Lee, Y. C. G., & Karpen, G. H. (2017). Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife*, *6*. https://doi.org/10.7554/elife.25762

Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex Chromosome Degeneration by Regulatory Evolution. *Current Biology*, *30*(15), 3001-3006.e3005. https://doi.org/10.1016/j.cub.2020.05.052

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663-666. https://doi.org/doi:10.1126/science.abj1813

Mahajan, S., Wei, K. H. C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biology*, *16*(7), e2006348. https://doi.org/10.1371/journal.pbio.2006348

Marand, A. P., Jansky, S. H., Zhao, H., Leisner, C. P., Zhu, X., Zeng, Z., Crisovan, E., Newton, L., Hamernik, A. J., Veilleux, R. E., Buell, C. R., & Jiang, J. (2017). Meiotic crossovers are associated with open chromatin and enriched with Stowaway transposons in potato. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1326-8

McVey, M., & Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics*, *24*(11), 529-538. https://doi.org/10.1016/j.tig.2008.08.007

Moretti, C., Vaiman, D., Tores, F., & Cocquet, J. (2016). Expression and epigenomic landscape of the sex chromosomes in mouse post-meiotic male germ cells. *Epigenetics & Chromatin*, *9*(1). https://doi.org/10.1186/s13072-016-0099-8

Nath, S., Welch, L. A., Flanagan, M. K., & White, M. A. (2022). Meiotic pairing and double-strand break formation along the heteromorphic threespine stickleback sex chromosomes. *Chromosome Research*. https://doi.org/10.1007/s10577-022-09699-0

Nguyen, A. H., & Bachtrog, D. (2021). Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. *PLoS Genetics*, *17*(4), e1009438. https://doi.org/10.1371/journal.pgen.1009438

Peichel, C. L., McCann, S. R., Ross, J. A., Naftaly, A. F. S., Urton, J. R., Cech, J. N., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., & White, M. A. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02097-x

Rodríguez Lorenzo, J. L., Hobza, R., & Vyskot, B. (2018). DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant Silene latifolia. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4936-y

Román, Á., Golz, J. F., Webb, A. A. R., Graham, I. A., & Haydon, M. J. (2020). Combining GAL4 GFP enhancer trap with split luciferase to measure spatiotemporal promoter activity in Arabidopsis. *The Plant Journal*, *102*(1), 187-198. https://doi.org/10.1111/tpj.14603

Shanfelter, A. F., Archambeault, S. L., & White, M. A. (2019). Divergent Fine-Scale Recombination Landscapes between a Freshwater and Marine Population of Threespine Stickleback Fish. *Genome Biology and Evolution*, *11*(6), 1552-1572. https://doi.org/10.1093/gbe/evz090

Spruce, C., Dlamini, S., Ananda, G., Bronkema, N., Tian, H., Paigen, K., Carter, G. W., & Baker, C. L. (2020). HELLS and PRDM9 form a pioneer complex to open chromatin at meiotic recombination hot spots. *Genes & Development*, *34*(5-6), 398-412. https://doi.org/10.1101/gad.333542.119

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., & Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, *353*(6294), 78-82. https://doi.org/doi:10.1126/science.aaf2403

White, M. A., Kitano, J., & Peichel, C. L. (2015). Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. *Molecular Biology and Evolution*, *32*(8), 1981-1995. https://doi.org/10.1093/molbev/msv078

Wucherpfennig, J. I., Howes, T. R., Au, J. N., Au, E. H., Roberts Kingman, G. A., Brady, S. D., Herbert, A. L., Reimchen, T. E., Bell, M. A., Lowe, C. B., Dalziel, A. C., & Kingsley, D. M. (2022). Evolution of stickleback spines through independent cis-regulatory changes at HOXDB. *Nature Ecology & Evolution*, *6*(10), 1537-1552. https://doi.org/10.1038/s41559-022-01855-3

Zhang, W., Wang, X., Yu, Q., Ming, R., & Jiang, J. (2008). DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Research*, *18*(12), 1938-1943. https://doi.org/10.1101/gr.078808.108

Zhou, Q., Ellison, C. E., Kaiser, V. B., Alekseyenko, A. A., Gorchakov, A. A., & Bachtrog, D. (2013). The Epigenome of Evolving Drosophila Neo-Sex Chromosomes: Dosage

Compensation and Heterochromatin Formation. *PLoS Biology*, *11*(11), e1001711. https://doi.org/10.1371/journal.pbio.1001711

APPENDICES

Supplemental tables

Supplemental Table 1. Uncharacterized genes enriched in spermatogonia cells

Cell type	Gene	Chromosome	Fold change (log2)	Percent of cells	Adjusted P-value
Early Spermatogonia	ENSGACG00000015423	chrII	0.660235532	0.19	7.53335E-95
Early Spermatogonia	ENSGACG00000004750	chrVI	1.273021376	0.463	8.106E-297
Early Spermatogonia	ENSGACG00000012927	chrVIII	2.431813574	0.973	0
Early Spermatogonia	ENSGACG00000001123	scaffold_129	2.076076396	0.918	0
Early Spermatogonia	ENSGACG00000014415	chrII	1.733711453	0.66	0
Early Spermatogonia	ENSGACG00000004882	chrXXI	0.85460714	0.238	1.466E-268
Early Spermatogonia	ENSGACG00000012951	chrVIII	0.864085175	0.279	8.5813E-226
Early Spermatogonia	ENSGACG00000016070	chrII	1.040140767	0.354	3.5637E-117
Early Spermatogonia	ENSGACG00000000546	scaffold_229	0.562109739	0.15	2.4673E-116

Supplemental Table 2. Gene ontology enrichment of cell clusters

Cell Cluster	Enrichment FDR	Genes	Pathway Genes	Fold Enrichment	Pathway
Early Spermatids	0.0000	95	984	2.0222	Organelle organization
Early Spermatids	0.0000	71	612	2.3529	RNA binding
Early Spermatids	0.0000	59	431	2.6521	Catalytic complex
Early Spermatids	0.0000	50	284	3.8066	Microtubule-based process
Early Spermatids	0.0000	48	286	3.4376	Microtubule cytoskeleton
Early Spermatids	0.0000	47	289	2.2189	Organelle envelope

Early Spermatids	0.0000	47	289	2.2189	Envelope
Early Spermatids	0.0000	47	414	2.0904	Establishment of localization in cell
Early Spermatids	0.0000	45	430	2.6593	ATP-dependent activity
Early Spermatids	0.0000	45	409	2.2771	Membrane-enclosed lumen
Early Spermatids	0.0000	45	409	2.2771	Organelle lumen
Early Spermatids	0.0000	45	409	2.2771	Intracellular organelle lumen
Early Spermatids	0.0001	44	432	2.0657	Cell projection organization
Early Spermatids	0.0000	43	378	2.4225	Nuclear protein-containing complex
Early Spermatids	0.0000	43	344	2.3145	Intracellular transport
Early Spermatids	0.0000	42	365	2.5534	Cell cycle
Early Spermatids	0.0001	39	234	2.1401	Mitochondrial envelope
Early Spermatids	0.0000	36	266	2.6454	Intracellular protein-containing complex
Early Spermatids	0.0002	35	214	2.1433	Mitochondrial membrane
Early Spermatids	0.0000	34	246	2.9621	Organelle assembly
Early Spermatids	0.0000	34	176	2.5885	Generation of precursor metabolites and energy
Early Spermatids	0.0005	33	308	2.0811	Supramolecular complex
Early Spermatids	0.0001	32	260	2.2725	Plasma membrane bounded cell projection
Early Spermatids	0.0001	31	260	2.3183	Macromolecule catabolic process
Early Spermatids	0.0001	31	311	2.2980	Chromosome
Early Spermatids	0.0000	30	153	4.0241	Cilium organization
Early Spermatids	0.0000	30	141	2.7859	Mitochondrial inner membrane
Early Spermatids	0.0000	30	152	2.6136	Organelle inner membrane
Early Spermatids	0.0001	30	247	2.3693	Cellular macromolecule catabolic process
Early Spermatids	0.0002	30	305	2.3047	Nuclear lumen
Early Spermatids	0.0000	29	111	5.6993	Cilium
Early Spermatids	0.0000	28	148	3.8790	Cilium assembly
Early Spermatids	0.0000	28	158	3.5316	Plasma membrane bounded cell projection assembly
Early Spermatids	0.0000	28	161	3.4797	Cell projection assembly
Early Spermatids	0.0013	25	202	2.2239	Polymeric cytoskeletal fiber
Early Spermatids	0.0000	24	159	3.2712	Cell cycle process
Early Spermatids	0.0000	24	150	3.0271	Microtubule
Early Spermatids	0.0001	24	186	2.6340	Protein catabolic process
Early Spermatids	0.0006	24	255	2.3861	ATP hydrolysis activity
Early Spermatids	0.0007	24	200	2.3583	Transferase complex
Early Spermatids	0.0000	23	137	4.0493	Microtubule-based movement
Early Spermatids	0.0004	23	172	2.4919	Cell division
Early Spermatids	0.0000	22	131	3.7942	Microtubule cytoskeleton organization
Early Spermatids	0.0000	22	75	3.6454	Electron transport chain
Early Spermatids	0.0001	22	97	2.8169	ATP metabolic process
Early Spermatids	0.0004	22	178	2.5822	Cellular protein catabolic process
Early Spermatids	0.0010	22	188	2.4145	Nucleoplasm

Early Spermatids	0.0000	21	90	3.0597	Energy derivation by oxidation of organic compounds
Early Spermatids	0.0008	21	176	2.4995	Proteolysis involved in cellular protein catabolic process
Early Spermatids	0.0010	21	263	2.4648	Chromosome organization
Early Spermatids	0.0000	20	67	4.4477	Protein folding
Early Spermatids	0.0000	20	63	3.7559	Aerobic respiration
Early Spermatids	0.0000	20	67	3.6742	Cellular respiration
Early Spermatids	0.0015	20	221	2.4495	MRNA metabolic process
Early Spermatids	0.0000	19	98	4.5875	Microtubule organizing center
Early Spermatids	0.0000	19	73	3.7340	Translation regulator activity
Early Spermatids	0.0000	19	125	3.3451	Mitotic cell cycle
Early Spermatids	0.0004	19	137	2.8169	Tubulin binding
Early Spermatids	0.0000	18	63	3.9003	Translation regulator activity, nucleic acid binding
Early Spermatids	0.0001	18	75	3.1043	Proton transmembrane transporter activity
Early Spermatids	0.0054	18	148	2.3402	Ribonucleoprotein complex
Early Spermatids	0.0000	17	60	3.9906	Translation factor activity, RNA binding
Early Spermatids	0.0001	17	62	3.2650	Electron transfer activity
Early Spermatids	0.0000	16	46	4.3616	Respirasome
Early Spermatids	0.0001	16	94	3.6543	Mitotic cell cycle process
Early Spermatids	0.0004	16	170	3.0730	MRNA processing
Early Spermatids	0.0016	16	115	2.7594	Microtubule binding
Early Spermatids	0.0090	16	133	2.3721	Ubiquitin-dependent protein catabolic process
Early Spermatids	0.0000	15	29	6.6716	Protein-containing complex localization
Early Spermatids	0.0000	15	48	3.9613	Nucleobase-containing compound transport
Early Spermatids	0.0005	15	91	3.1690	Peptidase complex
Early Spermatids	0.0077	15	82	2.4855	Proton transmembrane transport
Early Spermatids	0.0000	14	41	5.3777	Translation initiation factor activity
Early Spermatids	0.0000	14	42	5.1439	Unfolded protein binding
Early Spermatids	0.0000	14	45	4.7324	Translational initiation
Early Spermatids	0.0002	14	125	3.5851	Helicase activity
Early Spermatids	0.0001	13	34	4.2253	Oxidative phosphorylation
Early Spermatids	0.0002	13	74	3.7882	Proteasome complex
Early Spermatids	0.0009	13	79	3.3291	Endopeptidase complex
Early Spermatids	0.0012	13	43	3.2311	Oxidoreduction-driven active transmembrane transporter activity
Early Spermatids	0.0000	12	54	4.8290	Microtubule associated complex
Early Spermatids	0.0001	12	29	4.6095	Nucleic acid transport
Early Spermatids	0.0001	12	29	4.6095	RNA transport
Early Spermatids	0.0001	12	29	4.6095	Establishment of RNA localization
Early Spermatids	0.0001	12	31	4.4090	RNA localization
Early Spermatids	0.0078	12	127	2.8169	Ribonucleoprotein complex biogenesis
Early Spermatids	0.0000	11	32	6.1972	Cytoskeleton-dependent intracellular transport

Early Spermatids	0.0000	11	35	6.1972	Microtubule-based transport
Early Spermatids	0.0000	11	56	5.1643	Chromosomal region
Early Spermatids	0.0001	11	59	4.8925	RNA helicase activity
Early Spermatids	0.0001	11	59	4.8925	ATP-dependent activity, acting on RNA
Early Spermatids	0.0001	11	26	4.6479	MRNA transport
Early Spermatids	0.0003	11	57	4.2253	Organelle fission
Early Spermatids	0.0003	11	50	4.2253	Nuclear transport
Early Spermatids	0.0003	11	50	4.2253	Nucleocytoplasmic transport
Early Spermatids	0.0010	11	42	3.7183	Nucleoside triphosphate metabolic process
Early Spermatids	0.0010	11	55	3.7183	Spindle
Early Spermatids	0.0028	11	58	3.3199	Positive regulation of cellular protein metabolic process
Early Spermatids	0.0028	11	58	3.3199	Positive regulation of protein metabolic process
Early Spermatids	0.0021	10	37	3.6742	Nucleoside triphosphate biosynthetic process
Early Spermatocytes	0.0000	35	1311	7.3579	Non-membrane-bounded organelle
Early Spermatocytes	0.0000	35	1311	7.3579	Intracellular non-membrane-bounded organelle
Early Spermatocytes	0.0000	24	1767	3.3751	Cellular nitrogen compound biosynthetic process
Early Spermatocytes	0.0000	23	346	14.4761	Structural molecule activity
Early Spermatocytes	0.0000	23	793	6.3309	Organonitrogen compound biosynthetic process
Early Spermatocytes	0.0000	23	1806	3.2256	Cellular macromolecule biosynthetic process
Early Spermatocytes	0.0000	23	1814	3.2082	Macromolecule biosynthetic process
Early Spermatocytes	0.0000	22	150	22.0472	Ribosome
Early Spermatocytes	0.0000	22	296	14.5568	Translation
Early Spermatocytes	0.0000	22	299	14.3725	Peptide biosynthetic process
Early Spermatocytes	0.0000	22	326	13.2027	Amide biosynthetic process
Early Spermatocytes	0.0000	22	336	12.9763	Peptide metabolic process
Early Spermatocytes	0.0000	22	401	10.6613	Cellular amide metabolic process
Early Spermatocytes	0.0000	21	134	23.5613	Structural constituent of ribosome
Early Spermatocytes	0.0001	19	1665	3.2741	Protein-containing complex
Early Spermatocytes	0.0008	18	1817	2.8194	Purine ribonucleoside triphosphate binding
Early Spermatocytes	0.0009	18	1841	2.7731	Purine ribonucleotide binding
Early Spermatocytes	0.0010	18	1849	2.7566	Ribonucleotide binding
Early Spermatocytes	0.0010	18	1856	2.7485	Purine nucleotide binding
Early Spermatocytes	0.0018	18	1930	2.6059	Carbohydrate derivative binding
Early Spermatocytes	0.0004	17	1682	3.1113	Cellular component organization or biogenesis
Early Spermatocytes	0.0000	15	570	7.0378	Pyrophosphatase activity
Early Spermatocytes	0.0000	15	576	7.0378	Hydrolase activity, acting on acid anhydrides
Early Spermatocytes	0.0000	15	576	7.0378	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
Early Spermatocytes	0.0025	15	1596	2.8567	Cellular component organization
Early Spermatocytes	0.0000	14	542	6.9811	Nucleoside-triphosphatase activity
Early Spermatocytes	0.0001	14	984	4.4464	Organelle organization

Early Spermatocytes	0.0026	14	1407	2.9857	ATP binding
Early Spermatocytes	0.0030	14	1426	2.9312	Adenyl ribonucleotide binding
Early Spermatocytes	0.0030	14	1428	2.9253	Adenyl nucleotide binding
Early Spermatocytes	0.0000	12	430	11.1590	ATP-dependent activity
Early Spermatocytes	0.0000	10	38	35.5934	Ribosomal subunit
Early Spermatocytes	0.0000	10	67	32.2565	Protein folding
Early Spermatocytes	0.0000	10	148	19.1150	Ribonucleoprotein complex
Early Spermatocytes	0.0000	10	255	14.5381	ATP hydrolysis activity
Early spermatogonia	0.0000	56	1311	9.8658	Non-membrane-bounded organelle
Early spermatogonia	0.0000	56	1311	9.8658	Intracellular non-membrane-bounded organelle
Early spermatogonia	0.0000	49	150	40.5648	Ribosome
Early spermatogonia	0.0000	49	1767	5.7001	Cellular nitrogen compound biosynthetic process
Early spermatogonia	0.0000	49	1806	5.6769	Cellular macromolecule biosynthetic process
Early spermatogonia	0.0000	49	1814	5.6462	Macromolecule biosynthetic process
Early spermatogonia	0.0000	48	296	26.2366	Translation
Early spermatogonia	0.0000	48	299	25.9045	Peptide biosynthetic process
Early spermatogonia	0.0000	48	346	25.1098	Structural molecule activity
Early spermatogonia	0.0000	48	326	23.7959	Amide biosynthetic process
Early spermatogonia	0.0000	48	336	23.3880	Peptide metabolic process
Early spermatogonia	0.0000	48	401	19.2155	Cellular amide metabolic process
Early spermatogonia	0.0000	48	793	10.9436	Organonitrogen compound biosynthetic process
Early spermatogonia	0.0000	47	134	43.5612	Structural constituent of ribosome
Early spermatogonia	0.0008	19	1665	2.7413	Protein-containing complex
Early spermatogonia	0.0000	17	38	49.9852	Ribosomal subunit
Early spermatogonia	0.0000	17	148	26.8439	Ribonucleoprotein complex
Early spermatogonia	0.0000	16	612	6.8904	RNA binding
Early spermatogonia	0.0003	13	682	4.0753	Embryo development
Early spermatogonia	0.0012	11	551	4.0781	Immune system process
Early spermatogonia	0.0000	10	21	53.2930	Large ribosomal subunit
Late spermatids	0.0000	56	1311	4.5438	Non-membrane-bounded organelle
Late spermatids	0.0000	56	1311	4.5438	Intracellular non-membrane-bounded organelle
Late spermatids	0.0000	53	1767	2.8120	Cellular nitrogen compound biosynthetic process
Late spermatids	0.0000	50	793	5.1851	Organonitrogen compound biosynthetic process
Late spermatids	0.0000	47	1806	2.4835	Cellular macromolecule biosynthetic process
Late spermatids	0.0000	47	1814	2.4700	Macromolecule biosynthetic process
Late spermatids	0.0000	45	336	10.0304	Peptide metabolic process
Late spermatids	0.0000	45	401	8.2325	Cellular amide metabolic process
Late spermatids	0.0000	44	150	16.7305	Ribosome
Late spermatids	0.0000	44	296	11.0097	Translation
Late spermatids	0.0000	44	299	10.8695	Peptide biosynthetic process

Late spermatids	0.0000	44	326	9.9796	Amide biosynthetic process
Late spermatids	0.0000	42	346	10.1176	Structural molecule activity
Late spermatids	0.0000	41	134	17.4743	Structural constituent of ribosome
Late spermatids	0.0000	28	419	4.8698	Mitochondrion
Late spermatids	0.0007	26	988	2.3342	Organelle membrane
Late spermatids	0.0000	23	234	6.5591	Mitochondrial envelope
Late spermatids	0.0000	23	289	5.7925	Organelle envelope
Late spermatids	0.0000	23	289	5.7925	Envelope
Late spermatids	0.0000	21	214	6.6760	Mitochondrial membrane
Late spermatids	0.0053	21	671	2.2624	Oxidoreductase activity
Late spermatids	0.0000	19	141	9.3279	Mitochondrial inner membrane
Late spermatids	0.0000	19	152	8.6694	Organelle inner membrane
Late spermatids	0.0000	19	176	7.8394	Generation of precursor metabolites and energy
Late spermatids	0.0000	18	75	15.8663	Electron transport chain
Late spermatids	0.0001	18	612	3.5258	RNA binding
Late spermatids	0.0000	15	62	14.9170	Electron transfer activity
Late spermatids	0.0000	15	97	10.3887	ATP metabolic process
Late spermatids	0.0000	12	38	16.6218	Ribosomal subunit
Late spermatids	0.0000	12	43	15.5137	Oxidoreduction-driven active transmembrane transporter activity
Late spermatids	0.0000	12	63	12.9281	Aerobic respiration
Late spermatids	0.0000	12	67	12.5787	Cellular respiration
Late spermatids	0.0000	12	90	9.6961	Energy derivation by oxidation of organic compounds
Late spermatids	0.0000	12	148	8.7814	Ribonucleoprotein complex
Late spermatids	0.0000	12	130	7.6297	Primary active transmembrane transporter activity
Late spermatids	0.0013	12	304	3.7233	Active transmembrane transporter activity
Late spermatids	0.0000	11	34	20.3156	Oxidative phosphorylation
Late spermatids	0.0000	11	46	15.8010	Respirasome
Late spermatids	0.0058	11	307	3.3593	Chordate embryonic development
Late spermatids	0.0061	11	309	3.3330	Embryo development ending in birth or egg hatching
Late spermatids	0.0000	10	75	9.2344	Proton transmembrane transporter activity
Late spermatids	0.0000	10	82	8.6187	Proton transmembrane transport
Late spermatocytes	0.0000	104	1596	2.2698	Cellular component organization
Late spermatocytes	0.0000	87	1407	2.1490	ATP binding
Late spermatocytes	0.0000	87	1426	2.1147	Adenyl ribonucleotide binding
Late spermatocytes	0.0000	87	1428	2.1109	Adenyl nucleotide binding
Late spermatocytes	0.0000	77	984	2.7298	Organelle organization
Late spermatocytes	0.0000	48	612	2.7255	RNA binding
Late spermatocytes	0.0000	47	576	2.5624	Hydrolase activity, acting on acid anhydrides
Late spermatocytes	0.0000	47	687	2.2332	Cytoskeleton
Late spermatocytes	0.0000	46	570	2.5179	Pyrophosphatase activity

Late spermatocytes	0.0000	45	734	2.2892	Cellular component biogenesis
Late spermatocytes	0.0000	44	542	2.5406	Nucleoside-triphosphatase activity
Late spermatocytes	0.0000	43	430	4.1439	ATP-dependent activity
Late spermatocytes	0.0000	42	646	2.3555	Cellular component assembly
Late spermatocytes	0.0000	37	431	2.8933	Catalytic complex
Late spermatocytes	0.0000	34	284	4.8466	Microtubule-based proc.
Late spermatocytes	0.0000	32	365	3.3428	Cell cycle
Late spermatocytes	0.0000	32	432	2.5759	Cell projection organization
Late spermatocytes	0.0001	29	409	2.5439	Membrane-enclosed lumen
Late spermatocytes	0.0001	29	409	2.5439	Organelle lumen
Late spermatocytes	0.0001	29	409	2.5439	Intracellular organelle lumen
Late spermatocytes	0.0007	29	414	2.2421	Establishment of localization in cell
Late spermatocytes	0.0017	29	432	2.1222	Cytoskeleton organization
Late spermatocytes	0.0000	28	255	4.3052	ATP hydrolysis activity
Late spermatocytes	0.0001	27	344	2.5481	Intracellular transport
Late spermatocytes	0.0001	26	378	2.6552	Nuclear protein-containing complex
Late spermatocytes	0.0018	26	408	2.2237	Plasma membrane bounded cell projection organization
Late spermatocytes	0.0000	25	286	3.5269	Microtubule cytoskeleton
Late spermatocytes	0.0000	25	260	3.0275	Plasma membrane bounded cell projection
Late spermatocytes	0.0002	25	313	2.5723	Cell projection
Late spermatocytes	0.0001	23	266	2.9693	Intracellular protein-containing complex
Late spermatocytes	0.0001	23	311	2.9415	Chromosome
Late spermatocytes	0.0008	23	308	2.4980	Supramolecular complex
Late spermatocytes	0.0018	23	353	2.3488	Protein-containing complex subunit organization
Late spermatocytes	0.0000	22	263	4.0141	Chromosome organization
Late spermatocytes	0.0001	22	305	2.9229	Nuclear lumen
Late spermatocytes	0.0000	21	111	7.9826	Cilium
Late spermatocytes	0.0005	21	260	2.7111	Macromolecule catabolic proc.
Late spermatocytes	0.0000	20	153	4.9762	Cilium organization
Late spermatocytes	0.0001	20	246	3.1459	Organelle assembly
Late spermatocytes	0.0002	20	232	2.9429	Cellular protein-containing complex assembly
Late spermatocytes	0.0017	19	247	2.6263	Cellular macromolecule catabolic proc.
Late spermatocytes	0.0000	18	137	5.7284	Microtubule-based movement
Late spermatocytes	0.0000	18	148	4.6476	Cilium assembly
Late spermatocytes	0.0000	18	158	4.1749	Plasma membrane bounded cell projection assembly
Late spermatocytes	0.0000	18	161	4.1053	Cell projection assembly
Late spermatocytes	0.0000	18	150	4.0380	Microtubule
Late spermatocytes	0.0001	18	186	3.3287	Protein catabolic proc.
Late spermatocytes	0.0013	18	202	2.7676	Polymeric cytoskeletal fiber
Late spermatocytes	0.0003	17	188	3.2766	Nucleoplasm

Late spermatocytes	0.0000	16	67	5.9176	Protein folding
Late spermatocytes	0.0000	16	159	3.9809	Cell cycle proc.
Late spermatocytes	0.0000	15	91	5.0065	Peptidase complex
Late spermatocytes	0.0009	15	221	3.2073	MRNA metabolic proc.
Late spermatocytes	0.0018	15	178	3.0186	Cellular protein catabolic proc.
Late spermatocytes	0.0000	14	125	4.4554	Mitotic cell cycle
Late spermatocytes	0.0000	13	42	7.4124	Unfolded protein binding
Late spermatocytes	0.0000	13	74	5.9299	Proteasome complex
Late spermatocytes	0.0000	13	79	5.2323	Endopeptidase complex
Late spermatocytes	0.0001	13	170	4.3390	MRNA processing
Late spermatocytes	0.0002	11	94	4.7040	Mitotic cell cycle proc.
Late spermatocytes	0.0015	11	131	3.8597	Microtubule cytoskeleton organization
Late spermatocytes	0.0000	10	29	9.7746	Protein-containing complex localization
Late spermatocytes	0.0000	10	57	6.5164	Organelle fission
Late spermatocytes	0.0005	10	125	4.7188	Helicase activity
Late spermatocytes	0.0017	10	60	4.1468	Translation factor activity, RNA binding
Late spermatogonia	0.0000	99	1311	9.5706	Non-membrane-bounded organelle
Late spermatogonia	0.0000	99	1311	9.5706	Intracellular non-membrane-bounded organelle
Late spermatogonia	0.0000	85	1767	5.4781	Cellular nitrogen compound biosynthetic process
Late spermatogonia	0.0000	83	793	10.4437	Organonitrogen compound biosynthetic process
Late spermatogonia	0.0000	81	1806	5.1992	Cellular macromolecule biosynthetic process
Late spermatogonia	0.0000	81	1814	5.1713	Macromolecule biosynthetic process
Late spermatogonia	0.0000	79	336	20.9914	Peptide metabolic process
Late spermatogonia	0.0000	79	401	17.3155	Cellular amide metabolic process
Late spermatogonia	0.0000	78	150	35.6722	Ribosome
Late spermatogonia	0.0000	78	296	23.3327	Translation
Late spermatogonia	0.0000	78	299	23.0429	Peptide biosynthetic process
Late spermatogonia	0.0000	78	326	21.1995	Amide biosynthetic process
Late spermatogonia	0.0000	76	346	21.9077	Structural molecule activity
Late spermatogonia	0.0000	74	134	37.8457	Structural constituent of ribosome
Late spermatogonia	0.0000	37	612	8.5016	RNA binding
Late spermatogonia	0.0000	36	1665	2.9021	Protein-containing complex
Late spermatogonia	0.0005	28	1460	2.2843	Animal organ development
Late spermatogonia	0.0000	25	148	22.0198	Ribonucleoprotein complex
Late spermatogonia	0.0075	24	1682	2.0312	Cellular component organization or biogenesis
Late spermatogonia	0.0000	23	38	37.7223	Ribosomal subunit
Late spermatogonia	0.0000	20	682	3.4845	Embryo development
Late spermatogonia	0.0000	19	734	3.8619	Cellular component biogenesis
Late spermatogonia	0.0000	18	551	3.7223	Immune system process
Late spermatogonia	0.0000	17	288	6.4686	Hemopoiesis
Late spermatogonia	0.0000	17	294	6.3667	Immune system development

Late spermatogonia	0.0000	17	294	6.3667	Hematopoietic or lymphoid organ development
Late spermatogonia	0.0000	16	307	5.9454	Chordate embryonic development
Late spermatogonia	0.0000	16	309	5.8993	Embryo development ending in birth or egg hatching
Late spermatogonia	0.0000	13	21	38.6448	Large ribosomal subunit
Late spermatogonia	0.0000	12	127	16.7869	Ribonucleoprotein complex biogenesis
Late spermatogonia	0.0000	10	17	36.5868	Small ribosomal subunit
Late spermatogonia	0.0000	10	96	18.2934	Ribosome biogenesis
Late spermatogonia	0.0000	10	121	7.9271	Myeloid cell differentiation
Late spermatogonia	0.0007	10	365	4.7563	Cell cycle
Mid-late spermatocytes	0.0000	44	1682	2.6790	Cellular component organization or biogenesis
Mid-late spermatocytes	0.0000	43	1596	2.7239	Cellular component organization
Mid-late spermatocytes	0.0001	42	1817	2.2165	Purine ribonucleoside triphosphate binding
Mid-late spermatocytes	0.0001	42	1841	2.1814	Purine ribonucleotide binding
Mid-late spermatocytes	0.0001	42	1849	2.1689	Ribonucleotide binding
Mid-late spermatocytes	0.0001	42	1856	2.1627	Purine nucleotide binding
Mid-late spermatocytes	0.0002	42	1930	2.0541	Carbohydrate derivative binding
Mid-late spermatocytes	0.0000	37	984	3.8434	Organelle organization
Mid-late spermatocytes	0.0001	34	1407	2.4316	ATP binding
Mid-late spermatocytes	0.0001	34	1426	2.3889	Adenyl ribonucleotide binding
Mid-late spermatocytes	0.0001	34	1428	2.3843	Adenyl nucleotide binding
Mid-late spermatocytes	0.0008	31	1311	2.1996	Non-membrane-bounded organelle
Mid-late spermatocytes	0.0008	31	1311	2.1996	Intracellular non-membrane-bounded organelle
Mid-late spermatocytes	0.0002	20	570	3.2183	Pyrophosphatase activity
Mid-late spermatocytes	0.0002	20	576	3.2183	Hydrolase activity, acting on acid anhydrides
Mid-late spermatocytes	0.0002	20	576	3.2183	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
Mid-late spermatocytes	0.0000	19	432	4.3345	Cell projection organization
Mid-late spermatocytes	0.0007	18	542	3.0749	Nucleoside-triphosphatase activity
Mid-late spermatocytes	0.0035	18	734	2.6700	Cellular component biogenesis
Mid-late spermatocytes	0.0000	17	430	5.2824	ATP-dependent activity
Mid-late spermatocytes	0.0001	17	408	4.0313	Plasma membrane bounded cell projection organization
Mid-late spermatocytes	0.0033	17	646	2.7853	Cellular component assembly
Mid-late spermatocytes	0.0065	16	612	2.6949	RNA binding

Mid-late spermatocytes	0.0000	15	153	10.8133	Cilium organization
Mid-late spermatocytes	0.0000	14	284	6.5535	Microtubule-based process
Mid-late spermatocytes	0.0001	14	365	4.4265	Cell cycle
Mid-late spermatocytes	0.0000	13	148	9.7620	Cilium assembly
Mid-late spermatocytes	0.0000	13	158	8.6774	Plasma membrane bounded cell projection assembly
Mid-late spermatocytes	0.0000	13	161	8.5196	Cell projection assembly
Mid-late spermatocytes	0.0000	13	246	5.8572	Organelle assembly
Mid-late spermatocytes	0.0073	13	431	3.0427	Catalytic complex
Mid-late spermatocytes	0.0001	12	255	5.9251	ATP hydrolysis activity
Mid-late spermatocytes	0.0005	12	311	4.5056	Chromosome
Mid-late spermatocytes	0.0007	12	260	4.3253	Plasma membrane bounded cell projection
Mid-late spermatocytes	0.0031	12	313	3.6044	Cell projection
Mid-late spermatocytes	0.0000	10	111	14.4178	Cilium
Mid-late spermatocytes	0.0028	10	286	4.2910	Microtubule cytoskeleton
Mid-spermatocytes	0.0000	82	1311	7.2591	Non-membrane-bounded organelle
Mid-spermatocytes	0.0000	82	1311	7.2591	Intracellular non-membrane-bounded organelle
Mid-spermatocytes	0.0000	81	793	9.2357	Organonitrogen compound biosynthetic process
Mid-spermatocytes	0.0000	80	1767	4.6730	Cellular nitrogen compound biosynthetic process
Mid-spermatocytes	0.0000	76	296	20.8309	Translation
Mid-spermatocytes	0.0000	76	299	20.5672	Peptide biosynthetic process
Mid-spermatocytes	0.0000	76	326	18.8931	Amide biosynthetic process
Mid-spermatocytes	0.0000	76	336	18.5692	Peptide metabolic process
Mid-spermatocytes	0.0000	76	401	15.2564	Cellular amide metabolic process
Mid-spermatocytes	0.0000	76	1806	4.4212	Cellular macromolecule biosynthetic process
Mid-spermatocytes	0.0000	76	1814	4.3973	Macromolecule biosynthetic process
Mid-spermatocytes	0.0000	74	150	31.0206	Ribosome
Mid-spermatocytes	0.0000	72	346	19.0036	Structural molecule activity
Mid-spermatocytes	0.0000	71	134	33.3607	Structural constituent of ribosome
Mid-spermatocytes	0.0000	47	1665	3.3550	Protein-containing complex
Mid-spermatocytes	0.0000	32	612	6.8413	RNA binding
Mid-spermatocytes	0.0000	24	148	19.3621	Ribonucleoprotein complex
Mid-spermatocytes	0.0000	21	38	32.0685	Ribosomal subunit
Mid-spermatocytes	0.0000	21	682	3.2891	Embryo development
Mid-spermatocytes	0.0000	16	307	5.3868	Chordate embryonic development
Mid-spermatocytes	0.0000	16	309	5.3448	Embryo development ending in birth or egg hatching

Mid-spermatocytes	0.0007	16	419	3.0817	Mitochondrion
Mid-spermatocytes	0.0024	15	551	2.8130	Immune system process
Mid-spermatocytes	0.0000	14	160	8.0894	Cytosol
Mid-spermatocytes	0.0007	13	289	3.5862	Organelle envelope
Mid-spermatocytes	0.0007	13	289	3.5862	Envelope
Mid-spermatocytes	0.0000	12	21	32.0685	Large ribosomal subunit
Mid-spermatocytes	0.0000	12	130	8.1444	Primary active transmembrane transporter activity
Mid-spermatocytes	0.0004	12	288	4.2057	Hemopoiesis
Mid-spermatocytes	0.0004	12	214	4.2057	Mitochondrial membrane
Mid-spermatocytes	0.0004	12	279	4.1715	Homeostatic process
Mid-spermatocytes	0.0004	12	294	4.1379	Immune system development
Mid-spermatocytes	0.0004	12	294	4.1379	Hematopoietic or lymphoid organ development
Mid-spermatocytes	0.0005	12	304	4.0401	Active transmembrane transporter activity
Mid-spermatocytes	0.0008	12	234	3.7728	Mitochondrial envelope
Mid-spermatocytes	0.0000	11	141	5.9537	Mitochondrial inner membrane
Mid-spermatocytes	0.0001	11	152	5.5334	Organelle inner membrane
Mid-spermatocytes	0.0000	10	97	7.7742	ATP metabolic process
Somatic	0.0000	42	1767	3.8043	Cellular nitrogen compound biosynthetic process
Somatic	0.0000	40	1806	3.6182	Cellular macromolecule biosynthetic process
Somatic	0.0000	40	1814	3.5987	Macromolecule biosynthetic process
Somatic	0.0000	39	793	6.9242	Organonitrogen compound biosynthetic process
Somatic	0.0000	39	1311	5.4353	Non-membrane-bounded organelle
Somatic	0.0000	39	1311	5.4353	Intracellular non-membrane-bounded organelle
Somatic	0.0000	38	336	14.4133	Peptide metabolic process
Somatic	0.0000	38	401	11.8539	Cellular amide metabolic process
Somatic	0.0000	37	296	15.8332	Translation
Somatic	0.0000	37	299	15.6328	Peptide biosynthetic process
Somatic	0.0000	37	326	14.3604	Amide biosynthetic process
Somatic	0.0000	36	150	23.5610	Ribosome
Somatic	0.0000	34	134	24.9419	Structural constituent of ribosome
Somatic	0.0000	34	346	14.0976	Structural molecule activity
Somatic	0.0000	27	1665	3.0090	Protein-containing complex
Somatic	0.0000	22	612	7.2347	RNA binding
Somatic	0.0003	15	682	3.6281	Embryo development
Somatic	0.0002	14	551	4.0458	Immune system process
Somatic	0.0026	14	707	3.0947	Regulation of biological quality
Somatic	0.0000	13	279	6.8333	Homeostatic process
Somatic	0.0000	12	148	14.8347	Ribonucleoprotein complex
Somatic	0.0001	11	288	5.8279	Hemopoiesis
Somatic	0.0001	11	294	5.7369	Immune system development

Somatic	0.0001	11	294	5.7369	Hematopoietic or lymphoid organ development
Somatic	0.0000	10	77	15.8944	Myeloid cell homeostasis
Somatic	0.0000	10	88	14.5122	Homeostasis of number of cells
Somatic	0.0000	10	160	9.0211	Cytosol
Somatic	0.0004	10	307	5.1749	Chordate embryonic development