Evolutionary history of recombination on sex chromosomes and of sex-biased gene expression

Iulia Darolti

Department of Genetics, Evolution and Environment

University College London

PhD Supervisor: Professor Judith Mank

A thesis submitted for the degree of

Doctor of Philosophy

August, 2019

Declaration

I, Iulia Darolti, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Iulia Darolti

4

Abstract

Males and females within a species are often under different selection pressures, which can affect both gene sequence and expression. Sex differences in selection are predicted to precipitate sex chromosome formation, and the rate of recombination on sex chromosomes varies greatly across taxa. Examining the extent of recombination suppression on sex chromosomes across closely related species can offer insight into the forces shaping sex differences and sex chromosome evolution over time. Additionally, sex-biased genes are thought to encode sexually dimorphic traits and are therefore a useful way to examine the effect of sex-specific selection across the genome. In this thesis, I use whole genome and transcriptome sequencing data to characterise the structure and conservation of sex chromosome systems across related species. I combine this with patterns of sex-specific single nucleotide polymorphisms to uncover the degree of recombination suppression and divergence across poeciliid sex chromosomes, as well as explore the consequences of recombination arrest on gene expression patterns. Finally, I investigate the selective dynamics driving the expression and rate of sequence evolution of sex-biased genes.

Impact Statement

The focus of this thesis is the evolution of phenotypic differences between males and females, referred to as sexual dimorphism, and the major underlying mechanisms, sex chromosomes and differential gene expression. Sexual dimorphism is prevalent across the vast majority of animal species and has important evolutionary, biological and medical implications. Evidence suggests that the incidence and severity of diseases, the symptoms experienced and even the side effects of treatment can depend on sex. In spite of this, clinical research has generally focused on male individuals or, in cases where both sexes were included, results have not been analysed by sex. A recent study from the International Mouse Phenotyping Consortium has revealed that differences between the sexes significantly impact a large proportion of somatic traits, highlighting the importance of accounting for sexual dimorphism in biomedical research. Studying the mechanisms through which sexual dimorphism develops is fundamental to better understanding the extent and genetic implications of differences between the sexes.

Within the academic community, the research presented in this thesis has been disseminated at multiple international conferences, including The European Meeting for PhD Students in Evolutionary Biology (Gotland, Sweden, September 2016), the meeting of the European Society for Evolutionary Biology (Groningen, The Netherlands, August 2017), the Joint Congress on Evolutionary Biology (Montpellier, France, August 2018), The Evolution of Cooperation and Conflict Symposium (Uppsala, Sweden, May 2019) and the Evolution Conference

(Providence, USA, June 2019). Regarding the publication of this research in scientific journals, the work presented in Chapter 2 and Chapter 5 is published in the journals *Nature Communications* and *Molecular Ecology*, respectively, while the work in Chapter 3 is currently in press *in Proceedings of the National Academy of Sciences, USA*.

Beyond the scientific community, the research in this thesis has also been disseminated in the wider, non-academic community. The work on sexual dimorphism was shared and discussed with the members of the public as part of the public outreach initiative Soapbox Science (London, UK, May 2018). Moreover, this work has in part inspired the creation of the art exhibition entitled "Natural Creativity: Sex and Trickery" (Grant Museum of Zoology, London, UK, 19th October – 23rd December 2016), by Clara Lacy, a former artist in residence in the Mank lab.

Acknowledgements

I am extremely grateful to my supervisor, Judith Mank, for her guidance, invaluable advice, encouragement and all the opportunities given throughout my study. I have learned a lot from her about how to do good research and how to communicate it clearly. The constructive and exciting environment that she has created has made my PhD experience truly enjoyable.

I am thankful for the insightful feedback and contribution to my thesis from past and present members of the Mank lab. In particular, I very much thank Alison Wright for her enthusiasm, for teaching me essential bioinformatics and coding skills during my first year in the lab and for all her help in my work since then.

This journey would have been impossibly difficult without my family's love and reassurance. I am very grateful to my parents for all their advice over the years and for always believing in me, especially at times when I didn't. I thank my sister for keeping me light-hearted with her humour and for offering me much-needed distractions. I have had unwavering support throughout this process from my partner, Rareș Ciurba, and I cannot thank him enough for this. His eagerness to put me first and his endless patience amaze me every time. I also want to thank Alexandra Zegrean for her life-long friendship and for lending an ear when I needed one.

I owe a great deal to John Fitzpatrick and Tucker Gilman for introducing me to the world of research and for giving me the confidence to start on this path.

Table of Contents

12

List of figures

List of tables

Chapter 1. Introduction

1.1 Overview

In sexually reproducing species, males and females often have different reproductive optima and are subject to conflicting selection pressures (Andersson 1994; Arnqvist and Rowe 2005; Bateman 1948). In the process of maximising fitness in each sex, males and females of many species evolve striking phenotypic differences, referred to as sexual dimorphism (Darwin 1871). While primary sexual dimorphism of gonads and gametes directly influences reproduction, many somatic traits also have antagonistic fitness effects (Perry and Rowe 2015). As such, sexual dimorphism arguably accounts for the greatest breadth of intraspecific variation, manifesting across a broad range of phenotypes, including physiological, morphological and behavioural traits (Darwin 1871).

Despite the diversity and pervasiveness of sexual dimorphism across taxa, males and females respond to sex-specific selection using a largely shared genome. This can create strong intralocus sexual conflict, where an allele at a specific gene is beneficial to one sex but detrimental to the other (Bonduriansky and Chenoweth 2009; Parker and Partridge 1998). The constraint placed on the evolution of sex-specific phenotypes by shared genetic architecture can be resolved through multiple, non-mutually exclusive, mechanisms (Bonduriansky and Chenoweth 2009; Mank 2017), two of which are the focus of this thesis.

In some species, male and female genomes differ only by their sex chromosomes, which are unequally inherited between the two sexes. As such,

sex-limited (Y or W) chromosomes are thought to play an important role in facilitating sexual dimorphism and resolving sexual conflict by being a hotspot for sexually antagonistic genes (Rice 1984). However, species that lack sex-limited chromosomes or that do not rely on sex chromosomes for sex determination also exhibit a variety of phenotypic sex differences (Bachtrog et al. 2014; Eads et al. 2007; Mank et al. 2006). A second, arguably more prevalent, mechanism encoding for sexually dimorphic characteristics is the evolution of sex-specific gene regulation. Differential regulation of genes that are present in both sexes, known as sex-biased gene expression, is thought to account for a large fraction of the sexually selected traits observed in many species (Ellegren and Parsch 2007). Thus, understanding the evolution of sex chromosomes and of sex-biased gene expression is a powerful way to examine the effect of sex-specific selection across the genome.

1.2 Sex chromosomes

Chromosomal sex determination, either through male heterogametic (XY) or female heterogametic (ZW) systems, has evolved independently countless of times across predominantly diploid species (Bachtrog et al. 2014). Many unrelated sex chromosome systems share particular features, such as a unique inheritance mode and loss of recombination in the heterogametic sex, and as a result often follow common evolutionary divergence patterns that distinguish them from the rest of the genome (Bachtrog et al. 2011). Due to their unequal mode of segregation between males and females, sex chromosomes are

22

predicted to be a hotspot for sex-specific selection and sexually antagonistic genes (Connallon and Clark 2010; Rice 1984). Furthermore, sex chromosomes contribute to important evolutionary processes such as adaptation, speciation and genomic conflict (Kitano et al. 2009; Mank et al. 2014; Roberts et al. 2009). Therefore, the study of sex chromosomes is fundamental not just for understanding the selective forces driving the evolution of genetic sex determination but also for understanding sex-specific forces, the evolution of sexual conflict and the development of sexual dimorphism.

1.2.1 Sexual conflict and sex chromosome formation

The classic model of sex chromosome evolution (Bull 1983; Charlesworth et al. 2005; Rice 1987) posits that sex chromosomes originate from a pair of homologous, recombining, autosomes that have acquired, through translocation, mutation or duplication, a gene that acts as trigger for the sex determining pathway (Ohno 1967). Recombination suppression is predicted to arise in the heterogametic sex between the nascent sex chromosomes once this sex-determining locus becomes linked to a nearby gene with sexually antagonistic effects (Bull 1983; Fisher 1931). Linkage between these two loci leads to the formation of a so-called supergene, ensuring that the linked genes are inherited together in future generations and resolving the sexual conflict. Similar mechanisms of maintaining linkage between co-adapted gene complexes outside the sex chromosomes (Dobzhansky and Dobzhansky 1970) can be observed, for example, through recombination arrest over the region containing

the *S* locus that controls self-incompatibility in plants (Boyes et al. 1997), as well as over the social chromosome in fire ants (Wang et al. 2013).

Over time, selection can favour the further accumulation of loci with sex-specific benefits on the sex-limited chromosome, which in turn selects for the expansion of the non-recombining region. The loss of recombination triggers a host of non-adaptive processes (explained below), including the build-up of heterochromatin and deleterious mutations, that cause the once identical sex chromosomes to dramatically diverge from each other (Bachtrog 2013; Charlesworth and Charlesworth 2000). Stepwise expansion of the non-recombining region can generate different levels of divergence between gametologs, called strata (Bachtrog 2013; Charlesworth et al. 2005). It has long been assumed that the underlying mechanism for recombination arrest involves chromosomal inversions surrounding the linked loci (Charlesworth et al. 2005). Inversions would instantaneously prevent recombination throughout the inversed region and explain the evolutionary strata of divergence seen on the sex chromosomes of mammals (Lahn and Page 1999), fish (Natri et al. 2013; Roesti et al. 2013; Wright et al. 2017), birds (Handley et al. 2004; Wright et al. 2014), snakes (Vicoso et al. 2013a) and plants (Bergero et al. 2007; Hough et al. 2014; Nicolas et al. 2004). However, it is possible that inversions accumulate following sex chromosome divergence, as the loss of recombination through other means would abolish selection for gene order (Flot et al. 2013). Evidence from young sex chromosome systems indicates that divergence and recombination suppression is heterogeneous (Almeida et al. 2019; Bergero et al. 2013; Chibalina and Filatov 2011; Natri et al. 2013; Nicolas et al. 2004), which

24

contradicts the inversion mechanism. Moreover, medaka sex reversed females (XY) are still able to recombine according to their phenotypic sex (Kondo et al. 2001; Matsuda et al. 1999). These studies suggest that a universal mechanism of recombination arrest may not exist, as multiple processes can impact recombination frequencies on the sex chromosomes.

Furthermore, although the model of sex chromosome evolution as a form to resolve sexual conflict is a widely assumed mechanism, so far, there is little empirical evidence to support it (Franchini et al. 2018; Kitano et al. 2009; Roberts et al. 2009; Wright et al. 2017). This lack of evidence is due in large part to predominant focus on species with old sex chromosomes, where it can be difficult to determine whether the presence of loci with sex-specific fitness effects is the cause or, in fact, the outcome of sex chromosome evolution. More specifically, the non-recombining region provides a hotspot for sexually antagonistic loci to relocate to in order to benefit from linkage disequilibrium and resolve sexual conflict (Koerich et al. 2008). Alternatively, it is possible for genes on sex-limited chromosomes to acquire sex-specific functions following recombination suppression, as they adapt to their sex-specific environment (Zhou and Bachtrog 2012). As such, it remains unclear whether the presence of sexually antagonistic loci on the sex-limited chromosomes is the result of sexual conflict driving the evolution of sex chromosomes, or a consequence of sex-specific adaptation.

1.2.2 Consequences of sex chromosome recombination arrest

The lack of recombination between the sex chromosomes has widespread consequences not just at the chromosome level, but also across the entire

genome (Bachtrog 2013). A main consequence on the sex chromosomes is a reduction of their effective population size (N_e). N_e in this case reflects the copy number of each chromosome that is represented in the population, and thus influences the efficiency of selection and the strength of genetic drift (Charlesworth 2009). The non-recombining sex-limited Y and W chromosomes experience a reduction in N_e to a quarter of that of the autosomes, while their X and Z homologous regions, which are still able to recombine in the homomorphic sex, have their N_e reduced to three quarters of that of the autosomes. Therefore, the reduced N_e predisposes the Y and W chromosomes, and to a smaller degree the X and the Z chromosomes, to reduced purifying selection, increased genetic drift and hence a faster rate of accumulating deleterious mutations compared to the autosomes (Charlesworth 2009).

Moreover, the non-recombining regions are affected by a combination of processes such as background selection (purifying selection eliminating all linked loci if the detrimental effects of deleterious mutations outweigh the advantages of beneficial ones, Orr and Kim (1998)), selective sweeps (reduction of genetic variation around a beneficial mutation that has reached fixation due to positive selection, (Smith and Haigh 1974)), and Müller's ratchet (accumulation of deleterious mutations through stochasticity, (Müller 1932)), collectively referred to as Hill-Robertson effects (Comeron et al. 2008; Hill and Robertson 1966). As such, sex-limited chromosomes are predicted to accumulate overall fewer beneficial mutations but considerably more deleterious mutations compared to the rest of the genome, which in time can result in drastic gene deletions and loss of function.

26

The decay of gene activity on the Y or W leads to an imbalance in gene dose between the sexes (Malone et al. 2012; Sun et al. 2013), as the heterogametic sex becomes hemizygous (having a single copy of a gene) for many sex-linked loci, while the homogametic sex maintains two copies of all X- or Z-linked genes. Gene dose often, although not always, correlates with gene activity and expression which translates at the protein level (Guo et al. 1996; Malone et al. 2012). Genetic pathways that incorporate both autosomal and sex-linked genes are primarily affected by such imbalances in gene dose, with potential severe phenotypic consequences for the heterogametic sex. This process triggers strong selection for the evolution of a dosage compensation mechanism that will restore the balance in gene expression between the autosomes and the sex chromosomes in the heterogametic sex and between the sexes for X- or Z-linked genes (Ohno 1967). Through different mechanisms, dosage compensation in species such as mammals (Payer and Lee 2008; Pessia et al. 2012), *Drosophila* (Gupta et al. 2006; Larschan et al. 2011) and *Caenorhabditis elegans* (Ercan et al. 2007), has evolved at a "global" or "complete" level, balancing transcription across the entire sex chromosome. However, more commonly, in birds (Ellegren and Parsch 2007; Mank 2009; Zimmer et al. 2016); snakes (Vicoso et al. 2013a) and fish (Chen et al. 2014; Hale et al. 2018; Leder et al. 2010; White et al. 2015), dosage compensation is "incomplete" and acts on a gene-by-gene level, where dosage sensitive genes are compensated but average expression of the sex chromosomes is reduced in the heterogametic sex.

1.2.3   Diversity and turnover of sex chromosome systems

Across lineages, sex chromosome evolution is characterised by a striking diversity in the rates of divergence between the X and the Y (or the Z and the W) chromosomes (Bachtrog et al. 2011; Bachtrog et al. 2014). Species such as mammals (Lahn and Page 1999; Skaletsky et al. 2003), birds (Wright et al. 2014), *Drosophila* (Bachtrog 2013) and some snakes (Matsubara et al. 2006), possess old and heteromorphic, or highly differentiated, sex chromosomes. In these systems, the region of recombination suppression spans almost the entire length of the sex chromosomes, with only a small region still recombining, referred to as the pseudoautosomal region (PAR) (Bachtrog et al. 2014). However, there is a significant heterogeneity among clades, and even among species with shared sex chromosomes systems, in the spread of the non-recombining region, and the subsequent degree of sex chromosome divergence (Fujito et al. 2015; Sessions et al. 2016; Vicoso et al. 2013b). Age does not always reliably correlate with the extent of recombination suppression (Wright et al. 2016), as in some species, sex chromosomes maintain a largely homomorphic, undifferentiated, structure over long evolutionary periods (Ahmed et al. 2014; Stock et al. 2011; Vicoso and Bachtrog 2013; Vicoso et al. 2013b; Xu et al. 2018), while in others the two sex chromosomes are relatively young, yet profoundly distinct (Bergero et al. 2007).

Furthermore, certain taxonomic groups, such as teleost fish (Mank and Avise 2009; Mank et al. 2006; Ross et al. 2009; Volff and Schartl 2001), amphibians (Miura 2007), reptiles (Pennell et al. 2015) and insects (Vicoso and Bachtrog 2013), experience rapid sex chromosome turnover, where sex determination in related species transitions between male-heterogamety (XY)

28

and female-heterogamety (ZW). On the other hand, mammals and birds exhibit extreme conservatism, whereby all species share an XY (Bick and Jackson 1967) and, respectively, a ZW system (Fridolfsson et al. 1998). Theoretical work outlines multiple adaptive and non-adaptive evolutionary scenarios under which sex chromosome transitions could occur, including strong selection on a novel sexually antagonistic mutation (Roberts et al. 2009; Ser et al. 2010; Van Doorn and Kirkpatrick 2010), selection to reach optimal sex ratio (Beukeboom and Perrin 2014; Kozielska et al. 2006; Roberts et al. 2016), drift-induced sex chromosome transitions in the absence of selection (Bull and Charnov 1977; Veller et al. 2017) and transitions induced by the deleterious effects of highly degenerated sex chromosomes on the heterogametic sex (Blaser et al. 2013; Jeffries et al. 2018). Despite this extensive theory, empirical data testing these predictions is largely lacking, and many questions about sex chromosome transitions remain unanswered. Studies of young sex chromosome systems and comparisons between closely related species would shed light on the forces shaping the diversity and turnover of sex chromosome systems.

1.3 Sex-biased gene expression and sex-specific selection

With the exception of genes restricted to the sex-limited chromosome, males and females of the same species share an identical genome. While sexual dimorphism could be considered as predominantly the result of sex chromosome evolution, many dioecious species lack a sex-limited chromosome, or lack sex chromosomes altogether, yet still display clear sexual dimorphism (Bachtrog et

al. 2014; Eads et al. 2007; Mank et al. 2006). As such, sex chromosomes are clearly not necessary prerequisite for the evolution of sexual dimorphism, and they are likely not the only source of sexual dimorphism in species that do possess sex chromosomes. Indeed, although a link between sex chromosomes and a number of sexually dimorphic traits has been found in some systems (Kitano et al. 2009; Roberts et al. 2009; Sætre et al. 2003), many studies across a broad range of taxa do not find a disproportionately large association between sex chromosomes and sexually selected phenotypes (Fitzpatrick 2004; Husby et al. 2013; Mank et al. 2005).

Sex-specific traits have been extensively studied in relation to the differential regulation of genes expressed in both sexes, referred to as sex-biased gene expression (Ellegren and Parsch 2007; Mank 2017; Pointer et al. 2013; Ranz et al. 2003). Depending on the sex in which they are predominantly expressed, sex-biased genes can be divided into male-biased or female-biased, while genes showing similar expression between the sexes being referred to as unbiased. Sex-biased genes are thought to evolve in response to conflicting sex-specific selection pressures over optimal expression acting on the shared genetic content (Connallon and Knowles 2005) and are increasingly used to study the footprint of sex-specific selection within the genome (Dean et al. 2017; Gossmann et al. 2014; Mank 2017). The study of sex-biased genes is thus essential for testing the association between differential gene expression and the evolution of sexually dimorphic phenotypes, as well as for understanding how different selective forces shape the transcriptional profile of sexually dimorphic genes.

30

1.3.1 The ontogeny of sex-biased transcription

Sex-biased gene expression is prevalent in many sexual species, including mammals (Yang et al. 2006), birds (Harrison et al. 2015; Mank et al. 2008; Naurin et al. 2011; Pointer et al. 2013; Uebbing et al. 2013; Vicoso et al. 2013b; Wright et al. 2019b), fish (Böhne et al. 2014; Sharma et al. 2014; Small et al. 2009; Wright et al. 2018), insects (Ometto et al. 2010; Ranz et al. 2003; Whittle and Extavour 2017), nematodes (Albritton et al. 2014; Cutter and Ward 2004), fungi (Whittle and Extavour 2017; Whittle and Johannesson 2013) as well as higher plants (Gossmann et al. 2014; Lipinska et al. 2015). Extensive comparative studies in both animal and plant species have revealed that sex-biased genes vary in abundance and strength of sex bias across different tissues and developmental stages (Grath and Parsch 2016; Mank et al. 2010a; Perry et al. 2014; Robinson et al. 2014; Zemp et al. 2016; Zluvova et al. 2010).

While the majority of genes in vegetative, non-reproductive, tissues of plants are unbiased (Sanderson et al. 2019; Zemp et al. 2016), in animals, somatic genes vary in their extent of differential expression (Mank et al. 2008; Yang et al. 2006). Studies in mammals, birds and fish, show that differential gene expression in brain tissue is relatively low (Catalán et al. 2012; Mank et al. 2008; Reinius et al. 2008; Santos et al. 2008; Yang et al. 2006), while the liver is one of the somatic tissues with the highest proportion of sex-biased genes (Yang et al. 2006). Another commonly reported pattern is that, consistently across taxa, reproductive organs present the highest transcriptional dimorphism of all tissues (Mank 2017; Pointer et al. 2013; Robinson et al. 2014; Yang 2016; Zemp et al.

2014; Zluvova et al. 2010). This finding often correlates with the onset of reproductive maturity and is consistent with the reproductive role of sex-biased genes and the pronounced sexual dimorphism of reproductive organs.

The majority of studies on sexual dimorphism and sex-biased gene expression focus on adult individuals (Ellegren and Parsch 2007), as the evolutionary optima of males and females mainly diverge, and hence genomic conflict and sex-specific selection are stronger, at reproductive maturity (Badyaev 2002; Chippindale et al. 2001; Cox and Calsbeek 2009; Gibson et al. 2002). By contrast, juvenile individuals generally lack clear phenotypic sexual dimorphism, and it has therefore been suggested that conflicting sex-specific selection primarily manifests in adulthood (Chippindale et al. 2001; Gibson et al. 2002; Rice and Chippindale 2001).

Although most sexually dimorphic traits do indeed manifest in adults, some studies have shown that juveniles also express sex-biased genes, with sexually antagonistic traits and sex-specific developmental pathways being initiated earlier in life (Badyaev 2002; Hale et al. 2011; Ingleby et al. 2015; Khila et al. 2012; Magnusson et al. 2011; Perry et al. 2014; Zhao et al. 2011). Consistent with the expansion of sexual dimorphism later in development, sex-biased expression has also been shown to amplify from pre-reproductive stages to adult stages (Mank et al. 2010a; Perry et al. 2014). Differential expression of some genes can also be limited to specific developmental stages. For example, a number of genes in the chicken gonad have a sex-biased expression only at embryonic stages of development, and just a small proportion of genes are consistently sex-biased across adult and juvenile gonads (Mank et al. 2010a). As

32

such, it is becoming increasingly clear that analysing the onset of sex-biased gene expression, as well as the differences between the sexes in gene expression patterns across development, is important for understanding the variation in sexually dimorphic phenotypes and the selective forces that shape them.

1.3.2 Molecular evolution of sex-biased genes

Molecular evolutionary analyses indicate that different selective pressures can impact the rate of sequence evolution of sex-biased genes, and analysing these patterns can offer insight into the relative strength of male- versus female-specific selection (Ellegren and Parsch 2007). Compared to unbiased genes, sex-biased genes in reproductive tissues tend to have accelerated rates of divergence for protein-coding sequence, estimated through the ratio of nonsynonymous to synonymous substitutions (d_N/d_S) (Dean et al. 2017; Lipinska et al. 2015; Mank et al. 2010a; Perry et al. 2014; Sharma et al. 2014). Both adaptive and non-adaptive evolutionary processes can contribute to the faster rate of evolution of sex-biased genes, and multiple molecular analyses can be employed to distinguish between them. Estimates of d_N/d_S calculated from alignments of coding sequence can differentiate between positive selection or relaxed evolutionary constraint (d_N/d_S > 1), neutral evolution (d_N/d_S = 1) and purifying selection (d_N/d_S < 1). Additionally, the McDonald-Kreitman test can be used to compare rates of between-species divergence and within-species polymorphism (McDonald and Kreitman 1991). Genes subject to adaptive evolution are expected to have an excess of nonsynonymous divergence to

polymorphism ratio ($D_N/D_S > P_N/P_S$) (Fay et al. 2001; McDonald and Kreitman 1991).

Sex-biased genes may be under strong adaptive positive selection, which leads to rapid substitution of amino acids in these genes (Grath and Parsch 2016). In particular, male-biased genes expressed in reproductive tissues tend to be more numerous and to have higher expression and divergence rates than female-biased and unbiased genes (Assis et al. 2012; Grath and Parsch 2016; Harrison et al. 2015; Khaitovich et al. 2005). Moreover, studies comparing sex-biased expression profiles across the gonads of different species have reported high rates of turnover, gains or losses of sex-biased expression, particularly in the case of male-biased genes (Böhne et al. 2014; Harrison et al. 2015; Zhang et al. 2007). This has often been interpreted as the signature of sexual selection, particularly via sperm competition (Ellegren and Parsch 2007). For example, a phylogenetic comparative analysis of bird species subject to varying levels of sexual selection found a positive correlation between male sexual ornamentation and the turnover of testis-biased expression (Harrison et al. 2015). There are however some exceptions to this pattern. A few studies have reported either no difference in rates of divergence between male-biased and female-biased genes (Metta et al. 2006), or even the opposite pattern, with elevated rates of evolution for female-biased genes (Mank et al. 2007; Whittle and Johannesson 2013).

Faster rates of evolution for male-biased genes could also result from higher rates of transcription in the male germline or constraint on female-biased genes due to pleiotropy (Assis et al. 2012; Zhang et al. 2007). Non-adaptive evolutionary processes can also drive fast rates of sequence evolution of sex-

34

biased genes in some systems (Gershoni and Pietrokovski 2014), perhaps related to relaxed purifying selection (Hunt et al. 2011). Relaxed selective constraint can increase the rate of divergence of sex-biased genes through the accumulation of slightly deleterious mutations (Grath and Parsch 2016). These examples showcase the complexity of sex-biased gene expression and lead to questions about the relative role of drift, natural and sexual selection pressures in the evolution of sex-biased genes.

1.3.3 The genomic distribution of sex-biased genes

Given the unequal inheritance mode of sex chromosomes, theory predicts that sex-biased genes should have a non-random chromosomal distribution, and empirical studies across different organisms indeed report this. Specifically, male-biased genes are generally underrepresented on the X chromosome, whereas female-biased genes are found in excess (Khil et al. 2004; Meisel et al. 2012; Parisi et al. 2003; Reinke et al. 2004; Sturgill et al. 2007). In species with ZW chromosomes the opposite pattern is observed (Arunkumar et al. 2009; Kaiser and Ellegren 2006; Storchova and Divina 2006), however this is not consistent across all studies (Mank et al. 2010a; Mořkovský et al. 2010).

Different evolutionary processes can mediate the non-random distribution of sex-biased genes. One such mechanism relates to the resolution of sexual antagonism (Rice 1984). In male heterogametic systems, for example, the X chromosome spends two-thirds of its time in females while only one-third of its time in males. As a result, dominant alleles that are beneficial to females but detrimental to males are favoured to be X-linked. Similarly, there is an advantage

for recessive male-beneficial alleles to become X-linked as they are hemizygous in males. The benefits of X-linked sexually antagonistic alleles will vary, however, with the degree of homomorphism and the extent of recombination suppression between the X and the Y chromosomes (Oliver and Parisi 2004). As such, sexual antagonism can play a role in the distribution of sex-biased genes, which are often thought to represent loci with resolved sexual antagonism, and may cause the X chromosome to become "demasculinized", defined as a deficit of male-biased genes.

X-chromosome demasculinization can also be caused by meiotic X chromosome inactivation during male germline, which promotes movement of male-biased genes off the X chromosome in order to ensure that they are still expressed (Emerson et al. 2004; Hense et al. 2007; Sturgill et al. 2007; Vinckenbosch et al. 2006). This process however does not explain the underrepresentation of X-linked male-biased genes in somatic tissues (Sturgill et al. 2007; Vicoso and Charlesworth 2006).

The deficit of male-biased genes on the X chromosome may also be the result of Y chromosome degeneration and dosage compensation (Bachtrog et al. 2010; Meisel et al. 2012; Vicoso and Charlesworth 2009). It is more difficult for genes on the sex chromosomes to achieve a male-biased expression compared to genes on the autosomes if they are already hyper-expressed as a result of dosage compensation. Conversely, X-linked genes that still show a male-biased expression in highly heteromorphic sex chromosome systems may reflect an incomplete, limited, dosage compensation mechanism (Bachtrog et al. 2010). However, this mechanism cannot account for the observed deficit of Z-linked

36

female-biased genes in species with incomplete sex chromosome dosage compensation (Arunkumar et al. 2009; Kaiser and Ellegren 2006; Storchova and Divina 2006).

1.4 Thesis outline

Here, I combine whole genome and transcriptome sequencing data across multiple species to explore the evolution of sex chromosomes and of sex-biased gene expression and their role in sexual dimorphism.

The work in **Chapter 2**, incorporates coverage and polymorphism data from males and females in the common guppy, *Poecilia reticulata*, to test for the role of sexual antagonism on the evolution of sex chromosomes. The guppy sex chromosome shows very low levels of divergence and no perceptible loss of gene expression on the Y chromosome. Moreover, intra-specific variation in the extent of the sex chromosome non-recombining region correlates with the strength of sexual conflict.

In **Chapter 3**, I build on the findings of Chapter 2 in order to characterise the structure and conservation of sex chromosome systems across related poeciliid species. I combine this with patterns of sex-specific SNPs to uncover a striking heterogeneity across these species in the rates of sex chromosome divergence and recombination suppression. I then explore the consequences of

recombination arrest on gene expression patterns to reveal the first case of chromosome-wide dosage compensation in teleost fish.

In **Chapter 4**, I use pedigree data from poeciliid species to trace SNP inheritance patterns in order to distinguish between autosomal and sex-linked genes. Using this analysis, I identify shared and species-specific Y-linked sequences and estimate divergence rates for sex-linked genes. This analysis reveals that the high degree of sex chromosome homomorphy is maintained over long evolutionary time by occasional recombination events.

In **Chapter 5**, I investigate the selective dynamics driving the expression and rate of sequence evolution of sex-biased genes. Importantly, I draw parallels between selective forces acting in animal and plant species and uncover contrasting patterns in the rates of evolution of sex-biased genes in dioecious flowering plants relative to established patterns in animals. These findings suggest a role of stronger haploid selection in driving the molecular evolution of plant sex-biased loci.

In **Chapter 6**, I discuss the study results in each chapter in a broader context, with an emphasis on the evolution of sex chromosomes and of sex-biased gene expression, as well as suggest future research directions.

1.5 Statement of contribution

Chapter 2 was published as: Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE (2017) Convergent recombination suppression suggests a role of sexual selection in guppy sex chromosome formation. *Nature Communications* 8: 14251

In this project, I identified candidate Y-linked genes, investigated allele-specific expression patterns and tested for degeneration of Y gene activity. I also participated with all the co-authors in writing the final draft.

Chapter 3 is currently in press: Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR Larkin DM, Breden F, Mank JE (2019) Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *PNAS* in press

In this chapter, I designed the project in consultation with my co-authors, prepared all the samples for sequencing, performed all the analyses and wrote the first draft. I worked with all the co-authors on writing the final submitted and revised version.

Chapter 4 is in preparation for submission: Darolti I, Wright AE, Mank JE. Poeciliid sex chromosome integrity maintained by incomplete recombination suppression (in prep).

I designed the project in consultation with my co-authors, set up the pedigrees for the families used in the study, prepared all the samples for RNA-sequencing, performed all the data analyses, and wrote the first draft of the manuscript. I am working with all co-authors to write the final submitted version.

Chapter 5 was published as: Darolti I, Wright AE, Pucholt P, Berlin S, Mank JE (2018). Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*. *Molecular Ecology* 27: 694-708

I designed the project in consultation with my co-authors, performed all the analyses, wrote the first draft of the paper, and worked with the co-authors to polish this into the submitted and final revised version.

1.6 Glossary

Dosage compensation – mechanisms of gene expression regulation through which the balance in gene expression levels between the sex chromosomes and the autosomes in the heterogametic sex and between the sex chromosomes of the two sexes is restored

Dosage sensitive genes – genes that have a negative fitness effect when experiencing a change in dose or copy number

Effective population size – number of individuals contributing genes to next generations

Evolutionary strata – non-recombining regions with distinct levels of divergence between the sex chromosomes resulting from independent recombination suppression events

Gametologs – X- and Y-linked (or Z- and W-linked) homologous genes

Gene conversion – a mechanism of recombination, taking place during the process of double-strand break repair in meiosis, through which one allele changes into its homologous allele

Gene dosage – the number of copies of a gene within a genome

Hemizygosity – the lack of paired chromosomes in a diploid cell (e.g. the males in species with XY systems have only one copy of each sex chromosome and are therefore hemyzigous for the sex chromosomes)

Heterogametic sex – the sex of a species that possesses two distinct sex chromosomes (i.e. males in species with XY sex chromosomes and females in species with ZW sex chromosomes)

Heteromorphic sex chromosomes – a sex chromosome system in which the two chromosomes are substantially differentiated and one of the chromosomes has undergone significant degeneration in comparison to the other (e.g. the mammalian XY systems and the bird ZW systems)

Homogametic sex – the sex of a species that possesses two identical sex chromosomes (i.e. females in species with XY sex chromosomes and males in species with ZW sex chromosomes)

Homomorphic sex chromosomes – a sex chromosome system in which the two chromosomes are largely undifferentiated or have undergone no divergence or gene loss

Intralocus sexual conflict – form of antagonism resulting from different reproductive optima of males and females at a single locus

***K*-mers** – sequences of length *k*

Masculinization – excess of male-biased genes

Orthologs – genes in different species that are homologous as a result of descending from the same ancestral gene

Sex-limited – genes or chromosomes are sex-limited if they are expressed or present in only one of the sexes

1.7 Abbreviations

ASE – allele-specific expression

BLAST – basic local alignment tool

CI – confidence interval

ENC – effective number of codons

FC – fold change

MDS – multi-dimensional scaling

MSY – male specific Y

ncRNA – non-coding RNA

PAR – pseudoautosomal region

RPKM – reads per kilobase of transcript per million mapped reads

SNP – single nucleotide polymorphism

Chapter 2. Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation

2.1 Abstract

Sex chromosomes evolve once recombination is halted between a homologous pair of chromosomes. The dominant model of sex chromosome evolution posits that recombination is suppressed between emerging X and Y chromosomes in order to resolve sexual conflict. Here we test this model using whole genome and transcriptome resequencing data in the guppy, a model for sexual selection with many Y-linked colour traits. We show that although the nascent Y chromosome encompasses nearly half of the linkage group, there has been no perceptible degradation of Y chromosome gene content or activity. Using replicate wild populations with differing levels of sexually antagonistic selection for colour, we also show that sexual selection leads to greater expansion of the non-recombining region and increased Y chromosome divergence. These results provide empirical support for longstanding models of sex chromosome catalysis, and suggest an important role for sexual selection and sexual conflict in genome evolution.

2.2 Introduction

Sex chromosomes are typically thought to evolve as recombination is halted between a homologous pair of chromosomes in one sex. Although we have a detailed understanding of the evolutionary consequences of the loss of recombination for sex chromosome evolution (Bachtrog 2013; Bachtrog et al. 2011), we still do not understand the evolutionary forces acting to halt recombination in the first place. The dominant theoretical model for the early stages of sex chromosome evolution (Bull 1983; Fisher 1931; Rice 1987) predicts that recombination will be selected against in the region between a sex determining gene and a nearby locus with alleles of sex-specific effect. This theory, though prevalent, remains largely untested empirically, as most research has focused on older, highly divergent sex chromosome systems (Bachtrog et al. 2014; Wright et al. 2016), for which it is difficult to extrapolate the earliest stages and causes of divergence.

The sex chromosomes of the guppy (*Poecilia reticulata*) have been of interest for more than a century, following early reports that many sexually selected colour traits are passed through the patriline on the Y chromosome (Lindholm and Breden 2002; Winge 1927). These observations were central to the development of theories regarding the role of sexual conflict in recombination suppression and sex chromosome divergence (Bull 1983; Fisher 1931; Rice 1987). Colour is sexually antagonistic in guppies, as brightly coloured males are more attractive to females and more visible to predators, but brightly coloured females gain no fitness advantage and only suffer increased predation (Endler

1980; Houde and Endler 1990; Kemp et al. 2009). Therefore, in this system, current models of sex chromosome evolution predict that recombination would be selected against between the sex determining locus and linked loci involved in colouration. This process would shrink the pseudoautosomal region in favour of expanding X- and Y-specific regions, creating a male supergene on the Y chromosome containing multiple colouration loci and thereby resolving sexually antagonistic selection.

Even though the guppy sex chromosomes are a classic model for the study of sexual conflict and sex chromosome divergence, little is actually known about the pattern of divergence between the X and Y chromosomes. Recent linkage maps identified male recombination events restricted to the middle of chromosome 12 (Tripathi et al. 2009a), suggesting that the other half of the chromosome is functionally X- or Y-linked. Immunostaining of recombination nodules (Lisachov et al. 2015) was broadly concordant with recombination mapping, again suggesting that the X chromosome is split roughly in equal parts between X-specific and pseudoautosomal.

Recombination shows substantial local variation between males and females throughout the genomes of many organisms (Lenormand 2003; Wright et al. 2016) and identifying areas of restricted male recombination does not distinguish the sex chromosome from other areas where males simply do not recombine. However, the Y is morphologically distinguishable from the X chromosome (Nanda et al. 2014), and comparative genome hybridization of lab populations (Traut and Winking 2001) suggest that roughly half of the Y chromosome is male-specific. Because many vertebrate sex chromosomes show

46

progressive spread of the non-recombining region (Skaletsky et al. 2003; Vicoso et al. 2013a; Vicoso et al. 2013b; Wright et al. 2014), the large size of the guppy non-recombining region and male-specific regions suggest substantial divergence between the X and Y.

Recombination suppression between the X and Y chromosomes results in complete linkage of the male-specific region of the Y. The loss of recombination in this region typically limits the role of adaptive evolution and leads to strong background selection and linkage effects, causing loss of functional polymorphism in coding sequence over time (Bachtrog 2013). Roughly half of male colouration patterns are thought to be Y-linked (Lindholm and Breden 2002), and the remarkable diversity of male colour combinations implies an improbably large number of Y haplotypes maintained within populations for a sex chromosome system of at least intermediate age. Additionally, if recombination suppression really is driven by sexually antagonistic alleles (Bull 1983; Fisher 1931; Rice 1987), then we might expect recent but rapid spread of recombination suppression shortly after the emergence of sexual preferences for colour. Although sexually selected traits exist in many poeciliids, the vivid male colouration in *P. reticulata* is only shared by a few very close relatives (Meredith et al. 2011; Pollux et al. 2014), therefore, the expansions of the male-limited Y chromosome to engulf colouration loci might have occurred very recently.

Moreover, the degree of male colouration, and, therefore, the degree of sexual conflict over colour, varies substantially based on predation pressures. Across watersheds, downstream populations are typically associated with higher predation and males are far less colourful than upstream populations (Endler

1980; Endler 1984, 1995). Importantly, the proportion of colour patterns thought to be Y-linked varies between upstream and downstream populations (Gordon et al. 2012). The unusual gene content of the guppy sex chromosomes, therefore, makes it a uniquely powerful system for testing the role of sexual conflict and sexual selection in sex chromosome divergence.

In order to determine the degree of divergence between the X and Y chromosome in this species, we re-sequenced male and female genomes and transcriptomes of both laboratory and wild individuals. We find that the X and Y show sequence differentiation over nearly one half of the length of the chromosome, however, the divergence between the X and Y chromosome is remarkably subtle, indicating very low levels of divergence and likely recent origin of the sex chromosome. The large region of divergence is in contrast to reports of other nascent sex chromosome systems where the diverged region is highly restricted (Kamiya et al. 2012; Liu et al. 2004; Russell and Pannell 2015). Despite this young age, we detect evidence of Faster-X evolution in this region. Most importantly, we find convergent patterns of greater sex chromosome divergence in upstream populations, which experience substantially elevated sexual selection and sexual conflict, compared with downstream populations. Our results suggest that recombination suppression between the X and Y spread quickly in the recent history of this sex chromosome system, possibly driven by the presence of sexually antagonistic alleles related to sexual selection.

48

2.3 Materials and Methods

2.3.1 Sample collection

All samples were collected in accordance with national and institutional ethical guidelines. First, we sampled males and females from a single large, outbred laboratory population established in 1998 (Kotrschal et al. 2013). Tail samples were homogenized and stored in RNA later before RNA preparation, the remainder of each fish was stored in ethanol before DNA preparation.

Second, wild males were caught from three watersheds (Yarra, Quare, Aripo) in the Northern Range Mountains of Trinidad in February 2015 (see Sandkam et al. (2015) for description of the habitats). From each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Samples were collected and stored immediately in ethanol prior to DNA preparation.

2.3.2 Sequencing

Nucleic acids were extracted with RNAeasy Kit (Qiagen) and DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer protocols. The libraries were prepared and barcoded at The Wellcome Trust Centre for Human Genetics, University of Oxford using standard protocols. RNA was sequenced on an Illumina HiSeq 2500 resulting in on average 32 million 100 bp paired-end reads per sample. DNA was sequenced on an Illumina HiSeq 4000, resulting in on average 269 million 100 bp paired-end reads per individual sampled from a single

large, outbred laboratory population, and 123 million 100 bp paired-end reads per sample for individuals caught in the wild in Trinidad (Table S.2.1).

2.3.3 Quality trimming and filtering

DNA data were quality assessed using FastQC v0.11.4 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and quality trimmed using Trimmomatic v0.35 (Lohse et al. 2012). We filtered reads containing adaptor sequences and trimmed reads if the sliding window average Phred score over four bases was < 15 or if the leading/trailing bases had a Phred score < 3. Reads were removed post filtering if either read pair was < 50 bases in length. RNA-seq data was quality assessed and trimmed using the same criteria but with a minimum length threshold of 36 bases (Table S.2.1).

2.3.4 *De novo* genome assembly

Reads used to construct *de novo* genome assemblies were error corrected with Quake v0.3.5, specifying default settings and a k-mer length of 19 (Kelley et al. 2010) (Table S.2.2). Optimal k-mer length for *de novo* genome assemblies was estimated using kmergenie v1.6741 (Chikhi and Medvedev 2014).

We constructed a female de novo genome assembly with DNA-seq reads from two females using SOAPdenovo v2.04 (Luo et al. 2012) and specifying the multi-k-mer option with a starting k-mer of 37 and max k-mer of 55. All reads were used during both contig and scaffold assembly. During scaffolding (SOAPdenovo

50

scaff), the –F parameter was set to specify that gaps in scaffolds should be filled. Lastly, GapCloser was used to close gaps emerging during the scaffolding process. Sequences<1 Kb in length were filtered from the assembly (Table S.2.3).

2.3.5 Assigning chromosomal position

Guppy genes were downloaded from RefSeq (Guppy_female_1.0+MT, RefSeq assembly accession: GCF_000633615.1) and the longest isoform picked for each. Coding sequences were BLASTed against the *de novo* genome assembly using BLASTn v2.3.0 (Altschul et al. 1990) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score.

De novo scaffolds were assigned to the guppy reference chromosomes and oriented using the chromosomal location and start position of mapped guppy genes. If multiple genes mapped to a given scaffold, the scaffold was assigned to the reference chromosome that the majority of genes were located on. Specifically, at least 70% of genes mapping to a given scaffold must be located on the same chromosome in the reference genome otherwise the scaffold was discarded. The degree of concordance in assigned chromosome position using this approach is high (Table S.2.4), and only 320 scaffolds from the female genome assembly were discarded due to discordance between chromosomal locations.

2.3.6 Genomic coverage analysis

Male and female trimmed DNA-seq reads were separately mapped to the *de novo* genome assembly using Bwa v0.7.12 aln/sampe with default settings (Li and Durbin 2009). Uniquely mapped reads were extracted using grep 'XT:A:U' and soap.coverage v2.7.9 (http://soap.genomics.org.cn) was used to extract coverage of scaffolds in every individual. For each scaffold, coverage was defined as the total number of times each site was sequenced divided by the number of sites that were sequenced.

For lab populations, average coverage values were calculated for females and males separately. We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Male:female coverage was calculated for each scaffold as $\log_2$ (average male coverage) – $\log_2$ (average female coverage).

For upstream and downstream wild populations, coverage was estimated using Bwa v0.7.15 aln/sampe and the same pipeline as the lab populations. Average coverage was calculated for each gene separately across each population. To account for differences in sequencing depth across populations, the $\log_2$ coverage for each gene was normalized by the median $\log_2$ coverage of X chromosome ($\log_2$ coverage – median $\log_2$ coverage of X chromosome). Male:female coverage was estimated for each population relative to the normalized coverage of the female lab population.

2.3.7 Polymorphism analysis

Male and female trimmed DNA-seq reads from both wild and lab populations were separately mapped to the *de novo* genome assembly using Bowtie1 v1.1.2 (Langmead et al. 2009), specifying a maximum insert size for paired-end alignment of 1,400 and writing hits in map format. Map files were sorted by scaffold and bow2pro v0.1 (http://guanine.evolbio.mpg.de/) was used to generate a profile for each sample. Sites with coverage < 10 were excluded from the analysis and SNPs were called when a site had a major allele frequency of 0.3 times the site coverage. SNPs were only included in further analyses if they were located within genic regions (see Expression analysis method for detail on gene annotation). Average SNP density for each gene was calculated as sum(SNPs)/sum(no. of filtered sites). We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Genes were excluded if zero sites remained after filtering.

For lab populations, average SNP density was calculated separately for males and females. Male:female SNP density was calculated for each gene as $\log_2$ (average male SNP density) − $\log_2$ (average female SNP density).

For upstream and downstream wild populations, average SNP density was calculated for each gene separately across each population. To account for differences in overall genetic diversity across populations, the $\log_2$ SNP density for each gene was normalized by the median $\log_2$ SNP density of X chromosome ($\log_2$ SNP density − median $\log_2$ SNP density of X chromosome). Male:female

SNP density was estimated for each population relative to the normalized SNP density of the female lab population.

To calculate the probability that the convergence in patterns of SNP density across populations we observe is due to chance, we randomly sampled 10 Mb windows across the autosomes 1,000 times. For each window, we tested whether the upstream normalized male:female SNP density was greater than the downstream population in each river using a one-tailed Wilcoxon ranked sum test. We looked for windows where all three rivers had p values < 0.05 and the median SNP density in the lab population was greater than the 95% autosomal confidence interval.

2.3.8 Expression analysis

Male and female trimmed RNA-seq reads were separately mapped to the *de novo* genome assembly using HISAT2 v2.0.4 (Kim et al. 2015), suppressing unpaired and discordant alignments for paired reads and excluding reads from the sam output that failed to align. Reported alignments were tailored for transcript assemblers including StringTie.

Sam files were coordinate sorted using SAMtools v1.2 (Li et al. 2009) and converted to bam files. StringTie v1.2.3 (Pertea et al. 2015) was used to quantify gene expression and annotate the *de novo* assembly.

Specifically, StringTie was run on each sample with default settings and the output GTF files were merged. The combined GTF file was filtered to remove non-coding RNA (ncRNA) and transcripts less than 50 bp in length. Specifically,

54

transcript sequences were extracted using BEDTools getfasta (Quinlan and Hall 2010) and BLASTed to *Oryzias latipes* (MEDAKA1), *Gasterosteus aculeatus* (BROADS1), *Poecilia formosa* (PoeFor_5.1.2) and *Danio rerio* (GRCz10) ncRNA downloaded from Ensembl 84 (Flicek et al. 2014). Transcripts with blast hits to ncRNA were removed from the GTF file. StringTie was rerun on each sample and expression was only estimated for genes defined in the filtered GTF file. A minimum expression threshold of 2FPKM in at least half of the individuals of either sex was imposed. This final filtered data set (23,603 genes) was used in subsequent expression and polymorphism analyses.

Expression was normalized using EdgeR (Robinson et al. 2010). Sam files were name sorted using SAMtools and HTSeq count v0.6.1 (Anders et al. 2015) used to extract read counts for each gene. Genes were excluded if they were not located on scaffolds assigned to the guppy reference genome. In all, 13,306 genes remained after filtering. Expression was normalized using TMM (trimmed mean of m-values) in EdgeR and RPKM estimated for each gene. Individuals cluster transcriptomically by sex (Fig. S.2.4). Average RPKM for each gene was calculated separately for males and females. We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Male:female expression was calculated for each gene as $\log_2$ (average male RPKM)$-\log_2$ (average female RPKM).

We tested whether there was an enrichment of GO terms in the X-Y diverged region compared with the rest of the genome. *Danio rerio* (GRCz10) coding sequences were downloaded from Ensembl 84 (Flicek et al. 2014) and the longest isoform extracted for each gene. Longest isoforms were extracted for

our set of expressed guppy genes and BLASTed to *D. rerio* using BLASTn v2.3.0

(Altschul et al. 1990) with an e-value cutoff of $10e^{-10}$ and minimum percentage

identity of 30%. When genes mapped to multiple locations, the top blast hit was

chosen using the highest BLAST score. *D. rerio* orthologues were identified for

genes in the X-Y degenerate region (15–25 Mb) and compared with the

remainder of the genome using Gorilla (Eden et al. 2007; Eden et al. 2009).

2.3.9 Allele-specific expression analysis

I used RNA-seq data from a subset of four male and four female

individuals in order to investigate allele-specific expression patterns and Y gene

activity decay. I first mapped RNA-seq reads to the genome assembly using

HISAT2 and then used SAMtools to sort and bam convert the output files. I called

SNPs using SAMtools mpileup v1.3 (Koboldt et al. 2014) and parsed and filter

the output file using custom Python script. I filtered genes based on coverage,

imposing a minimum coverage of 10 in all individuals and a major allele frequency

of 10%. I kept genes that were heterozygous in all males and homozygous in all

females and for which all males shared one allele while the other allele was

shared with the female individuals. I tested for Y degeneration by comparing

expression for X- and Y-linked alleles in males.

2.3.10 Cluster analysis of expression data

Transcriptional similarity of normalized count data for female and male individuals was assessed using a multi-dimensional scaling plot (MDS) with default settings in EdgeR (Robinson et al. 2010). RPKM data was clustered using the R package Pheatmap and boostrap values calculated using Pvclust. UPGMA was used in the hierarchical cluster analysis and the distance matrix was computed using the Euclidean method.

2.3.11 Calculating moving averages

Moving averages of coverage, polymorphism and expression were calculated in R (R Core Team 2015) based on sliding window analyses using the roll_mean function. Ninety-five per cent confidence intervals for the moving average were calculated by randomly resampling (1,000 times, without replacement) autosomal scaffolds (coverage analysis) or genes (SNP density and expression analyses).

2.3.12 Faster-X analysis

Guppy transcript sequences were extracted using BEDTools getfasta (Quinlan and Hall 2010) and the longest isoform chosen for each of the 23,603 genes. Genes on genomic scaffolds without chromosomal locations were removed, leaving 13,306 genic sequences for the Faster-X analysis. *Oryzias latipes* (MEDAKA1), *Xiphophorus maculatus* (Xipmac4.4.2), *Poecilia formosa*

(PoeFor_5.1.2) were downloaded from Ensembl 84 (Flicek et al. 2014) and the longest transcript for each gene was identified. We determined orthology using reciprocal BLASTn v2.3.0 (Altschul et al. 1990) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score. In all, 7,382 reciprocal 1-1 orthologues across the four species were identified. We obtained open reading frames and protein coding sequence with BLASTx v2.3.0 with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30% using the approach in (Wright et al. 2015). Reciprocal orthologues with no BLASTx hits or a valid protein-coding sequence were excluded.

Reciprocal orthologues were aligned with PRANK v.140603 (Löytynoja and Goldman 2008) using the codon model and specifying the following guidetree: (((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*), *Oryzias latipes*). SWAMP v 31-03-14 (Harrison et al. 2014) was used to mask erroneous sequences in the alignments. Reciprocal orthologues were discarded if the alignment length was < 300 bp after removing gaps and masked sites. After this length filter, 5,349 reciprocal orthologues remained.

We used the branch model (model=2, nssites=0) in the CODEML package in PAML v4.8 (Yang 2007) to obtain divergence estimates using the following phylogeny ((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*, *Oryzias latipes*). The branch model was used to calculate mean d_N/d_S across the *Poecilia reticulata* branch. As mutational saturation and double hits can lead

to inaccurate divergence estimates (Axelsson et al. 2008) orthogroups were excluded if $d_S > 2$.

Orthologues were divided into genomic categories on the basis of their chromosomal location. For each category, mean d_N and mean d_S were calculated as the sum of the number of substitutions across all orthologues divided by the number of sites (d_N = sum D_N / sum N, d_S = sum D_S / sum S, where D_N and D_S are estimates of the number of nonsynonymous or synonymous substitutions and N and S are the number of nonsynonymous/synonymous sites). This approach prevents disproportionate weighting of shorter genes by avoiding the problems of infinitely high d_N/d_S estimates arising from sequences with extremely low d_S (Mank et al. 2007; Mank et al. 2009; Wright et al. 2015).

Significant differences in d_N, d_S and d_N/d_S between genomic categories were determined using permutation tests with 1,000 replicates. One-tailed tests were used to test for the Faster-X effect where we predict d_N/d_S is greater for X-linked gene relative to the autosomes. Two-tailed tests were used to test for differences in d_N and d_S. Bootstrapping with 1,000 repetitions was used to generate 95% confidence intervals.

2.3.13 Phylogenetic history of guppy populations

Using DNA-seq data, we reconstructed the phylogenetic relationships between the six wild populations. We mapped trimmed reads to the previously

sequenced guppy genome (Guppy_female_1.0+MT, RefSeq assembly accession: GCF_000633615.1) using Stampy v1.0.28 (Lunter and Goodson 2011) with a substitution rate of 0.01. After mapping, sam files were converted to bam and coordinate sorted using SAMtools v1.2 (Li et al. 2009) and then deduplicated using Picard tools v1.136 (http://broadinstitute.github.io/picard). Subsequently, we added read groups and merged libraries belonging to the same individual using Picard. We then called variants on all 24 individuals simultaneously using two independent methods (GATK and Platypus), and retained only SNPs called reliably with both methods and passing quality control filters.

As part of the GATK variant calling pipeline, v3.4.46 (McKenna et al. 2010), we first realigned reads around indels and recalibrated base quality scores. We then proceeded with variant calling using the HaplotypeCaller and GenotypeGVCFs tools. The second method we employed to call variants was Platypus v0.8.1 (Rimmer et al. 2014), which we ran in assembly mode, restricting calling to reads mapping to the 23 canonical chromosomes (that is, excluding those mapped to unplaced scaffolds).

After variant calling we removed indels, intersected the GATK and Platypus SNP sets, and applied stringent quality filtering. We removed singleton SNPs, multiallelic SNPs and SNPs failing the following quality thresholds: quality by depth >2, coverage >0.5x and <2x mean coverage, >2 reads for the alternative allele, mapping quality >40, allele bias Z score for mapping quality, base quality or read position <-1.96, or strand bias Fisher exact test p >0.05. We also removed

SNPs with missing genotype in any individual. This yielded 4.6 million high-quality SNPs.

Next, we used R package adegenet v2.0.1 (Jombart and Ahmed 2011) to construct a Euclidian distance matrix for the 24 individuals based either on all SNPs across the genome or on only the 72,623 SNPs between 15 and 25 MB on the X chromosome. We used the R package ape v3.5 (Paradis et al. 2004) to produce from each matrix a simple neighbour joining tree to visualize the genetic distance between the six populations, and performed 100 bootstrap iterations to assess support for each node.

2.4 Results

2.4.1 The structure of the guppy sex chromosomes

We first assembled the female genome using SOAPdenovo2, based on 480 million paired end reads from an outbred laboratory population. The assembly yielded 96,611 scaffolds, with an N50 of 11.3Kb and total length of 634.8 Mb, after a minimum length threshold of 1 Kb (Tables S.2.1-3). Guppy genes from the reference genome (Guppy_female_1.0+MT) were mapped to scaffolds in order to identify chromosomal positions, resulting in a final assembly of 19,206 ordered scaffolds oriented along the guppy chromosomes, with an N50 of 17.4 Kb and total length of 219.5 Mb (Tables S.2.3-4).

We then mapped male and female DNA-seq reads to our ordered scaffolds in order to identify regions of coverage difference between the sexes. Regions with longstanding recombination suppression in males will show

reduced mapping efficiency against the female genome assembly, as diverged sequence from the Y will no longer map to the X chromosome (Vicoso and Bachtrog 2015; Vicoso et al. 2013a; Vicoso et al. 2013b). Even with strict mapping thresholds (see 2.3 Methods) we could identify no large region of the genome with reduced coverage in males, which we would expect if a large portion of the Y was significantly diverged or degraded (Fig. S.2.1), and the overall distribution of coverage is largely symmetrical (Fig. S.2.2A). However, previous linkage maps have identified chromosome 12, which contains the sex determining gene, as the sex chromosome (Tripathi et al. 2009a), and this chromosome shows a slight shift in the distribution and has a significantly greater proportion of scaffolds with female-biased coverage than autosomes (Wilcoxon rank sum test p <0.001, Fig. S.2.2B, Fig. S.2.5). This suggests that recombination suppression between the X and Y chromosomes has led to very slight divergence between them.

If the Y has diverged, but not yet degraded significantly, we would expect to observe Y-specific single nucleotide polymorphisms (SNPs) in regions that retain substantial sequence similarity to the X, resulting in higher average male heterozygosity for the sex chromosomes (Hough et al. 2014; Muyle et al. 2012). When assessing all regions of the genome, we observe a shoulder of elevated SNP density in males (Fig. S.2.2C), due to significantly greater SNP density in males for the sex chromosomes compared with autosomal genes (Wilcoxon rank sum test p < 0.001, Fig. S.2.2D, Fig. S.2.5). When sex differences in coverage and SNP density are plotted together, the sex chromosome is a clear outlier to

the other chromosomes (Fig. 2.1), confirming low but significant levels of divergence.
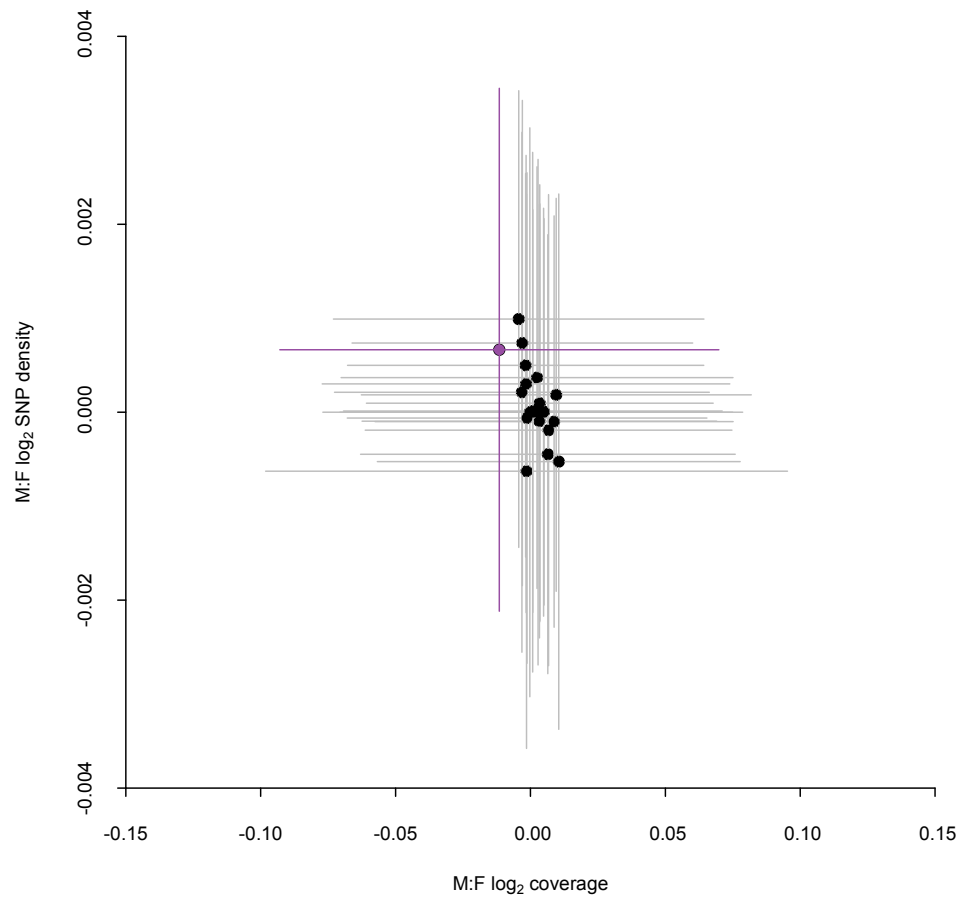
Fig. 2.1. Distribution of sex differences in coverage and SNP density for all chromosomes. The X chromosome is in purple. Horizontal and vertical lines denote interquartile ranges.

In order to determine the relative divergence between X and Y chromosomes, we plotted coverage and SNP density differences between males and female on our scaffolds against physical position on the guppy genome assembly. We detected significantly reduced male coverage outside the

autosomal 95% confidence interval from 22–25 Mb (Fig. 2.2). This region shows

the largest degree of X-Y sequence divergence and likely corresponds to the

oldest region of the sex chromosome (Stratum I). In contrast, between 15–25 Mb,

we detect significant elevation of male SNP density but no reduction in male

coverage, indicative of lower levels of X-Y divergence and suggesting that nearly

half of the sex chromosome has stopped recombining in males in the very recent

past (Stratum II, Fig. 2.3A, Fig. S.2.3).
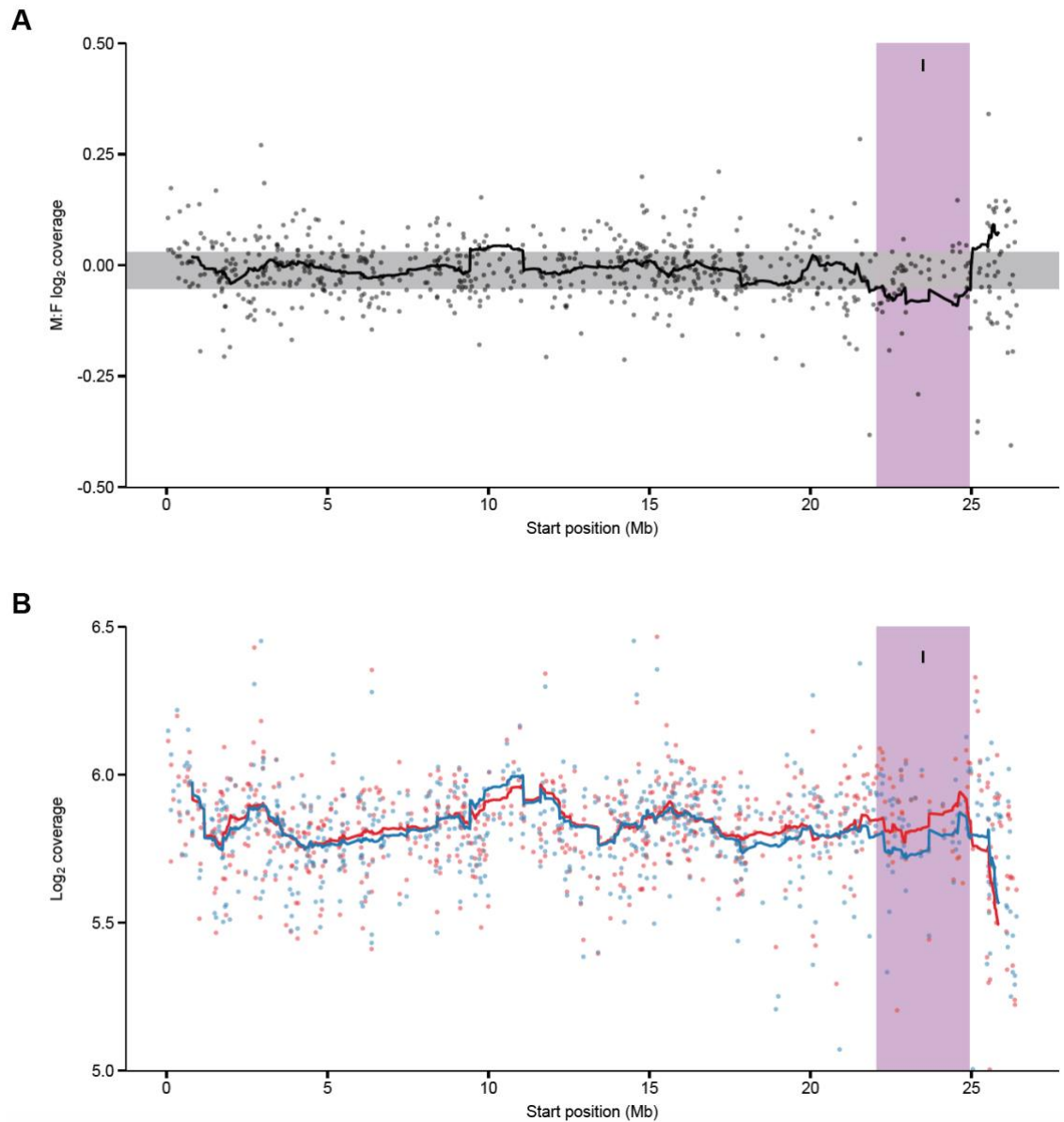
Fig. 2.2. Male and female coverage characteristics of guppy sex chromosome. (A) Moving average of coverage differences between male and female reads based on sliding window analysis (window size of 40 scaffolds). Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates in grey. (B) Male (blue) and female (red) coverage for the X chromosome. For both panels, dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22-25 Mb).

Our coverage and SNP analysis suggest that although male-specific SNPs have accumulated, the Y chromosome has not degenerated significantly. Because loss of gene activity often quickly follows loss of recombination on the Y chromosome (Bachtrog 2013), for each gene we plotted male and female expression level (RPKM) across the X chromosome. Our results show that the non-recombining region exhibits low levels of sexualization of gene content, with regions where the majority of genes exhibit female- or male-biased expression. However, there is no region of detectible loss of male gene activity, as would be expected with extensive Y chromosome decay (Fig. 2.3B, Fig. S.2.4). In contrast, in the region of the sex chromosome with the greatest coverage difference between males and females (Stratum I, 22–25 Mb), likely the area of greatest Y chromosome divergence, there is a slight excess of male-biased genes, indicating that this region of the Y chromosome has also not suffered any significant loss of gene activity. We tested for enrichment of GO terms for genes expressed in the X-Y diverged region (Strata I and II, 15–25 Mb) relative to the rest of the genome. However, there were no GO terms with an enrichment $p <$ 0.001.

Using SNP data in a subset of four males and four females, I identified 11 candidate Y-linked genes, all of which mapped to the non-recombining region of the sex chromosome. I tested for Y chromosome degeneration by comparing difference in expression between X- and Y-linked sequence in males. There was no significant reduction in Y/X expression between the non-recombining region and the PAR (p = 0.688, Wilcoxon rank sum test). I also found no significant difference when comparing male XY expression with male autosomal expression

66

(p = 0.225, Wilcoxon rank sum test) or with female XX expression (p = 0.819, Wilcoxon rank sum test). I confirmed that this result is not due to up-regulation of male X expression by also showing that male X expression is no different from female X expression (p = 0.833, Wilcoxon rank sum test) or male autosomal expression (p = 0.180, Wilcoxon rank sum test). Together, these results suggest lack of Y chromosome gene activity decay.

Fig. 2.3. Male and female SNP density and expression differences on guppy sex chromosome. (A) Moving average of male:female SNP density based on sliding window analysis (window size of 40 scaffolds). Ninety-five percent confidence intervals based on bootstrapping autosomal estimates are in grey. (B) Male (blue) and female (red) expression of genes along the X chromosome (window size of 40 genes). Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22-25 Mb) (see Fig. 2.2.), light purple indicates the region with less X-Y differentiation, where there is a significant excess of male SNPs (Stratum II, 15-22 Mb).

X chromosomes are predicted in many circumstances to show elevated rates of evolution (Charlesworth et al. 1987), and signatures of Faster-X evolution have been detected in old, heteromorphic sex chromosomes (Mank et al. 2010b; Meisel and Connallon 2013; Zhou and Bachtrog 2012). However, it is unclear whether a detectible signal of Fast-X would be expected in the early stages of sex chromosome evolution. We, therefore, compared rates of evolution for X-linked and autosomal coding sequence, and recovered a significant pattern of Faster-X in the guppy. X-linked d_N/d_S is greater though marginally non-significant for X-linked genes (86 genes, permutation test with 1,000 replicates, $p = 0.067$) relative to the autosomes (4,755 genes), due to a marginally significant increase in d_N (permutation test with 1000 replicates, $p = 0.014$) (Table S.2.6, Fig. S.2.5). This pattern is evident across both Strata I and II, indicating that low levels of sex chromosome divergence are sufficient to facilitate Faster-X processes.

2.4.2 Population variation in male colour and sexual conflict

Predation pressures vary substantially for natural guppy populations, with generally lower predation pressures upstream compared with downstream (Endler 1995). This has led to differences in female preference for male colouration (Houde and Endler 1990), with downstream males less vivid due to reduced female preferences and higher predation risks than upstream populations (Endler 1980; Endler 1984; Houde and Endler 1990). Upstream and downstream populations within watersheds are more closely related to each other than across watersheds (Fig. S.2.6A). Therefore, shifts in male colouration

have occurred independently in each watershed (Fraser et al. 2015), where downstream males are less colourful than upstream males.

Given the very recent origin of the guppy sex chromosomes, we might expect that if recombination suppression is indeed driven by sexual conflict over colour, there might be differences in the divergence of the sex chromosomes across different populations with more or less male colouration. In line with this prediction, there is evidence that different populations of wild guppies display different patterns of Y-linkage of colour traits (Gordon et al. 2015). We, therefore, examined patterns of sex-specific heterozygosity for upstream and downstream populations of wild guppies. We sampled three watersheds (Yarra, Quare, Aripo) and from each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Our results (Fig. 2.4, Table S.2.7) show that across replicate upstream populations, where males are more colourful, there is significantly greater divergence between the X and Y chromosomes than the ancestral downstream populations (Wilcoxon rank sum test between upstream and downstream populations across watersheds, Yarra p = 0.011, Quare p = 0.046, Aripo p = 0.017). Expansion of the non-recombining region and corresponding X-Y divergence has occurred repeatedly and independently across populations, as the phylogeny of these populations reveals that in each watershed, upstream populations are consistently derived from downstream populations (Fig. S.2.6B). By randomly sampling 10 Mb windows with 1,000 repetitions across the autosomes, we find that the probability of observing this convergence in SNP density across populations by chance is p

70

<0.004. In contrast, there are no differences in patterns of coverage between upstream and downstream populations in the area of greatest sex chromosome divergence (Stratum I, 22–25 Mb, Fig. S.2.7, Table S.2.8), indicating that X-Y divergence in this region predates the divergence of these wild population.
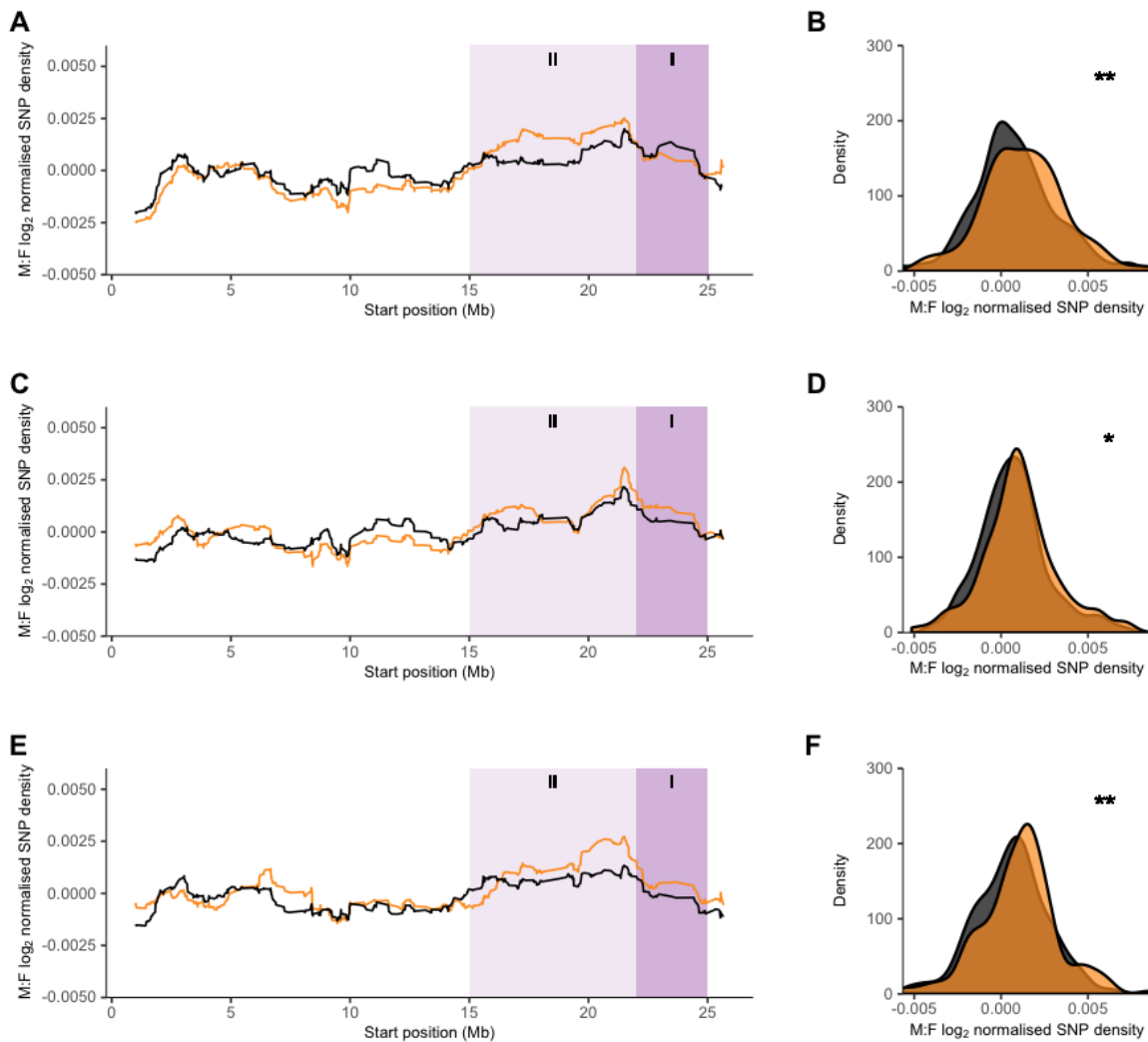
Fig. 2.4. Male:female SNP density for the X chromosome across upstream (orange) and downstream (black) guppy populations. (A, C and E) Moving average of normalised SNP density across the X chromosome based on sliding window analysis (window size of 40 genes) for Yarra (A), Quare (C) and Aripo (E) watersheds. Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in laboratory population males (Stratum I, 22-25 Mb) (see Fig. 2.2), light purple indicates the region with less X-Y differentiation, where there is a significant excess of male SNPs in laboratory populations (Stratum II, 15-22 Mb) (see Fig 2.3). (B, D and F) Distribution of sex differences in normalised SNP density for the X-Y diverged region (Strata I and II, 15-25 Mb) for Yarra (B), Quare (D) and Aripo (F) watersheds. [**]p value < 0.020, [*]p value < 0.050 based on permutation tests.

2.5 Discussion

Observations of Y-linkage for a large proportion of male colour patterns in guppies (Lindholm and Breden 2002; Winge 1927) helped form the nucleus of theories regarding the role of sexual conflict in sex chromosome formation (Bull 1983; Charlesworth et al. 2005; Rice 1987). Here we used individuals from natural and laboratory populations in conjunction with analysis of coverage, SNP and expression differences between males and females in this model system to test the role of sexual conflict in recombination suppression between the X and Y chromosomes. Our results suggest two regions of divergence on the sex chromosome. One region, likely the area of greatest Y chromosome divergence, is manifest with slightly reduced DNA coverage in males in a restricted region spanning 22–25 Mb. A larger region of more recent recombination suppression from 15–22 Mb is distinguishable only through the build-up of Y-specific SNPs. In both regions, although male-specific SNPs have accumulated on the Y, there is no evidence of large-scale decay of the Y chromosome or loss of gene activity observed in many older sex chromosome systems (Vicoso and Bachtrog 2015). Surprisingly, this region of divergence extends over nearly half of the sex chromosome, indicating that recombination has been suppressed over a large region very recently. The two strata we observe in guppies are consistent with step-wise patterns of sex chromosome formation observed in many other organisms, including mammals (Skaletsky et al. 2003), birds (Wright et al. 2014), *Silene* (Bergero et al. 2007), sticklebacks (White et al. 2015) and *Nothobranchius* (Reichwald et al. 2015). In the latter case, the authors

observed population-level variation in the youngest stratum, similar to what we observe in guppies, suggesting that strata can form independently within species.

Comparisons of coverage and SNP density between males and females, like the analyses we implement here, offer two complementary views of sex chromosome evolution. Coverage differences are expected in more diverged regions with significant Y chromosome degeneration. In contrast, sex-differences in SNP density, particularly in regions with elevated SNP density in the heterogametic sex, are expected in more diverged systems with little Y chromosome degeneration. However, implementing these approaches in young sex chromosome systems should be accompanied by information as to the location of the sex determining region, which has been previously mapped to the far end of chromosome 12 (Tripathi et al. 2009a). Ideally, Y-specific sequence data would be useful in verifying and dating stratum boundaries. However, this is complicated in our system due to the lack of complete lineage sorting of Y-specific SNPs, precluding the reconstruction of Y-specific sequence from our short-read data. In future work, long read RNA-seq data, optical mapping and other phasing approaches will be useful in confirming stratum boundaries and identifying Y-linked sequences. These data will also be important in determining whether inversions, which are often assumed to be involved in recombination suppression, are indeed the mechanism behind sex chromosome divergence.

Despite the limited sequence divergence between the X and Y chromosomes, we observe two evolutionary signatures that are typically only associated with heteromorphic sex chromosome systems. First, the X chromosome shows the early stages of sexualization for gene expression despite

74

limited evidence for degeneration in gene activity or content across the non-recombining Y chromosome. Previous evidence of sexualization comes from old, highly heteromorphic sex chromosome systems (Arunkumar et al. 2009; Meisel et al. 2012; Wright et al. 2012) and it was previously unclear how quickly sex-biased expression can accumulate after sex chromosome formation. Our results, therefore, indicate that sexualization of the X chromosome can occur very quickly after recombination is halted. Second, we detect a Faster-X effect, where X-linked coding sequence diverges more rapidly than the remainder of the genome. Until now, evidence for Faster-X was restricted to highly diverged sex chromosomes (Mank et al. 2010b; Meisel and Connallon 2013; Zhou and Bachtrog 2012), however, our results suggest that the Faster-X processes can accumulate rapidly following the loss of recombination. These findings have important consequences for the role of sex chromosomes in Haldane's rule (Haldane and Haldane 1922) and the Large-X effect in speciation (Masly and Presgraves 2007), and suggests that young or undifferentiated sex chromosomes may act as an important driver in the evolution of reproductive isolation (Dufresnes et al. 2016).

Most systems where sex chromosomes have formed recently (Kamiya et al. 2012; Liu et al. 2004; Russell and Pannell 2015), and even some older sex chromosome systems (Stock et al. 2011; Vicoso et al. 2013b), show restricted recombination in only a small region. The region of divergence extends over almost half of the sex chromosomes in the guppy, suggesting that recombination has been suppressed very quickly over a large region of the Y chromosome in guppies. This rapid spread of recombination suppression may have been driven

by the presence of sexually antagonistic alleles related to male colour on the proto-sex chromosome (Endler 1980; Houde and Endler 1990; Kemp et al. 2009). The high proportion of Y-linked colour patterns in guppies (Lindholm and Breden 2002; Winge 1927) is likely the product of rapid spread of recombination suppression between the X and Y chromosomes, which would resolve sexually antagonism by limiting colour expression to males.

Fish show remarkable variation in sex determination (Bachtrog et al. 2014; Mank et al. 2006) and rapid origin and turnover of sex chromosomes (Bachtrog et al. 2014; Devlin and Nagahama 2002). The tiger pufferfish has homomorphic sex chromosomes, where the sexes differ by only a single missense SNP (Kamiya et al. 2012), whereas a significant proportion of the sex chromosomes in sticklebacks are non-recombining (Kitano et al. 2009; White et al. 2015), and there has been substantial decay of gene activity on the Y chromosome. Although studies in related species are required to date the exact age of the sex chromosomes in *P. reticulata*, there is extensive sex chromosome turnover in the poeciliid clade (Devlin and Nagahama 2002; Mank et al. 2006), suggesting a recent origin of the sex chromosomes described here. This is consistent with expectations that the expansion of the Y-limited region was driven by sexual conflict over colouration, suggesting that Stratum II originated around the same period that male colouration emerged as a major component of female preference, likely <5Mya (Meredith et al. 2011; Pollux et al. 2014).

Our results suggest that this younger region of recombination suppression has expanded convergently in upstream populations as a consequence of increased sexual selection and sexual conflict over colouration (Endler 1984;

76

Houde and Endler 1990). We found the same convergent pattern of X and Y divergence between colourful upstream populations compared with the duller ancestral downstream populations over each of three replicate watersheds (Fig. 2.4). Upstream populations all showed greater divergence between the X and Y based on SNP density, and the region of significant SNP divergence extends over a larger region of the sex chromosomes. This accelerated divergence in upstream populations in each of the three watersheds has likely occurred independently, as populations within watersheds are well known to be monophyletic (Fraser et al. 2015). In support of this, our phylogenetic reconstruction reveals that in each watershed, upstream populations independently evolved from ancestral downstream populations. This suggests that sexual selection and sexual conflict over colour has driven greater Y divergence, consistent with longstanding theoretical predictions about the role of sexual antagonism in sex chromosome formation (Bull 1983; Charlesworth et al. 2005; Rice 1987). However, it is worth noting that our replicate upstream populations show some variation in the degree of differentiation, possibly due to demographic factors such as bottlenecks and recent expansions, date of colonization, rate of dispersal and gene flow between upstream and downstream populations, and effective population size, as well as stochastic processes.

Altogether, our results suggest that sexual conflict may be responsible for the remarkably rapid recent spread of recombination suppression to encompass colouration alleles within the Y chromosome. Moreover, our data are consistent with a role of sexual selection in accelerating divergence of the Y chromosome once recombination suppression is established.

Chapter 3. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers

3.1 Abstract

Once recombination is halted between the X and Y chromosome, sex chromosomes begin to differentiate and transition to heteromorphism. While there is a remarkable variation across clades in the degree of sex chromosome divergence, far less is known about the variation in sex chromosome differentiation within clades. Here, we combined whole genome and transcriptome sequencing data to characterise the structure and conservation of sex chromosome systems across Poeciliidae, the livebearing clade that includes guppies. We found that the *Poecilia reticulata* XY system is much older than previously thought, being shared not only with its sister species, *Poecilia wingei*, but also with *Poecilia picta*, which diverged roughly 20 mya. Despite the shared ancestry, we uncovered an extreme heterogeneity across these species in the proportion of the sex chromosome with suppressed recombination, and the degree of Y chromosome decay. The sex chromosomes in *P. reticulata* and *P. wingei* are largely homomorphic, with recombination in the former persisting over a substantial fraction. However, the sex chromosomes in *P. picta* are completely non-recombining and strikingly heteromorphic. Remarkably, the profound degradation of the ancestral Y chromosome in *P. picta* is counterbalanced by the evolution of functional chromosome-wide dosage compensation in this species,

the first such documented case in teleost fish. Our results offer important insight

into the initial stages of sex chromosome evolution and dosage compensation.

3.2 Introduction

Sex chromosome evolution is characterised by remarkable variation across lineages in the degree of divergence between the X and Y chromosomes (Bachtrog et al. 2011; Bachtrog et al. 2014). Derived from a pair of homologous autosomes, sex chromosomes begin to differentiate as recombination between them is supressed in the heterogametic sex over the region spanning a newly acquired sex-determining locus (Ohno 1967). The lack of recombination exposes the sex-limited Y chromosome to a range of degenerative processes that cause it to diverge in structure and function from the corresponding X chromosome, which still recombines in females (Bachtrog 2013; Charlesworth and Charlesworth 2000). Consequently, the sex chromosomes are expected to eventually transition from a homomorphic to a heteromorphic structure, supported by evidence from many of the old and highly differentiated systems found in mammals (Lahn and Page 1999; Skaletsky et al. 2003), birds (Wright et al. 2014), *Drosophila* (Bachtrog 2013) and snakes (Matsubara et al. 2006).

However, there is a significant heterogeneity among clades, and even among species with shared sex chromosome systems, in the spread of the non-recombining region, and the subsequent degree of sex chromosome divergence (Fujito et al. 2015; Sessions et al. 2016; Vicoso et al. 2013a). Age does not always reliably correlate with the extent of recombination suppression, as in some species the sex chromosomes maintain a largely homomorphic structure over long evolutionary periods (Ahmed et al. 2014; Stock et al. 2011; Vicoso et al. 2013a; Vicoso et al. 2013b; Xu et al. 2018), while in others the two sex

chromosomes are relatively young, yet profoundly distinct (Bergero et al. 2007). Comparing the structure and recombination patterns of sex chromosomes between closely related species is a powerful method to determine the forces shaping sex chromosome evolution over time.

Sex chromosome divergence can also lead to differences in X chromosome gene dose between males and females. Following recombination suppression, the Y chromosome undergoes gradual degradation of gene activity and content, leading to reduced gene dose in males (Charlesworth 1978; Charlesworth and Charlesworth 2000; Mank 2009). Genetic pathways that incorporate both autosomal and sex-linked genes are primarily affected by such imbalances in gene dose, with potential severe phenotypic consequences for the heterogametic sex (Malone et al. 2012). In some species, this process has led to the evolution of chromosome-level mechanisms to compensate for the difference in gene dose (Mank et al. 2011; Mullon et al. 2015). However, the majority of sex chromosome systems are associated with gene-by-gene level mechanisms, whereby dosage sensitive genes are compensated, but overall expression of the X chromosome is lower in males compared to females (Mank 2009; Mank 2013; Mullon et al. 2015).

As opposed to most mammals and birds, the sex chromosomes in many fish, lizard and amphibian species are characterised by a lack of heteromorphism, which has usually been attributed to processes such as sex chromosome turnover and sex reversal (Dufresnes et al. 2015; Ezaz et al. 2009; Ross et al. 2009; Sessions and Kezer 2012; Stock et al. 2011; Stock et al. 2013; Volff and Schartl 2001). As a result, closely related species from these taxonomic groups

often have a variety of sex chromosome systems found at different stages in evolution (Cnaani et al. 2008; Hillis and Green 1990; Nielsen et al. 2018; Ross et al. 2009). Alternatively, undifferentiated sex chromosomes in anolis lizards, for example, have been found to be the result of long-term conservation of a homomorphic ancestral system (Gamble et al. 2014). Additionally, global dosage compensation has not yet been found in fish, perhaps due to the transient nature of the sex chromosome systems and the general lack of heteromorphism in the group. However, incomplete dosage compensation, through a gene-by-gene regulation mechanism, may have evolved in sticklebacks (Leder et al. 2010; White et al. 2015), flatfish (Chen et al. 2014) and rainbow trout (Hale et al. 2018).

Poeciliid species have been the focus of many studies concerning sex determination (Volff and Schartl 2001). Moreover, many poeciliids exhibit sexual dimorphism, with some colour patterns and fin shapes controlled by sex-linked loci (Gordon et al. 2012; Lindholm and Breden 2002; Lindholm et al. 2004; Tripathi et al. 2009b; Winge 1927). The clade also has a diversity of genetic sex determination systems, with both male and female heterogametic sex chromosomes observed in different species (Schultheis et al. 2009; Traut and Winking 2001). Most work on poeciliid sex chromosome structure has focused on the *Poecilia reticulata* XY system, positioned on chromosome 12 (Tripathi et al. 2009a), which shows very low levels of divergence (Tripathi et al. 2009b; Wright et al. 2017). Although recombination is suppressed over almost half the length of the *P. reticulata* sex chromosome, there is little sequence differentiation between the X and Y chromosomes and no perceptible loss of Y-linked gene

activity in males (Wright et al. 2017). This low level of divergence suggests a recent origin of the sex chromosome system.

There is intra-specific variation in the extent of the non-recombining region within *P. reticulata*, correlated with the strength of sexual conflict (Wright et al. 2017). Additionally, although *P. reticulata* and its sister species, *P. wingei*, are thought to share an ancestral sex chromosome system (Morris et al. 2018; Nanda et al. 2014), there is some evidence for variation in Y chromosome divergence between these species (Morris et al. 2018). It is unclear whether the XY chromosomes maintain the same level of heteromorphism in other poeciliids (Nanda et al. 2014; Schultheis et al. 2009), or even whether they are homologous, to the sex chromosomes in *P. reticulata*.

Here I perform comparative genome and transcriptome analyses on multiple poeciliid species to test for conservation and turnover of sex chromosome systems and investigate patterns of sex chromosome differentiation in the clade. I find the XY system in *P. reticulata* to be older than previously thought, being shared with both *P. wingei* and *P. picta* and thus dating back to at least 20 mya. Despite the shared ancestry, I uncover an extreme heterogeneity across these species in the size of the non-recombining region, the sex chromosomes being largely homomorphic in *P. reticulata* and *P. wingei* while completely non-recombining and highly diverged across the entire chromosome in *P. picta*. Remarkably, although the Y chromosome in *P. picta* shows signatures of profound sequence degeneration, I observe equal expression of X-linked genes in males and females, which I find to be the result of dosage compensation

acting in this species. This is the first instance of chromosome-wide sex chromosome dosage compensation reported in a fish.

3.3 Materials and Methods

3.3.1 Sample collection and sequencing

We collected adult male and female individuals from four guppy species (*Poecilia wingei* from our laboratory population, *Poecilia picta* from Guyana, *Poecilia latipinna* and *Gambusia holbrooki* from Florida, USA). We chose these samples in order to obtain an even phylogenetic distribution. The species we assessed exhibit clear somatic dimorphisms, including colouration and size, in addition to gonadal differences. Most notably, females possess an enlarged abdomen and anal fin. In males, the anal fin is modified to form a gonopodium (i.e. and intromittent organ), which is clearly visible. Phenotypic sex was determined at the time of collection based on these measures. There were no intermediate or ambiguous individuals collected and sex was clearly visible and concordant across both somatic and phenotypic traits in all samples.

All samples were collected in accordance within ethical guidelines. *P. latipinna* and *G. holbrooki* were collected under Florida permit FNW17-10 and St. Mark's Refuge permit FF04RFSM00-17-09. *P. picta* was collected under permit from the Environmental Protection Agency of Guyana (Permit 120616 SP: 015). *P. wingei* was collected from our lab population, a colony of a strain maintained by a UK fish fancier.

From each species, we immediately stored head and tail samples from three males and three females in ethanol and, RNAlater, respectively. We extracted DNA from heads with the DNeasy Blood and Tissue Kit (Qiagen) and RNA from tails with the RNeasy Kit (Qiagen), following the manufacturer's instructions. Library preparation and sequencing were performed at The Wellcome Trust Centre for Human Genetics, University of Oxford, following standard protocols and using the Illumina HiSeq 4000 platform. Genomic DNA was used to construct paired-end (PE) sequencing libraries with short insert sizes (average insert size 500bp) and mate-pair (MP) libraries with long insert sizes (average insert size 2kb) for each individual. The Nextera Mate Pair Sample Preparation Kit was used for preparing mate-pair libraries. Using FastQC v0.11.4 (www.bioinformatics.babraham.ac.uk/projects/fastqc/), I assessed data quality and used Trimmomatic v0.36 (Lohse et al. 2012) to trim reads. For both DNA-seq and RNA-seq reads I removed adaptor sequences, regions of low Phred score (reads with average Phred score < 15 in sliding windows of four bases and reads with leading/trailing bases with a Phred score < 3) and short reads (if either read in a pair was shorter than 50bp).

3.3.2 Genome assembly

I first corrected the reads using Quake v0.3.5 (Kelley et al. 2010) and estimated the optimal assembly *k*-mer length using KmerGenie v1.6741 (Chikhi and Medvedev 2014). I then used SOAPdenovo v2.04 (Luo et al. 2012) to construct female *de novo* genome assemblies for *P. wingei*, *P. picta* and *G. holbrooki* and a male assembly for *P. latipinna*, using both the paired-end and

mate-pair reads (Table S2). The paired-end reads were used for both the contig and scaffolding steps of the assembly process, while the mate-pair reads were only used for scaffolding. Additionally, I used the SOAPdenovo GapCloser module to close the gaps resulting from the assembly scaffolding step. Finally, I removed sequences shorter than 1kb from the assemblies.

To improve assembly contiguity and reconstruct chromosomal fragments for each species, I followed the UCSC chains and nets pipeline from the kentUtils software suite (Kent et al. 2003) before employing the Reference-Assisted Chromosome Assembly (RACA) algorithm (Kim et al. 2013). The chains and nets pipeline is designed for building pairwise nucleotide alignments and bridging gaps between pairwise syntenic blocks to construct larger structures (Kent et al. 2003). A chain alignment represents an ordered pairwise sequence alignment between two species. A net alignment represents a collection of chains within a genome region, ordered in a hierarchical manner based on synteny scoring. The RACA algorithm incorporates the pairwise alignment files, together with read mapping information to identify syntenic fragments (regions which maintain sequence similarity and order) across the species used. RACA then estimates adjacency between syntenic fragments in each target genome to reconstruct predicted chromosome fragments (PCFs) for each target species (Kim et al. 2013).

First, for each species, I carried out DNA-seq read mappings to the *de novo* assemblies using Bowtie2 v2.3.3.1 (Langmead et al. 2009), reporting concordant mappings only (--no-discordant option) and using the appropriate mate orientations according to the insert size of the libraries (--fr option for short-insert libraries and –rf option for long-insert libraries). The resulting alignments

were converted into the RACA-specific input format (script available on the RACA website http://bioen-compbio.bioen.illinois.edu/RACA/).

I also obtained pairwise alignments using LASTZ (www.bx.psu.edu/%7Ersharris/lastz/; parameters C=0, E=30, H=2000, K=3000, L=3000, O=400, M=50) between a reference species (here I used the *X. hellerii* genome, obtained from NCBI GenBank Xiphophorus_hellerii-4.0, assembly accession GCA_003331165.1), the target species and an outgroup species (here I used the Medaka, *Oryzias latipes*, genome, obtained from GenBank ASM223467v1, assembly accession GCA_002234675.1). I then converted these alignments into chains and nets formats following the UCSC axtChain (-minScore=1000, -linearGap=medium), chainAntiRepeat, chainSort, chainPreNet and netSyntenic tools (Kent et al. 2003). The syntenic chains and nets fragments, together with the paired-end alignments, were used as input files for RACA (resolution=10000 for *P. picta* and *P. latipinna* and resolution=1000 for *P. wingei* and *G. holbrooki*). For each target species, RACA ordered and oriented target scaffolds into PCFs (Table S2), and I used this positional information of scaffolds in the genome for all further analyses.

3.3.3 Analysis of genomic coverage

For each species, using BWA v0.7.12 (Li and Durbin 2009), I mapped male and female paired-end DNA-seq reads to the *de novo* scaffolds with positional annotation from RACA, following the aln and sampe alignment steps, and extracted uniquely mapping reads. I then used soap.coverage v2.7.9 (http://soap.genomics.org.cn/) to calculate the coverage (number of times each

site was sequenced divided by the total number of sequenced sites) of each scaffold in each sample. For each scaffold, I calculated the male to female (M:F) fold change coverage as $\log_2$(average male coverage)–$\log_2$(average female coverage).

3.3.4 SNP density analysis

For each species, using Bowtie1 v1.1.2 (Langmead et al. 2009), I mapped male and female paired-end DNA-seq reads to the *de novo* scaffolds with positional annotation from RACA, generating map format output files. I sorted the map files by scaffold and converted them into profiles, which represent counts for each of the four nucleotide bases, for each individual using bow2pro v0.1 (http://guanine.evolbio.mpg.de/). For each site, I applied a minimum coverage threshold of 10 and called SNPs as sites with a major allele frequency of 0.3x the total site coverage. I obtained gene information through the expression analysis detailed below and for each gene I calculated the average SNP density as the number of SNPs divided by the number of filtered sites. I excluded SNPs outside of genic regions. For each gene I then calculated M:F fold change SNP density as $\log_2$(average male SNP density) – $\log_2$(average female SNP density).

3.3.5 Detecting sex chromosome non-recombining regions and strata of divergence

I used the fold change coverage and SNP density estimates to distinguish regions that are homologous and recombining between the sex chromosomes

88

from regions that show full or even partial sex chromosome divergence, and which are hence non-recombining. For each species, I generated 95% confidence intervals based on bootstrapping autosomal M:F coverage ratios and autosomal M:F SNP density ratios separately. For XY systems, I defined non-recombining, older strata of divergence as regions with a significant decrease in M:F coverage ratio outside the 95% confidence interval. In addition, I defined younger strata of divergence as regions with no reduction in male coverage but with a significant increase in M:F SNP density ratio outside the 95% confidence interval. Conversely, for ZW systems, a significant increase in M:F coverage ratio and a significant decrease in M:F SNP density ratio are expected for older and, respectively, younger regions of divergence.

3.3.6 Gene expression analysis

For each species, using HISAT2 v2.0.4 (Kim et al. 2015), I mapped male and female RNA-seq reads to scaffolds with positional annotation from RACA, reporting paired (--no-mixed) and concordant (--no-discordant) mappings only and tailoring the alignments for downstream transcript assembly (--dta). I used SAMtools to sort by coordinate and bam convert the sam output files. For each sample, I then used StringTie (Pertea et al. 2015) to obtain transcripts in a GTF file format, which I then merged to assemble a non-redundant set of transcripts for each species. Before further analyses, I filtered the merged GTF file for non-coding RNA (ncRNA) by using BEDTools getfasta (Quinlan and Hall 2010), extracted target transcript sequences and removed transcripts with BLAST hit to ncRNA sequences from *Poecilia formosa* (PoeFor_5.1.2), *Oryzias latipes*

(MEDAKA1), *Gasterosteus aculeatus* (BROADS1), and *Danio rerio* (GRCz10), obtained from Ensembl 84 (Flicek et al. 2014).

For each species, I estimated gene expression by extracting read counts for each gene using HTSeq-count (Anders et al. 2015) and the ncRNA filtered transcriptome. I only kept genes that were placed on scaffolds with positional information on PCFs. For these genes, I converted read counts to RPKM values with EdgeR (Robinson et al. 2010), normalised with TMM, and applied a minimum expression threshold of 2RPKM in half or more of the individuals in one sex. For each gene I then calculated M:F fold change expression as $\log_2$(average male expression)–$\log_2$(average female expression).

I identified sex-biased genes in EdgeR using a minimum of two-fold differential expression ($\log_2$ M:F RPKM >1 for male-biased genes and <-1 for female-biased genes) and a significant *p* value (p_{adj} <0.05 based on FDR correction for multiple testing, Benjamini and Hochberg (1995)).

I tested for an enrichment of GO terms in the non-recombining regions of the sex chromosomes relative to the rest of the genome in each species. I first extracted the longest isoform for each gene from the *Danio rerio* (GRCz10) coding sequences from Ensembl 84 (Flicek et al. 2014). I then BLASted longest isoforms from each of our target gene sets to the *D. rerio* sequences with BLASTn v2.3.0 (Altschul et al. 1990), using an e-value cutoff of $10e^{-10}$ and a minimum percentage identity of 30%. For genes with multiple alignment hits, I chose the top blast hit based on the highest BLAST score. I then compared *D. rerio* orthologues for genes in the non-recombining regions with those for genes in the rest of the genome using GOrilla (Eden et al. 2009).

3.3.7 *k*-mer analysis

In order to identify shared Y sequence across *P. reticulata*, *P. wingei* and *P. picta*, we followed a *k*-mer analysis method previously described in Morris et al. (2018). We have previously used this approach to successfully identify shared Y sequence between *P. reticulata* and *P. wingei* (Morris et al. 2018). Briefly, here we used the HAWK pipeline (Rahman et al. 2018) to count *k*-mers from paired-end DNA-seq reads and identify unique *k*-mers for each sex in each species. Across all the species, we then identified shared female unique *k*-mers and shared male unique *k*-mers, referred to as Y-mers (Morris et al. 2018).

3.3.8 Allele-specific expression (ASE) analysis

In order to estimate ASE patterns from RNA-seq data, I tailored previously published pipelines (Quinn et al. 2014; Zimmer et al. 2016). For each species, I called SNPs separately for males and females using SAMtools mpileup and VarScan (Koboldt et al. 2012), with parameters --min-coverage 2, --min-ave-qual 20, --min-freq-for-hom 0.90, and excluding triallelic SNPs and Ns. Additionally, I removed SNPs that were not located within genic regions from the final filtered gene dataset. To exclude potential sequencing errors from our SNP dataset, I applied coverage filtering thresholds (Quinn et al. 2014; Zimmer et al. 2016). Firstly, I set a minimum site coverage of 15 reads (the sum of major and minor alleles), as a power analysis indicated that at a minimum coverage of 15 reads we have a 78% power to detect a signal of allele-specific expression. Secondly, I applied a variable coverage filter that accounts for the change in the likelihood

of sequencing errors at different coverage levels (accounting for an error rate of 1 in 100 and a maximum coverage for a given site of 100,000 Quinn et al. (2014)). Lastly, to avoid the potential bias in our ASE estimations from the preferential assignment of reads to the reference allele (Stevenson et al. 2013), I removed clusters of more than 5 SNPs in 100 bp windows.

If genes have biallelic expression, meaning that alleles from both chromosomes are expressed at the same level, I expect a probability of around 0.5 of recovering reads from either chromosome. For each SNP in the final filtered dataset I tested for ASE by identifying significant deviations from the expected probability of 0.5 using a two-tailed binomial test (p <0.05). I corrected for multiple testing when running binomial tests on autosomal SNPs. Additionally, I called SNPs as ASE if a minimum of 70% of the reads stemmed from one of the chromosomes. I called genes as ASE if they had at least one SNP with a consistent ASE pattern across all heterozygous samples. I tested for significant differences in ASE patterns between the sexes and between the autosomes and the sex chromosomes using chi-square tests.

3.4 Results and Discussion

3.4.1 Comparative assembly of poeciliid sex chromosomes

We sequenced the genome and transcriptome of three male and three female individuals from each of the four target species (*P. wingei*, *P. picta*, *P. latipinna* and *Gambusia holbrooki*) (Table S.3.1), chosen to represent an even taxonomic distribution across Poeciliidae. For each species, we generated DNA-

seq with an average of 222 million 150bp paired-end reads (average insert size 500bp, resulting in an average of 76X coverage), and 77.8 million 150bp mate-pair reads (average insert size 2kb, averaging 22X coverage) per individual. We also generated on average 26.6 million 75bp paired-end RNA-seq reads for each individual.

Previous work on the sex chromosomes of these species showed evidence for male heterogametic systems in *P. wingei* (Nanda et al. 2014), *P. picta* (Lindholm et al. 2015) and *G. holbrooki* (Russo et al. 2009), and a female heterogametic system in *P. latipinna* (Sola et al. 1992; Sola et al. 1993). For each target species, I built scaffold-level *de novo* genome assemblies using SOAPdenovo (Luo et al. 2012) (Table S.3.2). Each assembly was constructed using the reads from the homogametic sex only in order to prevent co-assembly of X and Y reads. This allowed to later assess patterns of sex chromosome divergence based on differences between the sexes in read mapping efficiency to the genome (detailed below).

To obtain scaffold positional information for each species, I used the Reference-Assisted Chromosome Assembly (RACA) algorithm (Kim et al. 2013), which integrates comparative genomic data, through pairwise alignments between the genomes of a target, an outgroup (in this case *Oryzias latipes*) and a reference species (*Xiphophorus hellerii*), together with read mapping information from both sexes, to order target scaffolds into predicted chromosome fragments (Table S.3.2). RACA does not rely solely on sequence homology to the *X. hellerii* reference genome as a proxy for reconstructing the chromosomes in the target species, and instead incorporates read mapping and outgroup

information from *Oryzias latipes* (Kasahara et al. 2007) as well. This minimises mapping biases that might result from different degrees of phylogenetic similarity of our target species to the reference, *X. hellerii*. Using RACA, I reconstructed chromosomal fragments in each target genome and identified syntenic blocks (regions which maintain sequence similarity and order) across the chromosomes of the target and the reference species. This provided both comparison at sequence level for each target species with reference genome and positional information of scaffolds in chromosome fragments.

3.4.2 Extreme heterogeneity in sex chromosome differentiation patterns

For each target species, I used differences between males and females in genomic coverage and SNPs to identify non-recombining regions and strata of divergence. Additionally, I used published coverage and SNP density data in *P. reticulata* for comparative analyses (Wright et al. 2017).

In male heterogametic systems, non-recombining Y degenerate regions are expected to show a significantly reduced coverage in males compared to females, as males have only one X chromosome, compared to two in females. In contrast, autosomal and undifferentiated sex-linked regions have an equal coverage between the sexes. Thus, I defined older non-recombining strata of divergence as regions with a significantly reduced male to female coverage ratio compared to the autosomes.

Additionally, I used SNP densities in males and females to identify younger strata, representing earlier stages of sex chromosome divergence. In XY systems, regions that have stopped recombining more recently but which still retain high

94

sequence similarity between the X and the Y, show an increase in male SNP density compared to females, as Y reads, carrying Y-specific polymorphisms, still map to the homologous X regions. In contrast, I expect the opposite pattern of lower SNP density in males relative to females in regions of substantial Y degeneration as the X in males is effectively hemizygous (the Y copy is lost or exhibits substantial sequence divergence from the X orthology).

Previous studies have suggested a very recent origin of the *P. reticulata* sex chromosome system based on its large degree of homomorphism and the limited expansion of the Y-specific region (Nanda et al. 2014; Wright et al. 2017). Contrary to these expectations, my combined coverage and SNP density analysis indicates that *P. reticulata*, *P. wingei* and *P. picta* share the same sex chromosome system (Fig. 3.1; Fig. S.3.1 and Fig. S.3.2), revealing an ancestral system that dates back to at least 20 mya (Meredith et al. 2011). My findings suggest a far higher degree of sex chromosome conservation in this genus than we expected based on the small non-recombining region in *P. reticulata* in particular (Wright et al. 2017), and the high rate of sex chromosome turnover in fish in general (Mank et al. 2006; Pennell et al. 2018). By contrast, in the *Xiphophorous* and *Oryzias* genera sex chromosomes have evolved independently between sister species (Kondo et al. 2009; Volff and Schartl 2001), and there are even multiple sex chromosomes within *X. maculatus* (Orzack et al. 1980).
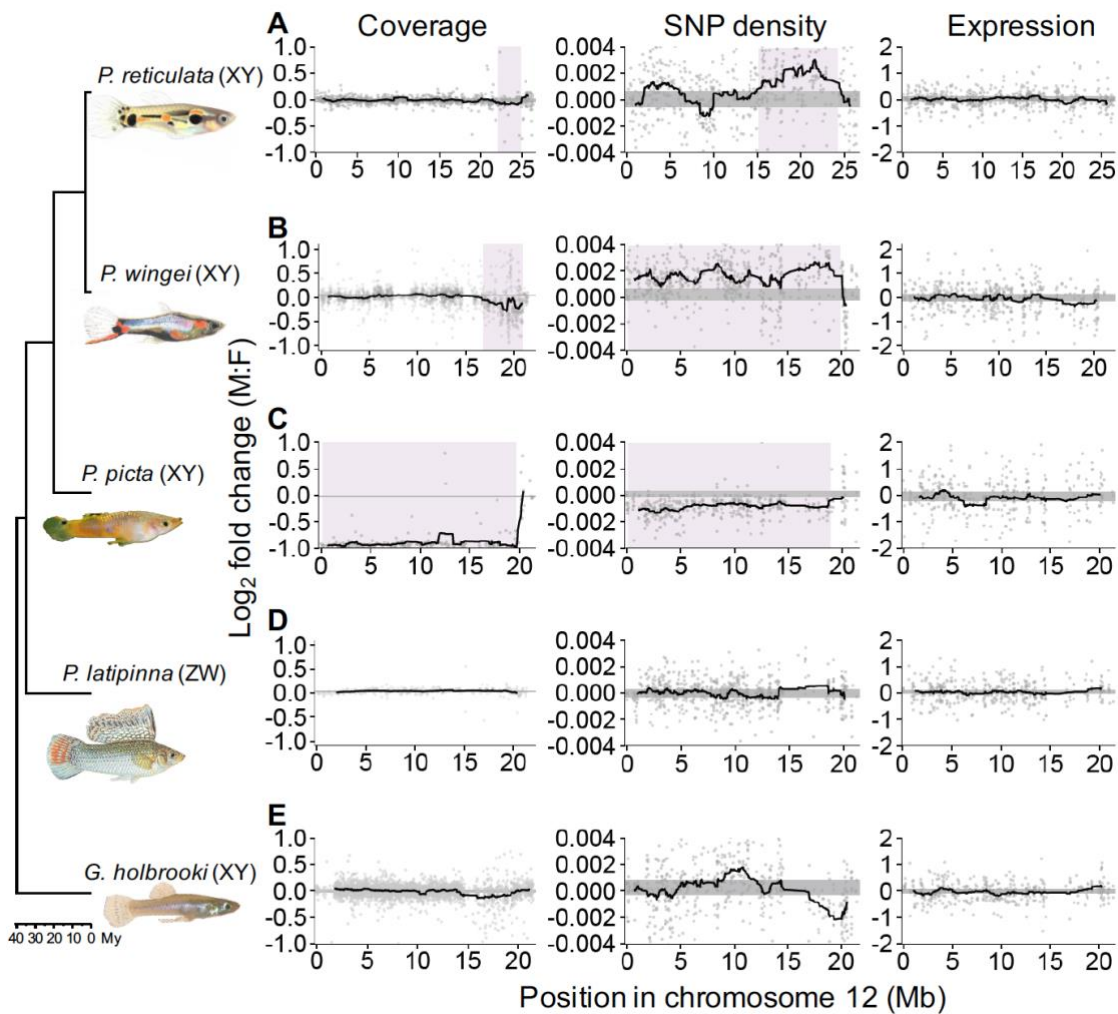
Fig. 3.1. Differences between the sexes in coverage, SNP density and expression across the guppy sex chromosome (*P. reticulata* chromosome 12) and syntenic regions in each of the target species. *X. hellerii* chromosome 8 is syntenic, and inverted, to the guppy sex chromosome. We used *X. hellerii* as the reference genome for our target chromosomal reconstructions. For consistency and direct comparison to *P. reticulata,* we used the *P. reticulata* numbering and chromosome orientation. Moving average plots show male to female differences in sliding windows across the chromosome in (A) *P. reticulata*, (B) *P. wingei*, (C) *P. picta*, (D) *P. latipinna* and (E) *G. holbrooki*. The 95% confidence intervals based on bootstrapping autosomal estimates are shown by the horizontal grey shaded areas. Highlighted in purple are the non-recombining regions of the *P. reticulata*, *P. wingei* and *P. picta* sex chromosomes, identified through a significant deviation from the 95% confidence intervals.

In addition to the unexpected conservation of this poeciliid sex chromosome system, I observe extreme heterogeneity in patterns of X-Y differentiation across the three species. The *P. wingei* sex chromosomes have a similar, yet more accentuated, pattern of divergence compared to *P. reticulata* (Fig. 3.1A and B). The non-recombining region appears to span the entire *P. wingei* sex chromosomes and, similar to *P. reticulata*, I can distinguish two evolutionary strata; an older stratum (17-20 Mb), showing significantly reduced male coverage, and a younger non-recombining stratum (0-17 Mb), as indicated by elevated male SNP density without a decrease in coverage (Fig. 3.1B). The old stratum has possibly evolved ancestrally to *P. wingei* and *P. reticulata*, as its size and estimated level of divergence appear to be conserved in the two species. The younger stratum, however, has expanded substantially in *P. wingei* relative to *P. reticulata* (Wright et al. 2017). These findings are consistent with the expansion of the heterochromatic block (Nanda et al. 2014) and the large-scale accumulation of repetitive elements on the *P. wingei* Y chromosome (Morris et al. 2018).

More surprisingly, however, is the pattern of sex chromosome divergence that I recover in *P. picta*, which shows an almost 2-fold reduction in male to female coverage across the entire length of the sex chromosomes relative to the rest of the genome (Fig. 3.1C). This indicates not only that the Y chromosome in this species is completely non-recombining with the X, but also that the Y chromosome has undergone significant degeneration. Consistent with the notion that genetic decay on the Y will produce regions that are effectively hemizygous, I also recover a significant reduction in male SNP density (Fig. 3.1C). A limited

pseudoautosomal region still remains at the far end of the chromosome, as both

the coverage and SNP density patterns in all three species suggest that

recombination persists in that area. As transitions from heteromorphic to

homomorphic sex chromosomes are not uncommon in fish and amphibians

(Pennell et al. 2018), it is also possible that the ancestral sex chromosome

resembles more the structure found in *P. picta* and that the sex chromosomes in

P. wingei and *P. reticulata* have undergone a transition to homomorphism.
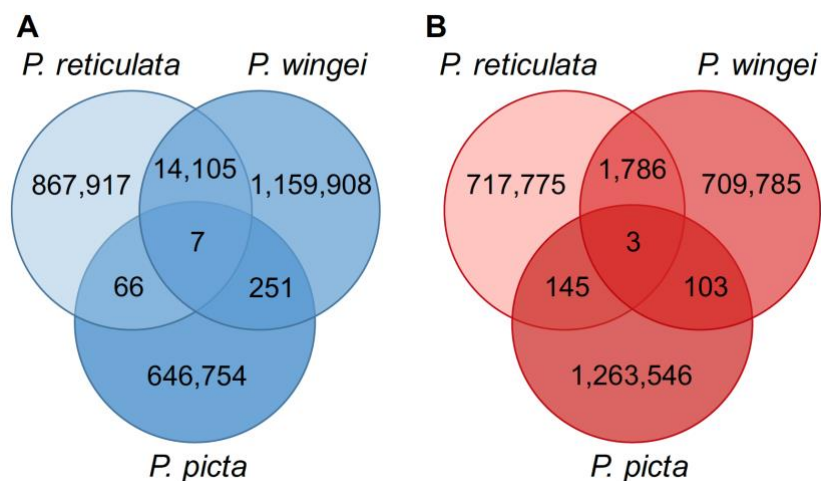
Fig. 3.2. Number of shared *k*-mers across *P. reticulata*, *P. wingei* and *P. picta*.
Species-specific and shared (A) male-unique *k*-mer (Y-mer) counts and (B) female-
unique *k*-mer counts.

In order to identify the ancestral Y region, we used *k*-mer analysis across

P. reticulata, P. wingei and *P. picta* which detects shared male-specific *k*-mers,

often referred to as Y-mers. Using this method, we have previously identified

many shared male-specific sequences between *P. reticulata* and *P. wingei*

(Morris et al. 2018) (Fig. 3.2). Curiously, here we recovered very few shared Y-

mers across all three species (Fig. 3.2), which suggests two possible scenarios in the evolution of *P. picta* sex chromosomes. It is possible that sex chromosome divergence began independently in *P. picta* compared to *P. reticulata* and *P. wingei*. Alternatively, the ancestral Y chromosome in *P. picta* may have been largely lost via deletion, resulting in either a very small Y chromosome, or an X0 system. To test for these alternative hypotheses, we reran the *k*-mer analysis in *P. picta* alone. We recovered almost twice as many female-specific *k*-mers than Y-mers in *P. picta* (Fig. 3.2), which indicates that much of the Y chromosome is indeed missing. This is consistent with the coverage analysis (Fig. 3.1C), which shows that male coverage of the X is half that of females, consistent with large-scale loss of homologous Y sequence.

I also used differences in coverage and SNP density between males and females to identify the sex chromosomes in the more distantly related *P. latipinna* and *G. holbrooki* (Fig. S.3.3 and Fig. S.3.4). The coverage could suggest that that the same region is the sex chromosome in both *P. reticulata* and *G. holbrooki,* as I found a small decrease in *G. holbrooki* male coverage towards the end of the chromosome (Fig. 3.1E). However, the SNP density patterns in *G. holbrooki* are not consistent with this finding, and therefore I cannot definitely conclude whether this is indeed the non-recombining region. Importantly, the sex chromosome appears to have evolved independently in *P. latipinna*, as in this species we cannot identify any areas of divergence or restricted recombination on the homologue of the *P. reticulata* sex chromosome (Fig. 3.1D). Lack of conservation of the sex chromosome system is not unexpected for *P. latipinna*, as this species has evolved a female heterogametic system. Regardless, the sex chromosomes

in *P. latipinna* and *G. holbrooki* are largely undifferentiated, and further linkage mapping of phenotypic sex is required to definitely determine the non-recombining region.

3.4.3 Y degeneration and dosage compensation in *Poecilia picta*

To investigate the extent of Y gene activity decay in our target species, I estimated allele-specific expression patterns (ASE) from RNA-seq data (Fig. 3.3A and 3.3B). If the X and Y chromosome are transcriptionally active at the same level, I expect a probability of around 0.5 of recovering reads from either chromosome. Conversely, regions of Y gene activity decay should be reflected by a significantly unbalanced contribution from the two sex chromosomes to the overall expression of heterozygous sites in males.

In *P. picta* I found that males and females have a similar proportion of autosomal genes with heterozygous sites (56% in females; 68% in males). However, out of the 363 sex-linked genes in *P. picta*, males have only 96 (27%) genes with Y-linked SNPs, while females have 177 (49%) genes with heterozygous sites. This likely indicates that many of the genes on the sex chromosomes in males are X-hemizygous. Of the 96 sex-linked genes with a transcriptionally active Y-linked copy, 77 show significant allele-specific expression in males (Fig. 3.3A). Indeed, my allelic differential expression analysis revealed that a significantly larger proportion of heterozygous sites show an ASE pattern on the sex chromosomes than on the autosomes in *P. picta* males ($\chi^2(1)=41.3710$, *p* < 0.001, Chi-square test; Fig. 3.3A). In contrast, in *P. picta* females and in both males and females of *P. reticulata* and *P. wingei*, the majority

100

of genes throughout the genome show equal transcription between the maternal and paternal chromosomes (Fig. 3.3A and B). These findings confirm extensive Y gene activity decay in *P. picta*.
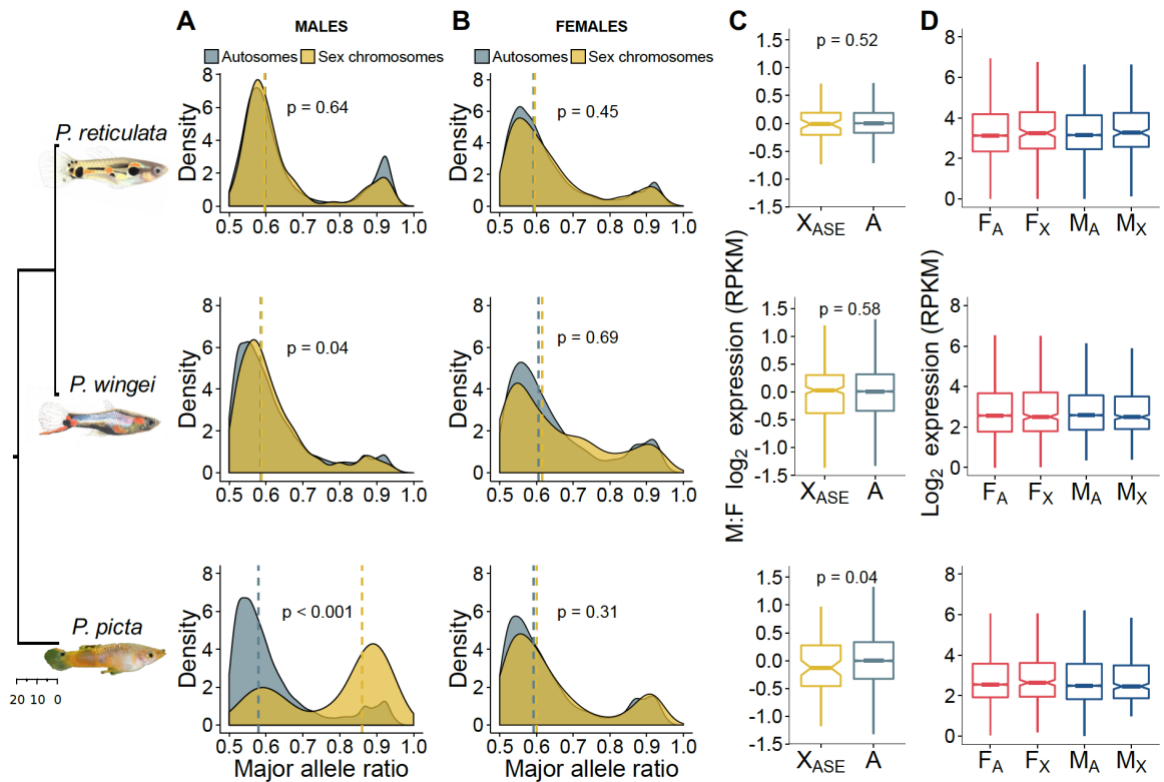
Fig. 3.3. Patterns of gene and allele-specific expression. Density plots show the distribution of the major allele frequency of autosomal (grey) and sex chromosome (yellow) genes in (A) males and (B) females of each species. Vertical dotted lines indicate median values and *p* values are based on Wilcoxon rank sum tests. (C) Boxplots show differences in $\log_2$ expression between the sexes (male:female) for autosomal genes (grey) and sex chromosome genes with an ASE pattern in males (yellow). *P* values are based on Wilcoxon rank sum tests. (D) Boxplots show average male (blue) and female (red) $\log_2$ expression for autosomes and the non-recombining region of the sex chromosomes in each species.

Given the profound degeneration of the Y chromosome in *P. picta*, I would expect the overall gene expression for the sex chromosomes to be roughly halved in males compared to females. Remarkably, in spite of these expectations, I found no significant reduction in average gene expression on the sex chromosomes in males when compared to either the sex chromosomes in females ($p = 0.09$, Wilcoxon signed rank test) or the autosomes in males ($p=0.94$, Wilcoxon rank sum test; Fig. 3.3D). This finding confirms the presence of a chromosome-wide dosage compensation mechanisms in *P. picta*, acting to counteract the imbalance in gene dose in males. Moreover, sex-linked genes with an ASE pattern in males show only a marginally significant decrease in male to female expression compared to autosomal genes ($p = 0.04$, Wilcoxon rank sum test; Fig. 3.3C). This likely represents a functional dosage compensation mechanism that is far more effective than the partial, localised dosage compensation recorded in other fish species (Chen et al. 2014; Leder et al. 2010).

I additionally tested for an enrichment of GO terms for genes in the non-recombining region of the sex chromosomes compared to genes expressed in the remainder of the genome, however, I found no GO terms with a significant ($p<0.001$) enrichment in any of the species. Examining differential expression patterns across each target genome revealed no evidence of sex chromosome sexualization, as levels of sex-biased gene expression are not significantly different between the autosomes and the sex chromosomes (Table S.3.3). However, I notice a slight deficit in male-biased genes on the *P. picta* sex chromosomes relative to the autosomes. This is possibly a consequence of Y degeneration and dosage compensation in this system, as it is more difficult for

genes on the sex chromosomes to achieve a male-biased expression compared to genes on the autosomes, given that they are already hyper-expressed as a result of dosage compensation (Vicoso and Charlesworth 2009).

3.5 Concluding remarks

My comparative analyses reveal a striking heterogeneity in the degree of recombination suppression and Y chromosome degeneration across poeciliid species with a shared sex chromosome system. Through multiple independent lines of evidence, including sequence coverage, *k*-mer analysis and ASE patterns, I show a profound degeneration of the Y chromosome in *P. picta*. Remarkably, the hemizygosity of the X in males has led to the evolution of a functional, chromosome-wide dosage compensation in this species, the first such instance found across teleost fish. My findings highlight the importance of comparative studies of sex chromosome differentiation within clades and suggest that fish may harbour extensive variation in sex chromosome evolution.

Chapter 4. Poeciliid sex chromosome integrity maintained by incomplete recombination suppression

4.1 Abstract

The loss of recombination triggers divergence between the sex chromosomes and promotes degeneration of the sex-limited chromosome. The sister species, *Poecilia reticulata* and *Poecilia wingei*, share a male-heterogametic sex chromosome system with restricted recombination across almost the entire length of the X and Y chromosomes. Despite of that, divergence between the X and the Y chromosome is very low and there is no perceptible decay of male gene activity. Here, I use RNA-seq data from multiple *P. reticulata* and *P. wingei* families to identify genes with sex-limited SNP inheritance. Phylogenetic tree analysis reveals that occasional recombination occurs between the sex chromosomes, as X- and Y-linked sequences cluster by species instead of by gametolog. This low-level recombination prevents the sex chromosomes from diverging and maintains Y integrity over substantial evolutionary time. In addition, recombination arrest appears to have expanded throughout the sex chromosomes more gradually instead of through a stepwise process.

104

4.2 Introduction

A common feature of sex chromosomes observed across a diverse array of taxa is the loss of recombination, which can encompass the majority or part of their length (Bachtrog et al. 2011; Bull 1983; Charlesworth et al. 2005). Non-recombining regions experience a reduction in the efficiency of selection to remove deleterious mutations as a result of decreased effective population size and accentuated Hill-Robertson effects (Charlesworth and Charlesworth 2000; Charlesworth et al. 2005). As a consequence, over time, initially identical sex chromosomes are expected to diverge from each other in gene content and nucleotide sequence and the sex-limited chromosome to degenerate (Bachtrog 2013; Charlesworth and Charlesworth 2000).

Recombination arrest can initially cover the region containing the sex-determining locus and subsequently expand over larger portions of the sex chromosomes (Charlesworth et al. 2005). In highly diverged sex chromosome systems, recombination is restricted to a small pseudoautosomal region (PAR) (Cortez et al. 2014; Zhou et al. 2014). The expansion of the non-recombining region can occur in a stepwise manner, through successive recombination suppression events. This results in regions with different levels of sequence divergence between the gametologs, referred to as evolutionary strata of divergence (Bachtrog 2013; Charlesworth et al. 2005; Lahn and Page 1999). Evidence of strata has been recorded in mammals (Lahn and Page 1999), fish (Natri et al. 2013; Roesti et al. 2013; Wright et al. 2017), birds (Handley et al.

2004; Wright et al. 2014), snakes (Vicoso et al. 2013a) and plants (Bergero et al. 2007; Hough et al. 2014; Nicolas et al. 2004; Wang et al. 2012).

Intra-chromosomal rearrangements are predicted to be the underlying mechanism for the evolution of discrete strata, as large-scale inversions on the sex-limited chromosome can instantaneously prevent recombination throughout the inverted region (Charlesworth et al. 2005). While patterns of sex chromosome divergence in older systems show the expected signatures of strata formation via inversions, namely clusters of gametologs with similar divergence estimates (Cortez et al. 2014; Handley et al. 2004; Lahn and Page 1999; Wright et al. 2014; Wright et al. 2012), newly evolved systems show less support for the classic model of sex chromosome evolution (Chibalina and Filatov 2011; Muyle et al. 2012; Roesti et al. 2013). In some younger systems, sequence divergence and recombination suppression between the gametologs are seen to occur heterogeneously across the length of the sex chromosomes, suggesting a gradual process instead of fixation of inversions on the sex-limited chromosome (Almeida et al. 2019; Bergero et al. 2013; Chibalina and Filatov 2011; Natri et al. 2013; Nicolas et al. 2004). Moreover, processes such as gene conversion or infrequent recombination events can prevent the sex-limited chromosome from degenerating (Stock et al. 2011; Stock et al. 2013), and may make it more difficult to clearly delineate between the different strata or between the non-recombining region and the PAR (Chibalina and Filatov 2011).

Previous studies have used RNA-seq data and analyses of SNP segregation patterns in families to isolate sex-linked genes from autosomal or pseudoautosomal genes and obtain gametologous sequences (Chibalina and

106

Filatov 2011; Hough et al. 2014; Muyle et al. 2012; Vicoso et al. 2013a; Wright et al. 2014). This approach allows to test for the presence of evolutionary strata by estimating divergence between gametologs and identifying clusters of sex-linked genes with different divergence rates (Lahn and Page 1999; Ross et al. 2005; Wright et al. 2014). While these analyses can be used to date recombination suppression events and distinguish between the strata of heteromorphic sex chromosomes, they are also useful to define the boundaries between the PAR and the non-recombining regions of younger, less differentiated, systems (Campos et al. 2017).

The common guppy (*Poecilia reticulata*) and its sister species the Endler's guppy (*Poecilia wingei*) share the same male-heterogametic sex chromosome system (Darolti et al. 2019; Morris et al. 2018). Our previous analyses of coverage and polymorphism data in males and females indicate that recombination suppression, while covering approximately half the length of the sex chromosomes in *P. reticulata*, has extended across almost the entire *P. wingei* system (Darolti et al. 2019; Wright et al. 2017). Another recent study reports differences in recombination patterns between guppy males and females, with crossover events being confined to the end of the sex chromosomes in males (Bergero et al. 2019). Despite the extensive recombination suppression, these systems are characterised by low levels of X-Y sequence divergence and no perceptible loss of Y gene activity and content (Darolti et al. 2019; Wright et al. 2017). In both species, there is evidence of two candidate evolutionary strata, however the differences in divergence between them are subtle (Darolti et al. 2019; Wright et al. 2017).

Here, I use the probability based approach, SEX-DETector (Muyle et al. 2016), to infer autosomal and sex-linked genes in *P. reticulata* and *P. wingei* based on RNA-seq data from families. Consistent with the expansion of the non-recombining region in *P. wingei*, I find significantly more sex-linked genes in this species than in *P. reticulata*. Sex-linked genes are predominantly distributed across the previously identified non-recombining regions, however there is no difference in X-Y sequence divergence between genes on the two evolutionary strata, or between genes on the PAR and the non-recombining region, suggesting that recombination arrest has evolved more gradually. A phylogenetic analysis of X- and Y-linked sequences reveals that the extensive homomorphism of the poeciliid sex chromosomes is maintained by incomplete recombination suppression.

4.3 Materials and Methods

4.3.1 Sample collection and sequencing

I sampled parents and F1 male and female offspring from three *P. reticulata* and two *P. wingei* within-species crosses. For generating each cross, I used a male and a virgin female either from a *P. reticulata* outbred laboratory population that descends from the high predation population of the Trinidadian river Quare (Kotrschal et al. 2013), or from our *P. wingei* laboratory population, derived from a strain maintained by a UK fish breeder. I only sampled families if the number of F1 progeny included at least five male and five female individuals,

as this is the minimum number of offspring per family required to reliably identify sex-linked genes using the SEX-DETector software.

From each individual, I obtained samples of adult tail tissue, which I preserved in RNAlater at -20°C prior to RNA preparation. I extracted RNA from these samples using the RNeasy Kit (Qiagen), following the manufacturer's instructions. Libraries were prepared at the SciLife Lab in Uppsala, Sweden, following standard protocol. RNA was sequenced on an Illumina HiSeq 2500 platform, resulting in, on average, 43 million 125 bp paired-end *P. reticulata* RNA-seq reads per sample and 41 million 125 bp paired-end *P. wingei* RNA-seq reads per sample. I assessed sample quality using FastQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and then removed adaptor sequences and trimmed reads with Trimmomatic v0.36 (Lohse et al. 2012). I trimmed regions with average Phred score < 15 in sliding windows of four bases, reads with Phred score < 3 for leading and trailing bases, as well as paired-end reads with either read pair shorter than 50 bp. Following trimming, I had, on average per sample, 29 million paired-end RNA-seq reads for *P. reticulata* and 28 million paired-end RNA-seq reads for *P. wingei* (Table S.4.1).

4.3.2 Constructing and filtering *de novo* transcriptome assemblies

I used the RNA-seq reads to construct a *de novo* transcriptome assembly for each species (Table S.4.2), following previously implemented methods (Bloch et al. 2018; Harrison et al. 2015; Wright et al. 2019b). I pooled all forward and reverse reads of each species and assembled them using Trinity v2.5.1 (Grabherr et al. 2011) with default parameters. I then filtered the assemblies to

109

remove redundancy and non-coding RNA. First, I used the Trinity align_and_estimate_abundance.pl script which maps RNA-seq reads to the transcriptomes using Bowtie2, suppressing unpaired and discordant alignments, and estimates transcript abundance for each sample using RSEM. I then selected the best isoform for each gene based on the highest average expression. In cases where multiple isoforms had the highest expression, I chose the longest isoform as the best isoform. I further filtered the assemblies for non-coding RNA by removing transcripts with a blast hit to *Poecilia formosa* (PoeFor_5.1.2) or *Oryzias latipes* (MEDAKA1) ncRNA sequences obtained from Ensembl 84 (Flicek et al. 2014). Lastly, I used Transdecoder v5.2.0 (http://transdecoder.github.io) with default parameters to remove transcripts missing an open-reading frame and transcripts with open-reading frames shorter than 150bp.

4.3.3 Assigning chromosomal position

I downloaded *P. reticulata* genes from NCBI RefSeq (Guppy_female_1.0+MT, RefSeq assembly accession: GCF_000633615.1) and identified the longest isoform for each gene. I BLASTed the best isoform sequences against the filtered *P. reticulata* and *P. wingei* transcriptome assemblies using BLASTn v2.3.0 (Altschul et al. 1990) with an e-value cutoff of $10e^{-10}$ and a 30% minimum percentage identity. For genes mapping to multiple *de novo* transcripts, I selected the top blast hit based on the highest BLAST score. I assigned positional information on *P. reticulata* chromosomes to *P. reticulata* and *P. wingei* transcripts, based on the chromosomal location of guppy genes. In a previous study, I have shown that the guppy reference genome has an inversion

110

on chromosome 12 (the sex chromosome) that is specific to the inbred strain on which the reference assembly was built (Darolti et al. 2019). Thus, when assigning *de novo* transcripts with positional information on the sex chromosome, I took into account the coordinates of the discovered inversion.

4.3.4 Inferring autosomal and sex-linked genes

In order to identify sex-linked genes I used the probability-based method SEX-DETector (Muyle et al. 2016), which analyses the genotypes of individuals in a cross (parents and male and female offspring) to infer the segregation mode of each contig. The model can distinguish between three segregation types: autosomal, sex-linked with both the X and the Y copies present and sex-linked with the X copy present but the Y copy lost or lowly expressed (X-hemizygous mode). For identifying sex-linked genes with X and Y alleles, SEX-DETector requires gametologs to co-assemble in one single transcript instead of separate transcripts. Thus, the SEX-DETector algorithm is best suited for species with sex chromosomes that have low or intermediate levels of divergence, as high levels of divergence prevent the co-assembly of X and Y alleles. To avoid assembly of X and Y copies of the same transcript into separate contigs, and thus wrongly inferring contigs as X-hemizygous, I used the recommended CAP3 software (Huang and Madan 1999) with a 90% minimum percent similarity between sequences to further assemble contigs together. Following this step, the final filtered transcriptome assemblies contained a total of 19,935 *P. reticulata* transcripts and 19,361 *P. wingei* transcripts (Table S.4.2).

Based on the SEX-DETector pipeline, I mapped reads onto the filtered assemblies using BWA v0.7.12 (Li and Durbin 2009) and then merged and sorted all libraries of each family separately using SAMtools v1.3.1 (Li et al. 2009). Using reads2snp v2.0 (http://kimura.univ-montp2.fr/PopPhyl/), I genotyped individuals at each locus before using SEX-DETector to infer, within each family, the segregation type for each transcript, using a minimum posterior segregation type probability of 0.8, and to construct predicted X and Y sequences for each sex-linked gene.

4.3.5 Phylogenetic analysis

In addition to the *P. reticulata* and *P. wingei de novo* transcript sequences, I downloaded *Oryzias latipes* (MEDAKA1), *Xiphophorus maculatus* (Xipmac4.4.2) and *Poecilia formosa* (PoeFor_5.1.2) transcripts from Ensembl 84 and identified the longest transcript for each gene. I determined orthology across all these species using a reciprocal BLASTn with an e-value cutoff of $10e^{-10}$ and a minimum percentage identity of 30%. I then used BLASTx to identify open reading frames in each orthologous group.

I conducted a phylogenetic analysis of X and Y-linked sequences together with their orthologs in outgroup species to investigate recombination suppression on the sex chromosomes. First, I aligned sequences using Prank v140603 (Löytynoja and Goldman 2008) and removed gaps in the alignments. I generated maximum likelihood phylogenetic trees with RAxML v8.2.12 (Stamatakis 2014), using the rapid bootstrap algorithm with the GTRGAMMA model and 100 bootstraps. I used the alignments of orthologous groups in which both the *P.*

reticulata and *P. wingei* genes were identified as sex-linked to generate a consensus phylogenetic tree with the majority rule consensus tree option in RAxML. I plotted all phylogenetic trees using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

4.3.6 Rates of evolution of sex-linked genes

For each identified sex-linked gene, I estimated synonymous pairwise X-Y divergence (dS_{XY}) using the *yn00* program in PAML v4.8 (Yang 2007), following the Yang and Nielsen method (Yang and Nielsen 2000). Within each species, I separated sex-linked genes into different categories based on their position on the sex chromosomes (PAR, non-recombining region, Stratum I, Stratum II) and tested for differences in dS_{XY} between the different categories and between the species using Wilcoxon rank sum tests in R.

I also estimated rates of evolution of X and Y sequences using the Prank multiple sequence alignments described above (same alignments I used to build the phylogenetic gene trees). I masked poorly aligned regions with SWAMP v 31-03-14 (Harrison et al. 2014). I then used branch model 2 from the CODEML package in PAML to estimate rates of nonsynonymous (d_N) and synonymous (d_S) substitutions for the X- and Y-linked branches in each tree. I used the inferred phylogeneic trees described above to generate an unrooted tree for each orthologous group, which is required in the CODEML analyses. I used 1,000 permutation test replicates to identify significant differences in d_N/d_S ratios between the X and Y sequences, between the gene categories and between the two species.

4.4 Results

I used genotypes of parents and offspring from multiple *P. reticulata* and *P. wingei* crosses to infer sex-linked genes following the SEX-DETector pipeline (Table 4.1 and 4.2). By mapping these genes to known *P. reticulata* transcripts and assigning them with a chromosomal position, I was able to verify that the majority of them are indeed found on the sex chromosome (Table 4.1 and 4.2), and these are the genes that I used in all subsequent analyses. Only three genes were inferred to be sex-linked in all the three *P. reticulata* families, while 142 genes showed signatures of sex-linkage in both *P. wingei* crosses. This is due to the fact that the sampled *P. reticulata* population is more outbred and has a higher genetic diversity than the sampled *P. wingei* population.

Table 4.1. Number of inferred *P. reticulata* sex-linked genes

Sex-linked genes	All families	Family 1	Family 2	Family 3
Total number	111	60	41	28
On the sex chromosomes	92 (83%)	54 (90%)	33 (81%)	21 (75%)
On the autosomes	13 (12%)	5 (8%)	3 (7%)	5 (18%)
On unplaced scaffolds	6 (5%)	1 (2%)	5 (12%)	2 (7%)

Table 4.2. Number of inferred *P. wingei* sex-linked genes

Sex-linked genes	All families	Family 1	Family 2
Total number	272	236	179
On the sex chromosomes	249 (92%)	219 (93%)	163 (91%)
On the autosomes	12 (4%)	7 (3%)	7 (4%)
On unplaced scaffolds	11 (4%)	10 (4%)	9 (5%)

Although *P. wingei* and *P. reticulata* share the same sex chromosome system (Darolti et al. 2019; Morris et al. 2018), I find substantially more sex-linked genes in *P. wingei*. Approximately 73% of the *P. reticulata* sex-linked genes are also found to be sex-linked in *P. wingei*, while only 27% of the *P. wingei* sex-linked genes are classified as sex-linked in *P. reticulata*. This confirms my previous finding that, in *P. wingei*, recombination suppression extends over a larger proportion of the sex chromosome than in *P. reticulata* (Darolti et al. 2019).
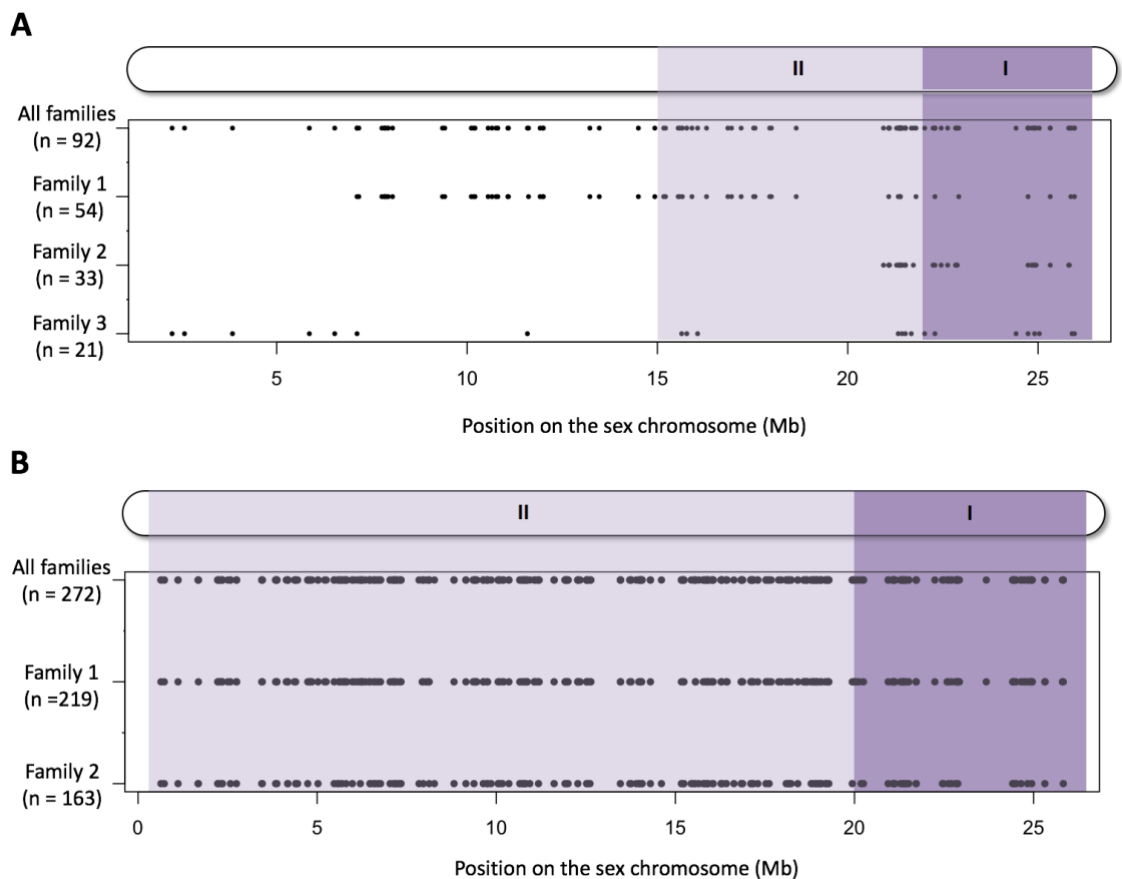
Fig. 4.1. Distribution of sex-linked genes on the sex chromosomes of *P. reticulata* (A) and *P. wingei* (B). The shaded purple regions indicate the previously identified non-recombining regions (Darolti et al. 2019; Wright et al. 2017). The dark purple area highlights the predicted older evolutionary stratum, where X-Y divergence is the greatest, while the light purple region highlights the younger non-recombining region.

Analysing the position of these genes along the sex chromosomes reveals that, indeed, *P. wingei* sex-linked genes are spread throughout the entire length of the sex chromosome, while *P. reticulata* sex-linked genes are more predominantly found towards the end of the chromosome (Fig. 4.1). When accounting for the total number of genes on the sex chromosomes, I find that there is a significant enrichment of sex-linked genes on the predicted non-recombining region of the *P. reticulata* sex chromosomes (χ^2=23.306, *p* < 0.001, Chi-square test). I recover the same result when testing for an enrichment of sex-linked genes on either Stratum II (χ^2=15.498, *p* < 0.001, Chi-square test) or Stratum I (χ^2=19.789, *p* < 0.001, Chi-square test) in comparison to the predicted pseudoautosomal region. There is no significant difference in the number of sex-linked genes between the two strata of divergence in either of the two species (χ^2=0.751, *p* = 0.386, Chi-square test in *P. reticulata* and χ^2=0.008, *p* = 0.927, Chi-square test in *P. wingei*). Approximately 10% of the genes on the predicted *P. reticulata* pseudoautosomal region (0-15 Mb) are inferred to be sex-linked, suggesting that recombination suppression events expand beyond the previously identified non-recombining region, but have not resulted in significant sequence divergence of the X and Y chromosomes.

I used genes identified as sex-linked in both *P. reticulata* and *P. wingei*, together with orthologs in *P. formosa*, *X. maculatus* and *O. latipes*, in a phylogenetic analysis to investigate recombination suppression on the sex chromosomes. I found five sex-linked genes showing clustering of X- and Y-linked sequences by gametolog type instead of by species, suggesting that these

116

genes have stopped recombining before the two species diverged (Fig. 4.2A-E). On the *P. reticulata* sex chromosome, one of these genes is located in the predicted Stratum I region, while the other four are found in Stratum II. Three of these four genes are positioned close to the boundary between the two strata (21-22Mb), suggesting that the *P. reticulata* old non-recombining region is potentially larger than previously estimated based on sequence divergence alone. On the *P. wingei* sex chromosome, all but one of these genes are found on the older non-recombining region. However, in the majority of sex-linked genes, gametologs cluster by species (Fig. 4.2F), which is indicative of either more recent, or incomplete, recombination suppression.
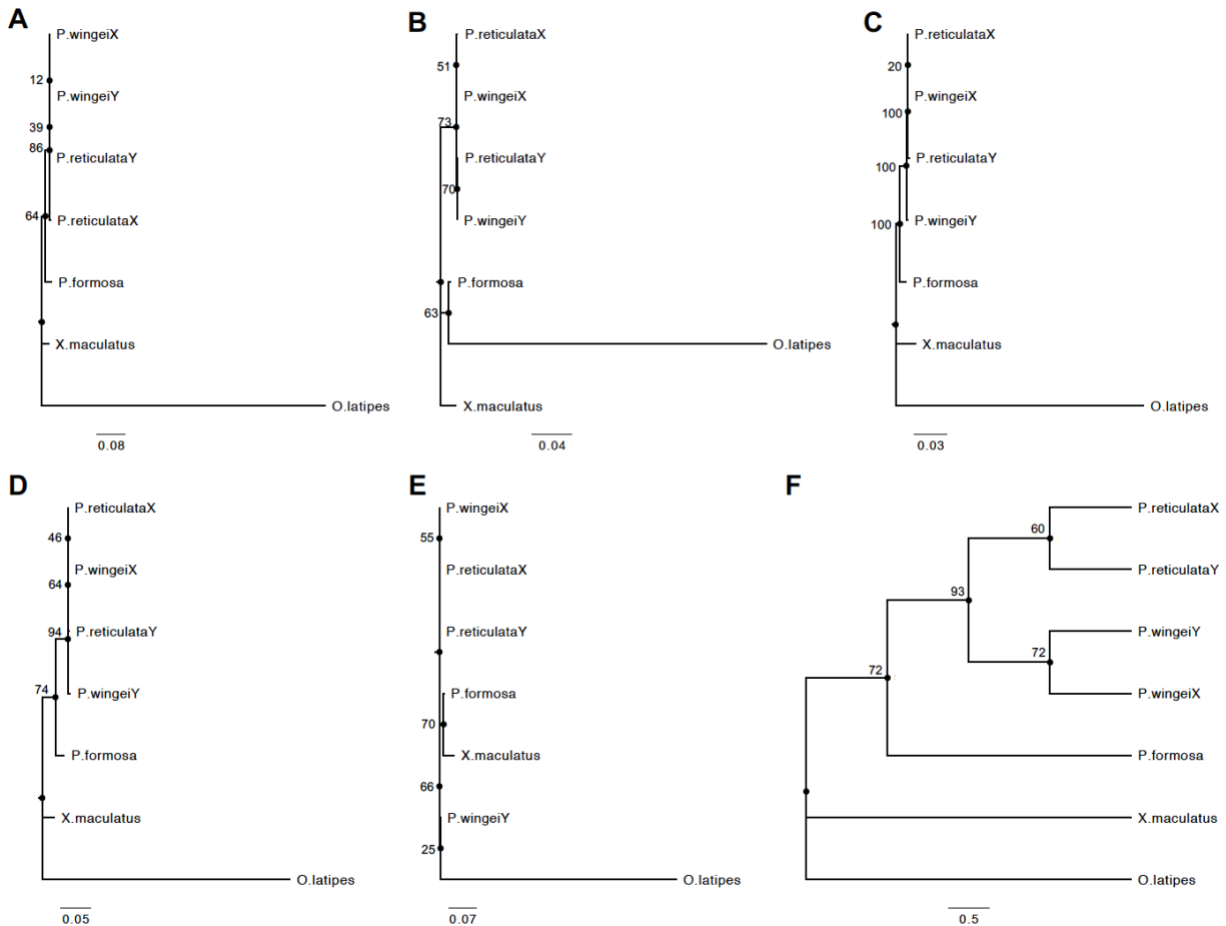
Fig. 4.2. Phylogenetic gene trees for _P. reticulata_ and _P. wingei_ X- and Y-linked sequences. (A-E) Shown are the phylogenetic trees of the five sex-linked genes in which the X and Y sequences cluster by gametologs instead of by species. (F) Majority consensus tree based on alignments of all genes identified as sex-linked in both species. Numbers at each node represent bootstrap values based on 100 permutations.

118

I estimated the rate of synonymous substitutions (dS$_{XY}$) between each pair of X- and Y-linked sequences to test for differences in divergence between sex-linked genes in different sex chromosome regions (PAR, non-recombining region, Stratum II, Stratum I) and between the two species (Fig. 4.3). I found no correlation between pairwise synonymous divergence of sex-linked genes and their position on the sex chromosome in either *P. reticulata* (R^2=0.02, p = 0.279) or *P. wingei* (R^2=0.04, p = 0.095). There is also no significant difference in dS$_{XY}$ between the non-recombining region and the PAR (Wilcoxon rank sum test p=0.4), or between either of the Strata and the PAR (Wilcoxon rank sum tests p=0.7 for Stratum I and p=0.5 for Stratum II) in *P. reticulata*. On average, sex-linked genes in *P. wingei* have higher dS$_{XY}$ than those in *P. reticulata*, however this difference is marginally non-significant (median *P. wingei* dS$_{XY}$ = 0.0054, median *P. reticulata* dS$_{XY}$ = 0.0045; Wilcoxon rank sum tests p = 0.059). There were no significant differences even when excluding genes with a dS$_{XY}$ of 0.

I also tested whether Y-linked sequences have accumulated more deleterious mutations than X-linked sequences by estimating rates of nonsynonymous (d_N) and synonymous (d_S) substitutions and calculating average (d_N/d_S) for each gametolog branch. Overall, Y-linked sequences in both species show a higher d_N/d_S compared to X-linked sequences, however this is not significant (Table S.4.3). These results are in line with our previous finding that transcription rates for genes on the sex chromosomes are equal between the sexes in these two species, and that there is no evidence of Y degeneration (Darolti et al. 2019). Consistently, SEX-DETector inferred only three and,

respectively, no X-hemyzigous genes in *P. reticulata* and *P. wingei*. All three *P. reticulata* XO genes are located in the older non-recombining region.

A

B
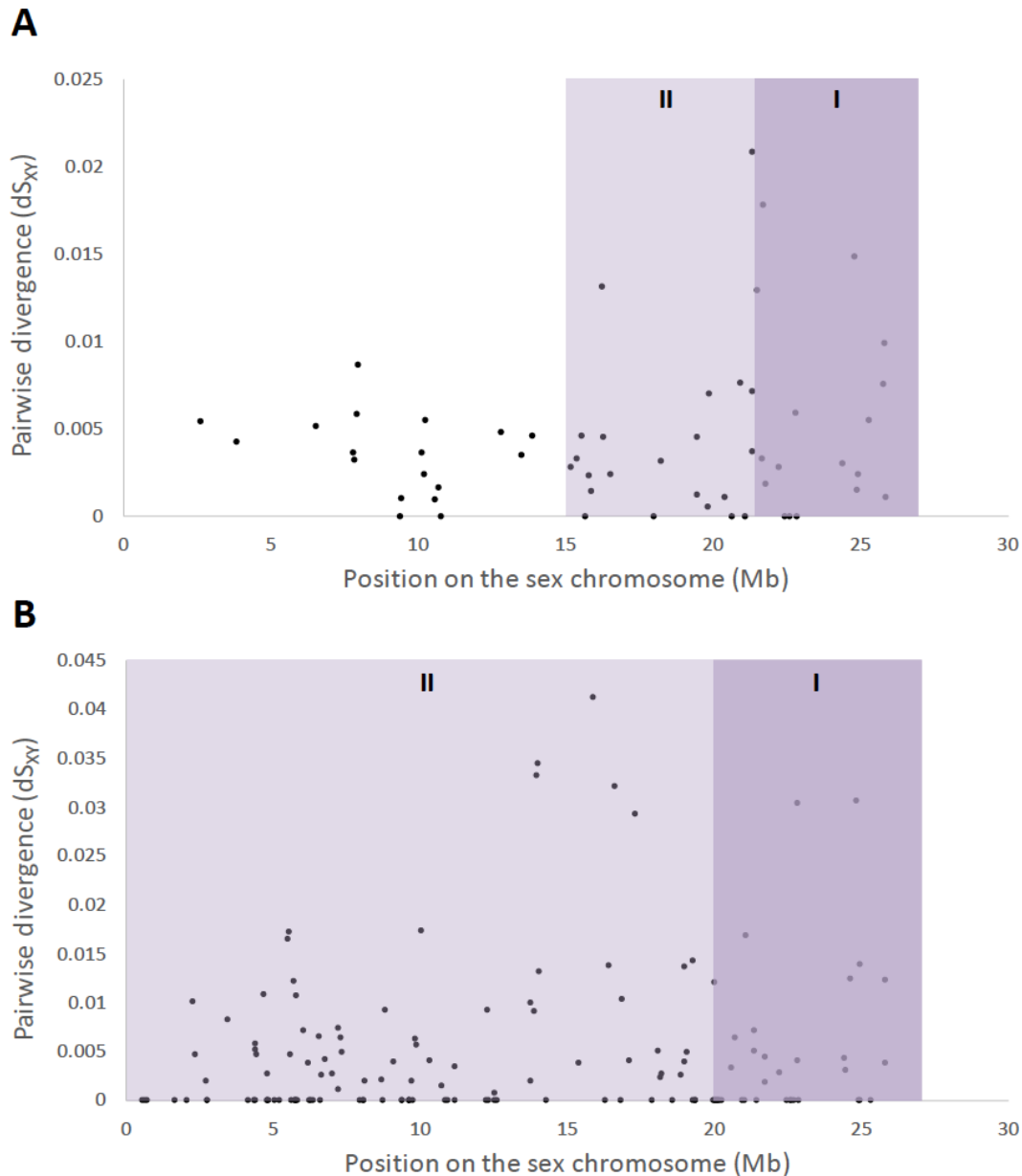
Fig. 4.3. Pairwise synonymous divergence (*d*S$_{XY}$) of *P. reticulata* (A) and *P. wingei* (B) sex-linked genes across the sex chromosomes. The shaded purple regions indicate the previously identified non-recombining regions (Darolti et al. 2019; Wright et al. 2017). The dark purple area highlights the predicted older stratum of divergence, while the light purple are highlights the younger non-recombining region.

4.5 Discussion

4.5.1 Infrequent recombination events maintain Y chromosome integrity

Complete recombination suppression over portions of the sex chromosomes is expected over time to favour the accumulation of deleterious mutations and lead to loss of gene activity and function on the sex-limited chromosome (Bachtrog 2013). In spite of these predictions, *P. reticulata* and *P. wingei* share a sex chromosome system that is largely non-recombining, yet which shows little X-Y sequence divergence and no reduction of male gene activity and content (Darolti et al. 2019; Wright et al. 2017). These findings lead to questions about the mechanisms maintaining Y sequence integrity in these species.

One mechanism preventing sex chromosome degeneration is incomplete recombination suppression (Perrin 2009). Infrequent X-Y recombination in males or in sex reversed females has been shown to be sufficient to maintain a high level of sequence similarity between gametologs (Stock et al. 2011; Stock et al. 2013). Using a phylogenetic analysis of sex-linked genes in *P. reticulata* and *P. wingei* together with orthologs in outgroup species, I show that Y chromosomes in *P. reticulata* and *P. wingei* have a higher sequence similarity to their homologous X regions than to each other (Fig. 4.2F). In the complete absence of recombination, I expect X- and Y-linked sequences to cluster by gametolog instead of by species, however I see this occurring in only five sex-linked genes. These results suggest that occasional recombination on the poeciliid sex chromosomes is preventing their degeneration.

Interestingly, *Poecilia picta*, a close relative, appears to share the same sex chromosome system, however, in contrast to *P. reticulata* and *P. wingei*, the Y chromosome is highly degenerated in *P. picta* (Darolti et al. 2019). Identifying and analysing sex-linked genes in *P. picta* would be helpful in order to determine whether the sex chromosome system is indeed shared across all three species and would offer insight into the processes driving the observed inter-specific variation in sex chromosome recombination suppression and differentiation.

An alternative mechanisms acting to prevent the neutral degradation of the sex-limited chromosome is gene conversion between gametologs (Rosser et al. 2009; Slattery et al. 2000; Trombetta et al. 2009; Wright et al. 2014). Gene conversion events happen in meiosis when one allele is converted into its homolog in the process of double-strand DNA break repair (Chen et al. 2007). Testing for the presence of gene conversion generally relies on identifying stretches of X-Y identical sequence delimited by variable sites (Wright et al. 2014). Unfortunately, here, the high similarity between the identified X- and Y-linked sequences in *P. reticulata* and *P. wingei* prevents such analyses.

4.5.2   Recombination arrest shows signs of gradual expansion

Our recent analyses of coverage and polymorphism data in males and females indicate that the sex chromosomes in *P. wingei* are more diverged than in *P. reticulata* (Darolti et al. 2019; Wright et al. 2017). Indeed, here I find that the extent of sex-linkage is greater in *P. wingei* compared to *P. reticulata*, although the non-recombining region in the latter seems to be wider than we initially predicted, as we find sex-linked genes present in the previously defined

122

pseudoautosomal region. We have also estimated that sex chromosomes of different guppy wild populations vary in their degree of X-Y divergence (Wright et al. 2017). Recent linkage mapping failed to find evidence supporting this (Bergero et al. 2019), however this may have been due to insufficient statistical power to detect differences in rare X-Y recombination events (Wright et al. 2019a).

Here, the results of my sex-linkage analysis in *P. reticulata* clearly show that there is persistent X-Y recombination throughout the length of the sex chromosome that prevents Y chromosome degeneration, despite the significant age of the sex chromosome system (Darolti et al. 2019). It would not be unlikely for the frequency or chromosomal location of these persistent recombination events to vary across different wild populations, resulting in the observed intra-specific variation in sex chromosome differentiation. Analyses of guppy sex-linked sequences from natural populations would offer insight into the evolutionary dynamics of recombination suppression.

We have shown that the sex chromosomes in both species carry signatures of evolutionary strata of divergence (Darolti et al. 2019; Wright et al. 2017). Here, I find five sex-linked genes that show clustering of X- and Y-linked sequences by gametolog, confirming recombination arrest before the divergence of species. As these genes are located either in the predicted older stratum of divergence or close to its boundary, it allows me to define the sex chromosome region spanning 20-25 Mb as being the ancestral, more differentiated, non-recombining region.

Estimates of pairwise synonymous divergence of sex-linked genes and rates of evolution of X- versus Y-linked sequences indicate no detectable

difference in divergence between the different sex chromosome regions (Fig. 4.3, Table S.4.3). This suggests a more gradual expansion of recombination suppression instead of a stepwise process resulting from inversions. However, discrete evolutionary strata and clear boundaries between the non-recombining region and the PAR may be masked by the stochastic differences between genes in X-Y divergence resulting from occasional recombination events (Chibalina and Filatov 2011).

Chapter 5. Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*

5.1 Abstract

The relative rate of evolution for sex-biased genes has often been used as a measure of the strength of sex-specific selection. In contrast to studies in a wide variety of animals, far less is known about the molecular evolution of sex-biased genes in plants, particularly in dioecious angiosperms. Here, we investigate the gene expression patterns and evolution of sex-biased genes in the dioecious plant *Salix viminalis*. We observe lower rates of sequence evolution for male-biased genes expressed in the reproductive tissue compared to unbiased and female-biased genes. These results could be partially explained by the lower codon usage bias for male-biased genes leading to elevated rates of synonymous substitutions compared to unbiased genes. However, the stronger haploid selection in the reproductive tissue of plants, together with pollen competition, would also lead to higher levels of purifying selection acting to remove deleterious variation. Future work should focus on the differential evolution of haploid- and diploid-specific genes to understand the selective dynamics acting on these loci.

5.2 Introduction

Many species show a wealth of phenotypic differences between the sexes. However, apart from genes on sex chromosomes, males and females share the same genome, and sexually dimorphic traits are therefore thought to arise as a result of differential regulation of genes occurring in both sexes (Ellegren and Parsch 2007; Mank 2017; Pointer et al. 2013; Ranz et al. 2003), often referred to as sex-biased gene expression. Sex-biased genes are thought to evolve in response to conflicting sex-specific selection pressures over optimal expression acting on this shared genetic content (Connallon and Knowles 2005) and are increasingly used to study the footprint of sex-specific selection within the genome (Dean et al. 2017; Gossmann et al. 2014; Mank 2017).

In contrast to animals, where sexual dimorphism is more frequent, only a small percentage (~5%) of flowering plants are dioecious (Renner 2014; Robinson et al. 2014), where individuals have exclusively male or female reproductive organs. The majority (~90%) of angiosperms are hermaphroditic (Ainsworth 2000; Barrett and Hough 2013), where flowers are bisexual, while another small fraction are monoecious, where separate flowers within the same plant carry different reproductive organs (Renner 2014). Despite being rare, dioecy has evolved in flowering plants many times independently (Charlesworth 2002) and is distributed across the majority of angiosperm higher taxa (Helibuth 2000; Käfer et al. 2017).

Although sexual dimorphism is generally more extensive in animal species, male and female dioecious flowering plants also undergo conflicts over trait

optima and are subject to natural and sexual selection leading to a range of phenotypic sexual differences (Barrett and Hough 2013). Studies of differential male and female gene expression patterns in plants (Muyle et al. 2017) indicate that sex-biased gene expression plays a role in the evolution of sexual dimorphism in morphological (e.g., anther and ovule development pathways in asparagus, Harkess et al. (2015)), physiological (e.g., salinity tolerance in poplars, Jiang et al. (2012)) and ecological traits (e.g., response to fungal infection in *Silene latifolia*, Zemp et al. (2015)).

Extensive studies in plants and animals have shown that genes with sex-biased expression vary in abundance across different developmental stages and tissues (Grath and Parsch 2016; Perry et al. 2014; Robinson et al. 2014; Zemp et al. 2016; Zluvova et al. 2010). Evolutionary dynamics analyses also indicate that different evolutionary pressures impact the rate of sequence evolution of sex-biased genes; for example, sex-biased genes in reproductive tissues tend to have different rates of protein evolution compared to unbiased genes (Dean et al. 2017; Lipinska et al. 2015; Mank et al. 2010a; Perry et al. 2014; Sharma et al. 2014). In animal systems, where the rates of sequence divergence of sex-biased genes have been studied more widely, male-biased genes in many species, including *Drosophila* and adult birds, tend to be more numerous and to have higher expression and divergence rates (Assis et al. 2012; Grath and Parsch 2016; Harrison et al. 2015; Khaitovich et al. 2005) compared to female-biased and unbiased genes. This has often been interpreted as the signature of sexual selection, particularly sperm competition (Ellegren and Parsch 2007). However, studies in other organisms have reported elevated rates of evolution in female-

biased genes (Mank et al. 2010a; Whittle and Johannesson 2013), leading to questions about the relationship between rates of evolution and sexual selection. In *Arabidopsis*, genes expressed in pollen have lower rates of evolution (Gossmann et al. 2014). Moreover, non-adaptive evolutionary processes have been shown to drive the fast rates of sequence evolution observed in sex-biased genes in some systems (Gershoni and Pietrokovski 2014; Harrison et al. 2015) perhaps related to relaxed purifying selection (Hunt et al. 2011).

Sexual selection in flowering plants is also thought to be strong (Moore and Pannell 2011), acting on gene expression patterns predominantly through pollen competition. Male gametophytic tissue in *Arabidopsis thaliana* and rice has been shown to express a higher proportion of recently evolved genes compared to other tissues (Cui et al. 2015). Some of these young genes possess essential pollen-specific functions, suggesting a role for pollen competition in facilitating de novo gene development. As male-biased mutation is thought to be strong due to the elevated numbers of germ cell divisions in male cells (Whittle and Johnston 2003), pollen competition, in this case, was suggested to counteract the potentially negative effects of higher mutation rates present in male gametophytes (Cui et al. 2015). Similarly, younger genes in the gametophyte of *A. thaliana*, rice and soya bean were also found to have higher rates of evolution compared to genes in the sporophytic tissue, however to varying degrees in males and females (Gossmann et al. 2016). Suggested reasons for these findings concerned the lower tissue complexity, and hence lower genetic interaction, in the gametophyte as well as differences between the sexes.

128

Plants additionally differ from animals in having a longer haploid phase in their life cycle, suggesting that haploid selection may act more forcefully to remove mildly deleterious recessive variation in pollen-expressed genes. Previous work on *A. thaliana* showed that the predominance of selfing, and similarly the intragametophytic selfing in moss species (Szovenyi et al. 2014), leads to the more effective purging of mildly deleterious recessive variation (Gossmann et al. 2014). In the obligate outcrossing plant *Capsella grandiflora*, pollen-specific genes, but not sperm-enriched genes, evolve under both stronger purifying and positive selection compared to genes from sporophytic tissues (Arunkumar et al. 2013). These findings are indicative of a potential combined effect of haploid selection and pollen competition acting in pollen-specific cells, whereas selective pressures are expected to be more relaxed for sperm-specific genes as there is no competition between them (Arunkumar et al. 2013).

These studies make it increasingly clear that many evolutionary forces shape the sequence evolution of sex-biased genes, including sexual selection through sperm competition (Ellegren and Parsch 2007), haploid selection and natural selection (Ingvarsson 2010). Particularly in plants, in order to understand the relative contribution of these forces, it is important to study rates of evolution in species with different levels of gamete competition, motivating studies on outcrossing dioecious species.

The basket willow, *Salix viminalis*, is a dioecious woody angiosperm (Cronk et al. 2015), belonging, together with other willow and poplar (*Populus*) species, to the Salicaceae family. *S. viminalis* is characterised by rapid seed development and growth (Ghelardini et al. 2014); it is both insect- and wind-

pollinated (Peeters and Totland 1999); and it has a recently evolved ZW sex chromosome system (Pucholt et al. 2017). Willow and poplar species have reproductive structures characterised by clusters of unisexual inflorescences referred to as catkins (Fig. 5.1). Flowers in male willow catkins present a reduced number of stamens with anthers and filaments; however, they lack a vestigial ovary, indicating floral reduction compared to other related non-catkin-bearing dioecious species (Cronk et al. 2015; Fisher 1928). Flowers in female willow catkins contain pistils with style, stigma and an ovary. However, they also show a high degree of floral reduction as there is an absence of staminodes and, similarly to male catkins, they lack a perianth with petals and sepals (Cronk et al. 2015; Fisher 1928; Karrenberg et al. 2002), potentially with a role in facilitating wind pollination (Karrenberg et al. 2002).
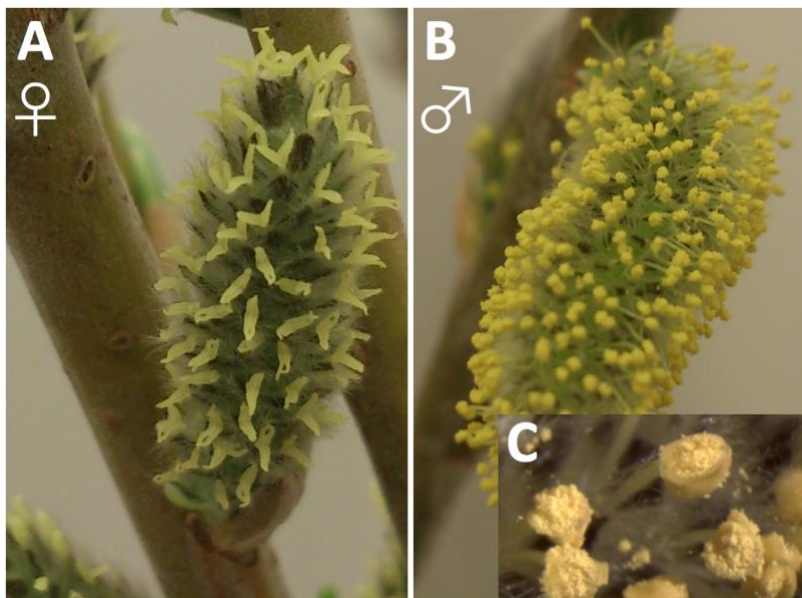
Fig. 5.1. Physical appearance of adult *Salix viminalis* catkins. (A) Female catkin with protruding pistillate flowers. (B) Male catkins with protruding staminate flowers. (C) Anthers of male catkins abundant in pollen grains.

My study of gene expression patterns in male and female *S. viminalis* individuals begins to explore the selective forces acting on sex-biased gene evolution in dioecious plants. I analysed sex-biased gene expression patterns in *S. viminalis* from two different tissues, vegetative (leaf) and sex-specific reproductive (catkin) tissue. I found the reproductive tissue to be more transcriptionally dimorphic and identified overall higher expression levels for male-biased genes than for female-biased genes, consistent with previous studies (Grath and Parsch 2016). Interestingly, however, I found that in catkin, male-biased genes on the autosomes and the pseudoautosomal region have significantly lower rates of sequence divergence than both unbiased and female-biased genes. Similarly, female-biased genes show lower rates of sequence evolution in comparison with unbiased genes; however, the difference is not significant. I could not detect any significant differences in the proportion of genes evolving under positive selection between either male-biased or female-biased genes and unbiased genes. The low rates of male-biased sequence evolution could be partly explained by the higher rate of silent mutations in male-biased genes resulting from lower codon usage bias. However, haploid selection would also be expected in this tissue to exert a stronger purifying force to remove deleterious recessive mutations.

5.3 Materials and Methods

5.3.1 Sample collection and sequencing

We obtained RNA-seq data from leaves and catkins from three female (78021, 78195, 78183) and three male (81084, T76, Hallstad 1-84) *S. viminalis* accessions (Pucholt et al. 2017). These accessions represent unrelated germplasm samples collected in Europe and Western Russia that were subsequently planted in a field archive near Uppsala, Sweden, where they were part of the *S. viminalis* association mapping population (Berlin et al. 2014; Hallingback et al. 2016). As previously described (Pucholt et al. 2017), stem cuttings were collected in the field and transferred to a growth chamber with 22°C constant temperature and 18 hr day length. After seven and thirteen days, respectively, fully developed adult catkins and young leaves were collected from each accession. RNA from each accession and tissue was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich Co. LLC) following variant B of the instructions provided by the manufacturer and including an on-column DNase treatment step. One RNA-seq library for each sample was prepared from 1 μg total RNA using the TruSeq stranded mRNA sample preparation kit (Cat# RS-122-2101/2102, Illumina Inc.) including polyA selection. The library preparation was carried out according to the manufacturer's protocol (#15031047, rev E). Sequencing was performed on an Illumina HiSeq 2500 instrument with paired-end 125 bp read length, v4 sequencing chemistry, and all twelve libraries were pooled and sequenced on three lanes. Preparation of the RNA-seq libraries and

sequencing were performed at the SNP&SEQ Technology Platform in Uppsala, Sweden.

We recovered an average of 42 million 125-bp paired-end reads per sample. After assessing data quality with FastQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), I used Trimmomatic v0.36 (Lohse et al. 2012) to remove adaptor sequences and trim the reads, removing regions where the average Phred score in sliding windows of four bases was <15 as well as reads for which the leading/trailing bases had a Phred score <3. Following trimming, I removed paired-end reads where either read pair was <50 bp (Table S.5.1), resulting in an average of 30 million paired-end reads per sample.

5.3.2 Expression analysis

I mapped RNA-seq reads to the de novo male genome assembly (Pucholt et al. 2017) using HISAT2 v2.0.4 (Kim et al. 2015), filtering reads with unpaired (-no-mixed option) or discordant (-no-discordant option) alignments. To generate a reference transcriptome, I sorted and converted alignment output sam files into bam files using SAMtools v1.2 (Li et al. 2009) and extracted gene coordinates for each sample using StringTie v1.2.4 (Pertea et al. 2015) with default parameters. I then merged output GTF files of all samples to obtain a non-redundant set of transcript coordinates and used BEDTools getfasta to extract sequences (Quinlan and Hall 2010). I filtered ncRNA by BLASTing transcript sequences to

the *Arabidopsis thaliana* ncRNA (Ensembl Plants 32; Flicek et al. (2014)) using BLASTn and an e-value cutoff of $10e^{-10}$.

I extracted read alignments for transcripts in each sample and tissue separately from the filtered transcriptome reference using StringTie and obtained read counts using HTSeq v.0.6.1 (Anders et al. 2015). I estimated RPKM values using EdgeR (Robinson et al. 2010) in R (R Core Team 2015) and filtered transcripts for a minimum expression threshold of 2 RPKM in at least half of the individuals in one sex (in this case, at least two of the three individuals per each sex) as per previous similar studies (Harrison et al. 2015; Pointer et al. 2013). I only retained transcripts with positional information on annotated chromosomes (Pucholt et al. 2017) for further analysis and normalised separately for each tissue using TMM in EdgeR.

I performed hierarchical clustering of average gene expression for genes expressed in both tissues with bootstrap resampling (1,000 replicates) in the R package Pvclust v.2.0 (R Core Team 2015; Suzuki and Shimodaira 2006). I generated a heatmap of $\log_2$ average male and female expression in the two tissues using the R package Pheatmap v.1.0.7 (Kolde 2012; R Core Team 2015).

I identified sex-biased expression based on a minimum of twofold differential expression ($\log_2$ M:F RPKM >1 for male-biased expression and <−1 for female-biased expression) and a significant p value (p_{adj} <0.05 following FDR correction for multiple testing Benjamini and Hochberg (1995)) in EdgeR.

5.3.3 Sequence divergence analysis

Additional to *S. viminalis*, I obtained coding sequences for *P. trichocarpa* from Ensembl Plants 32 (Flicek et al. 2014), *Populus tremula* and *Populus tremuloides* from PopGenIE (Sundell et al. 2015) and *Salix suchowensis* (http://115.29.234.170/willow/, Dai et al. (2014)). I identified the longest transcript for each gene in each species, and I used a reciprocal BLASTn with an e-value cut-off of $10e^{-10}$ and a minimum percentage identity of 30% to identify orthologs. I used BLASTx to obtain open reading frames of the identified orthologous groups, which I aligned with Prank v140603 (Löytynoja and Goldman 2008), using the rooted tree ((*Salixviminalis*, *Salixsuchowensis*), (((*Populustremula*, *Populutremuloides*), *Populustrichocarpa*)). Finally, I removed gaps from the alignments.

To ensure the accurate calculation of divergence estimates, I masked poorly aligned regions with SWAMP v 31-03-14 (Harrison et al. 2014). I employed a two-step masking approach, first using a shorter window size to exclude sequencing errors causing short stretches of nonsynonymous substitutions and then a large window size to remove alignment errors caused by variation in exon splicing (Harrison et al. 2014). Specifically, I first masked regions with more than seven nonsynonymous substitutions in a sliding window scan of 15 codons, followed by a second masking where more than two nonsynonymous substitutions were present in a sliding window scan of four codons. To choose these thresholds, I imposed a range of masking criteria on our data set and conducted the branch-site test on these test data sets. I manually observed the alignment of genes with the highest log likelihood scores to choose the most

efficient and appropriate masking criteria. I subsequently removed genes where the alignment (after removal of gaps and masked regions) was <300 bp, which likely represent incomplete sequences. This resulted in 7,631 1:1 orthologs.

I tested the robustness of the 1:1 orthologs data set by separately inferring orthologous groups using OrthoMCL (Li et al. 2003), an approach with higher specificity (Altenhoff and Dessimoz 2009). I first obtained protein sequences for *P. trichocarpa* from Ensembl Plants 32 (Flicek et al., 2014) and for *P. tremula* and *P. tremuloides* from PopGenIE (Sundell et al., 2015). As the protein sequences for *S. suchowensis* were not available, I used AUGUSTUS (Stanke and Morgenstern, 2005) to obtain the protein sequences for *S. suchowensis* and *S. viminalis* from the coding sequences of the longest transcripts of their genes. I ran AUGUSTUS with default parameters, *Arabidopsis thaliana* as the reference species and with the option of only predicting complete genes. I used the protein sequences for the five species to estimate 1:1 orthologs using OrthoMCL (Li et al., 2003), following the user specifications.

As OrthoMCL relies on the Markov Clustering algorithm, it is useful in identifying cases of co-orthology (a duplicate of a gene in one species that is orthologous to a gene in another species) within the total 1:1 orthologous groups identified. By excluding these co-orthologous groups, I recovered fewer 1:1 orthologs (1,346 after filtering for polymorphism and divergence data); however, the divergence results were consistent with my broader data set based on reciprocal blast (Table S.5.2). As such, I concluded that the reciprocal best-hit approach was appropriate to use in this case.

I further used branch model 2 (model = 2, nssites = 0, fixomega = 0, omega = 0.4) from the CODEML package in PAML v4.8 (Yang 2007) to obtain divergence estimates and calculate mean d_N/d_S specifically for the *S. viminalis* branch using the unrooted tree ((*Salixviminalis*, *Salixsuchowensis*), *Populustrichocarpa*, *Populustremula*, *Populustremuloides*). Mutation-saturated sites did not have an effect on the resulting divergence estimates as none of the orthologs had d_S >2 (Axelsson et al. 2008). In addition, I also obtained omega values for each sex-bias gene category by running the CODEML branch model 2 in PAML separately on the concatenated sequences of all genes in each gene category. This approach reduces the influence of codon bias in estimating rates of divergence (Bierne and Eyre-Walker 2003).

Based on their genomic location in the *S. viminalis* genome (Pucholt et al. 2017), I divided orthologs into two groups, orthologs on the autosomes (including the pseudoautosomal region of the Z chromosome) and orthologs on the Z-linked non-recombining region. Because genes on sex chromosomes can exhibit accelerated rates of evolution (Charlesworth et al. 1987), and this may be more often due to non-adaptive processes on Z chromosomes (Mank et al. 2010b; Wright et al. 2015), I analysed rates of evolution separately for autosomal and Z-linked loci. I calculated mean d_N (the number of nonsynonymous substitutions over nonsynonymous sites) and mean d_S (the number of synonymous substitutions over synonymous sites) separately for each group of orthologs as the ratio of the sum of the number of substitutions across all orthologs in that group, resulted from PAML, to the number of sites (d_N = sum D_N/sum N; d_S = sum D_S/sum S). By calculating mean d_N and d_S through this method, the

issue of infinitely high d_N/d_S estimates arising from low d_S sequences and skew from short sequences is avoided (Mank et al. 2007). I used bootstrapping with 1,000 replicates to determine the 95% confidence intervals and pairwise comparisons with 1,000 permutation test replicates to identify significant differences in d_N, d_S and d_N/d_S between the categories.

5.3.4 Polymorphism analysis

I obtained polymorphism data by mapping the RNA-seq reads to the reference genome assembly using STAR aligner v2.5.2b (Dobin et al. 2013) in the two-pass mode and with default parameters, retaining uniquely mapping reads only. I conducted SNP calling using SAMtools mpileup and VarScan v2.3.9 mpileup2snp (Koboldt et al. 2012). I ran SAMtools mpileup with a maximum read depth of 10,000,000 and minimum base quality of 20 for consistency with VarScan minimum coverage filtering. The base alignment quality (BAQ) adjustment was disabled in SAMtools as it imposes a too stringent adjustment of base quality scores (Koboldt et al. 2014). I ran VarScan mpileup2snp with minimum coverage of 20, minimum of three supporting reads, minimum average quality of 20, minimum variant allele frequency of 0.15, minimum frequency for homozygote of 0.85, strand filter on and p value of 0.05. I defined valid SNPs as sites with a coverage ≥20 in at least half of the individuals in one sex (minimum of two of the three individuals in a sex) and a minor allele frequency ≥0.20, identifying a total of 235,106 SNPs. I identified whether SNPs were synonymous or nonsynonymous by matching them to the reading frame.

138

As the divergence and polymorphism analyses use different filtering criteria, I ensured the two data sets were comparable by identifying a set of codons where all sites pass the filtering criteria for both analyses. I only kept codons where (i) all sites pass the minimum coverage threshold of 20 in at least half of the individuals in one sex, (ii) there are no alignment gaps following Prank alignment, and (iii) there were no ambiguity data (Ns) following SWAMP masking. In further analyses, I only used genes with both divergence and polymorphism information. This ensures that the number of synonymous (S) and nonsynonymous sites (N) is identical across divergence and polymorphism analyses, and therefore suitable for McDonald–Kreitman tests. I have therefore used the same number of nonsynonymous (N) and synonymous (S) sites in our calculations of d_N, p_N and, respectively, d_S and p_S.

We calculated mean p_N (number of nonsynonymous polymorphisms over nonsynonymous sites) and mean p_S (number of synonymous polymorphisms over synonymous sites) for each gene category as the ratio of the sum of the number of polymorphisms to the sum of the number of sites (p_N = sum P_N/sum N; p_S = sum P_S/sum S).

5.3.5 Analysis of synonymous codon usage bias

I estimated codon usage bias using codonW (http://codonw.sourceforge.net) through the effective number of codons (ENC) (Wright 1990). The ENC measure determines the degree to which the entire genetic code is used in each gene, ENC values ranging from 20 (indicating extreme bias, where only one codon is used for one amino acid) to 61 (indicating

no bias, where all amino acids are represented equally by all possible codons) (Wright 1990). This measure is not biased by the different lengths of the coding regions being analysed, and as such, it has been shown to be more reliable than other commonly used methods of estimating codon usage bias (Comeron and Aguadé 1998). I calculated the effective number of codons for all the genes with divergence and polymorphism data (Table 5.2).

5.3.6 Tests of positive selection

To identify genes evolving under adaptive evolution, I used the McDonald-Kreitman test (McDonald and Kreitman 1991), which contrasts the ratio of nonsynonymous and synonymous substitutions with polymorphisms. For each gene, I used a 2×2 contingency table and a Fisher's exact test in R to test for deviations from neutrality using numbers of nonsynonymous and synonymous substitutions (D_N, D_S) and polymorphisms (P_N, P_S). As the McDonald–Kreitman test lacks power with low table cell counts, I excluded genes from the analysis if, within the contingency table, the sum over any row or column was less than six (Andolfatto 2008; Begun et al. 2007). For genes with significant deviations in D_N, D_S, P_N and P_S, a higher nonsynonymous-to-synonymous substitutions ratio relative to polymorphisms ratio ($d_N/d_S > p_N/p_S$) represented a signature of positive selection. I then tested for significant differences between sex-biased and unbiased genes in the proportion of genes with signatures of positive selection using Fisher's exact test.

For each gene category, I also used the divergence and polymorphism data to calculate the average direction of selection (DoS) statistic (Stoletzki and

140

Eyre-Walker 2011). For each gene, I calculated DoS as the difference between the proportion of nonsynonymous substitutions and the proportion of nonsynonymous polymorphisms $(DoS = D_N/(D_N + D_S) - P_N/(P_N + P_S))$, where positive DoS values indicate positive selection, a value of zero indicates neutral evolution while negative values indicate purifying selection and segregating deleterious mutations (Stoletzki and Eyre-Walker 2011). Additional to the McDonald–Kreitman test, I also used the DoS statistic to test, using Fisher's exact test, for differences in the proportion of fixed nonsynonymous sites and nonsynonymous polymorphisms.

5.4 Results

5.4.1 Gene expression in catkin and leaf

I mapped RNA-seq reads from two tissues, catkin (reproductive tissue) and leaf (vegetative tissue), of male and female *S. viminalis* individuals to the genome assembly yielding an average of 30 million read mappings per sample after quality control and trimming (Table S.5.1). Following expression filtering, I recovered 8,186 significantly expressed genes in catkin and 7,638 significantly expressed genes in leaf.

I first assessed transcriptional similarity across tissues and sexes using hierarchical clustering of gene expression levels (Fig. 5.2). I found that the reproductive tissue is more transcriptionally dimorphic than the vegetative tissue, consistent with studies in many other species (Jiang and Machado 2009; Mank et al. 2008; Pointer et al. 2013; Yang 2016). Expression for male catkin clusters

most distantly from both male and female expressions in leaf. I identified 3,567 genes (43% of all filtered catkin genes) showing sex-biased expression in catkin ($\log_2$ fold change >1 or <-1, $p_{adj} <0.05$), compared to expression in the vegetative tissue, where I identified only seven (0.09%) leaf sex-biased genes (Fig. 5.3). Even with a more relaxed fold change threshold for defining differentially expressed genes ($\log_2$ fold change >0.5 or <-0.5, $p_{adj} < 0.05$), I still could not identify any additional leaf sex-biased genes. There are also no shared sex-biased genes between the two tissues.
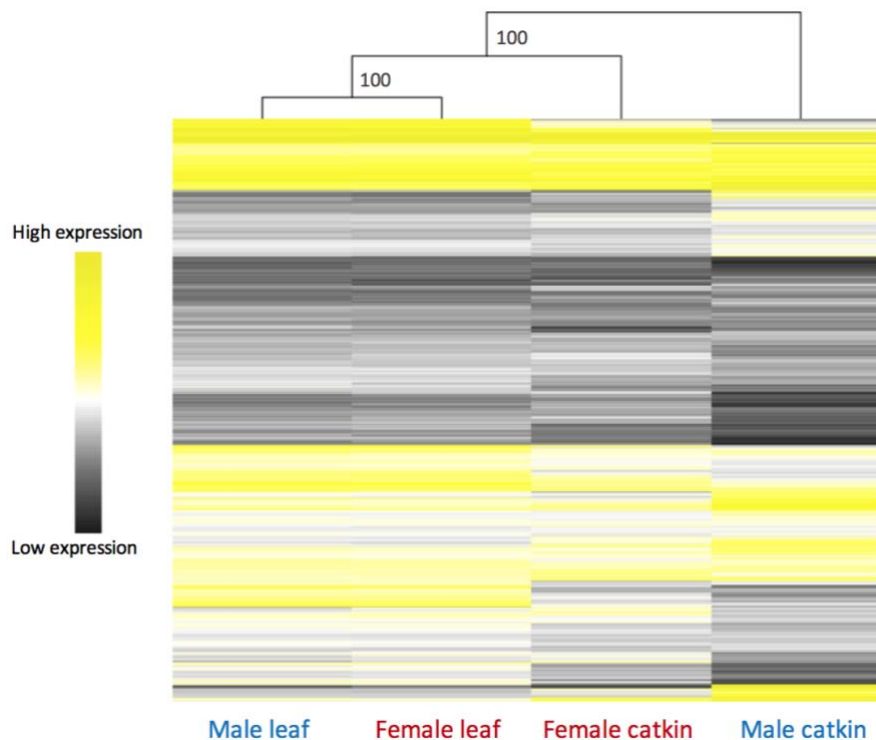
Fig. 5.2. Heatmap and hierarchical clustering of average male (blue) and female (red) gene expression in catkin and leaf. The heatmap represents all the filtered genes expressed in both tissues (7,257). Hierarchical gene clustering is based on Euclidean distance with average linkage for $\log_2$ RPKM expression for each gene. Numbers at nodes represent the 1,000 replicates percentage bootstrap results.

5.4.2 Dynamics of catkin sex-biased gene expression

Although female-biased genes (n = 1,820) are slightly more numerous than male-biased genes (n = 1,747), the magnitude of differential expression ($\log_2$ FC) for male-biased genes is significantly greater than that for female-biased genes (Wilcoxon rank sum test $p < 0.001$). Average male expression for male-biased genes is significantly higher than average female expression for female-biased genes (Fig. 5.3, Wilcoxon rank sum test $p < 0.001$), although male expression for female-biased genes is significantly lower than female expression for male-biased genes (Fig. 5.3, Wilcoxon rank sum test $p < 0.001$).
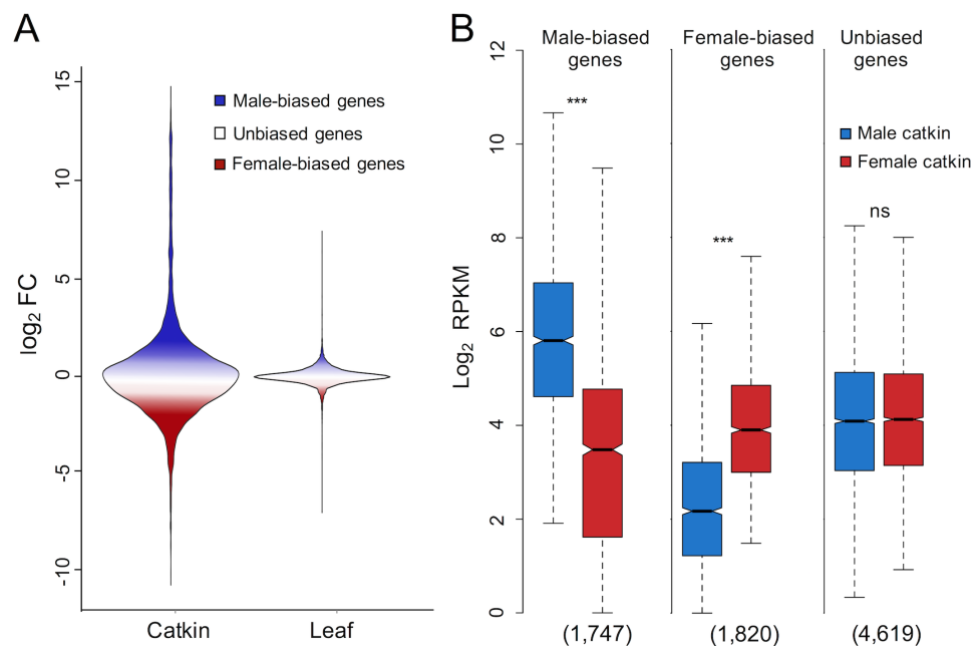
Fig. 5.3. Sex-biased gene expression in *Salix viminalis*. (A) Proportion and range of differentially expressed and unbiased genes in catkin and leaf. (B) Comparison between male and female average expression for sex-biased and unbiased genes in catkin. Numbers in brackets represent the number of genes in each category. Significant differences between male and female expression based on Wilcoxon rank sum tests are denoted (ns= non-significant, ***$p < 0.001$).

I grouped sex-biased genes based on different fold change thresholds and compared average male and female catkin expression for the genes in each category. This analysis suggests that catkin male-biased genes may arise from increased expression in males and decreased expression in females (Fig. 5.4). For female-biased genes, however, there is a decreasing trend in male expression with increasing fold change thresholds but a constant female expression across all thresholds (Fig. 5.4), suggesting that female bias results primarily from downregulation of male expression.

The paucity of sex-biased genes in the leaf tissue makes it a useful comparison to further assess the sex-specific changes that give rise to male- and female-biased genes. I therefore used leaf expression as the putative ancestral expression state. For the subset of catkin sex-biased genes that also had expression in the leaf tissue, I determined the difference in expression between catkin and leaf across the same fold change thresholds used in Fig. 5.4. For male-biased genes in the catkin, I found significant differences between catkin and leaf expression in both sexes, although to a lesser extent in females (Fig. S.5.1). On the other hand, for catkin female-biased genes, I also observe large differences in male expression between catkin and leaf samples; however, I found little to no female expression changes between the two tissues (Fig. S.5.1).
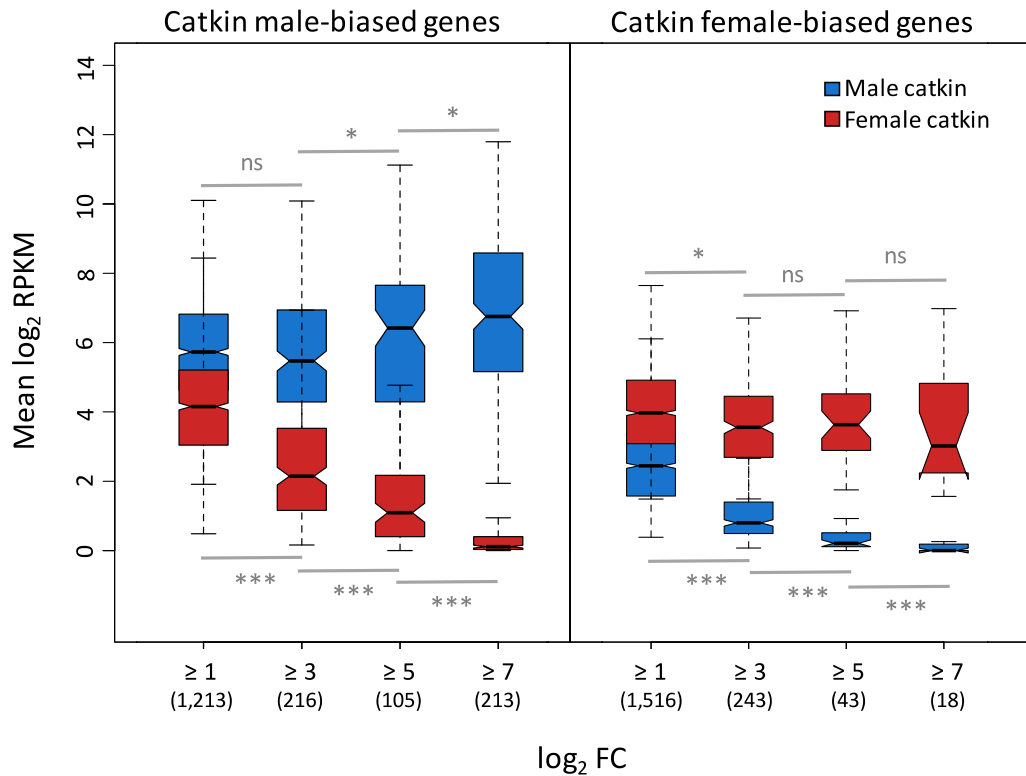
Fig. 5.4. Average male and female catkin gene expression at different sex-bias fold change thresholds for all catkin male-biased and female-biased genes. Numbers in brackets represent the number of genes in each fold change category. Significance level is based on Wilcoxon rank sum tests (ns = non-significant, $^*p < 0.05$, $^{***}p < 0.001$).

I further divided catkin sex-biased genes into autosomal (including the pseudoautosomal region of the sex chromosomes) and Z-linked genes. On the autosomes, I found 3,536 sex-biased genes (1,728 male-biased and 1,808 female-biased genes). On the non-recombining region of the Z chromosome, I found only 31 sex-biased genes (19 male-biased and 12 female-biased genes); however, considering the narrow region of recombination suppression between the sex chromosomes (3.5–8.8 Mbp) (Pucholt et al. 2017), these sex-biased genes represent 44% of the total identified gene content in the non-recombining sex-chromosome region.

145

Table 5.1. Divergence and polymorphism estimates for catkin gene categories on autosomes and the non-recombining Z region

Tissue	Location	Category[a]	n Genes[b]	d_N (95% CI) sig.[c]	d_S (95% CI) sig.[c]	d_N/d_S (95% CI) sig.[c]	p_N (95% CI) sig.[c]	p_S (95% CI) sig.[c]	p_N/p_S (95% CI) sig.[c]	DoS sig.[d]
		UB	1,754	0.0030 (0.0028 - 0.0031)	0.0135 (0.0130 - 0.0141)	0.2204 (0.2101 - 0.2311)	0.0027 (0.0025 - 0.0028)	0.0109 (0.0103 - 0.0114)	0.2456 (0.2328 - 0.2584)	-0.0495
	Autosomes and recombining Z	MB	674	**0.0032 (0.0030 - 0.0035) p = 0.012**	**0.0162 (0.0147 - 0.0187) p < 0.001**	**0.1951 (0.1769 - 0.2157) p < 0.001**	**0.0029 (0.0026 - 0.0032) p = 0.022**	**0.0116 (0.0108 - 0.0124) p = 0.042**	0.2491 (0.2260 - 0.2727) p = 0.682	-0.0346 p = 0.627
		FB	732	0.0031 (0.0029 - 0.0033) p = 0.094	**0.0149 (0.0141 - 0.0158) p < 0.001**	0.2095 (0.1938 - 0.2256) p = 0.082	**0.0030 (0.0028 - 0.0033) p = 0.002**	**0.0121 (0.0113 - 0.0131) p < 0.001**	0.2477 (0.2293 - 0.2666) p = 0.774	-0.0375 p = 0.916
Catkin		UB	12	0.0032 (0.0024 - 0.0043)	0.0102 (0.0056 - 0.0145)	0.3130 (0.2141 - 0.5498)	0.0015 (0.0007 - 0.0032)	0.0045 (0.0022 - 0.0078)	0.3407 (0.1778 - 0.5588)	0.0800
	Non-recombining Z	MB	3	0.0029 (0.0 - 0.0140) p = 0.378	0.0143 (0.0091 - 0.0396) p = 0.730	0.2019 (0.0 - 0.3533) p = 0.244	0.0029 (0.0 - 0.0210) p = 0.396	0.0104 (0.0039 - 0.0505) p = 0.084	0.2781 (0.0 - 0.4151) p = 0.082	0.0088 p = 0.563
		FB	4	0.0032 (0.0021 - 0.0037) p = 0.964	0.0100 (0.0061 - 0.0207) p = 0.926	0.3172 (0.1649 - 0.4996) p = 0.948	**0.0045 (0.0008 - 0.0082) p = 0.026**	**0.0138 (0.0055 - 0.0229) p < 0.001**	0.3243 (0.0845 - 0.4657) p = 0.882	0.0770 p = 0.761

[a] Unbiased (UB), male-biased (MB) and female-biased (FB) genes.

[b] Number of genes with both divergence and polymorphism data.

[c] p values based on 1,000 replicates permutation tests comparing male-biased and female-biased genes with unbiased genes. Significant p values (< 0.05) are shown in bold.

[d] p values from Wilcoxon nonparametric tests comparing male-biased and female-biased genes with unbiased genes. Significant p values (< 0.05) are shown in bold.

5.4.3 Rates of evolution

I compared the overall ratios of nonsynonymous-to-synonymous nucleotide substitutions (d_N/d_S) between catkin and leaf and found no significant differences between the two ($p = 0.476$, significance based on permutation tests with 1,000 replicates). I also did not find a significant difference in the evolution of unbiased genes between the two tissues ($p = 0.056$ from permutation tests with 1,000 replicates), likely influenced by the large overlap of genes between them (97% of catkin unbiased genes represent 58% of the unbiased genes expressed in leaf). I found too few significantly sex-biased genes in the leaf tissue to make any statistical comparisons of rates of sequence evolution between catkin and leaf sex-biased genes.

I also compared the ratio of d_N/d_S between sex-biased and unbiased genes in catkin to test for differences in the rate of evolutionary divergence. Interestingly, I found that on autosomes, although male-biased genes have more amino acid substitutions than both unbiased and female-biased genes, as shown by significantly higher d_N values, d_N/d_S for male-biased genes was significantly lower, indicating slower rates of functional evolution relative to unbiased (Table 5.1; Table S.5.2) and female-biased genes ($p < 0.001$, significance based on permutation tests with 1,000 replicates). Similar results were obtained when we estimated d_N/d_S from a data set of 1:1 orthologs that excluded cases of co-orthology (Table S.5.2), as well as from omega values resulting from running CODEML branch model 2 in PAML on concatenated sequences of genes in each sex-bias gene category (Table S.5.3). This lower d_N/d_S ratio is caused in part by

a disproportionate increase in synonymous substitutions compared to nonsynonymous substitutions, causing the relationship between d_N and d_S in male-biased genes to lie further away from direct proportionality than in the case of unbiased genes (Fig. S.5.2).

Female-biased autosomal loci also show the same pattern as male-biased genes relative to unbiased genes; however, this result was is significant (Table 5.1; Table S.5.2). On the non-recombining Z, male-biased genes also show lower rates of evolution compared to unbiased genes; however, this finding is not significant, likely due to the small sample size of male-biased genes (n = 3). In contrast, female-biased Z-linked loci show accelerated rates of evolution in comparison with male-biased Z-linked genes ($p < 0.001$, significance based on permutation tests with 1,000 replicates).

Highly expressed genes are often observed to exhibit lower d_N/d_S values (Cherry 2010; Drummond et al. 2005; Pál et al. 2001; Slotte et al. 2011); therefore, to determine whether expression level might explain my results, I divided sex-biased and unbiased genes into quartiles based on overall expression. As expected, I found that as gene expression level increases, the rate of sequence divergence decreases, and this holds true for both sex-biased and unbiased genes (Fig. S.5.3). To further investigate the effect of expression level on the variation in rates of sequence divergence between sex-bias categories, I used a multiple regression analysis to predict d_N/d_S results based on expression level and degree of sex-bias. For defining the degree of sex-bias, I classed genes into five groups, highly female-biased genes (FC≤−3), lowly female-biased genes (−3<FC≤−1), unbiased genes (−1<FC<1), lowly male-biased genes (1≤FC< 3)

148

and highly male-biased genes (FC≥3). I found a significant negative relationship between d_N/d_S values and both average $\log_2$ RPKM expression level ($\beta = -.03$, $p < .001$) and degree of sex-bias ($\beta = -.04$, $p = .014$). There was no significant effect of the interaction between expression level and degree of sex-bias on d_N/d_S results, suggesting that any differences in the rates of sequence evolution due to sex-bias are independent of the gene expression level for each sex-bias category. Despite these results, the adjusted r^2 was very low ($r^2 = 0.01$), indicating that other factors, such as purifying or haploid selection, largely explain the vast majority of sequence divergence results.

I also estimated average levels of synonymous codon usage bias for sex-biased and unbiased genes to determine whether this could explain the differences in the rates of synonymous substitutions between the gene categories. Stronger codon usage bias has been associated with higher gene expression as selective forces act to increase translational efficiency (Duret 2002; Ingvarsson 2010). Codon bias has also been shown to differ between differentially expressed genes, with male-biased genes undergoing weaker codon usage bias than female-biased (Magnusson et al. 2011; Mank et al. 2008) however, this varied across different developmental stages (Whittle et al. 2007) and unbiased genes (Hambuch and Parsch 2005). Additionally, greater codon bias has been estimated for genes with lower rates of synonymous substitutions (Urrutia and Hurst 2001).

I estimated codon usage bias for genes in each category through the effective number of codons (ENC), where stronger codon bias is indicated by lower ENC values. The differences in codon bias between the different gene

categories are subtle, and the gene frequency spectra for all categories are distributed towards the higher end of the effective number of codons (ENC), hence lower codon usage bias (Fig. S.5.4). However, male-biased genes have a significantly lower codon usage bias than both unbiased (Table 5.2) and female-biased genes ($p < 0.001$, significance based on permutation tests with 1,000 replicates). These findings, together with the higher rates of synonymous substitutions in male-biased genes compared to unbiased and female-biased genes, indicate weaker purifying selection on silent mutations in male-biased genes (Sharp and Li 1987).

Table 5.2. Codon usage bias for catkin sex-bias gene categories

Tissue	Location	Catgeory	n Genes[a]	ENC[b] sig.[c]
Catkin	Autosomes and recombining Z	Unbiased	1,754	52.15
		Male-biased	674	52.71 (**$p < 0.001$**)
		Female-biased	732	52.20 ($p = 0.588$)

[a] Number of genes with both divergence and polymorphism data.

[b] Average effective number of codons for each gene category.

[c] p values based on 1,000 replicates permutation test comparing male-biased and female-biased genes relative to unbiased genes. Significant p values (<0.05) are shown in bold.

I used polymorphism data to calculate the ratio of nonsynonymous-to-synonymous polymorphisms (p_N/p_S). Sex-biased genes on both autosomes and the non-recombining Z region have significantly higher nonsynonymous and synonymous polymorphism levels compared to unbiased genes; however, the p_N/p_S ratio is not significantly different in either of the comparisons (Table 5.1).

150

To distinguish between the selective pressures acting on sequence evolution, I used the McDonald–Kreitman test of selection, comparing the ratios of d_N/d_S to p_N/p_S for each gene category. Following filtering, I recovered six unbiased, one male-biased and two female-biased genes showing signatures of positive selection (Table 5.3). However, there is no significant difference in the proportion of genes evolving under positive selection between either of the gene categories (Table 5.3, significance denoted in table). Because the McDonald–Kreitman test is extremely conservative, I also assessed selection pressures on sex-biased genes using the direction of selection test (Stoletzki and Eyre-Walker 2011). Through the DoS statistic, I recovered 681 unbiased, 262 male-biased and 282 female-biased genes under putative positive selection (DoS > 0), yet, consistent with the McDonald–Kreitman test, I found no significant differences in the proportion of genes evolving under positive selection (Fisher's exact test $p > 0.9$ for both female-biased and male-biased genes in comparison with unbiased genes). Taken together, the divergence and polymorphism analyses, through tests of positive selection, suggest that the lower rates of sequence evolution seen in male-biased genes could be due to purifying selection acting to remove deleterious recessive mutations.

Table 5.3. McDonald-Kreitman test of selection

Tissue	Location	Category	n Genes[a]	Positive selection[b] sig.[c]
Catkin	Autosomes and recombining Z	Unbiased	1,766	6
		Male-biased	677	1 ns
		Female-biased	736	2 ns

[a] Number of genes with both divergence and polymorphism data.

[b] Number of genes with significant positive selection indicated by significant deviations in D_N, D_S, P_N and P_S and $d_N/d_S > p_N/p_S$.

[c] Significance based on Fisher's exact test comparing sex-biased to unbiased genes (ns = non-significant).

5.5 Discussion

The evolution of sex-biased gene sequence has been extensively analysed in animal systems. In contrast, far less is known about the evolution of sex-biased genes in plants in general and in dioecious angiosperms in particular. Previous work in *A. thaliana*, an annual and largely selfing hermaphroditic species, found low rates of evolution in pollen-expressed genes, although with evidence of a higher proportion of sites under positive selection (Gossmann et al. 2014). This could be the result of the greater haploid selection in plants; however, it could also be, at least partially, the result of the selfing mating system in this species, which leads to the purging of recessive deleterious variation. Similarly, in the self-incompatible close relative of *A. thaliana*, *C. grandiflora*, a larger fraction of pollen-specific genes was found to evolve under strong purifying selection and to also exhibit faster protein evolution rates compared to

152

sporophytic genes (Arunkumar et al. 2013). This is suggested to be the result of both higher pollen competition and the haploid nature of the pollen-specific tissue.

Here, I investigate the evolution of sex-biased genes in *S. viminalis*, a perennial dioecious (obligate outcrossing) species with partial wind pollination. Similarly to *C. grandiflora* (Kao and McCubbin 1996), *S. viminalis* theoretically experiences far higher levels of pollen competition than *A. thaliana*, particularly inter-male competition. Although I might expect the high levels of sperm competition in *S. viminalis* to produce higher rates of protein evolution for male-biased genes, I observe the opposite. Moreover, in contrast to work in *C. grandiflora* (Arunkumar et al. 2013), I did not find evidence of a high proportion of male-biased genes under positive selection.

The observed dynamics of sex-biased gene expression in *S. viminalis* is consistent with previous reports in a wide range of species. Equivalent to studies on somatic and reproductive tissues in animal systems (Mank 2017; Pointer et al. 2013; Yang 2016), I found that the reproductive tissue is far more transcriptionally dimorphic than the vegetative tissue (Fig. 5.2 and 5.3). Additionally, in plant species in particular, very few studies have been able to identify any significant sex-biased genes in nonreproductive tissues (Robinson et al. 2014; Zemp et al. 2014; Zluvova et al. 2010). I also found that, in catkin, male-biased genes are expressed at significantly higher levels and have a higher magnitude of sex-bias than female-biased genes (Fig. 5.3). The level of sex-biased gene expression found in the *S. viminalis* reproductive tissue is markedly lower than that in animal species (Jiang and Machado 2009; Pointer et al. 2013), consistent with the significantly higher degree of sexual dimorphism in animal systems. On the other

hand, I found a larger percentage of sex-biased genes compared to several plant and algae species with low levels of sexual dimorphism (Harkess et al. 2015; Lipinska et al. 2015; Zemp et al. 2016). This is indicative of higher intersexual morphological differences in the *S. viminalis* reproductive tissue, which is consistent with previous descriptions of the structural differences between male and female catkins (Cronk et al. 2015).

Contrary to findings from the dioecious *Silene latifolia* (Zemp et al. 2016), however similarly to reports from animal and algae systems (Lipinska et al. 2015; Perry et al. 2014), my results indicate that sex-biased gene expression has likely evolved as an outcome of expression changes in males (Fig. S.5.1). This would also explain why catkin male samples are more transcriptionally different than catkin female samples with respect to leaf samples (Fig. S.2). These results suggest that ancestral intralocus sexual conflict may have been more detrimental to males, leading to the evolution of sex-biased gene expression in order to resolve such conflicts.

Additionally, although not statistically significant, I found that male-biased genes have higher p_N/p_S values compared to both unbiased and female-biased genes, which is in stark contrast to divergence results where I found male-biased genes to have significantly lower d_N/d_S values. Given that perturbations in population size can alter estimates of polymorphism (Pool and Nielsen 2007; Tajima 1989), it is difficult to assess the causes of the contrasting results between d_N/d_S and p_N/p_S estimates for sex-biased genes. Nevertheless, divergence estimates are less sensitive to demographic fluctuations and I more

strongly rely on this measurement in my analyses of evolutionary rates of sex-biased genes.

Sex-biased genes in willow exhibit higher expression levels than unbiased genes, and highly expressed male-biased and female-biased genes have significantly lower rates of evolution than unbiased and lowly expressed sex-biased genes (Fig. S.5.3). The fact that highly expressed genes evolve more slowly could be due to a range of different reasons, which are still highly debated (Drummond et al. 2005). The structural or functional features of the proteins they encode (Drummond et al. 2005), high pleiotropic constraints acting on the genes (Pál et al. 2001) as well as gene conversion events (Petes and Hill 1998) have all been suggested as potential mechanisms through which highly expressed genes could have lower rates of sequence evolution. Although the high expression of many sex-biased genes in *S. viminalis* may partially explain their slower rates of evolution, my analysis reveals a very weak correlation between expression level and rate of evolution, indicating that, in this case, expression level does not largely explain the low rates of sex-biased gene evolution.

It is interesting that the lower d_N/d_S values of male-biased genes are associated with an overall increase in synonymous mutations relative to nonsynonymous mutations (Fig. S.5.2). This, plus the observation that male-biased genes experience lower levels of codon usage bias (Table 5.2), could suggest that the d_N/d_S results are influenced by different levels of codon usage across gene expression categories. Different selection forces are thought to lead to codon usage bias, such as positive selection for preferred synonymous mutations (mutations that lead to preferred codons) and purifying selection acting

on unfavourable mutations, preventing a decrease in the frequency of preferred codons (Hershberg and Petrov 2008). Despite previous expectations that selection acting at synonymous sites is weak (Akashi 1995; Hershberg and Petrov 2008), several studies suggest that a range of selection strengths, spanning from weak to strong selection, influence the evolution of synonymous mutations, and hence codon usage bias measures (Hershberg and Petrov 2008; Lawrie et al. 2013). However, although differential codon bias across expression categories has the potential to influence my d_N/d_S estimates, my additional PAML analysis (Table S.5.3) indicates that this is not likely to be the case.

Similar to the findings from *A. thaliana* and C*. grandiflora*, the unusual rates of evolution of sex-biased genes in *S. viminalis* could also be explained by the differential selection pressures acting on diploid versus haploid life stages. Haploid selection (Joseph and Kirkpatrick 2004) is more effective at removing recessive deleterious mutations than selection in the diploid life stages, where dominant alleles can mask the effects of deleterious recessive alleles (Kondrashov and Crow 1991). Although all predominantly diploid organisms pass through both haploid and diploid phases, animal species employ different mechanisms through which selection on the haploid stage is minimised (Otto et al. 2015). Not only can aneuploid spermatids still be potentially viable (Lindsley and Grell 1969), indicating limited haploid expression, but studies in mice have shown that genetically haploid spermatids evade haploid selection by sharing gene products through cytoplasmic bridges (Erickson 1973), becoming thus phenotypically diploid (Braun et al. 1989).

156

Haploid selection is far more extensive in plants due to both the larger proportion of the life cycle spent in the haploid phase and active gene transcription, which has been observed in gametes, particularly in pollen (Otto et al. 2015). In addition to haploid selection, male gametophytes in angiosperm species are under strong sexual selection pressures (Erbar 2003; Snow and Spira 1996), particularly in outcrossing species. Mechanisms of sexual selection in angiosperms include pollen tube and pistil interactions and pollen competition over ovules, which is exacerbated in outcrossing species (Bernasconi et al. 2004).

It is important to note that the reduced floral structure and microscopic nature of the catkin (Cronk et al. 2015) makes it nearly impossible to separate haploid from diploid reproductive tissue in this species. However, our catkin preparations are highly enriched for haploid cells (Fig. 5.1), when compared to the vegetative samples. I expect that rates of evolution for purely haploid sex-biased tissue to be even lower than what I observe if haploid selection is indeed the primary cause of the slower rates of evolution.

Apart from insect pollen dispersal, willows also have wind-dispersed pollination (Peeters and Totland 1999) and experience high levels of pollen competition. The observed patterns of gene sequence evolution in *S. viminalis* support the notion that pollen competition in conjunction with haploid selection produces greater levels of purifying selection on male-biased genes. This would remove deleterious variation and lead to significantly slower rates of functional gene sequence evolution. Interestingly, the algae *Ectocarpus*, a species where sex-biased genes are subject almost entirely to haploid selection, shows accelerated rates of evolution for both male- and female-biased genes

(Lipinska et al. 2015). This suggests that haploid selection may not be the only force that influences the rate of evolution of sex-biased genes in haploid cells. Indeed, data from haploid-specific genes (pollen-specific genes in *S. viminalis*) would help to more precisely determine the degree to which the currently observed lower rates of evolution of male-biased genes can be explained by haploid selection or other factors such as expression breath (Arunkumar et al. 2013; Gossmann et al. 2014; Szovenyi et al. 2013).

In summary, my findings are generally consistent with previous reports on patterns of sex-bias gene expression in plant and animal species. However, distinct forces may differentiate rates of sex-biased gene sequence evolution between animal and plant systems. The reduction in haploid selection in animals may limit the power of purifying selection to remove mildly deleterious variation, particularly when it is largely recessive. In *S. viminalis,* I observe reduced rates of evolution for male-biased genes, consistent with increased purifying selection from the extended haploid phase. Even though male-biased genes show relaxed levels of codon bias, this does not seem to be a major driver of the reduced rate of evolution. Future work should focus on investigating the differences in the relative strength of haploid versus diploid selection in dioecious angiosperm species in shaping the evolution of sex-biased genes.

Chapter 6. Discussion

Males and females are often subject to contrasting selection pressures, despite sharing the majority of their genes (Arnqvist and Rowe 2005). Sex-specific evolutionary forces can act on a large proportion of the genome, affecting genome structure, gene sequence and transcription rates (Connallon and Clark 2010). The evolution of sex chromosomes and of sex-biased gene expression are two main routes to resolving conflict between male- and female-specific selection. In this thesis, I combine genomic and transcriptomic data from males and females across multiple related poeciliid species to characterise the structure and conservation of different sex chromosome systems and assess the dynamics of sex chromosome recombination suppression and divergence. I additionally explore the selective forces driving the evolution of sex-biased gene expression, contrasting between selective regimes in animals and plants.

6.1 Dynamics of sex chromosome recombination suppression and degeneration across poeciliid species

The classic model of sex chromosome evolution posits that arrested recombination between the sex chromosomes is driven by sexual antagonism, as linkage between the sex-determining locus and a nearby gene with sex-specific benefits can resolve sexual conflict and allow males and females to reach their respective reproductive optima (Bull 1983; Charlesworth et al. 2005; Fisher 1931; Rice 1987). Both theoretical and empirical studies have shown that

sexually antagonistic mutations are favoured to accumulate on the sex chromosomes (Charlesworth 1978; Chippindale et al. 2001; Gibson et al. 2002; Rice 1987; Rice 1992), however whether this promotes sex chromosome differentiation or is, in fact, the result of recombination suppression remains unclear. Additionally, sexual antagonism is predicted to facilitate sex chromosome turnover. Establishing a new sex chromosome system can be achieved by linking a novel or existing sex determining locus to an autosomal sexually antagonistic gene (Van Doorn and Kirkpatrick 2010). In cichlids for example, a novel ZW sex determining system containing a sexually antagonistic gene with benefits in females has invaded a population with an ancestral male sex determining system (Roberts et al. 2016). However, in this case also, it is unclear if the establishment of the novel sex chromosome system predates or is a result of the W-linkage of the sexually antagonistic locus.

Here, we provide support for the sexual antagonism model of sex chromosome evolution by analysing the structure and differentiation of sex chromosomes across guppy populations experiencing different degrees of sexual conflict for colour (Chapter 2). Female preference for bright male colouration is stronger in low-predation populations (Breden and Stoner 1987), leading to higher sexual conflict and greater levels of male colouration (Endler 1980). By analysing patterns of recombination suppression on the sex chromosomes of high and low predation populations we show that the strength of sexual antagonism correlates with the extent of sex chromosome recombination suppression (Chapter 2, Wright et al. (2017)). Specifically, low predation populations which experience stronger sexual conflict have sex chromosomes

160

that recombine less and that are more differentiated than high predation populations (Chapter 2, Wright et al. (2017)).

In addition to the intra-specific variation in guppy sex chromosome differentiation, my study also reveals an extreme heterogeneity in patterns of sex chromosome divergence across related poeciliid species. My comparative analysis of sex chromosomes in *P. reticulata*, *P. wingei* and *P. picta* indicate that these species share the same male heterogametic system, yet show substantial variation in the extent of recombination suppression and Y chromosome degeneration (Chapter 3, Darolti et al. (2019)). More specifically, the sex chromosomes in *P. reticulata* and *P. wingei* are largely homomorphic, while in *P. picta* they are highly diverged and degenerated. Remarkably, I demonstrate that the profound Y decay in *P. picta* has led to the evolution of chromosome-wide dosage compensation in this species, the first such documented case in teleost fish. The evolution of dosage compensation has the potential to further create conflict between the sexes over optimal gene expression for the homomorphic sex chromosome (Mank et al. 2014). In this case, however, I find that sexual conflict is avoided as dosage compensation is achieved through the upregulation of the active X chromosome in males, without an impact on female gene expression.

The variation between species in the extent of recombination arrest and degeneration of the shared poeciliid system leads to questions about the evolutionary forces driving sex chromosome differentiation in some but not other species. Evidence from numerous species shows that recombination suppression and degeneration does not always correlate with the age of the sex

chromosomes (Ahmed et al. 2014; Stock et al. 2011; Vicoso et al. 2013a; Vicoso et al. 2013b; Xu et al. 2018), and that different mechanisms, including rare recombination (Stock et al. 2011; Stock et al. 2013), gene conversion (Trombetta et al. 2009; Wright et al. 2014) and sexual antagonism (Vicoso et al. 2013b), can be underlying this effect. My analysis of sex-linked genes in *P. reticulata* and *P. wingei* confirms that persistent X-Y recombination takes place throughout the length of the sex chromosomes, preventing Y chromosome degeneration despite the significant age of the sex chromosome system (Chapter 4).

Taken together, my results reveal a substantial intra- and inter-specific variation in poeciliid sex chromosome recombination suppression and degeneration. These findings highlight the dynamic nature of sex chromosomes and offer important insight into the evolutionary processes of sex chromosome formation and degeneration.

<u>Future study direction</u>

While I demonstrate that, despite being largely non-recombining, the sex chromosomes in *P. reticulata* and *P. wingei* maintain high X-Y sequence similarity through rare recombination events, it remains to be determined why the *P. picta* sex chromosomes are so degenerate. Inferring sex-linked genes through pedigree analysis in *P. picta* and identifying orthology to *P. reticulata* and *P. wingei* sequence would make it possible to confirm whether the sex chromosome system has indeed formed ancestrally to these species or instead evolved independently. Analyses of sex-linked sequence and recombination suppression in *P. picta* would also allow to determine the forces affecting Y chromosome

162

integrity in this species. The three poeciliid species could also experience different strength or focus of sexual conflict, which could contribute to species-specific variation in sex chromosome differentiation. Identifying regions of sexual conflict within the genome would help test this hypothesis.

6.2 Contrasting evolutionary forces in plants and animals shape distinct patterns of sex-biased gene expression

The evolution of sexual dimorphism is primarily facilitated by the regulation of sex-specific gene expression (Grath and Parsch 2016). Sex-biased genes in some animal and plant species share multiple characteristics, including abundance in adult and reproductive tissues, predominant distribution across sex chromosomes and excess of male-biased genes (Grath and Parsch 2016). Recent studies however are beginning to uncover the emerging pattern that distinct selective forces in animals and plants shape the evolution of sex-biased genes differently.

Analysing gene expression data in reproductive and vegetative tissues of males and females from the dioecious plant, *Salix viminalis*, I was able to assess the presence and evolution of sex-biased genes (Chapter 5, Darolti et al. (2018)). Consistent with studies comparing sex-biased gene expression in reproductive and somatic tissues in animals (Harrison et al. 2015; Mank et al. 2010a; Perry et al. 2014), I show that the catkin reproductive tissue is highly transcriptionally dimorphic, while for the vegetative tissue, gene expression levels in males and females remain undifferentiated. Also in line with findings from animal studies

(Grath and Parsch 2016), I show that male-biased genes have a higher overall expression and a greater degree of sex-bias than female-biased genes. Together with the finding that differential expression of genes results mainly from expression changes in males, this suggests that the evolution of sex-biased genes in *S. viminalis* is to a large extent driven by selection on males, possibly through sexual selection and sexual conflict.

Similar to animals, sexual selection in flowering plants is thought to be strong, acting on gene expression patterns predominantly through pollen competition (Moore and Pannell 2011). Plant species relying primarily on wind pollination are predicted to express stronger sexual dimorphism than insect-pollinated species. Indeed, *S. viminalis* and its relative *Populus balsamifera* are both wind-pollinated and show a more pronounced sex-biased gene expression (~36% and ~44% of genes in *P. balsamifera* and, respectively, *S. viminalis* are sex-biased) than the insect-pollinated species *Silene latifolia* (~17% of genes show sex-biased expression) (Darolti et al. 2018; Sanderson et al. 2019; Zemp et al. 2016). Compared to animal species, however, rates of sex-biased gene expression appear to be lower in plants, likely the result of a higher degree of sexual dimorphism in animals (Barrett and Hough 2013). While mate choice in animals has a large influence on the development of sexual dimorphism (Smith 1991), plant sexual interactions are more indirect, and thus female mate choice is more limited (Moore and Pannell 2011), although in some species it can still take place through selective embryo abortion following fertilization or incompatibility selection (Barrett and Hough 2013). In addition, the time since the evolution of dioecy and the establishment of separate sexes is substantially

164

shorter in flowering plants compared to animals, contributing to the differences in the extent of sexual dimorphism observed between animals and plants (Barrett and Hough 2013).

While the rates of sequence divergence of sex-biased genes have been extensively studied in animals, we know far less about the evolution of sex-biased genes in dioecious plants. In *S. viminalis*, I discover that male-biased genes evolve slower than female-biased and unbiased genes (Chapter 5, Darolti et al. (2018)), a result that is in stark contrast to the typically faster rates of evolution of sex-biased, and in particular of male-biased, genes observed in animal species (Grath and Parsch 2016). This is an emerging pattern in plant species, as other recent studies also report lower rates of sequence divergence for sex-biased genes relative to unbiased genes (Cossard et al. 2018; Sanderson et al. 2019).

A key difference in the development of plants and animals is that plants spend a large portion of their life cycle in a haploid phase where gene transcription is active in gametes, particularly in pollen (Otto et al. 2015). The observed slow evolution of sex-biased genes in plants could be the result of haploid selection, and indeed, pollen male-biased genes have been found to experience strong purifying selection acting on deleterious recessive mutations (Arunkumar et al. 2013). By contrast, animal species have substantially reduced selection on the haploid stage (Otto et al. 2015).

Future study direction

It is becoming increasing clear that multiple factors, such as developmental processes and mating systems, can affect the rate of evolution of

sex-biased genes differently between plants and animals. The catkin *S. viminalis* samples used in my study are abundant in pollen grains (Chapter 5), however, a comparison between diploid and haploid sex-biased genes would be required to confirm the effects of haploid selection on the rates of evolution of *S. viminalis* male-biased genes. With evolutionary drivers such as haploid selection potentially having a major impact on plant, but not animal, gamete gene expression, future studies should also explore the relative contribution of different selective forces to the evolution of sex-biased genes.

In addition, in my study I used expression in the unbiased leaf tissue as a proxy for putative ancestral expression state. I compared male and female expression for genes present in both leaf and catkin tissues to assess whether the transition to the sex-biased state has evolved as a result of expression changes in males. Similarly, a study on *S. latifolia*, investigated the direction of gene expression changes that result in sex-biased expression by comparing the dioecious *S. latifolia* with a related gynodiecious species, where female and hermaphrodite individuals co-exist (Zemp et al. 2016). Ideally, however, to test the hypothesis that sexual selection pressure in males is driving the evolution of sex-biased gene expression, a comparative analysis on a dioecious species with a monoecious outgroup should be performed.

Appendix

Table S.2.1. Sequencing information for each sample

Sample	Source	Method	Coverage	Raw paired reads	Paired reads after trimming	% removed
Female_4	Lab pop	DNA-seq	79X	276,700,698	253,668,178	8.32
Female_7	Lab pop	DNA-seq	75X	262,392,382	233,869,784	10.87
Male_11	Lab pop	DNA-seq	68X	238,377,854	217,155,394	8.90
Male_8	Lab pop	DNA-seq	85X	298,121,974	271,223,610	9.02
Female_1	Lab pop	RNA-seq	-	31,614,139	30,942,700	2.12
Female_4	Lab pop	RNA-seq	-	42,162,172	41,258,427	2.14
Female_7	Lab pop	RNA-seq	-	29,059,783	28,520,214	1.86
Female_13	Lab pop	RNA-seq	-	32,392,057	31,628,760	2.36
Male_5	Lab pop	RNA-seq	-	26,029,449	25,434,019	2.29
Male_8	Lab pop	RNA-seq	-	31,955,977	31,182,426	2.42
Male_11	Lab pop	RNA-seq	-	36,271,770	35,430,424	2.32
Male_17	Lab pop	RNA-seq	-	30,579,016	29,853,906	2.37
Male_6	Lab pop	RNA-seq	-	31,303,333	30,559,027	2.38
Male_9	Lab pop	RNA-seq	-	37,546,710	36,718,647	2.21
Male_15	Lab pop	RNA-seq	-	29,875,402	29,258,994	2.06
Male_18	Lab pop	RNA-seq	-	39,603,241	38,785,961	2.06
Male_12	Lab pop	RNA-seq	-	32,568,129	31,793,985	2.38
Male_2	Lab pop	RNA-seq	-	23,196,956	22,777,671	1.81
Male_14	Lab pop	RNA-seq	-	31,859,225	31,110,939	2.35
Male_1	Wild AD	DNA-seq	37X	130,231,185	118,932,936	8.68
Male_3	Wild AD	DNA-seq	29X	101,814,846	93,738,591	7.93
Male_4	Wild AD	DNA-seq	30X	104,256,689	96,147,514	7.78
Male_5	Wild AD	DNA-seq	41X	1438,00,633	131,847,549	8.31
Male_12	Wild YU	DNA-seq	36X	126,989,240	114,251,020	10.03
Male_13	Wild YU	DNA-seq	40X	140,058,828	129,438,072	7.58
Male_14	Wild YU	DNA-seq	30X	103,948,475	95,917,845	7.73
Male_15	Wild YU	DNA-seq	29X	102,834,035	93,981,379	8.61
Male_16	Wild QD	DNA-seq	36X	126,993,935	115,367,896	9.15
Male_17	Wild QD	DNA-seq	33X	117,037,503	107,793,725	7.90
Male_19	Wild QD	DNA-seq	37X	129,116,621	118,530,011	8.20
Male_20	Wild QD	DNA-seq	30X	105,910,625	97,094,279	8.32
Male_32	Wild AU	DNA-seq	38X	133,405,487	121,078,915	9.24
Male_33	Wild AU	DNA-seq	36X	126,870,348	116,645,902	8.06
Male_34	Wild AU	DNA-seq	34X	120,555,304	111,191,193	7.77
Male_35	Wild AU	DNA-seq	33X	116,057,513	105,723,853	8.90
Male_41	Wild YD	DNA-seq	40X	139,172,741	127,519,332	8.37
Male_42	Wild YD	DNA-seq	37X	129,566,748	119,662,352	7.64
Male_43	Wild YD	DNA-seq	27X	94,884,948	87,529,008	7.75
Male_44	Wild YD	DNA-seq	30X	103,984,825	94,569,988	9.05
Male_51	Wild QU	DNA-seq	28X	98,581,145	89,880,179	8.83
Male_52	Wild QU	DNA-seq	36X	125,273,377	112,674,653	10.06
Male_53	Wild QU	DNA-seq	31X	107,398,826	99,357,018	7.49

| Male_54 | Wild QU | DNA-seq | 46X | 160,156,505 | 147,797,565 | 7.72 |

A = Aripo watershed, Y = Yarra watershed, Q = Quare watershed
U = Upstream population, D = Downstream population

Table S.2.2. Error corrected reads for de novo genome assembly

Sex	Source	Sequencing	No. of individuals	Coverage	Error corrected paired reads
Female	Lab pop	DNA-seq	2	137X	479,641,121

Table S.2.3. Female de novo genome assembly statistics

Pre 1kb length filter				Post 1kb length filter				Oriented scaffolds			
No.	N50 (Kb)	N90 (Kb)	Length (Mb)	No.	N50 (Kb)	N90 (bp)	Length (Mb)	No.	N50 (Kb)	N90 (Kb)	Length (Mb)
2,361,160	5.4	0.1	963.6	96,611	11.3	2.8	634.8	19,206	17.4	5.6	219.5

Table S.2.4. Assignment of chromosomal position

Reference genome	Genes	Genes mapped to assembly	Scaffolds with mapped genes	Scaffolds discarded due to mapping discordance	Oriented scaffolds	% with multiple mapped genes all from the same reference chromosome
Guppy	25,694	25,460	19,526	320	19,206	92%

Table S.2.5. Coverage and SNP density

	Autosomes				X chromosome			
	No.	M $\log_2$ median	F $\log_2$ median	M:F $\log_2$ median	No.	M $\log_2$ median	F $\log_2$ median	M:F $\log_2$ median
Coverage (Wilcoxon rank sum test p-value)	17,353	5.8505	5.8477	0.0024	709	5.8273 (0.0224)	5.8324 (< 0.001)	-0.0115 (<0.001)
SNP density (Wilcoxon rank sum test p-value)	13,286	0.0023	0.0022	0.0000	555	0.0030 (<0.001)	0.0023 (0.837)	0.0007 (<0.001)

Wilcoxon rank sum test between autosomal and X chromosome medians

Table S.2.6. Faster-X effect

	Autosomes	X-Y diverged region	Stratum II	Stratum I
No.	4755	86	70	16
d_N/d_S	0.091	0.107	0.105	0.106
(95% CI)	0.088-0.094	0.088-0.129	0.084-0.131	0.067-0.172
(*p* value)		0.067	0.107	0.249
d_N	0.003	0.004	0.004	0.006
(95% CI)	0.003-0.003	0.003-0.005	0.003-0.004	0.004-0.009
(*p* value)		**0.012**	0.182	**0.002**
d_S	0.034	0.037	0.034	0.057
(95% CI)	0.033-0.034	0.033-0.042	0.030-0.039	0.043-0.069
(*p* value)		0.1	0.77	**<0.001**

X-Y diverged region refers to Stratum II (15-22Mb) & Stratum I (> 22Mb)
P-values calculated relative to the autosomes using permutation tests with 1000 replicates. Confidence intervals calculated using bootstrapping with 1000 replicates.

Table S.2.7. Normalised SNP densities in the X-Y diverged regions across upstream and downstream guppy populations

	Downstream			Upstream		
River	No.	M $\log_2$ median	M:F $\log_2$ median	No.	M $\log_2$ median	M:F $\log_2$ median
Yarra (Wilcoxon rank sum test *p* value)	197	-0.0001	0.0006	194	0.0006 (0.0002)	0.0013 (0.0105)
Quare (Wilcoxon rank sum test *p* value)	196	-0.0001	0.0006	196	0.0001 (0.0090)	0.0009 (0.0342)
Aripo (Wilcoxon rank sum test *p* value)	195	-0.0000	0.0007	197	0.0002 (0.0039)	0.0013 (0.0209)

X-Y diverged region refers to Stratum II (15-22Mb) & Stratum I (> 22Mb)
Wilcoxon rank sum test between downstream and upstream medians

Table S.2.8. Normalised coverage in Stratum I of the sex chromosome across upstream and downstream guppy populations

River	Downstream			Upstream		
	No.	M log$_2$ median	M:F log$_2$ median	No.	M log$_2$ median	M:F log$_2$ median
Yarra (Wilcoxon rank sum test *p* value)	59	0.0183	-0.0484	59	0.0065 (0.7345)	-0.0330 (0.8548)
Quare (Wilcoxon rank sum test *p* value)	59	0.0365	-0.0289	59	0.0037 (0.5011)	-0.0444 (0.2494)
Aripo (Wilcoxon rank sum test *p* value)	59	0.0292	-0.0519	59	0.0241 (0.6129)	-0.0329 (0.9314)

Wilcoxon rank sum test between downstream and upstream medians
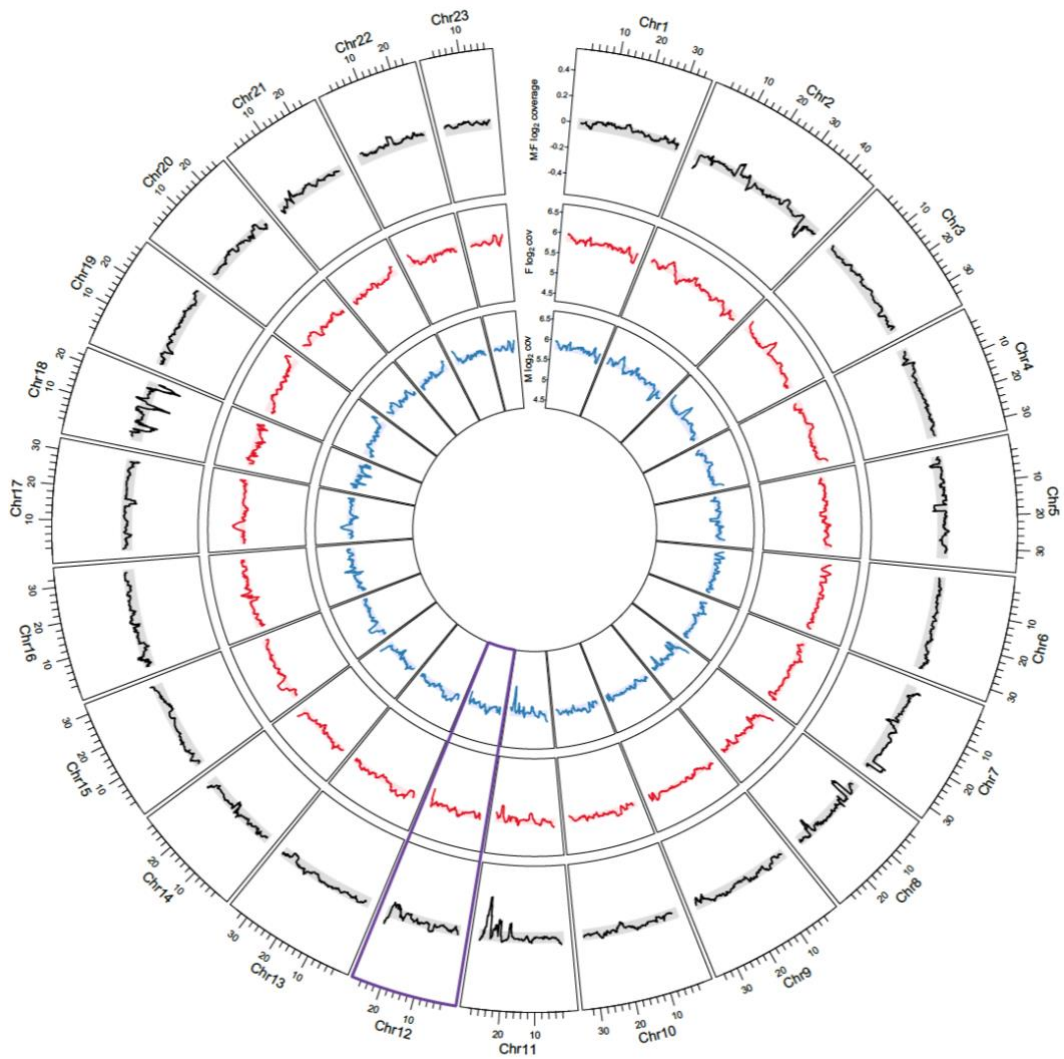
Fig. S.2.1. Circos plot of male and female coverage for oriented scaffolds, with moving averages based on window sizes of 40 scaffolds. For each chromosome, male coverage (blue), female coverage (red) and male:female coverage (black) is shown. 95% confidence intervals based on bootstrapping autosomal estimates are shown for male coverage (light blue), female coverage (pink) and male:female coverage (grey). The X chromosome, which contains the sex determining gene, is highlighted in purple.
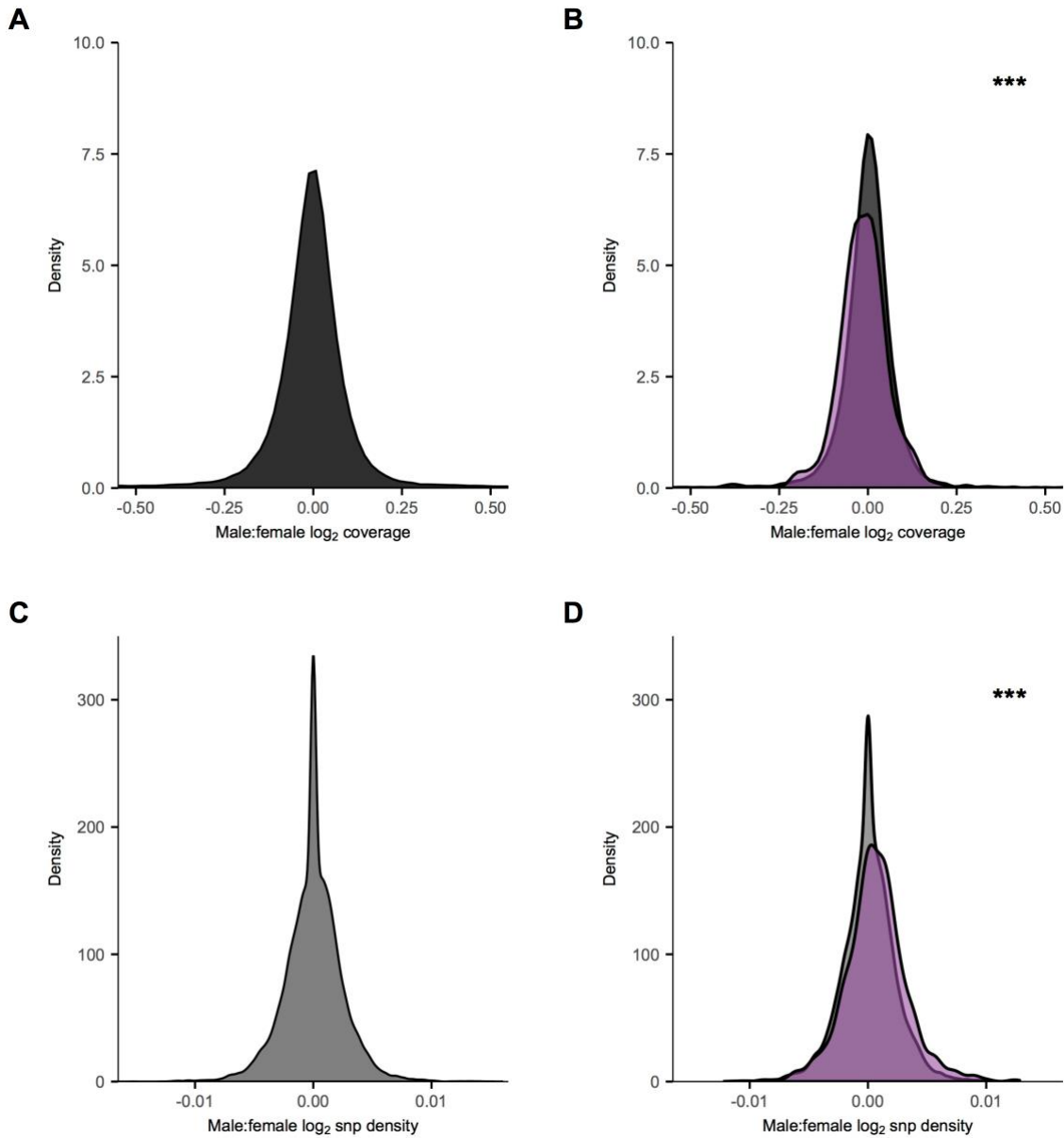
Fig. S.2.2. Distribution of male:female coverage (A and B) and SNP density (C and D) for assembled scaffolds. Panel A. Coverage differences for all scaffolds greater than 1Kb in length. Panel. B. Coverage differences for autosomal (grey) and the X chromosome (purple) scaffolds. Panel C. Differences in male and female SNP density for all scaffolds greater than 1Kb in length. Panel D. Differences in male and female SNP density for autosomes (grey) and the X chromosome (purple). *** p value < 0.001.
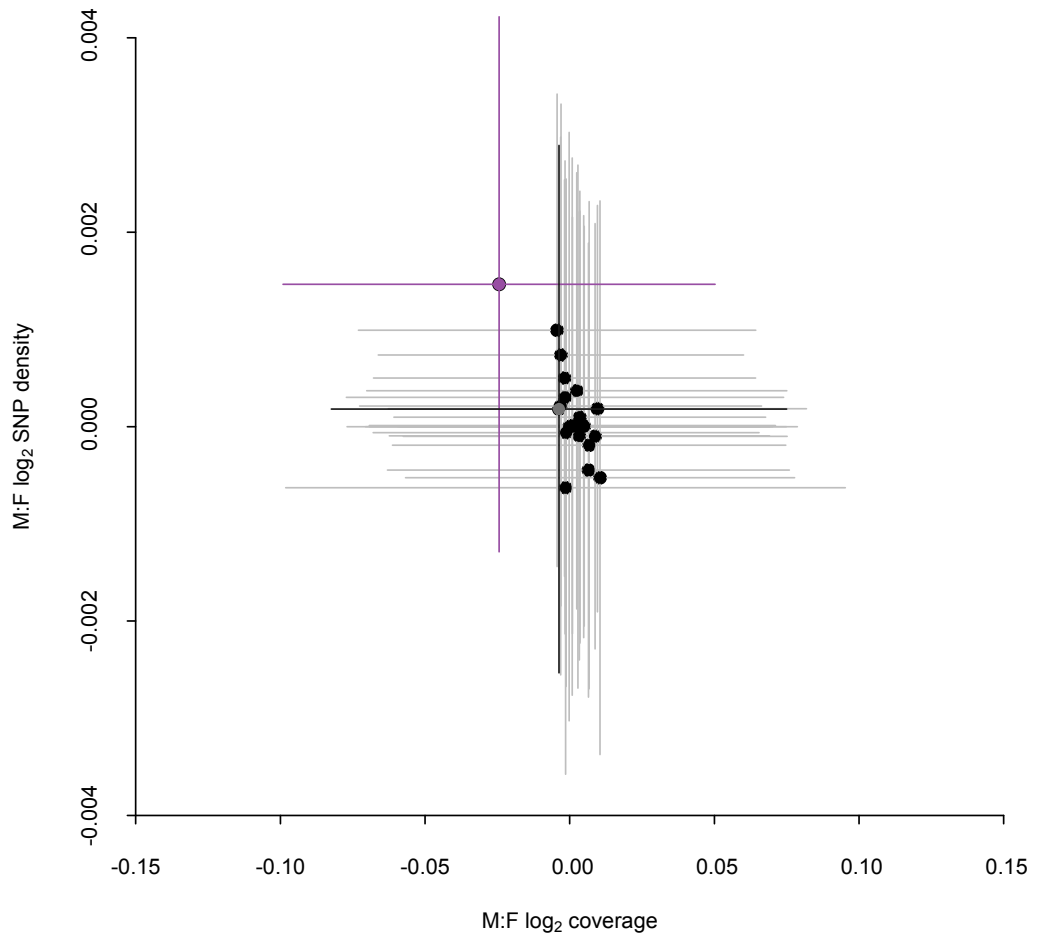
Fig. S.2.3. Distribution of sex differences in coverage and SNP density for all chromosomes. The X-Y diverged region (Strata I & II, 15 – 25 Mb) is in purple and PAR (<15 Mb) is in grey. Horizontal and vertical lines denote interquartile ranges.

173
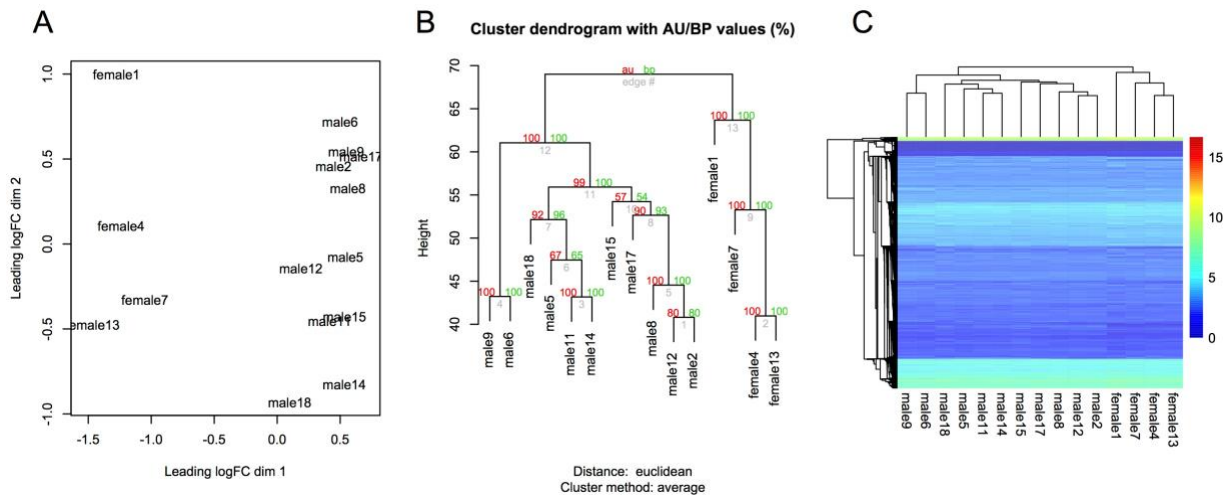
Fig. S.2.4. Cluster analysis of expression data. Panel A. MDS plot of normalised count data. Panel B. Cluster dendrogram of normalised $\log_2$ RPKM values. Approximately unbiased p-values are shown in red and bootstrap probability values are shown in green. Panel C. Clustered heatmap of normalised $\log_2$ RPKM values.
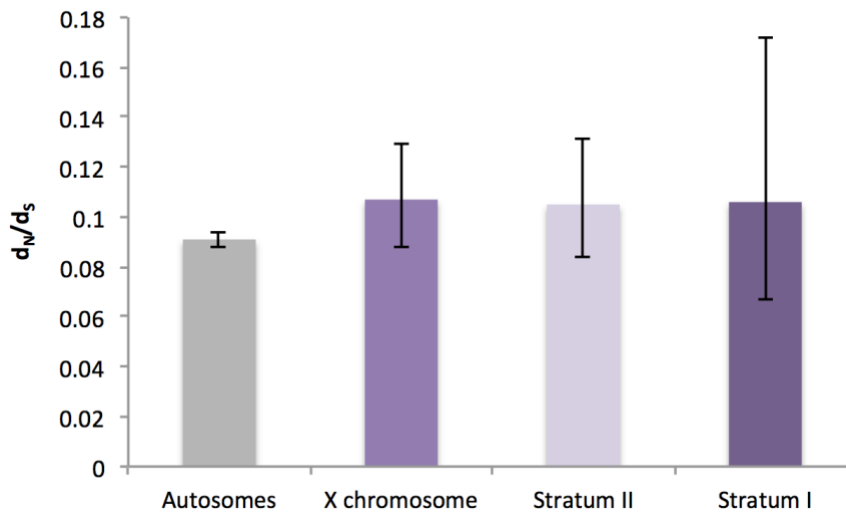
Fig. S.2.5. Estimates of mean d_N/d_S for the autosomes and the X chromosome. 95% confidence intervals were calculated by bootstrapping with 1000 replicates. X chromosome refers only to the X-Y region (Strata I & II, 15 – 25 Mb) and does not include genes in the PAR.

174
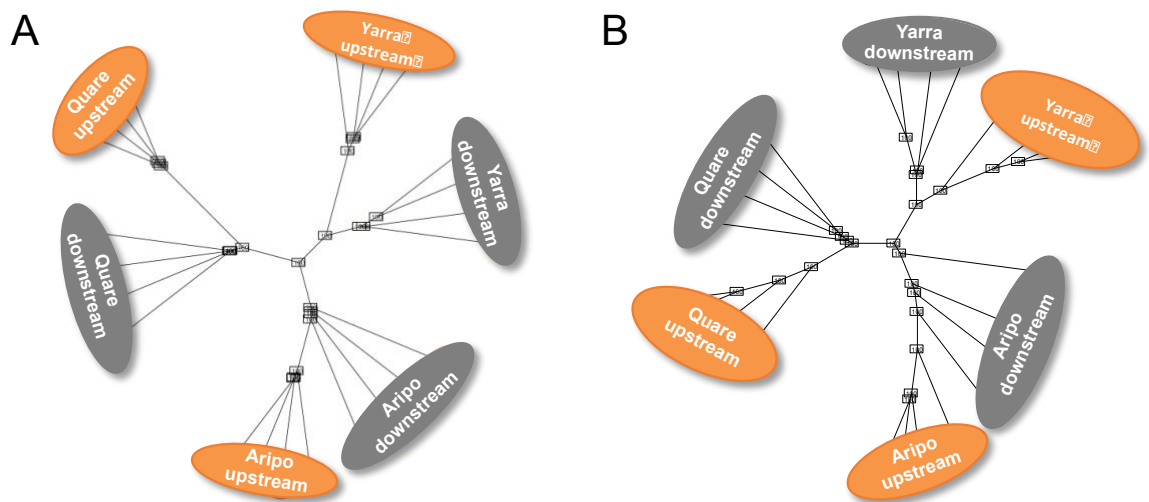
Fig. S.2.6. Phylogeny of upstream (orange) and downstream (grey) guppy populations across three watersheds (Yarra, Quare, Aripo) in Trinidad. The analysis is based on (A) all 4.6 million SNPs across the whole genome or (B) 72,623 SNPs located on the X-Y diverged region (Strata I & II, 15 – 25 Mb). Neighbor joining trees were constructed from Euclidian distances between individuals. Bootstrap support is indicated at each node.
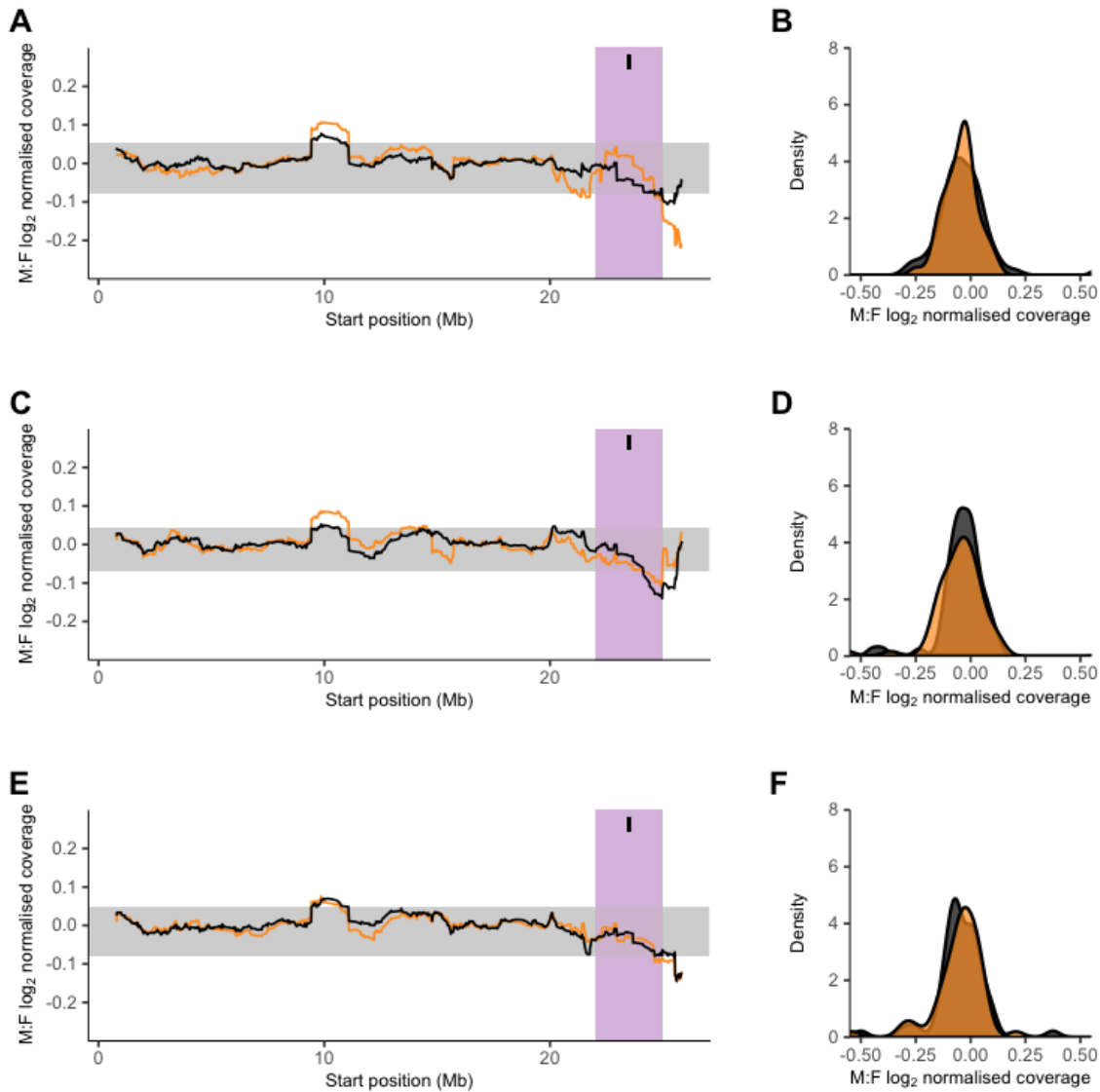
Fig. S.2.7. Male:female coverage for the X chromosome across upstream (orange) and downstream (black) guppy populations. Panels A. C. and E. Moving averages of normalised coverage across the X chromosome based on sliding window analysis (window size of 40 scaffolds) for Yarra (panel A), Quare (panel C) and Aripo (panel E) watersheds. 95% confidence intervals based on bootstrapping autosomal estimates are in grey. Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in laboratory population males (Stratum I, 22-25 Mb) (see Fig. 2.2), light purple indicates the region with less X-Y differentiation, where there a significant excess of male SNPs in laboratory populations (Stratum II, 15 – 22 Mb). Panels B. D. and F. Distribution of sex differences in normalised coverage for the oldest region of the X chromosome (Stratum I, 22 – 25 Mb) for Yarra (panel B), Quare (panel D) and Aripo (panel F) watershed.

Table S.3.1. Sequencing results for each sample

Species (Treatment)	Sample no. (Sex)	Paired reads after trimming	% kept after trimming	Coverage
Poecilia wingei (DNA-seq PE)	291 (F)	222,019,309	97.7	77X
	292 (F)	209,095,391	92.6	72X
	293 (F)	244,778,587	92.3	85X
	294 (M)	221,308,140	92.9	76X
	295 (M)	245,199,642	93.3	85X
	296 (M)	214,802,737	93.2	74X
Poecilia picta (DNA-seq PE)	247 (F)	201,783,529	92.5	70X
	248 (F)	248,146,529	93.4	86X
	265 (F)	251,440,989	93.2	87X
	266 (M)	264,471,289	93.4	91X
	267 (M)	209,266,241	93.3	72X
	268 (M)	213,098,477	93.7	74X
Poecilia latipinna (DNA-seq PE)	269 (F)	242,950,245	93.6	83X
	270 (F)	186,547,462	92.7	64X
	271 (F)	194,577,608	92.7	67X
	272 (M)	235,795,174	93.3	81X
	289 (M)	229,757,997	93.4	79X
	290 (M)	232,391,653	93.0	80X
Gambusia holbrooki (DNA-seq PE)	241 (F)	217,994,173	93.8	75X
	242 (F)	193,263,881	93.5	67X
	243 (F)	229,309,343	93.3	79X
	244 (M)	195,792,613	93.4	68X
	245 (M)	194,586,542	93.6	67X
	246 (M)	220,591,540	93.4	76X
Poecilia wingei (DNA-seq MP)	013 (F)	80,809,424	58.0	23X
	014 (F)	76,562,926	58.1	22X
	015 (F)	77,120,163	58.5	22X
	016 (M)	75,360,153	56.4	22X
	018 (M)	80,705,804	57.9	23X
	019 (M)	83,808,049	58.8	24X
Poecilia picta (DNA-seq MP)	013 (F)	81,263,670	57.7	23X
	014 (F)	75,174,083	56.9	22X
	015 (F)	86,920,083	57.1	25X
	016 (M)	73,917,330	56.4	21X
	018 (M)	79,696,940	56.0	23X
	019 (M)	76,727,662	57.1	22X
Poecilia latipinna (DNA-seq MP)	002 (F)	87,479,612	56.1	25X
	004 (F)	87,085,262	56.8	25X
	005 (F)	54,308,904	56.4	16X
	006 (M)	78,744,655	57.0	23X

	007 (M)	84,406,439	54.2	24X
	012 (M)	88,707,007	58.9	26X
Gambusia holbrooki (DNA-seq MP)	002 (F)	82,118,221	66.3	23.6
	004 (F)	76,472,890	55.6	22.0
	005 (F)	77,475,370	54.2	22.3
	006 (M)	66,891,462	56.4	19.2
	007 (M)	72,014,055	56.5	20.7
	012 (M)	63,635,368	56.8	18.3
Poecilia wingei (RNA-seq)	201 (F)	35,176,172	94.0	-
	202 (F)	47,040,049	94.4	-
	203 (F)	48,558,664	94.4	-
	265 (M)	44,255,632	94.2	-
	266 (M)	41,375,146	94.2	-
	267 (M)	42,277,857	93.9	-
Poecilia picta (RNA-seq)	282 (F)	33,616,549	93.9	-
	284 (F)	43,438,223	94.3	-
	285 (M)	45,953,612	94.3	-
	286 (M)	39,836,450	94.0	-
	287 (M)	43,314,678	94.1	-
	302 (F)	48,435,135	94.0	-
Poecilia latipinna (RNA-seq)	228 (F)	48,056,489	94.3	-
	229 (F)	34,836,324	94.3	-
	230 (M)	35,640,155	94.7	-
	231 (M)	34,564,529	93.8	-
	232 (M)	34,774,385	93.9	-
	288 (F)	50,234,040	94.0	-
Gambusia holbrooki (RNA-seq)	204 (F)	38,909,731	94.5	-
	205 (F)	44,717,526	94.9	-
	206 (F)	45,915,199	97.7	-
	207 (M)	46,496,039	94.1	-
	208 (M)	42,781,352	94.3	-
	281 (M)	39,993,511	93.1	-

Table S.3.2. Assembly statistics

Species	Total assembly length (Mb)	N50 (kb)	No. *de novo* scaffolds	No. RACA Predicted Chromosome Fragments
P. wingei	795.5	14.6	120,169	400
P. picta	782.2	150.6	9,640	201
P. latipinna	787.5	90.3	13,851	255
G. holbrooki	617.8	6.1	137,790	27

Table S.3.3. Differential gene expression results

Species	Categories	Autosomes + PAR	Non-recombining region	Chi-square test
P. reticulata	Total genes	13,075	231	
	Sex-biased	531 (4.0%)	11 (4.8%)	$X^2(1) = 0.1183, p = 0.73$
	Male-biased	337 (2.6%)	7 (3.0%)	$X^2(1) = 0.0439, p = 0.83$
	Female-biased	194 (1.5%)	4 (1.7%)	$X^2(1) = 0.0009, p = 0.97$
P. wingei	Total genes	12,066	472	
	Sex-biased	775 (6.4%)	34 (7.2%)	$X^2(1) = 0.2889, p = 0.59$
	Male-biased	346 (2.9%)	16 (3.4%)	$X^2(1) = 0.2546, p = 0.61$
	Female-biased	429 (3.6%)	18 (3.8)	$X^2(1) = 0.0255, p = 0.87$
P. picta	Total genes	10,706	363	
	Sex-biased	2,176 (20.3%)	77 (21.2%)	$X^2(1) = 0.0729, p = 0.79$
	Male-biased	929 (8.7%)	29 (7.9.%)	$X^2(1) = 0.1070, p = 0.74$
	Female-biased	1,247 (11.6%)	48 (13.2%)	$X^2(1) = 0.5320, p = 0.47$

A

B

Poecilia wingei

XY

Fig. S.3.1. Coverage and SNP density differences between the sexes (male:female) for *P. wingei* scaffolds placed by RACA on the reference *X. hellerii* chromosomes. (A) Average coverage and SNP density fold change for each chromosome. Shown in blue is *X. hellerii* chromosome 8, which is syntenic to the guppy sex chromosome (*P. reticulata* chromosome 12), and constitutes the sex chromosome in *P. wingei.* Interquartile ranges are represented by the vertical and horizontal lines. (B) Circos plot showing moving average of $\log_2$ M:F coverage (outer ring) and $\log_2$ M:F SNP density (inner ring) fold change across each chromosome. Highlighted in blue is the XY sex chromosome in *P. wingei*. Horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the genome, excluding the sex chromosome.

A

B

Poecilia picta

XY

182
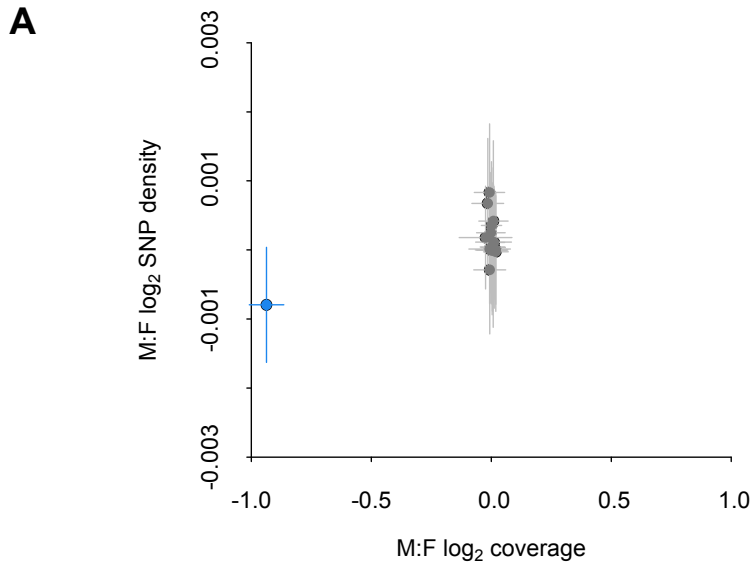
Fig. S.3.2. Coverage and SNP density differences between the sexes (male:female) for *P. picta* scaffolds placed by RACA on the reference *X. hellerii* chromosomes. (A) Average coverage and SNP density fold change for each chromosome. Shown in blue is *X. hellerii* chromosome 8, which is syntenic to the guppy sex chromosome (*P. reticulata* chromosome 12), and constitutes the sex chromosome in *P. picta.* Interquartile ranges are represented by the vertical and horizontal lines. (B) Circos plot showing moving average of $\log_2$ M:F coverage (outer ring) and $\log_2$ M:F SNP density (inner ring) fold change across each chromosome. Highlighted in blue is the XY sex chromosome in *P. picta*. Horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the genome, excluding the sex chromosome.

A

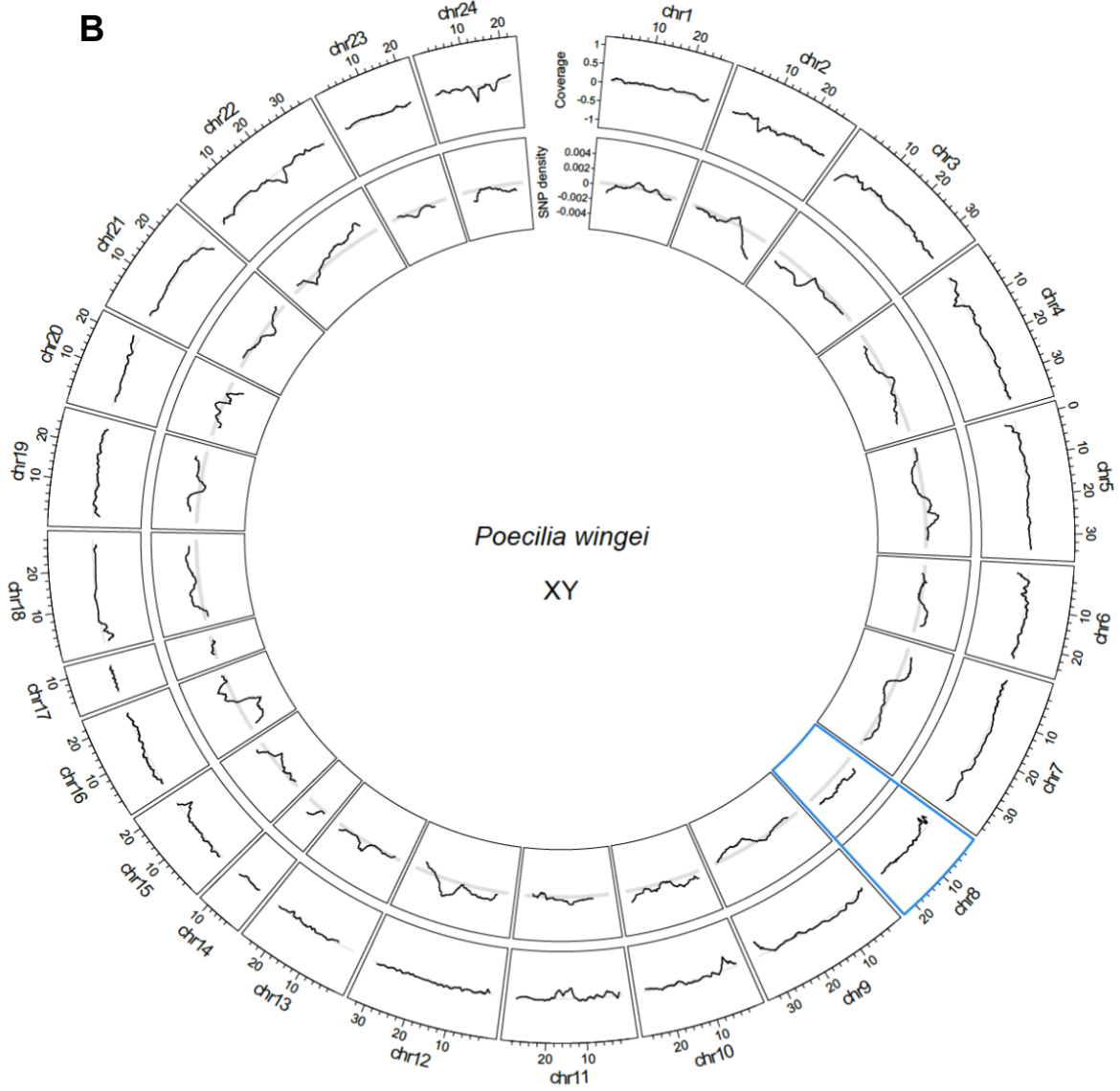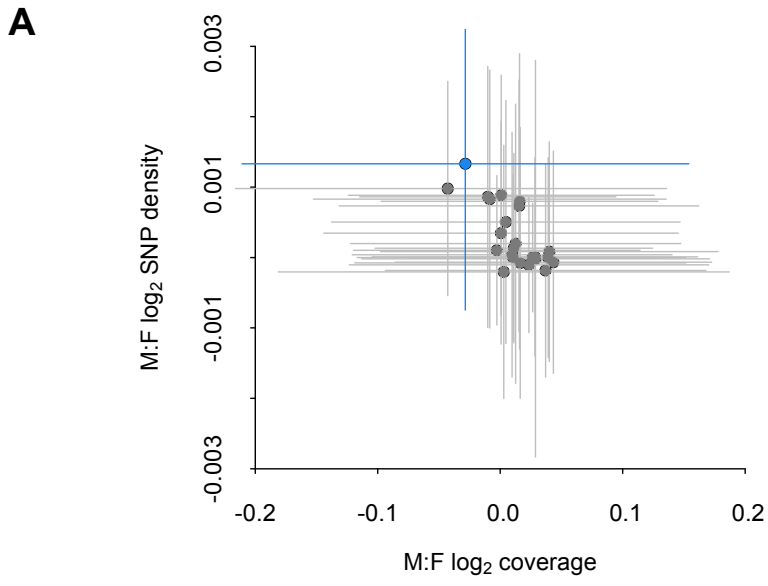**B**

Poecilia latipinna

ZW
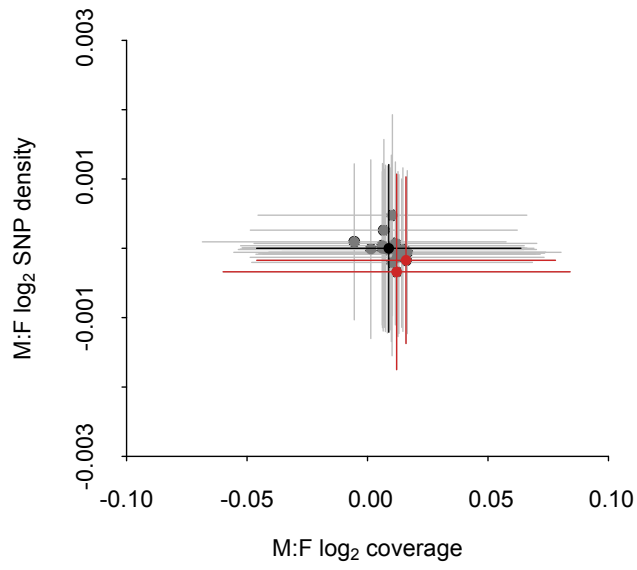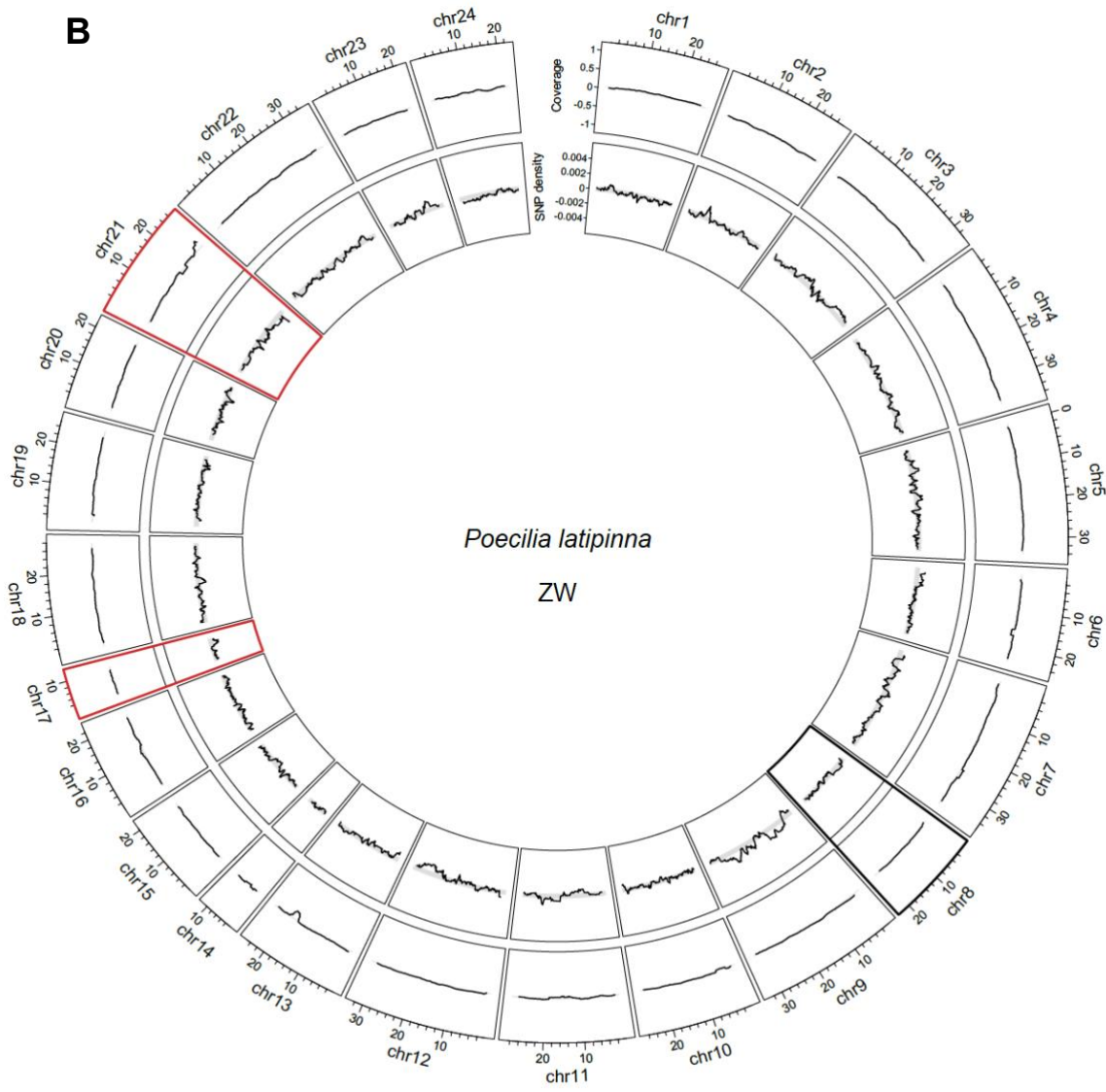
Fig. S.3.3. Coverage and SNP density differences between the sexes (male:female) for *P. latipinna* scaffolds placed by RACA on the reference *X. hellerii* chromosomes. (A) Average coverage and SNP density fold change for each chromosome. Shown in red are chromosomes 17 and 21, ZW sex chromosome candidates for *P. latipinna*. Chromosome 8, which is syntenic to the guppy sex chromosome (*P. reticulata* chromosome 12), is shown in black. Interquartile ranges are represented by the vertical and horizontal lines. (B) Circos plot showing moving average of $\log_2$ M:F coverage (outer ring) and $\log_2$ M:F SNP density (inner ring) fold change across each chromosome. Highlighted in red are the *P. latipinna* ZW sex chromosome candidates, as identified in (A). Horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the genome, excluding the sex chromosome candidates.
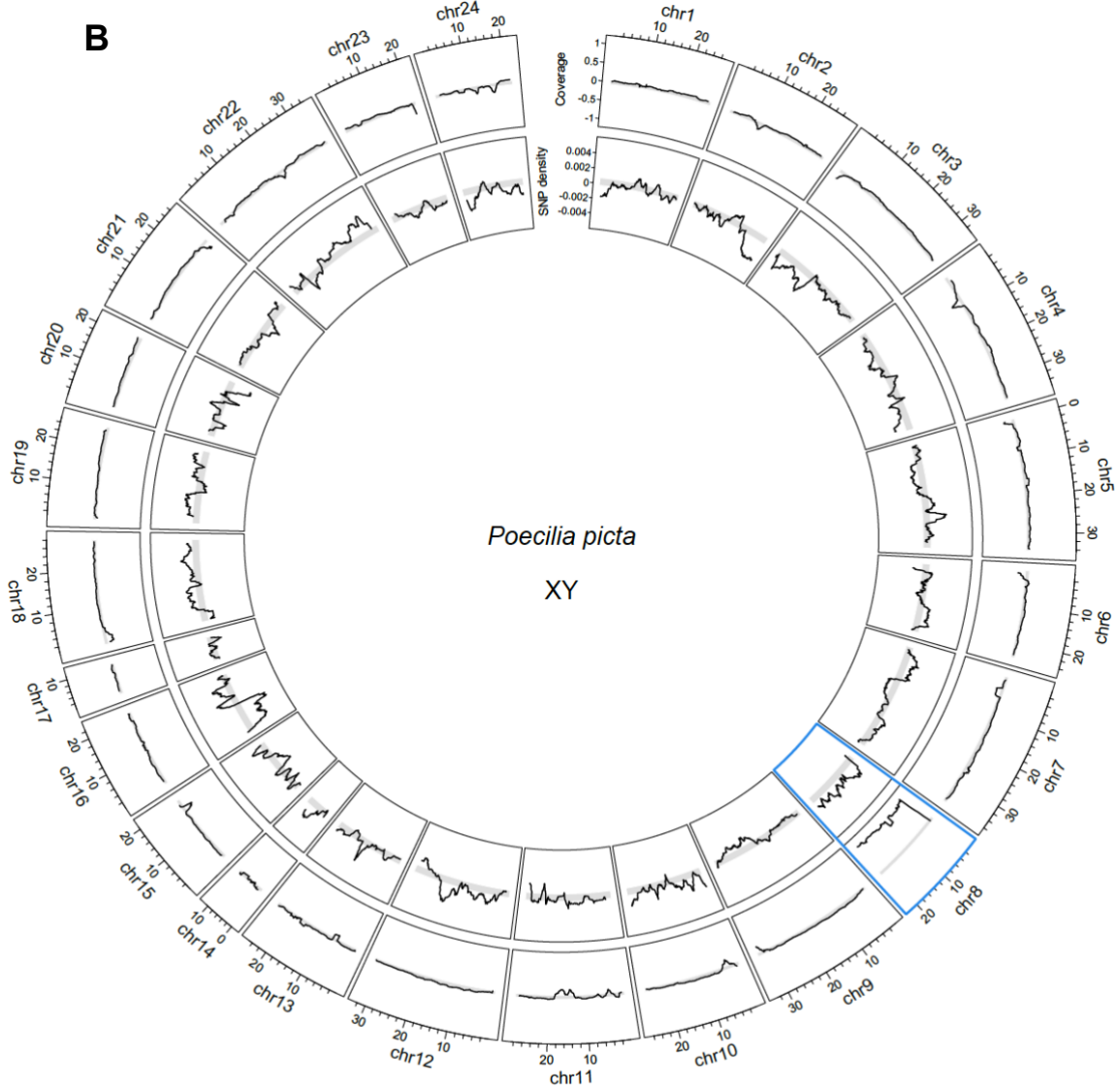
A

B

Gambusia holbrooki

XY
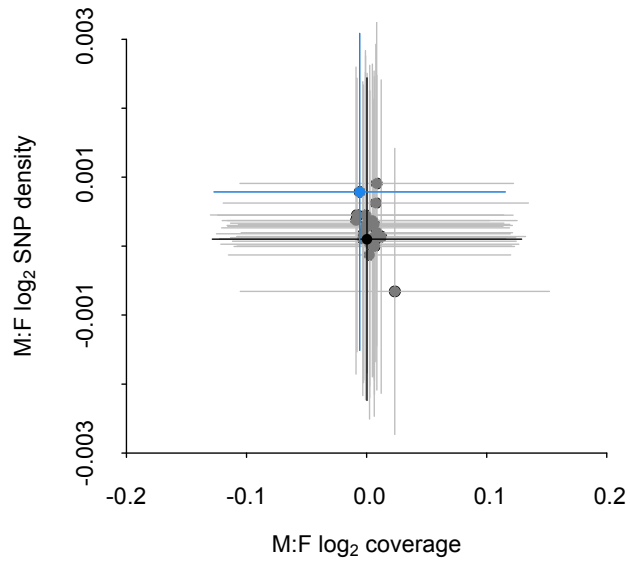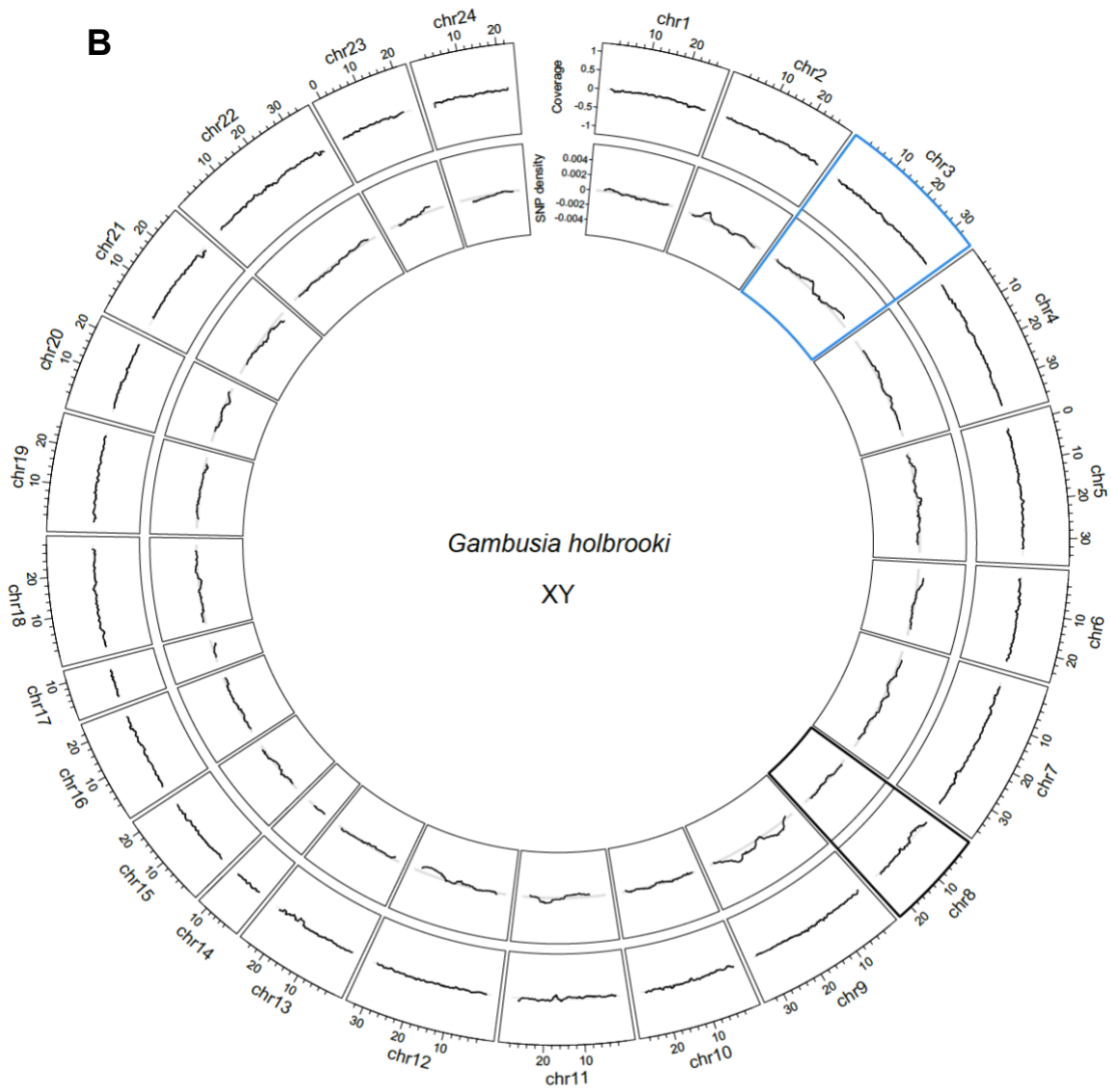
Fig. S.3.4. Coverage and SNP density differences between the sexes (male:female) for *G. holbrooki* scaffolds placed by RACA on the reference *X. hellerii* chromosomes. (A) Average coverage and SNP density fold change for each chromosome. Shown in blue is chromosome 3, an XY sex chromosome candidate for *G. holbrooki*. Chromosome 8, which is syntenic to the guppy sex chromosome (*P. reticulata* chromosome 12), is shown in black. Interquartile ranges are represented by the vertical and horizontal lines. (B) Circos plot showing moving average of $\log_2$ M:F coverage (outer ring) and $\log_2$ M:F SNP density (inner ring) fold change across each chromosome. Highlighted in blue is a *G. holbrooki* XY sex chromosome candidate, as identified in (A). Horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the genome, excluding the sex chromosome candidate.

Table S.4.1. Sequencing information for each sample

Species (Treatment)	Family Number	Sample	Raw paired reads (Million)	Paired reads after trimming (Million)	% kept
P. reticulata (RNA-seq PE)	Family 1	Sire	42.3	29.1	68.8
		Dam	34.7	21.8	62.8
		Off. Fem. 1	29.5	20.1	68.1
		Off. Fem. 2	135.4	88.2	65.1
		Off. Fem. 3	27.9	23.3	83.5
		Off. Fem. 4	33.4	20.8	62.3
		Off. Fem. 5	25.2	17.6	69.8
		Off. Mal. 1	18.6	12.3	66.1
		Off. Mal. 2	20.3	14.0	69.0
		Off. Mal. 3	42.8	27.2	63.6
		Off. Mal. 4	47.6	35.7	75.0
		Off. Mal. 5	59.2	43.3	73.1
	Family 2	Sire	52.2	33.8	64.8
		Dam	86.6	58.4	67.4
		Off. Fem. 1	24.5	16.4	66.9
		Off. Fem. 2	35.8	23.4	65.4
		Off. Fem. 3	53.5	36.7	68.6
		Off. Fem. 4	27.7	18.9	68.2
		Off. Fem. 5	57.4	35.9	62.5
		Off. Mal. 1	72.7	46.7	64.2
		Off. Mal. 2	19.5	12.7	65.1
		Off. Mal. 3	76.3	51.7	67.8
		Off. Mal. 4	33.1	24.4	73.7
		Off. Mal. 5	25.4	16.6	65.4
	Family 3	Sire	28.1	23.5	83.6
		Dam	33.2	19.8	59.6
		Off. Fem. 1	35.2	21.6	61.4
		Off. Fem. 2	23.5	15.5	66.0
		Off. Fem. 3	50.2	31.1	62.0
		Off. Fem. 4	97.9	63.4	64.8
		Off. Fem. 5	30.5	20.6	67.5
		Off. Mal. 1	39.9	26.8	67.2
		Off. Mal. 2	45.1	26.9	59.6
		Off. Mal. 3	26.0	22.3	85.8
		Off. Mal. 4	30.6	25.6	83.7
		Off. Mal. 5	35.7	22.7	63.6

P. wingei (RNA-seq PE)	Family 1	Sire	25.4	15.3	60.2
		Dam	92.0	58.8	63.9
		Off. Fem. 1	29.9	22.7	75.9
		Off. Fem. 2	30.6	21.2	69.3
		Off. Fem. 3	32.1	24.3	75.7
		Off. Fem. 4	26.1	17.7	67.8
		Off. Fem. 5	31.6	23.8	75.3
		Off. Mal. 1	34.0	24.9	73.2
		Off. Mal. 2	22.0	15.4	70.0
		Off. Mal. 3	58.7	41.9	71.4
		Off. Mal. 4	48.1	34.1	70.9
		Off. Mal. 5	43.9	27.3	62.2
	Family 2	Sire	54.1	35.7	66.0
		Dam	52.0	35.0	67.3
		Off. Fem. 1	25.6	16.3	63.7
		Off. Fem. 2	31.4	27.9	88.9
		Off. Fem. 3	54.8	29.1	53.1
		Off. Fem. 4	26.0	16.7	64.2
		Off. Fem. 5	29.1	19.3	66.3
		Off. Mal. 1	48.2	30.1	62.5
		Off. Mal. 2	46.1	30.0	65.1
		Off. Mal. 3	71.2	49.0	68.8
		Off. Mal. 4	36.1	24.9	69.0
		Off. Mal. 5	34.2	23.6	69.0

Table S.4.2. *De novo* transcriptome assembly statistics

		Before filtering	After best isoform selection	After ncRNA filter	After ORF filter	After CAP3
P. reticulata	No. transcripts	490,973	249,752	249,528	21,141	19,935
	N50	2,511	1,050	1,048	1,713	1,821
	Median length	548	357	357	1,041	1,092
P. wingei	No. transcripts	408,978	209,871	209,682	20,340	19,361
	N50	2,730	1,150	1,148	1,749	1,845
	Median length	554	355	355	1,074	1,119

Table S.4.3. Divergence estimates for *P. reticulata* and *P. wingei* X- and Y-linked gametologs

Category	d_N (95% CI)	sig.[*]	d_S (95% CI)	sig.[*]	d_N/d_S (95% CI)	sig.[*]
X *P. reticulata*	0.0002 (0.0000–0.0008)		0.0045 (0.0027-0.0086)		0.0410 (0.0000–0.1395)	
		0.914		0.106		0.102
Y *P. reticulata*	0.0004 (0.0002–0.0006)		0.0044 (0.0031–0.0063)		0.0869 (0.0444–0.1273)	
X *P. wingei*	0.0003 (0.0001–0.0006)		0.0050 (0.0033–0.0074)		0.0651 (0.0126–0.1489)	
		0.464		0.978		0.528
Y *P. wingei*	0.0003 (0.0001–0.0006)		0.0045 (0.0026–0.0077)		0.0722 (0.0357–0.1324)	

Significance represents *p* values based on 1,00 replicates permutation tests.

Table S.5.1. Sequencing and quality trimming information for each sample

Sample	Raw paired reads	Paired reads following trimming	% reads kept
78021_female_catkin	48,479,402	37,549,325	77.45
78183_female_catkin	32,434,747	23,739,771	73.19
78195_female_catkin	44,612,862	31,372,609	70.32
81084_male_catkin	36,261,721	26,799,852	73.91
Hallstad1-84_male_catkin	47,186,936	34,726,235	73.59
T76_male_catkin	49,433,319	35,279,702	71.37
78021_female_leaf	28,577,122	20,738,684	72.57
78183_female_leaf	42,774,423	31,112,678	72.74
78195_female_leaf	46,605,756	33,091,822	71.00
81084_male_leaf	42,739,744	29,762,943	69.64
Hallstad1-84_male_leaf	42,597,909	30,932,365	72.61
T76_male_leaf	45,688,342	31,098,461	68.07

Table S.5.2. Divergence and polymorphism estimates for orthologs resulting from the OrthoMCL pipeline

Tissue	Location	Category[a]	n Genes[b]	d_N (95% CI) sig.[c]	d_S (95% CI) sig.[c]	d_N/d_S (95% CI) sig.[c]	p_N (95% CI) sig.[c]	p_S (95% CI) sig.[c]	p_N/p_S (95% CI) sig.[c]	DoS sig.[d]
Catkin	Autosomes and recombining Z	UB	738	0.0030 (0.0028 - 0.0033)	0.0130 (0.0122 - 0.0140)	0.2323 (0.2143 - 0.2518)	0.0028 (0.0026 - 0.0031)	0.0110 (0.0102 - 0.0119)	0.2571 (0.2337 - 0.2808)	-0.0495
		MB	257	0.0031 (0.0028 - 0.0035) P = 0.598	**0.0150 (0.0132 - 0.0170) P = 0.002**	**0.2065 (0.1782 - 0.2423) P = 0.038**	0.0030 (0.0026 - 0.0034) P = 0.306	0.0114 (0.0102 - 0.0128) P = 0.452	0.2633 (0.2269 - 0.3058) P = 0.688	-0.0346 P = 0.826
		FB	341	0.0031 (0.0028 - 0.0035) P = 0.458	**0.0143 (0.0131 - 0.0157) P = 0.004**	0.2170 (0.1941 - 0.2444) P = 0.132	**0.0032 (0.0028 - 0.0036) P = 0.014**	0.0117 (0.0105 - 0.0130) P = 0.096	0.2708 (0.2458 - 0.3012) P = 0.282	-0.0375 P = 0.086
	Non-recombining Z	UB	5	0.0030 (0.0016 - 0.0043)	0.0072 (0.0025 - 0.0106)	0.4093 (0.2318 - 0.8811)	0.0006 (0.0002 - 0.0010)	0.0031 (0.0012 - 0.0052)	0.1995 (0.0579 - 0.6606)	0.0800
		MB	3	0.0029 (0.0 - 0.0140) P = 0.756	**0.0143 (0.0091 - 0.0396) P < 0.001**	**0.2019 (0.0 - 0.3533) P < 0.001**	**0.0029 (0.0 - 0.0210) P < 0.001**	**0.0104 (0.0039 - 0.0505) P < 0.001**	0.2781 (0.0 - 0.4151) P = 0.592	0.0088 P = 0.786
		FB	2	0.0036 (0.0033 - 0.0037) P = 0.228	**0.0147 (0.0094 - 0.0345) P < 0.001**	0.2458 (0.0966 - 0.3981) P = 0.182	**0.0082 (0.0033 - 0.0100) P < 0.001**	**0.0229 (0.0226 - 0.0238) P < 0.001**	0.3563 (0.14 - 0.4398) P = 0.434	0.0770 P = 0.857

[a] Unbiased (UB), male-biased (MB) and female-biased (FB) genes.

[b] Number of genes with both divergence and polymorphism data.

[c] P values based on 1,000 replicates permutation tests comparing male-biased and female-biased genes with unbiased genes. Significant P values (< 0.05) are shown in bold.

[d] P values from Wilcoxon nonparametric tests comparing male-biased and female-biased genes with unbiased genes. Significant P values (< 0.05) are shown in bold.

Table S.5.3. Comparison of divergence results from two methods of estimating d_N/d_S

Tissue	Location	Category[a]	n Genes[b]	Ratio of the sum of the number of substitutions to the number of sites			Concatenated sequences		
				d_N (95% CI) sig.[c]	d_S (95% CI) sig.[c]	d_N/d_S (95% CI) sig.[c]	d_N (95% CI) sig.[c]	d_S (95% CI) sig.[c]	d_N/d_S (95% CI) sig.[c]
Catkin	Autosomes and recombining Z	UB	1,754	0.0030 (0.0028 - 0.0031)	0.0135 (0.0130 - 0.0141)	0.2204 (0.2101 - 0.2311)	0.0030	0.0132	0.2232
		MB	674	**0.0032 (0.0030 - 0.0035)** **$p = 0.012$**	**0.0162 (0.0147 - 0.0187)** **$p < 0.001$**	**0.1951 (0.1769 - 0.2157)** **$p < 0.001$**	0.0032	0.0158	0.2017
		FB	732	0.0031 (0.0029 - 0.0033) $p = 0.094$	**0.0149 (0.0141 - 0.0158)** **$p < 0.001$**	0.2095 (0.1938 - 0.2256) $p = 0.082$	0.0031	0.0146	0.2128

[a] Unbiased (UB), male-biased (MB) and female-biased (FB) genes.

[b] Number of genes with both divergence and polymorphism data.

[c] P values based on 1,000 replicates permutation tests comparing male-biased and female-biased genes with unbiased genes. Significant p values (< 0.05) are shown in bold.
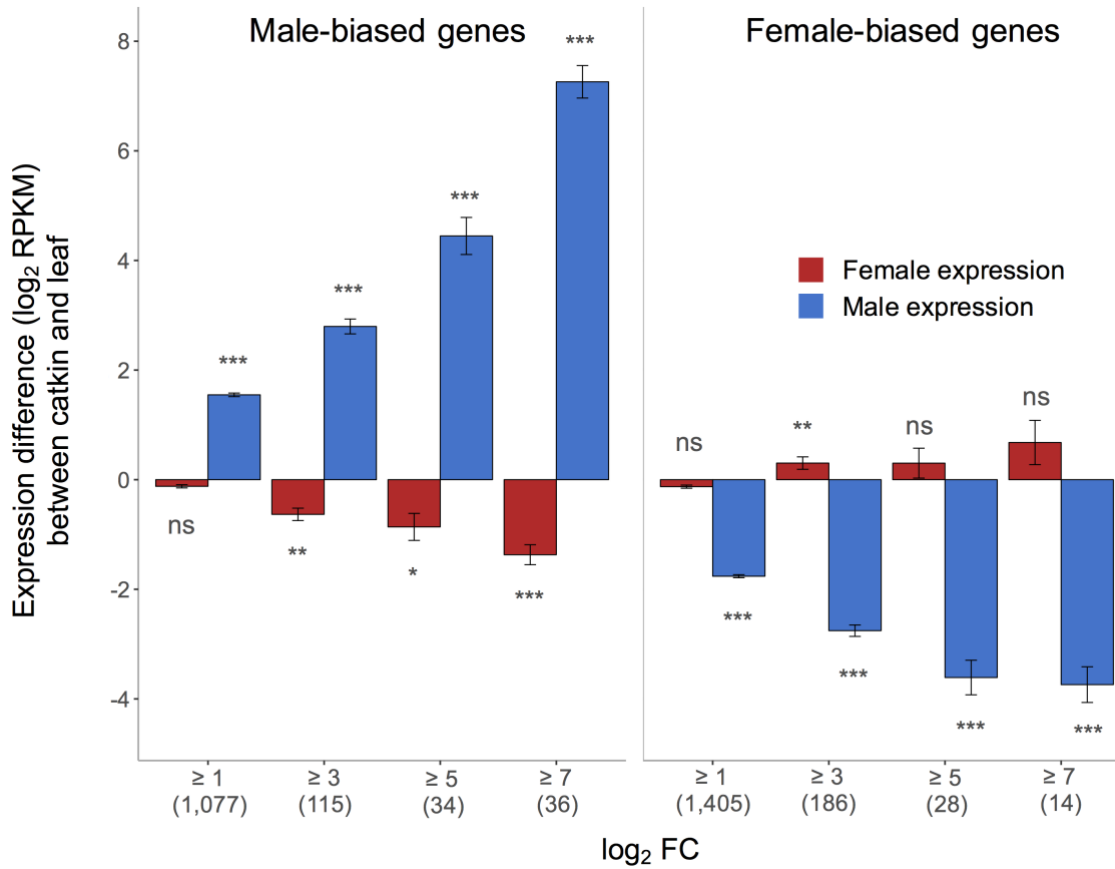
Fig. S.5.1. Expression differences between catkin and leaf samples (with standard errors of the mean) at different sex-bias fold change thresholds for male-biased and female-biased catkin genes. Male expression differences are shown in blue while female expression differences in red. Positive values indicate genes with higher expression in catkin than in leaf while negative values correspond to genes with higher expression in leaf compared to catkin. Significant differences in gene expression between the two tissues were determined through Wilcoxon rank sum tests (ns = non-significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).
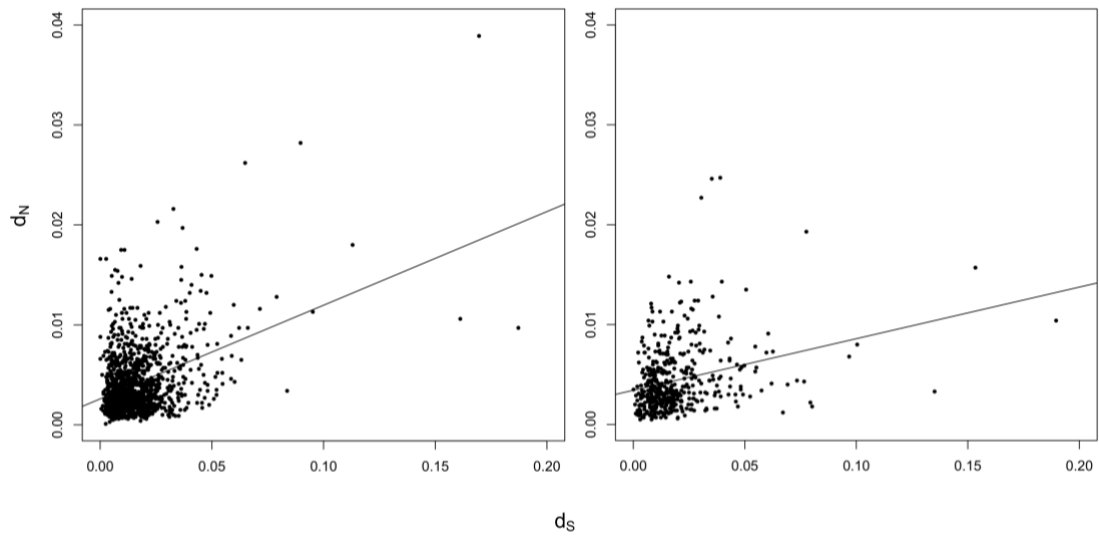
Fig. S.5.2. Relationship between the rate of synonymous (d_S) and nonsynonymous (d_N) substitutions for unbiased and male-biased genes.
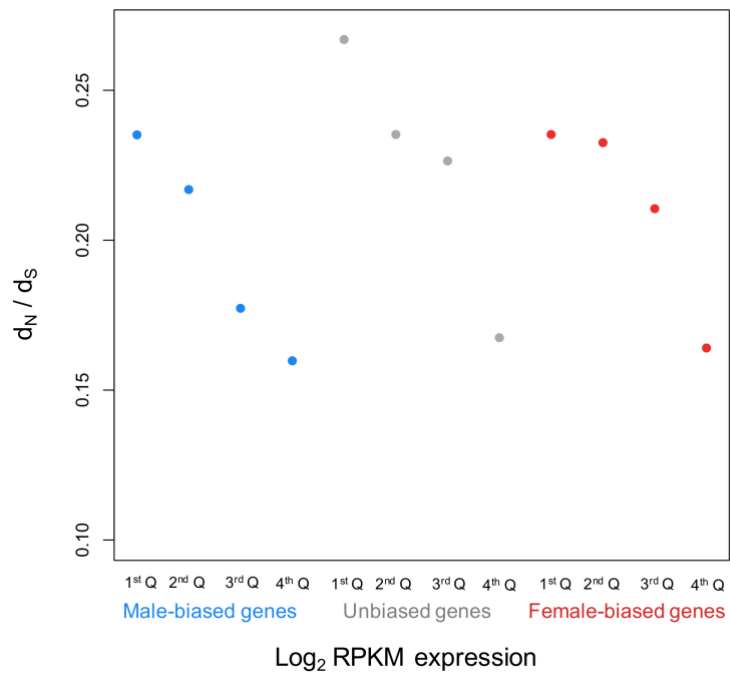
Fig. S.5.3. The ratio of nonsynonymous to synonymous nucleotide substitutions for male-biased, unbiased and female-biased genes. Both sex-biased and unbiased genes are divided into four equal quartiles of expression, where 1st quartile has the lowest expression and 4th quartile has the highest expression.
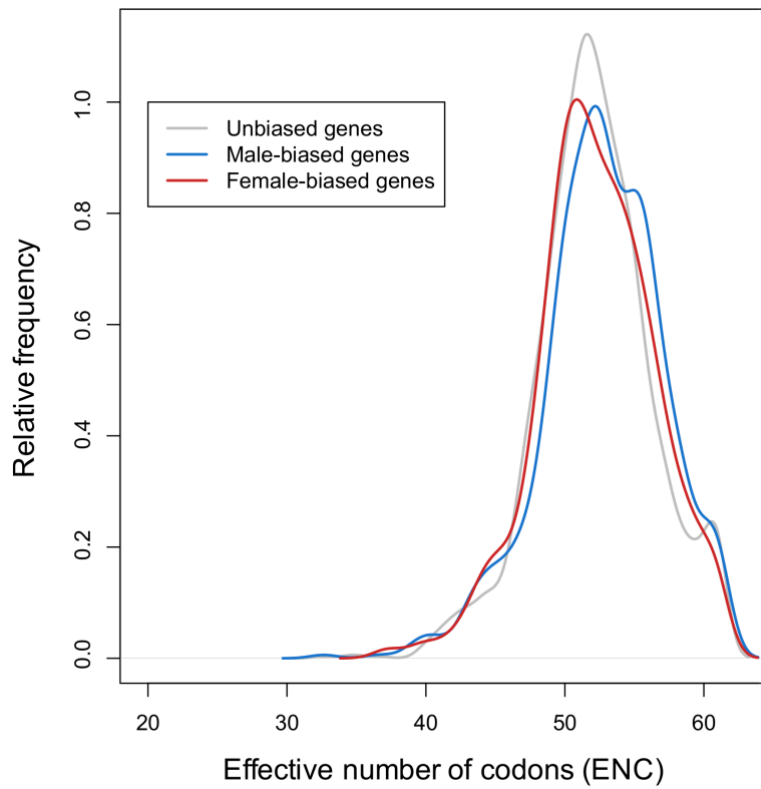
Figure S.5.4. Density distribution of the effective number of codons (ENC) for unbiased (grey), male-biased (blue) and female-biased (red) genes. ENC is inversely related to the level of codon usage bias, ranging from 20 (extreme bias) to 61 (no bias).

Key publications

Wright AE, **Darolti I**, Bloch NI, Oostra V, Sandkam BA, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE. (2017). Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nature Communications. 8:14251.

Darolti I, Wright AE, Pucholt P, Berlin S, Mank JE. Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*. (2018). Molecular Ecology. 27: 694-708.

Morris J, **Darolti I**, Bloch NI, Wright AE, Mank JE. (2018). Shared and species-specific patterns of nascent Y chromosome evolution in two guppy species. Genes. 9: 238.

Darolti I, Wright AE, Sandkam BA, Morris J, Bloch NI, Farré M, Fuller RC, Bourne GR, Larkin DM, Breden F, Mank JE (In press.) Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. Proceedings of the National Academy of Sciences, USA.

Wright AE, **Darolti I**, Bloch NI, Oostra V, Sandkam BA, Buechel SD, Kolm N, Breden F, Vicoso B, Mank JE (2019). On the power to detect rare recombination events. Proceedings of the National Academy of Sciences, USA. 116: 12607-12608.

Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation

Alison E. Wright[1], Iulia Darolti[1], Natasha I. Bloch[1], Vicencio Oostra[1], Ben Sandkam[2], Severine D. Buechel[3], Niclas Kolm[3], Felix Breden[2], Beatriz Vicoso[4] & Judith E. Mank[1]

Sex chromosomes evolve once recombination is halted between a homologous pair of chromosomes. The dominant model of sex chromosome evolution posits that recombination is suppressed between emerging X and Y chromosomes in order to resolve sexual conflict. Here we test this model using whole genome and transcriptome resequencing data in the guppy, a model for sexual selection with many Y-linked colour traits. We show that although the nascent Y chromosome encompasses nearly half of the linkage group, there has been no perceptible degradation of Y chromosome gene content or activity. Using replicate wild populations with differing levels of sexually antagonistic selection for colour, we also show that sexual selection leads to greater expansion of the non-recombining region and increased Y chromosome divergence. These results provide empirical support for longstanding models of sex chromosome catalysis, and suggest an important role for sexual selection and sexual conflict in genome evolution.

[1] Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. [2] Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6. [3] Department of Zoology, Stockholm University, Svante Arrheniusväg 18 B, Stockholm 106 91, Sweden. [4] Institute of Science and Technology, Am Campus 1A, Klosterneuburg 3400, Austria. Correspondence and requests for materials should be addressed to A.E.W. (email: alison.e.wright@ucl.ac.uk).

ex chromosomes are typically thought to evolve as recombination is halted between a homologous pair of chromosomes in one sex. Although we have a detailed understanding of the evolutionary consequences of the loss of recombination for sex chromosome evolution[1,2], we still do not understand the evolutionary forces acting to halt recombination in the first place. The dominant theoretical model for the early stages of sex chromosome evolution[3–5] predicts that recombination will be selected against in the region between a sex determining gene and a nearby locus with alleles of sex-specific effect. This theory, though prevalent, remains largely untested empirically, as most research has focused on older, highly divergent sex chromosome systems[6,7], for which it is difficult to extrapolate the earliest stages and causes of divergence.

The sex chromosomes of the guppy (Poecilia reticulata) have been of interest for more than a century, following early reports that many sexually selected colour traits are passed through the patriline on the Y chromosome[8,9]. These observations were central to the development of theories regarding the role of sexual conflict in recombination suppression and sex chromosome divergence[3–5]. Colour is sexually antagonistic in guppies, as brightly coloured males are more attractive to females and more visible to predators, but brightly coloured females gain no fitness advantage and only suffer increased predation[10–12]. Therefore, in this system, current models of sex chromosome evolution predict that recombination would be selected against between the sex determining locus and linked loci involved in colouration. This process would shrink the pseudoautosomal region in favour of expanding X- and Y-specific regions, creating a male supergene on the Y chromosome containing multiple colouration loci and thereby resolving sexually antagonistic selection.

Even though the guppy sex chromosomes are a classic model for the study of sexual conflict and sex chromosome divergence, little is actually known about the pattern of divergence between the X and Y chromosomes. Recent linkage maps identified male recombination events restricted to the middle of chromosome 12 (ref. 13), suggesting that the other half of the chromosome is functionally X- or Y-linked. Immunostaining of recombination nodules[14] was broadly concordant with recombination mapping, again suggesting that the X chromosome is split roughly in equal parts between X-specific and pseudoautosomal.

Recombination shows substantial local variation between males and females throughout the genomes of many organisms[7,15] and identifying areas of restricted male recombination does not distinguish the sex chromosome from other areas where males simply do not recombine. However, the Y is morphologically distinguishable from the X chromosome[16], and comparative genome hybridization of lab populations[17] suggest that roughly half of the Y chromosome is male-specific. Because many vertebrate sex chromosomes show progressive spread of the non-recombining region[18–21], the large size of the guppy non-recombining region and male-specific regions suggest substantial divergence between the X and Y.

Recombination suppression between the X and Y chromosomes results in complete linkage of the male-specific region of the Y. The loss of recombination in this region typically limits the role of adaptive evolution and leads to strong background selection and linkage effects, causing loss of functional polymorphism in coding sequence over time[1]. Roughly half of male colouration patterns are thought to be Y-linked[8], and the remarkable diversity of male colour combinations implies an improbably large number of Y haplotypes maintained within populations for a sex chromosome system of at least intermediate age. Additionally, if recombination suppression really is driven by sexually antagonistic alleles[3–5], then we might expect recent

but rapid spread of recombination suppression shortly after the emergence of sexual preferences for colour. Although sexually selected traits exist in many Poeciliids, the vivid male colouration in P. reticulata is only shared by a few very close relatives[22,23], therefore, the expansions of the male-limited Y chromosome to engulf colouration loci might have occurred very recently.

Moreover, the degree of male colouration, and, therefore, the degree of sexual conflict over colour, varies substantially based on predation pressures. Across watersheds, downstream populations are typically associated with higher predation and males are far less colourful than upstream populations[24–26]. Importantly, the proportion of colour patterns thought to be Y-linked varies between upstream and downstream populations[27]. The unusual gene content of the guppy sex chromosomes, therefore, makes it a uniquely powerful system for testing the role of sexual conflict and sexual selection in sex chromosome divergence.

In order to determine the degree of divergence between the X and Y chromosome in this species, we resequenced male and female genomes and transcriptomes of both laboratory and wild individuals. We find that the X and Y show sequence differentiation over nearly one half of the length of the chromosome, however, the divergence between the X and Y chromosome is remarkably subtle, indicating very low levels of divergence and likely recent origin of the sex chromosome. The large region of divergence is in contrast to reports of other nascent sex chromosome systems where the diverged region is highly restricted[28–30]. Despite this young age, we detect evidence of Faster-X evolution in this region. Most importantly, we find convergent patterns of greater sex chromosome divergence in upstream populations, which experience substantially elevated sexual selection and sexual conflict, compared with downstream populations. Our results suggest that recombination suppression between the X and Y spread quickly in the recent history of this sex chromosome system, possibly driven by the presence of sexually antagonistic alleles related to sexual selection.

Results

The structure of the guppy sex chromosomes. We first assembled the female genome using SOAPdenovo2, based on 480 million paired end reads from an outbred laboratory population. The assembly yielded 96,611 scaffolds, with an N50 of 11.3 Kb and total length of 634.8 Mb, after a minimum length threshold of 1 Kb (Supplementary Tables 1–3). Guppy genes from the reference genome (Guppy_female_1.0 + MT) were mapped to scaffolds in order to identify chromosomal positions, resulting in a final assembly of 19,206 ordered scaffolds oriented along the guppy chromosomes, with an N50 of 17.4 Kb and total length of 219.5 Mb (Supplementary Tables 3 and 4).

We then mapped male and female DNA-seq reads to our ordered scaffolds in order to identify regions of coverage difference between the sexes. Regions with longstanding recombination suppression in males will show reduced mapping efficiency against the female genome assembly, as diverged sequence from the Y will no longer map to the X chromosome[19,20,31]. Even with strict mapping thresholds (see methods) we could identify no large region of the genome with reduced coverage in males, which we would expect if a large portion of the Y was significantly diverged or degraded (Supplementary Fig. 1), and the overall distribution of coverage is largely symmetrical (Supplementary Fig. 2A). However, previous linkage maps have identified chromosome 12, which contains the sex determining gene, as the sex chromosome[13], and this chromosome shows a slight shift in the distribution and has a significantly greater
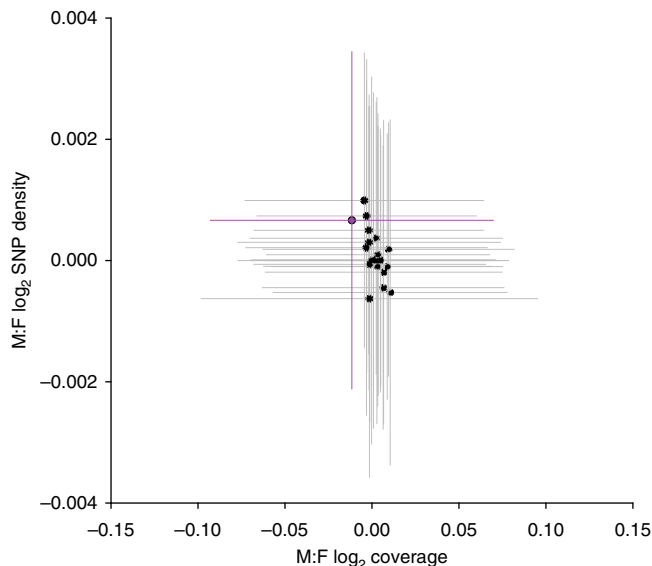
Figure 1 | Distribution of sex differences in coverage and SNP density for all chromosomes. The X chromosome is in purple. Horizontal and vertical lines denote interquartile ranges.

proportion of scaffolds with female-biased coverage than autosomes (Wilcoxon rank sum test $P < 0.001$, Supplementary Fig. 2B, Supplementary Table 5). This suggests that recombination suppression between the X and Y chromosomes has led to very slight divergence between them.

If the Y has diverged, but not yet degraded significantly, we would expect to observe Y-specific single nucleotide polymorphisms (SNPs) in regions that retain substantial sequence similarity to the X, resulting in higher average male hetero-zygosity for the sex chromosomes[32,33]. When assessing all regions of the genome, we observe a shoulder of elevated SNP density in males (Supplementary Fig. 2C), due to significantly greater SNP density in males for the sex chromosomes compared with autosomal genes (Wilcoxon rank sum test $P < 0.001$, Supplementary Fig. 2D, Supplementary Table 5). When sex differences in coverage and SNP density are plotted together, the sex chromosome is a clear outlier to the other chromosomes (Fig. 1), confirming low but significant levels of divergence.

In order to determine the relative divergence between X and Y chromosomes, we plotted coverage and SNP density differences between males and female on our scaffolds against physical position on the guppy genome assembly. We detected significantly reduced male coverage outside the autosomal 95% confidence interval from 22–25 Mb (Fig. 2). This region shows the largest degree of X-Y sequence divergence and likely corresponds to the oldest region of the sex chromosome (Stratum I). In contrast, between 15–25 Mb, we detect significant elevation of male SNP density but no reduction in male coverage, indicative of lower levels of X-Y divergence and suggesting that nearly half of the sex chromosome has stopped recombining in males in the very recent past (Stratum II) (Fig. 3a, Supplementary Fig. 3).

Our coverage and SNP analysis suggest that although male-specific SNPs have accumulated, the Y chromosome has not degenerated significantly. Because loss of gene activity often quickly follows loss of recombination on the Y chromosome[1], for each gene we plotted male and female expression level (RPKM) across the X chromosome. Our results show that the non-recombining region exhibits low levels of sexualization of

gene content, with regions where the majority of genes exhibit female- or male-biased expression. However, there is no region of detectable loss of male gene activity, as would be expected with extensive Y chromosome decay (Fig. 3b, Supplementary Fig. 4). In contrast, in the region of the sex chromosome with the greatest coverage difference between males and females (Stratum I, 22–25 Mb), likely the area of greatest Y chromosome divergence, there is a slight excess of male-biased genes, indicating that this region of the Y chromosome has also not suffered any significant loss of gene activity. We tested for enrichment of GO terms for genes expressed in the X-Y diverged region (Strata I and II, 15–25 Mb) relative to the rest of the genome. However, there were no GO terms with an enrichment $P < 0.001$.

X chromosomes are predicted in many circumstances to show elevated rates of evolution[34], and signatures of Faster-X evolution have been detected in old, heteromorphic sex chromosomes[35–37]. However, it is unclear whether a detectable signal of Fast-X would be expected in the early stages of sex chromosome evolution. We, therefore, compared rates of evolution for X-linked and autosomal coding sequence, and recovered a significant pattern of Faster-X in the guppy. X-linked d_N/d_S is greater though marginally non-significant for X-linked genes (86 genes, permutation test with 1,000 replicates, $P = 0.067$) relative to the autosomes (4,755 genes), due to a marginally significant increase in d_N (permutation test with 1000 replicates, $P = 0.014$) (Supplementary Table 6, Supplementary Fig. 5). This pattern is evident across both Strata I and II, indicating that low levels of sex chromosome divergence are sufficient to facilitate Faster-X processes.

Population variation in male colour and sexual conflict. Predation pressures vary substantially for natural guppy populations, with generally lower predation pressures upstream compared with downstream[26]. This has led to differences in female preference for male colouration[11], with downstream males less vivid due to reduced female preferences and higher predation risks than upstream populations[24,25,38]. Upstream and downstream populations within watersheds are more closely related to each other than across watersheds (Supplementary Fig. 6A). Therefore, shifts in male colouration have occurred independently in each watershed[39], where downstream males are less colourful than upstream males.

Given the very recent origin of the guppy sex chromosomes, we might expect that if recombination suppression is indeed driven by sexual conflict over colour, there might be differences in the divergence of the sex chromosomes across different populations with more or less male colouration. In line with this prediction, there is evidence that different populations of wild guppies display different patterns of Y-linkage of colour traits[40]. We, therefore, examined patterns of sex-specific heterozygosity for upstream and downstream populations of wild guppies. We sampled three watersheds (Yarra, Quare, Aripo) and from each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Our results (Fig. 4, Supplementary Table 7) show that across replicate upstream populations, where males are more colourful, there is significantly greater divergence between the X and Y chromosomes than the ancestral downstream populations (Wilcoxon rank sum test between upstream and downstream populations across watersheds, Yarra $P = 0.011$, Quare $P = 0.046$, Aripo $P = 0.017$). Expansion of the non-recombining region and corresponding X-Y divergence has occurred repeatedly and independently across populations, as the phylogeny of these populations reveals that in each watershed,
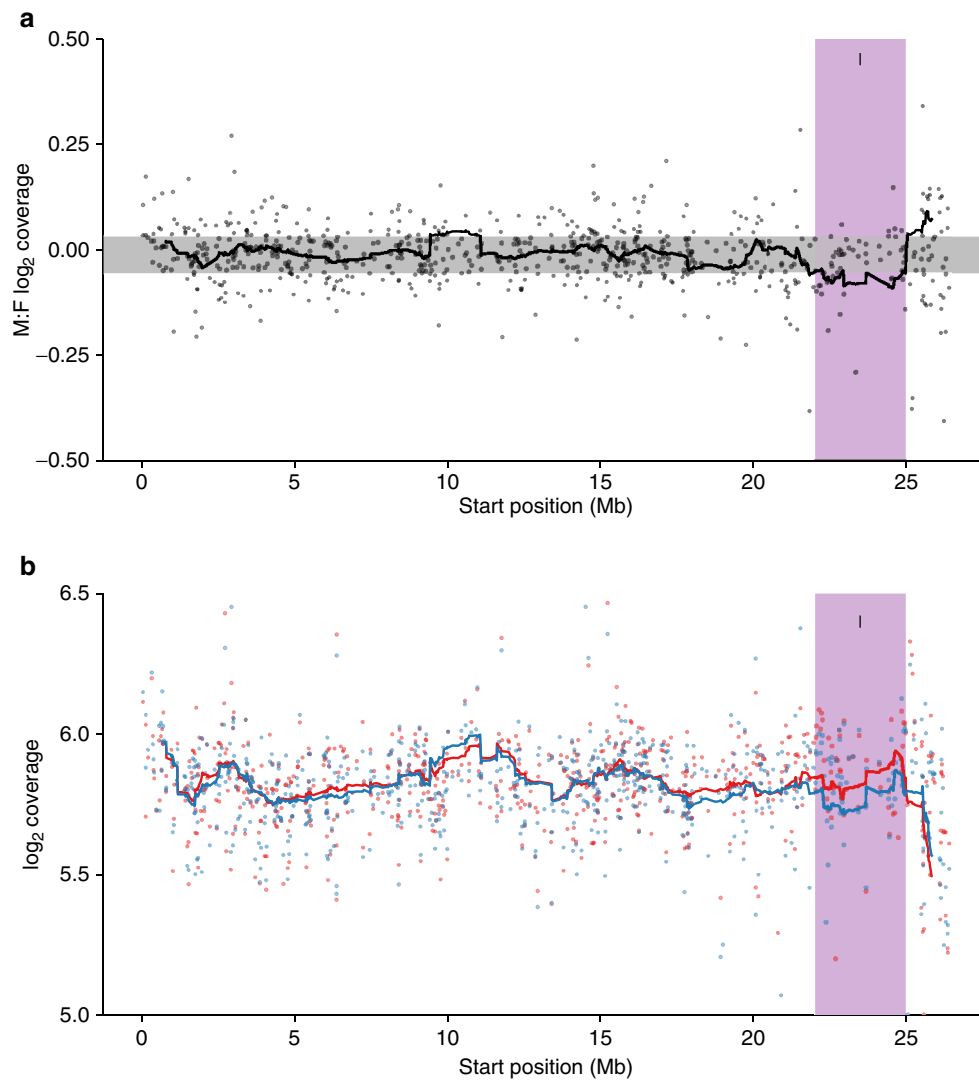
Figure 2 | Male and female coverage characteristics of guppy sex chromosome. (**a**) Moving average of coverage differences between male and female reads based on sliding window analysis (window size of 40 scaffolds). Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. (**b**) Male (blue) and female (red) coverage for the X chromosome. For both panels, dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22–25 Mb).

upstream populations are consistently derived from downstream populations (Supplementary Fig. 6B). By randomly sampling 10 Mb windows with 1,000 repetitions across the autosomes, we find that the probability of observing this convergence in SNP density across populations by chance is $P < 0.004$. In contrast, there are no differences in patterns of coverage between upstream and downstream populations in the area of greatest sex chromosome divergence (Stratum I, 22–25 Mb, Supplementary Fig. 7, Supplementary Table 8), indicating that X-Y divergence in this region predates the divergence of these wild populations.

Discussion

Observations of Y-linkage for a large proportion of male colour patterns in guppies[8,9] helped form the nucleus of theories regarding the role of sexual conflict in sex chromosome formation[3–5]. Here we used individuals from natural and laboratory populations in conjunction with analysis of coverage, SNP and expression differences between males and females in this model system to test the role of sexual conflict in recombination

suppression between the X and Y chromosomes. Our results suggest two regions of divergence on the sex chromosome. One region, likely the area of greatest Y chromosome divergence, is manifest with slightly reduced DNA coverage in males in a restricted region spanning 22–25 Mb. A larger region of more recent recombination suppression from 15–22 Mb is distinguishable only through the build-up of Y-specific SNPs. In both regions, although male-specific SNPs have accumulated on the Y, there is no evidence of large-scale decay of the Y chromosome or loss of gene activity observed in many older sex chromosome systems[31,41]. Surprisingly, this region of divergence extends over nearly half of the sex chromosome, indicating that recombination has been suppressed over a large region very recently. The two strata we observe in guppies are consistent with step-wise patterns of sex chromosome formation observed in many other organisms, including mammals[42], birds[21], *Silene*[43], sticklebacks[44] and *Nothobranchius*[45]. In the latter case, the authors observed population-level variation in the youngest stratum, similar to what we observe in guppies, suggesting that strata can form independently within species.
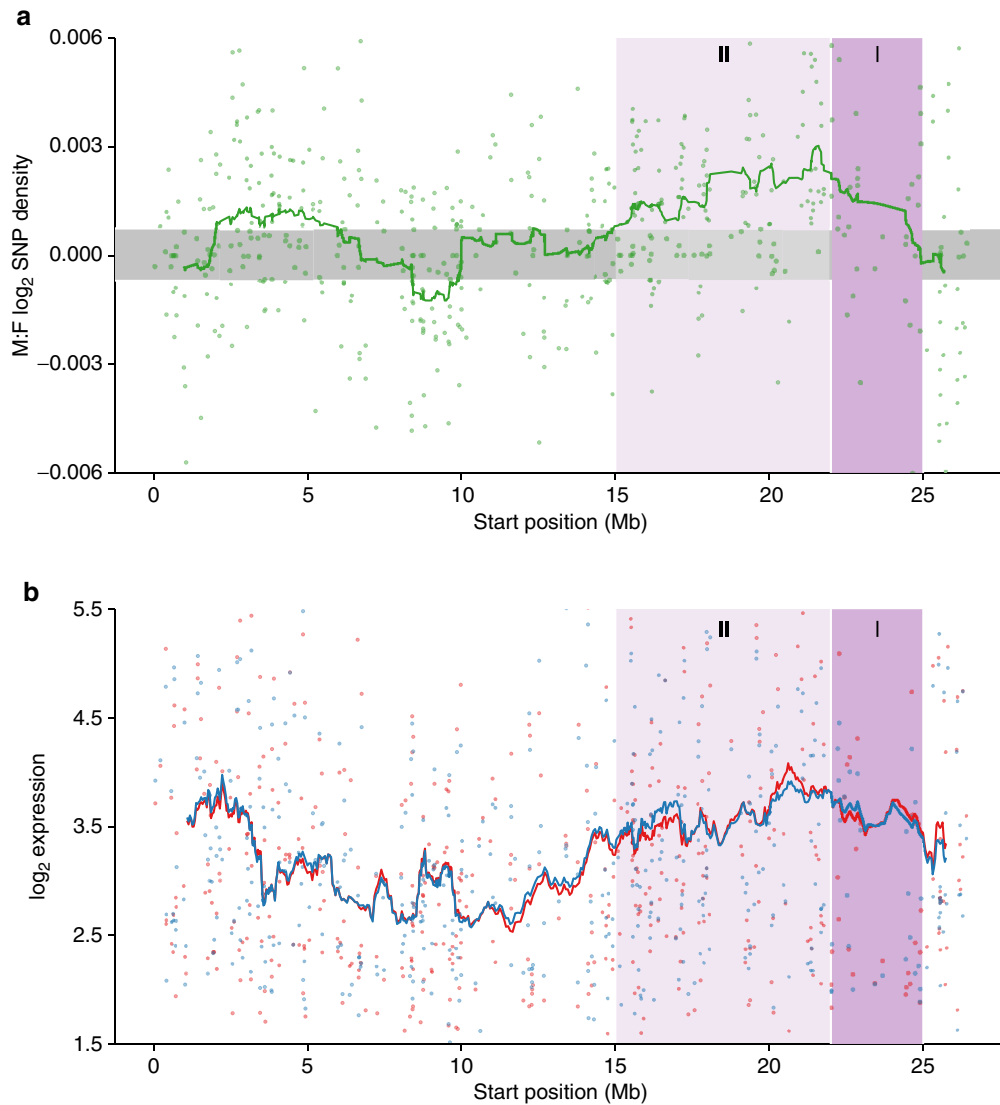
Figure 3 | Male and female SNP density and expression differences on guppy sex chromosome. (a) Moving average of male:female SNP density based on sliding window analysis (window size of 40 scaffolds). Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. (b) Male (blue) and female (red) expression of genes along the X chromosome (window size of 40 genes). Dark purple indicates the region of the sex chromosomes with the greatest X-Y sequence divergence, where coverage is significantly less in males (Stratum I, 22–25 Mb) (see Fig. 2), light purple indicates the region with less X-Y differentiation, where there is a significant excess of male SNPs (Stratum II, 15–22 Mb).

Comparisons of coverage and SNP density between males and females, like the analyses we implement here, offer two complementary views of sex chromosome evolution. Coverage differences are expected in more diverged regions with significant Y chromosome degeneration. In contrast, sex-differences in SNP density, particularly in regions with elevated SNP density in the heterogametic sex, are expected in more diverged systems with little Y chromosome degeneration. However, implementing these approaches in young sex chromosome systems should be accompanied by information as to the location of the sex determining region, which has been previously mapped to the far end of chromosome 12 (ref. 13). Ideally, Y-specific sequence data would be useful in verifying and dating stratum boundaries. However, this is complicated in our system due to the lack of complete lineage sorting of Y-specific SNPs, precluding the reconstruction of Y-specific sequence from our short-read data. In future work, long read RNA-seq data, optical mapping and other phasing approaches will be useful in confirming stratum

boundaries and identifying Y-linked sequences. These data will also be important in determining whether inversions, which are often assumed to be involved in recombination suppression, are indeed the mechanism behind sex chromosome divergence.

Despite the limited sequence divergence between the X and Y chromosomes, we observe two evolutionary signatures that are typically only associated with heteromorphic sex chromosome systems. First, the X chromosome shows the early stages of sexualization for gene expression despite limited evidence for degeneration in gene activity or content across the non-recombining Y chromosome. Previous evidence of sexualization comes from old, highly heteromorphic sex chromosome systems[46–48] and it was previously unclear how quickly sex-biased expression can accumulate after sex chromosome formation. Our results, therefore, indicate that sexualization of the X chromosome can occur very quickly after recombination is halted. Second, we detect a Faster-X effect, where X-linked coding sequence diverges more rapidly than the remainder of the
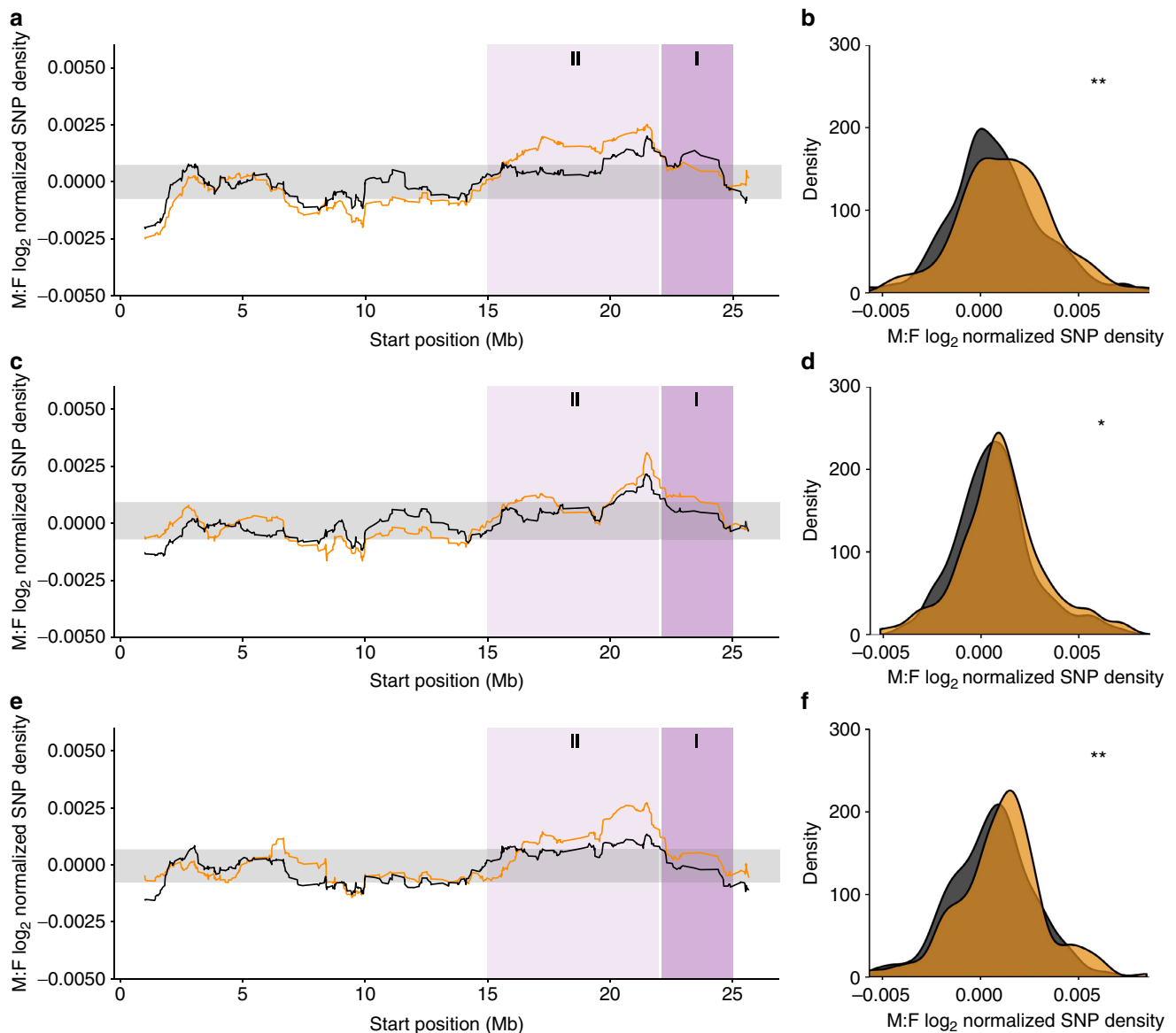
Figure 4 | Male:female SNP density for the X chromosome across upstream (orange) and downstream (black) guppy populations. (**a**,**c** and **e**) Moving averages of normalized SNP density across the X chromosome based on sliding window analysis (window size of 40 genes) for Yarra (**a**), Quare (**c**) and Aripo (**e**) watersheds. Ninety-five per cent confidence intervals based on bootstrapping autosomal estimates are in grey. Dark purple indicates the region of the sex chromosomes with the greatest X–Y sequence divergence, where coverage is significantly less in laboratory population males (Stratum I, 22–25 Mb) (see Fig. 2), light purple indicates the region with less X–Y differentiation, where there is a significant excess of male SNPs in laboratory populations (Stratum II, 15–22 Mb) (see Fig. 3). (**b**,**d** and **f**) Distribution of sex differences in normalized SNP density for the X–Y diverged region (Strata I and II, 15–25 Mb) for Yarra (**b**), Quare (**d**) and Aripo (**f**) watersheds. ******P-value < 0.020, *P-value < 0.050 based on permutation tests.

genome. Until now, evidence for Faster-X was restricted to highly diverged sex chromosomes[35–37], however, our results suggest that the Faster-X processes can accumulate rapidly following the loss of recombination. These findings have important consequences for the role of sex chromosomes in Haldane's rule[49] and the Large-X effect in speciation[50], and suggests that young or undifferentiated sex chromosomes may act as an important driver in the evolution of reproductive isolation[51].

Most systems where sex chromosomes have formed recently[28–30], and even some older sex chromosome systems[20,52], show restricted recombination in only a small region. The region of divergence extends over almost half of the

sex chromosomes in the guppy, suggesting that recombination has been suppressed very quickly over a large region of the Y chromosome in guppies. This rapid spread of recombination suppression may have been driven by the presence of sexually antagonistic alleles related to male colour on the proto-sex chromosome[10–12]. The high proportion of Y-linked colour patterns in guppies[8,9] is likely the product of rapid spread of recombination suppression between the X and Y chromosomes, which would resolve sexually antagonism by limiting colour expression to males.

Fish show remarkable variation in sex determination[6,53] and rapid origin and turnover of sex chromosomes[6,54]. The tiger pufferfish has homomorphic sex chromosomes, where

the sexes differ by only a single missense SNP[28], whereas a significant proportion of the sex chromosomes in sticklebacks are non-recombining[44,55], and there has been substantial decay of gene activity on the Y chromosome. Although studies in related species are required to date the exact age of the sex chromosomes in *P. reticulata*, there is extensive sex chromosome turnover in the poeciliid clade[54,56], suggesting a recent origin of the sex chromosomes described here. This is consistent with expectations that the expansion of the Y-limited region was driven by sexual conflict over colouration, suggesting that Stratum II originated around the same period that male colouration emerged as a major component of female preference, likely <5Mya (refs 22,23).

Our results suggest that this younger region of recombination suppression has expanded convergently in upstream populations as a consequence of increased sexual selection and sexual conflict over colouration[11,25]. We found the same convergent pattern of X and Y divergence between colourful upstream populations compared with the duller ancestral downstream populations over each of three replicate watersheds (Fig. 4). Upstream populations all showed greater divergence between the X and Y based on SNP density, and the region of significant SNP divergence extends over a larger region of the sex chromosomes. This accelerated divergence in upstream populations in each of the three watersheds has likely occurred independently, as populations within watersheds are well known to be monophyletic[39]. In support of this, our phylogenetic reconstruction reveals that in each watershed, upstream populations independently evolved from ancestral downstream populations. This suggests that sexual selection and sexual conflict over colour has driven greater Y divergence, consistent with longstanding theoretical predictions about the role of sexual antagonism in sex chromosome formation[3–5]. However, it is worth noting that our replicate upstream populations show some variation in the degree of differentiation, possibly due to demographic factors such as bottlenecks and recent expansions, date of colonization, rate of dispersal and gene flow between upstream and downstream populations, and effective population size, as well as stochastic processes.

Altogether, our results suggest that sexual conflict may be responsible for the remarkably rapid recent spread of recombination suppression to encompass colouration alleles within the Y chromosome. Moreover, our data are consistent with a role of sexual selection in accelerating divergence of the Y chromosome once recombination suppression is established.

Methods

Sample collection. All samples were collected in accordance with national and institutional ethical guidelines. First, we sampled males and females from a single large, outbred laboratory population established in 1998 (ref. 57). Tail samples were homogenized and stored in RNA later before RNA preparation, the remainder of each fish was stored in ethanol before DNA preparation.

Second, wild males were caught from three watersheds (Yarra, Quare, Aripo) in the Northern Range Mountains of Trinidad in February 2015 (see ref. 58 for description of the habitats). From each watershed, four males were caught from an upstream population and four males were caught from a downstream population. Samples were collected and stored immediately in ethanol prior to DNA preparation.

Sequencing. Nucleic acids were extracted with RNAeasy Kit (Qiagen) and DNeasy Blood and Tissue Kit (Qiagen) using the manufacturer protocols. The libraries were prepared and barcoded at The Wellcome Trust Centre for Human Genetics, University of Oxford using standard protocols. RNA was sequenced on an Illumina HiSeq 2500 resulting in on average 32 million 100 bp paired-end reads per sample. DNA was sequenced on an Illumina HiSeq 4000, resulting in on average 269 million 100 bp paired-end reads per individual sampled from a single large, outbred laboratory population, and 123 million 100 bp paired-end reads per sample for individuals caught in the wild in Trinidad (Supplementary Table 1).

Quality trimming and filtering. DNA data were quality assessed using FastQC v0.11.4 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and quality trimmed using Trimmomatic v0.35 (ref. 59). We filtered reads containing adaptor sequences and trimmed reads if the sliding window average Phred score over four bases was <15 or if the leading/trailing bases had a Phred score <3. Reads were removed post filtering if either read pair was <50 bases in length. RNA-seq data was quality assessed and trimmed using the same criteria but with a minimum length threshold of 36 bases (Supplementary Table 1).

De novo genome assembly. Reads used to construct *de novo* genome assemblies were error corrected with Quake v0.3.5, specifying default settings and a kmer length of 19 (ref. 60) (Supplementary Table 2). Optimal kmer length for *de novo* genome assemblies was estimated using kmergenie v1.6741 (ref. 61).

We constructed a female *de novo* genome assembly with DNA-seq reads from two females using SOAPdenovo v2.04 (ref. 62) and specifying the multi-kmer option with a starting kmer of 37 and max kmer of 55. All reads were used during both contig and scaffold assembly. During scaffolding (SOAPdenovo scaff), the –F parameter was set to specify that gaps in scaffolds should be filled. Lastly, GapCloser was used to close gaps emerging during the scaffolding process. Sequences <1 Kb in length were filtered from the assembly (Supplementary Table 3).

Assigning chromosomal position. Guppy genes were downloaded from RefSeq (Guppy_female_1.0 + MT, RefSeq assembly accession: GCF_000633615.1) and the longest isoform picked for each. Coding sequences were BLASTed against the *de novo* genome assembly using BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score.

De novo scaffolds were assigned to the guppy reference chromosomes and oriented using the chromosomal location and start position of mapped guppy genes. If multiple genes mapped to a given scaffold, the scaffold was assigned to the reference chromosome that the majority of genes were located on. Specifically, at least 70% of genes mapping to a given scaffold must be located on the same chromosome in the reference genome otherwise the scaffold was discarded. The degree of concordance in assigned chromosome position using this approach is high (Supplementary Table 4), and only 320 scaffolds from the female genome assembly were discarded due to discordance between chromosomal locations.

Genomic coverage analysis. Male and female trimmed DNA-seq reads were separately mapped to the *de novo* genome assembly using Bwa v0.7.12 aln/sampe with default settings[64]. Uniquely mapped reads were extracted using grep 'XT:A:U' and soap.coverage v2.7.9 (http://soap.genomics.org.cn) was used to extract coverage of scaffolds in every individual. For each scaffold, coverage was defined as the total number of times each site was sequenced divided by the number of sites that were sequenced.

For lab populations, average coverage values were calculated for females and males separately. We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Male:female coverage was calculated for each scaffold as $\log_2$(average male coverage) – $\log_2$(average female coverage).

For upstream and downstream wild populations, coverage was estimated using Bwa v0.7.15 aln/sampe and the same pipeline as the lab populations. Average coverage was calculated for each gene separately across each population. To account for differences in sequencing depth across populations, the $\log_2$ coverage for each gene was normalized by the median $\log_2$ coverage of X chromosome ($\log_2$ coverage—median $\log_2$ coverage of X chromosome). Male:female coverage was estimated for each population relative to the normalized coverage of the female lab population.

Polymorphism analysis. Male and female trimmed DNA-seq reads from both wild and lab populations were separately mapped to the *de novo* genome assembly using Bowtie1 v1.1.2 (ref. 65), specifying a maximum insert size for paired-end alignment of 1,400 and writing hits in map format. Map files were sorted by scaffold and bow2pro v0.1 (http://guanine.evolbio.mpg.de/) was used to generate a profile for each sample. Sites with coverage <10 were excluded from the analysis and SNPs were called when a site had a major allele frequency of 0.3 times the site coverage. SNPs were only included in further analyses if they were located within genic regions (see Expression analysis method for detail on gene annotation). Average SNP density for each gene was calculated as sum(SNPs) divided by sum(no. of filtered sites). We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Genes were excluded if zero sites remained after filtering.

For lab populations, average SNP density was calculated separately for males and females. Male:female SNP density was calculated for each gene as $\log_2$(average male SNP density) – $\log_2$(average female SNP density).

For upstream and downstream wild populations, average SNP density was calculated for each gene separately across each population. To account for differences in overall genetic diversity across populations, the $\log_2$ SNP density for each gene was normalized by the median $\log_2$ SNP density of X chromosome ($\log_2$ SNP density—median $\log_2$ SNP density of X chromosome). Male:female

SNP density was estimated for each population relative to the normalized SNP density of the female lab population.

To calculate the probability that the convergence in patterns of SNP density across populations we observe is due to chance, we randomly sampled 10 Mb windows across the autosomes 1,000 times. For each window, we tested whether the upstream normalized male:female SNP density was greater than the downstream population in each river using a one-tailed Wilcoxon ranked sum test. We looked for windows where all three rivers had P-values < 0.05 and the median SNP density in the lab population was greater than the 95% autosomal confidence interval.

Expression analysis. Male and female trimmed RNA-seq reads were separately mapped to the *de novo* genome assembly using HISAT2 v2.0.4 (ref. 66), suppressing unpaired and discordant alignments for paired reads and excluding reads from the sam output that failed to align. Reported alignments were tailored for transcript assemblers including StringTie.

Sam files were coordinate sorted using SAMtools v1.2 (ref. 67) and converted to bam files. StringTie v1.2.3 (ref. 67) was used to quantify gene expression and annotate the *de novo* assembly.

Specifically, StringTie was run on each sample with default settings and the output GTF files were merged. The combined GTF file was filtered to remove non-coding RNA (ncRNA) and transcripts less than 50 bp in length. Specifically, transcript sequences were extracted using bedtools getfasta[68] and BLASTed to *Oryzias latipes* (MEDAKA1), *Gasterosteus aculeatus* (BROADS1), *Poecilia formosa* (PoeFor_5.1.2) and *Danio rerio* (GRCz10) ncRNA downloaded from Ensembl 84 (ref. 69). Transcripts with blast hits to ncRNA were removed from the GTF file. StringTie was rerun on each sample and expression was only estimated for genes defined in the filtered GTF file. A minimum expression threshold of 2FPKM in at least half of the individuals of either sex was imposed. This final filtered data set (23,603 genes) was used in subsequent expression and polymorphism analyses.

Expression was normalized using EdgeR[70]. Sam files were name sorted using SAMtools and HT-seq count v0.6.1 (ref. 71) used to extract read counts for each gene. Genes were excluded if they were not located on scaffolds assigned to the guppy reference genome. In all, 13,306 genes remained after filtering. Expression was normalized using TMM (trimmed mean of m-values) in EdgeR and RPKM estimated for each gene. Individuals cluster transcriptomically by sex (Supplementary Fig. 4). Average RPKM for each gene was calculated separately for males and females. We added 1 to each value to avoid infinitely high numbers associated with $\log_2 0$. Male:female expression was calculated for each gene as $\log_2$ (average male RPKM)—$\log_2$(average female RPKM).

We tested whether there was an enrichment of GO terms in the X–Y diverged region compared with the rest of the genome. *Danio rerio* (GRCz10) coding sequences were downloaded from Ensembl 84 (ref. 69) and the longest isoform extracted for each gene. Longest isoforms were extracted for our set of expressed guppy genes and BLASTed to *D. rerio* using BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score. *D. rerio* orthologues were identified for genes in the X–Y degenerate region (15–25 Mb) and compared with the remainder of the genome using GOrilla[72,73].

Cluster analysis of expression data. Transcriptional similarity of normalized count data for female and male individuals was assessed using a multi-dimensional scaling plot (MDS) with default settings in EdgeR[74]. RPKM data was clustered using the R package pheatmap and boostrap values calculated using pvclust. UPGMA was used in the hierarchical cluster analysis and the distance matrix was computed using the Euclidean method.

Calculating moving averages. Moving averages of coverage/polymorphism/expression were calculated in R[75] based on sliding window analyses using the roll_mean function. Ninety-five per cent confidence intervals for the moving average were calculated by randomly resampling (1,000 times, without replacement) autosomal scaffolds (coverage analysis) or genes (SNP density and expression analyses).

Faster-X analysis. Guppy transcript sequences were extracted using bedtools getfasta[68] and the longest isoform chosen for each of the 23,603 genes. Genes on genomic scaffolds without chromosomal locations were removed, leaving 13,306 genic sequences for the Faster-X analysis. *Oryzias latipes* (MEDAKA1), *Xiphophorus maculatus* (Xipmac4.4.2), *Poecilia formosa* (PoeFor_5.1.2) were downloaded from Ensembl 84 (ref. 69) and the longest transcript for each gene was identified. We determined orthology using reciprocal BLASTn v2.3.0 (ref. 63) with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30%. When genes mapped to multiple locations, the top blast hit was chosen using the highest BLAST score. In all, 7,382 reciprocal 1-1 orthologues across the four species were identified. We obtained open reading frames and protein coding sequence with BLASTx v2.3.0 with an e-value cutoff of $10e^{-10}$ and minimum percentage identity of 30% using the approach in Wright et al.[76] Reciprocal orthologues with no BLASTx hits or a valid protein-coding sequence were excluded.

Reciprocal orthologues were aligned with PRANK v.140603 (ref. 77) using the codon model and specifying the following guidetree; (((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*), *Oryzias latipes*). SWAMP v 31-03-14 (ref. 78) was used to mask erroneous sequences in the alignments. Reciprocal orthologues were discarded if the alignment length was < 300 bp after removing gaps and masked sites. After this length filter, 5,349 reciprocal orthologues remained.

We used the branch model (model = 2, nssites = 0) in the CODEML package in PAML v4.8 (ref. 79) to obtain divergence estimates using the following phylogeny; ((*Poecilia reticulata*, *Poecilia formosa*), *Xiphophorus maculatus*, *Oryzias latipes*). The branch model was used to calculate mean d_N/d_S across the *Poecilia reticulata* branch. As mutational saturation and double hits can lead to inaccurate divergence estimates[80] orthogroups were excluded if $d_S > 2$.

Orthologues were divided into genomic categories on the basis of their chromosomal location. For each category, mean d_N and mean d_S were calculated as the sum of the number of substitutions across all orthologues divided by the number of sites (d_N = sum D_N/sum N, d_S = sum D_S/sum S, where D_N and D_S are estimates of the number of nonsynonymous or synonymous substitutions and N and S are the number of nonsynonymous/synonymous sites). This approach prevents disproportionate weighting of shorter genes by avoiding the problems of infinitely high d_N/d_S estimates arising from sequences with extremely low d_S (refs 76,81,82).

Significant differences in d_N, d_S and d_N/d_S between genomic categories were determined using permutation tests with 1,000 replicates. One-tailed tests were used to test for the Faster-X effect where we predict d_N/d_S is greater for X-linked gene relative to the autosomes. Two-tailed tests were used to test for differences in d_N and d_S. Bootstrapping with 1,000 repetitions was used to generate 95% confidence intervals.

Phylogenetic history of guppy populations. Using DNA-seq data, we reconstructed the phylogenetic relationships between the six wild populations. We mapped trimmed reads to the previously sequenced guppy genome (Guppy_female_1.0 + MT, RefSeq assembly accession: GCF_000633615.1) using Stampy v1.0.28 (ref. 83) with a substitution rate of 0.01. After mapping, sam files were converted to bam and coordinate sorted using SAMtools v1.2 (ref. 84) and then deduplicated using Picard tools v1.136 (ref. 85). Subsequently, we added read groups and merged libraries belonging to the same individual using Picard. We then called variants on all 24 individuals simultaneously using two independent methods (GATK and Platypus), and retained only SNPs called reliably with both methods and passing quality control filters.

As part of the GATK variant calling pipeline, v3.4.46 (ref. 86), we first realigned reads around indels and recalibrated base quality scores. We then proceeded with variant calling using the HaplotypeCaller and GenotypeGVCFs tools. The second method we employed to call variants was Platypus v0.8.1 (ref. 87), which we ran in assembly mode, restricting calling to reads mapping to the 23 canonical chromosomes (that is, excluding those mapped to unplaced scaffolds).

After variant calling we removed indels, intersected the GATK and Platypus SNP sets, and applied stringent quality filtering. We removed singleton SNPs, multiallelic SNPs and SNPs failing the following quality thresholds: quality by depth > 2, coverage $> 0.5x$ and $< 2x$ mean coverage, > 2 reads for the alternative allele, mapping quality > 40, allele bias Z score for mapping quality, base quality or read position < -1.96, or strand bias Fisher exact test $P > 0.05$. We also removed SNPs with missing genotype in any individual. This yielded 4.6 million high-quality SNPs.

Next, we used R package adegenet v2.0.1 (ref. 88) to construct a Euclidian distance matrix for the 24 individuals based either on all SNPs across the genome or on only the 72,623 SNPs between 15 and 25 MB on the X chromosome. We used the R package ape v3.5 (ref. 89) to produce from each matrix a simple neighbour joining tree to visualize the genetic distance between the six populations, and performed 100 bootstrap iterations to assess support for each node.

References

1. Bachtrog, D. Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. *Nat. Rev. Genet.* **14,** 113–124 (2013).
2. Bachtrog, D. et al. Are all sex chromosomes created equal? *Trends Genet.* **27,** 350–357 (2011).
3. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, 1983).
4. Fisher, R. A. The evolution of dominance. *Biol. Rev.* **6,** 345–368 (1931).
5. Rice, W. R. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41,** 911–914 (1987).
6. Bachtrog, D. et al. Sex determination: why so many ways of doing it? *PLoS Biol.* **12,** e1001899 (2014).
7. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7,** 12087 (2016).

8. Lindholm, A. & Breden, F. Sex chromosomes and sexual selection in poeciliid fishes. *Am. Nat.* **160**, S214–S224 (2002).

9. Winge, M. The location of eighteen genes in *Lebistes reticulata*. *J. Genet.* **18**, 1–43 (1927).

10. Endler, J. A. Natural selection on color patterns in Poecilia reticulata. *Evolution* **34**, 76–91 (1980).

11. Houde, A. E. & Endler, J. A. Correlated evolution of female mating preferences and male color patterns in the guppy, Poecilia reticulata. *Science* **248**, 1405–1408 (1990).

12. Kemp, D. J., Reznick, D. N., Grether, G. F. & Endler, J. A. Predicting the direction of ornament evolution in Trinidadian guppies (Poecilia reticulata). *Proc. R. Soc. B-Biol. Sci.* **276**, 4335–4343 (2009).

13. Tripathi, N., Hoffmann, M., Weigel, D. & Dreyer, C. Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (Poecilia reticulata). *Genetics* **182**, 365–374 (2009).

14. Lisachov, A. P., Zadesenets, K. S., Rubtsov, N. B. & Borodin, P. M. Sex chromosome synapsis and recombination in male guppies. *Zebrafish* **12**, 174–180 (2015).

15. Lenormand, T. The evolution of sex dimorphism in recombination. *Genetics* **163**, 811–822 (2003).

16. Nanda, I. *et al.* Sex chromosome polymorphism in guppies. *Chromosoma* **123**, 373–383 (2014).

17. Traut, W. & Winking, H. Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res.* **9**, 659–672 (2001).

18. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–U822 (2003).

19. Vicoso, B., Emerson, J. J., Zektser, Y., Manajan, S. & Bachtrog, D. Comparative sex chromosome divergence in snakes: differentiation and lack of global dosage compensation. *PLoS Biol.* **11**, e1001643 (2013).

20. Vicoso, B., Kaiser, V. B. & Bachtrog, D. Sex-biased gene expression at homomorphic sex chromosomes in Emus and its implications for sex chromosome evolution. *Proc. Natl Acad. Sci., USA* **110**, 6453–6458 (2013).

21. Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A. & Mank, J. E. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* **68**, 3281–3295 (2014).

22. Meredith, R. W., Pires, M. N., Reznick, D. N. & Springer, M. S. Molecular phylogenetic relationships and the evolution of the placenta in Poecilia (Micropoecilia) (Poeciliidae: Cyprinodontiformes). *Mol. Phylogenet. Evol.* **55**, 631–639 (2010).

23. Pollux, B. J. A., Meredith, R. W., Springer, M. S., Garland, T. & Reznick, D. N. The evolution of the placenta drives shift in sexual selection in livebearing fish. *Nature* **513**, 233–236 (2014).

24. Endler, J. A. Nateral selection on color patterns in Poecilia reticulata. *Evolution* **34**, 76–91 (1980).

25. Endler, J. A. Natural and sexual selection on color patterns in Poeciliid fishes. *Env. Biol. Fish.* **9**, 173–190 (1983).

26. Endler, J. A. Multiple trait coevolution and environmetnal gradients in guppies. *Trends Ecol. Evol.* **10**, 22–29 (1995).

27. Gordon, S. P., Lopez-Sepulcre, A. & Reznick, D. N. Predation-associated differences in sex linkage of wild guppy coloration. *Evolution* **66**, 912–918 (2012).

28. Kamiya, T. *et al.* A Trans-Species missense SNP in Amhr2 Is Associated with Sex determination in the Tiger Pufferfish, Takifugu rubripes (Fugu). *Plos Genet.* **8**, e1002798 (2012).

29. Liu, Z. Y. *et al.* A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**, 348–352 (2004).

30. Russell, J. R. W. & Pannell, J. R. Sex determination in dioecious Mercurialis annua and its close diploid and polyploid relatives. *Heredity* **114**, 262–271 (2015).

31. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *Plos Biol.* **13**, e1002078 (2015).

32. Hough, J., Hollister, J. D., Wang, W., Barrett, S. C. H. & Wright, S. I. Genetic degeneration of old and young Y chromosomes in the flowering plant Rumex hastatulus. *Proc. Natl Acad. Sci. USA* **111**, 7713–7718 (2014).

33. Muyle, A. *et al.* Rapid de novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. *Plos Biol.* **10** (2012).

34. Charlesworth, B., Coyne, J. A. & Barton, N. H. The relative rates of evolution of sex-chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).

35. Zhou, Q. & Bachtrog, D. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* **337**, 341–345 (2012).

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

37. Mank, J. E., Vicoso, B., Berlin, S. & Charlesworth, B. Effective population size and the faster-x effect: empirical results and their interpretation. *Evolution* **64**, 663–674 (2010).

38. Houde, A. E. & Endler, J. A. Correlate evolution of female mating preferences and male color patterns in the guppy Poecilia reticulata. *Science* **248**, 1405–1408 (1990).

39. Fraser, B. A., Kunstner, A., Reznick, D. N., Dreyer, C. & Weigel, D. Population genomics of natural and experimental populations of guppies (Poecilia reticulata). *Mol. Ecol.* **24**, 389–408 (2015).

40. Gordon, S. P. *et al.* Selection analysis on the rapid evolution of a secondary sexual trait. *Proc. R. Soc. B-Biol. Sci.* **282**, 20151244 (2015).

41. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *Plos Biol.* **13**, 22 (2015).

42. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).

43. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia:* Evidence from new sex-linked genes. *Genetics* **175**, 1945–1954 (2007).

44. White, M., Kitano, J. & Peichel, C. L. Purifying selection maintains dosage sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol. Biol. Evol.* **32**, 1981–1995 (2015).

45. Reichwald, K. *et al.* Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* **163**, 1527–1538 (2015).

46. Arunkumar, K. P., Mita, K. & Nagaraju, J. The silkworm Z chromosome is enriched in testis-specific genes. *Genetics* **182**, 493–501 (2009).

47. Meisel, R. P., Malone, J. H. & Clark, A. G. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res.* **22**, 1255–1265 (2012).

48. Wright, A. E., Moghadam, H. K. & Mank, J. E. Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics* **192**, 1433-+ (2012).

49. Haldane, J. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**, 101–109 (1922).

50. Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *Plos Biol.* **5**, 1890–1898 (2007).

51. Dufresnes, C. *et al.* Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. *Sci. Rep.* **6**, 21029 (2016).

52. Stock, M. *et al.* Ever-young sex chromosomes in European tree frogs. *Plos Biol.* **9**, e1001062 (2011).

53. Mank, J. E., Promislow, D. E. L. & Avise, J. C. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87**, 83–93 (2006).

54. Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364 (2002).

55. Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079–1083 (2009).

56. Mank, J. E., Promislow, D. E. L. & Avise, J. C. Evolution of alternative sex determining mechanisms in teleost fishes. *Biol. J. Linn. Soc.* **87**, 83–93 (2006).

57. Kortrschal, A. *et al.* Experimental evidence for costs and benefits of evolving a larger brain. *Curr. Biol.* **23**, 168–171 (2013).

58. Sandkam, B., Young, C. M. & Breden, F. Beauty in the eye of the beholder: colour vision is tuned to mate preference in the Trinidadian Guppy (Poecilia reticulata). *Mol. Ecol.* **24**, 596–609 (2016).

59. Lohse, M. *et al.* RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **40**, W622–W627 (2012).

60. Kelley, D. R., Schatz, M. C. & Salzberg, S. L. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* **11**, R116 (2010).

61. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**, 31–37 (2014).

62. Luo, R. B. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).

63. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. basic local alignment search tooL. *J. Mol. Biol.* **215**, 403–410 (1990).

64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

65. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

66. Kim, D., Landmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–U121 (2015).

67. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

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

69. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** D749–D755 (2014).

70. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

71. Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

72. Eden, E., Lipson, D., Yogev, S. & Yakhini, Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* **3,** 508–522 (2007).

73. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48 (2009).

74. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 138–140 (2010).

75. R Core Team. R: A language and environment for statistical computing: R Foundation for Statistical Computing. https://www.R-project.org/ (2015).

76. Wright, A. E. *et al.* Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Mol. Ecol.* **24,** 1218–1235 (2015).

77. Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102,** 10557–10562 (2005).

78. Harrison, P. W., Jordan, G. E. & Montgomery, S. H. SWAMP: Sliding Window Alignment Masker for PAML. *Evol. Bioinform.* **10,** 197–204 (2014).

79. Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).

80. Axelsson, E. *et al.* Natural selection in avian protein-coding genes expressed in brain. *Mol. Ecol.* **17,** 3008–3017 (2008).

81. Mank, J. E., Axelsson, E. & Ellegren, H. Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res.* **17,** 618–624 (2007).

82. Mank, J. E., Nam, K. & Ellegren, H. Faster-Z evolution is predominantly due to genetic drift. *Mol. Biol. Evol.* **27,** 661–670 (2010).

83. Lunter, G. & Goodson, M. Stampy: A statistical algorithm for sensitive and fast mapping of illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).

84. Li, H. *et al.* Sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

85. Broad Institute. Picard toolkit. http://broadinstitute.github.io/picard.

86. McKenna, A. *et al.* The genome analysis toolkit: A Mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

87. Rimmer, A. *et al.* Intrating mapping, assembly and haplotype based approaches for calling variance in clinical sequencing applications. *Nat. Genet.* **46,** 912–918 (2014).

88. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27,** 3070–3071 (2011).

89. Paradis, E., Claude, J. & Strimmer, K. APE: Analysis of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–290 (2004).

Acknowledgements

Author contributions

J.E.M. conceived of the study, and J.E.M. and A.E.W. designed the experiments. N.K., S.D.B., B.S. and F.B. collected samples used for sequencing in this study. A.E., I.D., N.I.B., V.O. and B.V. analysed the data. All authors contributed to writing the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wright, A. E. *et al.* Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* **8,** 14251 doi: 10.1038/ncomms14251 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ORIGINAL ARTICLE

Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*

Iulia Darolti[1] (iD) | Alison E. Wright[1,2] | Pascal Pucholt[3,4] (iD) | Sofia Berlin[3]* | Judith E. Mank[1,5]*

[1]Department of Genetics, Evolution and Environment, University College London, London, UK

[2]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

[3]Department of Plant Biology, Linnean Centre for Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[4]Array and Analysis Facility, Department of Medical Science, Uppsala University, Uppsala, Sweden

[5]Department of Organismal Biology, Uppsala University, Uppsala, Sweden

Correspondence
Iulia Darolti, Department of Genetics, Evolution and Environment, University College London, London, UK.
Email: iulia.darolti.15@ucl.ac.uk

Funding information
Helge Ax:son Johnsons Stiftelse; Vetenskapsrådet, Grant/Award Number: 2011-3544; European Research Council, Grant/Award Number: 260233, 680951; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/M009513/1

Abstract

The relative rate of evolution for sex-biased genes has often been used as a measure of the strength of sex-specific selection. In contrast to studies in a wide variety of animals, far less is known about the molecular evolution of sex-biased genes in plants, particularly in dioecious angiosperms. Here, we investigate the gene expression patterns and evolution of sex-biased genes in the dioecious plant *Salix viminalis*. We observe lower rates of sequence evolution for male-biased genes expressed in the reproductive tissue compared to unbiased and female-biased genes. These results could be partially explained by the lower codon usage bias for male-biased genes leading to elevated rates of synonymous substitutions compared to unbiased genes. However, the stronger haploid selection in the reproductive tissue of plants, together with pollen competition, would also lead to higher levels of purifying selection acting to remove deleterious variation. Future work should focus on the differential evolution of haploid- and diploid-specific genes to understand the selective dynamics acting on these loci.

KEYWORDS
codon usage bias, dioecious angiosperms, sex-biased gene expression, sexual selection

1 | INTRODUCTION

Many species show a wealth of phenotypic differences between the sexes (Parsch & Ellegren, 2013). However, apart from genes on sex chromosomes, males and females share the same genome, and sexually dimorphic traits are therefore thought to arise as a result of differential regulation of genes occurring in both sexes (Ellegren & Parsch, 2007; Mank, 2017; Pointer, Harrison, Wright, & Mank, 2013;

Ranz, Castillo-Davis, Meiklejohn, & Hartl, 2003), often referred to as sex-biased gene expression. Sex-biased genes are thought to evolve in response to conflicting sex-specific selection pressures over optimal expression acting on this shared genetic content (Connallon & Knowles, 2005) and are increasingly used to study the footprint of sex-specific selection within the genome (Dean et al., 2017; Gossmann, Schmid, Grossniklaus, & Schmid, 2014; Mank, 2017).

In contrast to animals, where sexual dimorphism is more frequent, only a small percentage (~5%) of flowering plants are dioecious (Renner, 2014; Robinson et al., 2014), where individuals have

*Joint senior authors.

exclusively male or female reproductive organs. The majority (~90%) of angiosperms are hermaphroditic (Ainsworth, 2000; Barrett & Hough, 2013), where flowers are bisexual, while another small fraction are monoecious, where separate flowers within the same plant carry different reproductive organs (Renner, 2014). Despite being rare, dioecy has evolved in flowering plants many times independently (Charlesworth, 2002) and is distributed across the majority of angiosperm higher taxa (Heilbuth, 2000; Käfer, Marais, & Pannell, 2017).

Although sexual dimorphism is generally more extensive in animal species, male and female dioecious flowering plants also undergo conflicts over trait optima and are subject to natural and sexual selection leading to a range of phenotypic sexual differences (Barrett & Hough, 2013). Studies of differential male and female gene expression patterns in plants (Muyle, Shearn, & Marais, 2017) indicate that sex-biased gene expression plays a role in the evolution of sexual dimorphism in morphological (e.g., anther and ovule development pathways in asparagus, Harkess et al., 2015), physiological (e.g., salinity tolerance in poplars, Jiang et al., 2012) and ecological traits (e.g., response to fungal infection in *Silene latifolia*, Zemp, Tavares, & Widmer, 2015).

Extensive studies in plants and animals have shown that genes with sex-biased expression vary in abundance across different developmental stages and tissues (Grath & Parsch, 2016; Perry, Harrison, & Mank, 2014; Robinson et al., 2014; Zemp et al., 2016; Zluvova, Zak, Janousek, & Vyskot, 2010). Evolutionary dynamics analyses also indicate that different evolutionary pressures impact the rate of sequence evolution of sex-biased genes; for example, sex-biased genes in reproductive tissues tend to have different rates of protein evolution compared to unbiased genes (Dean et al., 2017; Lipinska et al., 2015; Mank, Nam, Brunström, & Ellegren, 2010a; Perry et al., 2014; Sharma et al., 2014). In animal systems, where the rates of sequence divergence of sex-biased genes have been studied more widely, male-biased genes in many species, including *Drosophila* and adult birds, tend to be more numerous and to have higher expression and divergence rates (Assis, Zhou, & Bachtrog, 2012; Grath & Parsch, 2016; Harrison et al., 2015; Khaitovich et al., 2005) compared to female-biased and unbiased genes. This has often been interpreted as the signature of sexual selection, particularly sperm competition (Ellegren & Parsch, 2007). However, studies in other organisms have reported elevated rates of evolution in female-biased genes (Mank et al., 2010a; Whittle & Johannesson, 2013), leading to questions about the relationship between rates of evolution and sexual selection. In *Arabidopsis*, genes expressed in pollen have lower rates of evolution (Gossmann et al., 2014). Moreover, nonadaptive evolutionary processes have been shown to drive the fast rates of sequence evolution observed in sex-biased genes in some systems (Gershoni & Pietrokovski, 2014; Harrison et al., 2015) perhaps related to relaxed purifying selection (Hunt et al., 2011).

Sexual selection in flowering plants is also thought to be strong (Moore & Pannell, 2011), acting on gene expression patterns predominantly through pollen competition. Male

gametophytic tissue in *Arabidopsis thaliana* and rice has been shown to express a higher proportion of recently evolved genes compared to other tissues (Cui et al., 2015). Some of these young genes possess essential pollen-specific functions, suggesting a role for pollen competition in facilitating de novo gene development. As male-biased mutation is thought to be strong due to the elevated numbers of germ cell divisions in male cells (Whittle & Johnston, 2003), pollen competition, in this case, was suggested to counteract the potentially negative effects of higher mutation rates present in male gametophytes (Cui et al., 2015). Similarly, younger genes in the gametophyte of *A. thaliana*, rice and soya bean were also found to have higher rates of evolution compared to genes in the sporophytic tissue, however to varying degrees in males and females (Gossmann, Saleh, Schmid, Spence, & Schmid, 2016). Suggested reasons for these findings concerned the lower tissue complexity, and hence lower genetic interaction, in the gametophyte as well as differences between the sexes.

Plants additionally differ from animals in having a longer haploid phase in their life cycle, suggesting that haploid selection may act more forcefully to remove mildly deleterious recessive variation in pollen-expressed genes. Previous work on *A. thaliana* showed that the predominance of selfing, and similarly the intragametophytic selfing in moss species (Szövényi et al., 2014), leads to the more effective purging of mildly deleterious recessive variation (Gossmann et al., 2014). In the obligate outcrossing plant *Capsella grandiflora*, pollen-specific genes, but not sperm-enriched genes, evolve under both stronger purifying and positive selection compared to genes from sporophytic tissues (Arunkumar, Josephs, Williamson, & Wright, 2013). These findings are indicative of a potential combined effect of haploid selection and pollen competition acting in pollen-specific cells, whereas selective pressures are expected to be more relaxed for sperm-specific genes as there is no competition between them (Arunkumar et al., 2013).

These studies make it increasingly clear that many evolutionary forces shape the sequence evolution of sex-biased genes, including sexual selection through sperm competition (Ellegren & Parsch, 2007), haploid selection and natural selection (Ingvarsson, 2010). Particularly in plants, in order to understand the relative contribution of these forces, it is important to study rates of evolution in species with different levels of gamete competition, motivating studies on outcrossing dioecious species.

The basket willow, *Salix viminalis*, is a dioecious woody angiosperm (Cronk, Needham, & Rudall, 2015), belonging, together with other willow and poplar (*Populus*) species, to the Salicaceae family. *S. viminalis* is characterized by rapid seed development and growth (Ghelardini et al., 2014); it is both insect- and wind-pollinated (Peeters & Totland, 1999); and it has a recently evolved ZW sex chromosome system (Pucholt, Wright, Conze, Mank, & Berlin, 2017). Willow and poplar species have reproductive structures characterized by clusters of unisexual inflorescences referred to as catkins (Figure 1). Flowers in male willow catkins present a reduced number of stamens with anthers and filaments; however, they lack a vestigial
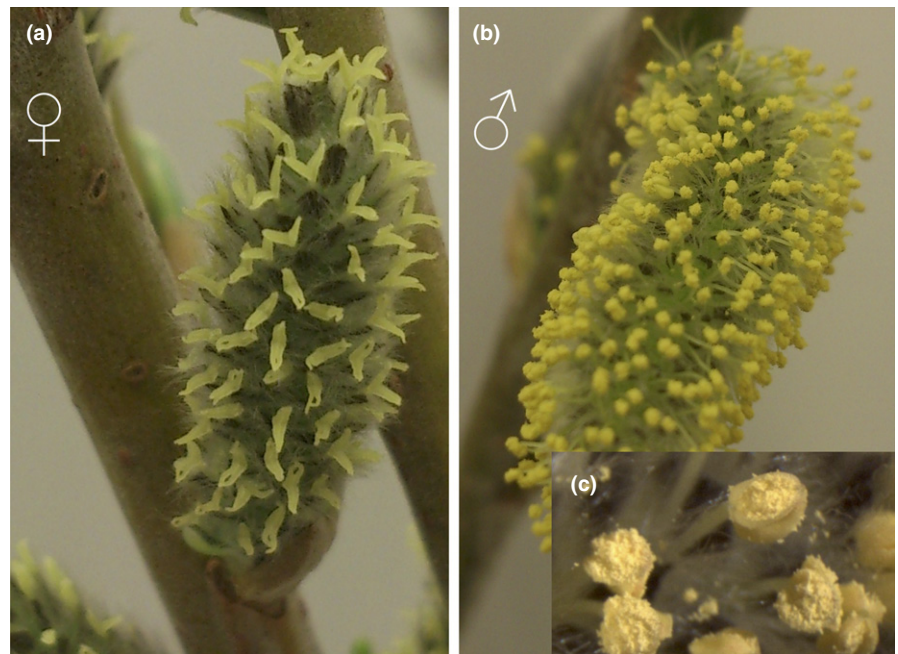
FIGURE 1 Physical appearance of adult *S. viminalis* catkins. (a) Female catkins with protruding pistillate flowers. (b) Male catkins with protruding staminate flowers. (c) Anthers of male catkins abundant in pollen grains

ovary, indicating floral reduction compared to other related non-catkin-bearing dioecious species (Cronk et al., 2015; Fisher, 1928). Flowers in female willow catkins contain pistils with style, stigma and an ovary. However, they also show a high degree of floral reduction as there is an absence of staminodes and, similarly to male catkins, they lack a perianth with petals and sepals (Cronk et al., 2015; Fisher, 1928; Karrenberg, Kollmann, & Edwards, 2002), potentially with a role in facilitating wind pollination (Karrenberg et al., 2002).

Our study of gene expression patterns in male and female *S. viminalis* individuals begins to explore the selective forces acting on sex-biased gene evolution in dioecious plants. We analysed sex-biased gene expression patterns in *S. viminalis* from two different tissues, vegetative (leaf) and sex-specific reproductive (catkin) tissue. We found the reproductive tissue to be more transcriptionally dimorphic and identified overall higher expression levels for male-biased genes than for female-biased genes, consistent with previous studies (Grath & Parsch, 2016). Interestingly, however, we found that in catkin, male-biased genes on the autosomes and the pseudoautosomal region have significantly lower rates of sequence divergence than both unbiased and female-biased genes. Similarly, female-biased genes show lower rates of sequence evolution in comparison with unbiased genes; however, the difference is not significant. We could not detect any significant differences in the proportion of genes evolving under positive selection between either male-biased or female-biased genes and unbiased genes. The low rates of male-biased sequence evolution could be partly explained by the higher rate of silent mutations in male-biased genes resulting from lower codon usage bias. However, haploid selection would also be expected in this tissue to exert a stronger purifying force to remove deleterious recessive mutations.

2 | MATERIALS AND METHODS

2.1 | Sample collection and sequencing

We obtained RNA-seq data from leaves and catkins from three female (78021, 78195, 78183) and three male (81084, T76, Hallstad 1-84) *S. viminalis* accessions (Pucholt et al., 2017; reads are deposited in the European Nucleotide Archive under Accession no. PRJEB15050). These accessions represent unrelated germplasm samples collected in Europe and Western Russia that were subsequently planted in a field archive near Uppsala, Sweden, where they were part of the *S. viminalis* association mapping population (Berlin et al., 2014; Hallingbäck et al., 2016). As previously described (Pucholt et al., 2017), stem cuttings were collected in the field and transferred to a growth chamber with 22°C constant temperature and 18 hr day length. After seven and thirteen days, respectively, fully developed adult catkins and young leaves were collected from each accession. RNA from each accession and tissue was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich Co. LLC) following variant B of the instructions provided by the manufacturer and including an on-column DNase treatment step. One RNA-seq library for each sample was prepared from 1 μg total RNA using the TruSeq stranded mRNA sample preparation kit (Cat# RS-122-2101/2102, Illumina Inc.) including polyA selection. The library preparation was carried out according to the manufacturer's protocol (#15031047, rev E). Sequencing was performed on an Illumina HiSeq2500 instrument with paired-end 125 bp read length, v4 sequencing chemistry, and all twelve libraries were pooled and sequenced on three lanes. Preparation of the RNA-seq libraries and sequencing were performed at the SNP&SEQ Technology Platform in Uppsala, Sweden.

We recovered an average of 42 million 125-bp paired-end reads per sample. After assessing data quality with FASTQC v0.11.3 (http://

www.bioinformatics.babraham.ac.uk/projects/fastqc/), we used TRIMMOMATIC v0.36 (Lohse et al., 2012) to remove adaptor sequences and trim the reads, removing regions where the average Phred score in sliding windows of four bases was <15 as well as reads for which the leading/trailing bases had a Phred score <3. Following trimming, we removed paired-end reads where either read pair was <50 bp (Table S1), resulting in an average of 30 million paired-end reads per sample.

2.2 | Expression analysis

We mapped RNA-seq reads to the de novo male genome assembly (Pucholt et al., 2017) using HISAT2 v2.0.4 (Kim, Langmead, & Salzberg, 2015), filtering reads with unpaired (-no-mixed option) or discordant (-no-discordant option) alignments. To generate a reference transcriptome, we sorted and converted alignment output sam files into bam files using SAMTOOLS v1.2 (Li et al., 2009) and extracted gene coordinates for each sample using STRINGTIE v1.2.4 (Pertea et al., 2015) with default parameters. We then merged output GTF files of all samples to obtain a nonredundant set of transcript coordinates and used BEDTOOLS getfasta to extract sequences (Quinlan & Hall, 2010). We filtered ncRNA by BLASTing transcript sequences to the *Arabidopsis thaliana* ncRNA (Ensembl Plants 32; Flicek et al., 2014) using BLASTN and an e-value cut-off of 1×10^{-10}.

We extracted read alignments for transcripts in each sample and tissue separately from the filtered transcriptome reference using STRINGTIE and obtained read counts using HTSEQ v.0.6.1 (Anders, Pyl, & Huber, 2015). RPKM values were estimated using EDGER (Robinson, McCarthy, & Smyth, 2010) in R (R core team 2015) and transcripts filtered for a minimum expression threshold of 2 RPKM in at least half of the individuals in one sex (in this case, at least two of the three individuals per each sex) as per previous similar studies (Harrison et al., 2015; Pointer et al., 2013). We only retained transcripts with positional information on annotated chromosomes (Pucholt et al., 2017) for further analysis and normalized separately for each tissue using TMM in EDGER.

We performed hierarchical clustering of average gene expression for genes expressed in both tissues with bootstrap resampling (1,000 replicates) in the R package PVCLUST v.2.0 (R Core Team, 2015; Suzuki & Shimodaira, 2006). We generated a heatmap of $\log_2$ average male and female expression in the two tissues using the R package PHEATMAP v.1.0.7 (Kolde, 2012; R Core Team, 2015).

We identified sex-biased expression based on a minimum of twofold differential expression ($\log_2$ M:F RPKM > 1 for male-biased expression and < −1 for female-biased expression) and a significant p value ($p_{adj} < .05$ following FDR correction for multiple testing (Benjamini & Hochberg, 1995)) in EDGER.

2.3 | Sequence divergence analysis

Additional to *S. viminalis*, we obtained coding sequences for *P. trichocarpa* from Ensembl Plants 32 (Flicek et al., 2014), *Populus tremula* and *Populus tremuloides* from PopGenIE (Sundell et al., 2015) and *Salix suchowensis* (http://115.29.234.170/willow/ (Dai et al., 2014)). The longest transcript for each gene was identified in all species, and a reciprocal BLASTN with an e-value cut-off of 1×10^{-10} and a minimum percentage identity of 30% was used to identify orthologs. We used BLASTX to obtain open reading frames of the identified orthologous groups, which we aligned with PRANK v140603 (Löytynoja & Goldman, 2008), using the rooted tree ((*Salixviminalis*, *Salixsuchowensis*), ((*Populustremula*, *Populustremuloides*), *Populustrichocarpa*)). Gaps were removed from the alignments.

To ensure the accurate calculation of divergence estimates, poorly aligned regions were masked with SWAMP v 31-03-14 (Harrison, Jordan, & Montgomery, 2014). We employed a two-step masking approach, first using a shorter window size to exclude sequencing errors causing short stretches of nonsynonymous substitutions and then a large window size to remove alignment errors caused by variation in exon splicing (Harrison et al., 2014). Specifically, we first masked regions with more than seven nonsynonymous substitutions in a sliding window scan of 15 codons, followed by a second masking where more than two nonsynonymous substitutions were present in a sliding window scan of four codons. To choose these thresholds, we imposed a range of masking criteria on our data set and conducted the branch-site test on these test data sets. We manually observed the alignment of genes with the highest log likelihood scores to choose the most efficient and appropriate masking criteria. We subsequently removed genes where the alignment (after removal of gaps and masked regions) was < 300 bp, which likely represent incomplete sequences. This resulted in 7,631 1:1 orthologs.

We tested the robustness of the 1:1 orthologs data set (Supporting Information) by separately inferring orthologous groups using ORTHOMCL (Li, Stoeckert, & Roos, 2003), an approach with higher specificity (Altenhoff & Dessimoz, 2009). As ORTHOMCL relies on the Markov Clustering algorithm, it is useful in identifying cases of coorthology (a duplicate of a gene in one species that is orthologous to a gene in another species) within the total 1:1 orthologous groups identified. By excluding these co-orthologous groups, we recovered fewer 1:1 orthologs (1,346 after filtering for polymorphism and divergence data); however, the divergence results were consistent with our broader data set based on reciprocal BLAST (Table S2). As such, we concluded that the reciprocal best-hit approach was appropriate to use in this case.

We further used branch model 2 (model = 2, nssites = 0, fixomega = 0, omega = 0.4) from the CODEML package in PAML v4.8 (Yang, 2007) to obtain divergence estimates and calculate mean d_N/d_S specifically for the *S. viminalis* branch using the unrooted tree ((*Salixviminalis*, *Salixsuchowensis*), *Populustrichocarpa*, *Populustremula*, *Populustremuloides*). Mutation-saturated sites did not have an effect on the resulting divergence estimates as none of the orthologs had $d_S > 2$ (Axelsson et al., 2008). In addition, we also obtained omega values for each sex-bias gene category by running the CODEML branch model 2 in PAML separately on the concatenated sequences of all genes in each gene category. This approach reduces the influence of codon bias in estimating rates of divergence (Bierne & Eyre-Walker, 2003).

Based on their genomic location in the *S. viminalis* genome (Pucholt et al., 2017), we divided orthologs into two groups, orthologs on the autosomes (including the pseudoautosomal region of the Z chromosome) and orthologs on the Z-linked nonrecombining region. Because genes on sex chromosomes can exhibit accelerated rates of evolution (Charlesworth, Coyne, & Barton, 1987), and this may be more often due to nonadaptive processes on Z chromosomes (Mank, Vicoso, Berlin, & Charlesworth, 2010b; Wright et al., 2015), we analysed rates of evolution separately for autosomal and Z-linked loci. Mean d_N (the number of nonsynonymous substitutions over nonsynonymous sites) and mean d_S (the number of synonymous substitutions over synonymous sites) were calculated separately for each group of orthologs as the ratio of the sum of the number of substitutions across all orthologs in that group, resulted from PAML, to the number of sites (d_N = sum D_N/sum N; d_S = sum D_S/sum S). By calculating mean d_N and d_S through this method, the issue of infinitely high d_N/d_S estimates arising from low d_S sequences and skew from short sequences is avoided (Mank, Hultin-Rosenberg, Axelsson, & Ellegren, 2007). Bootstrapping with 1,000 replicates was used to determine the 95% confidence intervals. Pairwise comparisons with 1,000 permutation test replicates were used to identify significant differences in d_N, d_S and d_N/d_S between the categories.

2.4 | Polymorphism analysis

We obtained polymorphism data by mapping the RNA-seq reads to the reference genome assembly using STAR aligner v2.5.2b (Dobin et al., 2013) in the two-pass mode and with default parameters, retaining uniquely mapping reads only. We conducted SNP calling using SAMTOOLS mpileup and VARSCAN v2.3.9 mpileup2snp (Koboldt et al., 2012). We ran SAMTOOLS mpileup with a maximum read depth of 10,000,000 and minimum base quality of 20 for consistency with VARSCAN minimum coverage filtering. The base alignment quality (BAQ) adjustment was disabled in SAMTOOLS as it imposes a too stringent adjustment of base quality scores (Koboldt, Larson, & Wilson, 2014). We ran VARSCAN mpileup2snp with minimum coverage of 20, minimum of three supporting reads, minimum average quality of 20, minimum variant allele frequency of 0.15, minimum frequency for homozygote of 0.85, strand filter on and p value of .05. We defined valid SNPs as sites with a coverage ≥ 20 in at least half of the individuals in one sex (minimum of two of the three individuals in a sex) and a minor allele frequency ≥ 0.20, identifying a total of 235,106 SNPs. We identified whether SNPs were synonymous or nonsynonymous by matching them to the reading frame.

As the divergence and polymorphism analyses use different filtering criteria, we ensured the two data sets were comparable by identifying a set of codons where all sites pass the filtering criteria for both analyses. We only kept codons where (i) all sites pass the minimum coverage threshold of 20 in at least half of the individuals in one sex, (ii) there are no alignment gaps following PRANK alignment, and (iii) there were no ambiguity data (N_s) following SWAMP masking. Only genes with both divergence and polymorphism information were used in further analyses. This ensures that the number of

synonymous (S) and nonsynonymous sites (N) is identical across divergence and polymorphism analyses, and therefore suitable for McDonald–Kreitman tests. We have therefore used the same number of nonsynonymous (N) and synonymous (S) sites in our calculations of d_N, p_N and, respectively, d_S and p_S.

We calculated mean p_N (number of nonsynonymous polymorphisms over nonsynonymous sites) and mean p_S (number of synonymous polymorphisms over synonymous sites) for each gene category as the ratio of the sum of the number of polymorphisms to the sum of the number of sites (p_N = sum P_N/sum N; p_S = sum P_S/sum S).

2.5 | Analysis of synonymous codon usage bias

Codon usage bias was estimated using CODONW (http://codonw.sourceforge.net) through the effective number of codons (ENC) (Wright, 1990). The ENC measure determines the degree to which the entire genetic code is used in each gene, ENC values ranging from 20 (indicating extreme bias, where only one codon is used for one amino acid) to 61 (indicating no bias, where all amino acids are represented equally by all possible codons) (Wright, 1990). This measure is not biased by the different lengths of the coding regions being analysed, and as such, it has been shown to be more reliable than other commonly used methods of estimating codon usage bias (Comeron & Aguadé, 1998). The effective number of codons was calculated for all the genes with divergence and polymorphism data (Table 2).

2.6 | Tests of positive selection

To identify genes evolving under adaptive evolution, we used the McDonald–Kreitman test (McDonald & Kreitman, 1991), which contrasts the ratio of nonsynonymous and synonymous substitutions with polymorphisms. For each gene, we used a 2 × 2 contingency table and a Fisher's exact test in R to test for deviations from neutrality using numbers of nonsynonymous and synonymous substitutions (D_N, D_S) and polymorphisms (P_N, P_S). As the McDonald–Kreitman test lacks power with low table cell counts, genes were excluded from the analysis if, within the contingency table, the sum over any row or column was less than six (Andolfatto, 2008; Begun et al., 2007). For genes with significant deviations in D_N, D_S, P_N and P_S, a higher nonsynonymous-to-synonymous substitutions ratio relative to polymorphisms ratio ($d_N/d_S > p_N/p_S$) represented a signature of positive selection. We then tested for significant differences between sex-biased and unbiased genes in the proportion of genes with signatures of positive selection using Fisher's exact test.

For each gene category, we also used the divergence and polymorphism data to calculate the average direction of selection (DoS) statistic (Stoletzki & Eyre-Walker, 2011). DoS was calculated for each gene as the difference between the proportion of nonsynonymous substitutions and the proportion of nonsynonymous polymorphisms ($DoS = D_N/(D_N + D_S) - P_N/(P_N + P_S)$), where positive DoS values indicate positive selection, a value of zero indicates neutral evolution while negative values indicate purifying selection and segregating deleterious mutations (Stoletzki & Eyre-Walker, 2011). Additional to

the McDonald–Kreitman test, we also used the DoS statistic to test, using Fisher's exact test, for differences in the proportion of fixed nonsynonymous sites and nonsynonymous polymorphisms.

3 | RESULTS

3.1 | Gene expression in catkin and leaf

RNA-seq reads from two tissues, catkin (reproductive tissue) and leaf (vegetative tissue), of male and female *S. viminalis* individuals were mapped to the genome assembly yielding an average of 30 million read mappings per sample after quality control and trimming (Table S1). Following expression filtering, we recovered 8,186 significantly expressed genes in catkin and 7,638 significantly expressed genes in leaf.

We first assessed transcriptional similarity across tissues and sexes using hierarchical clustering of gene expression levels (Figure 2). We found that the reproductive tissue was more transcriptionally dimorphic than the vegetative tissue, consistent with studies in many other species (Jiang & Machado, 2009; Mank, Hultin-Rosenberg, Webster, & Ellegren, 2008; Pointer et al., 2013; Yang, Zhang, & He, 2016). Expression for male catkin clustered most distantly from both male and female expressions in leaf. We identified 3,567 genes (43% of all filtered catkin genes) showing sex-biased expression in catkin ($\log_2$ fold change > 1 or < −1, p_{adj} < .05), compared to expression in the vegetative tissue, where we identified only seven (0.09%) leaf sex-biased genes (Figure 3). Even with a more relaxed fold change threshold for defining differentially expressed genes ($\log_2$ fold change > 0.5 or < −0.5, p_{adj} < .05), we still could not

FIGURE 2 Heatmap and hierarchical clustering of average male (blue) and average female (red) gene expression in catkin and leaf. The heatmap represents all the filtered genes expressed in both tissues (7,257). Hierarchical gene clustering is based on Euclidean distance with average linkage for $\log_2$ RPKM expression for each gene. Numbers at nodes represent the 1,000 replicates percentage bootstrap results

identify any additional leaf sex-biased genes. There were also no shared sex-biased genes between the two tissues.

3.2 | Dynamics of catkin sex-biased gene expression

Although female-biased genes (n = 1,820) were slightly more numerous than male-biased genes (n = 1,747), the magnitude of differential expression ($\log_2$ FC) for male-biased genes was significantly greater than that for female-biased genes (Wilcoxon rank sum test p < .001). Average male expression for male-biased genes was significantly higher than average female expression for female-biased genes (Figure 3, Wilcoxon rank sum test p < .001), although male expression for female-biased genes was significantly lower than female expression for male-biased genes (Figure 3, Wilcoxon rank sum test p < .001).

We grouped sex-biased genes based on different fold change thresholds and compared average male and female catkin expression for the genes in each category. This analysis suggests that catkin male-biased genes may arise from increased expression in males and decreased expression in females (Figure 4). For female-biased genes, however, there is a decreasing trend in male expression with increasing fold change thresholds but a constant female expression across all thresholds (Figure 4), suggesting that female bias results primarily from downregulation of male expression.

The paucity of sex-biased genes in the leaf tissue makes it a useful comparison to further assess the sex-specific changes that give rise to male- and female-biased genes. We therefore used leaf expression as the putative ancestral expression state. For the subset of catkin sex-biased genes that also had expression in the leaf tissue, we determined the difference in expression between catkin and leaf across the same fold change thresholds used in Figure 4. For male-biased genes in the catkin, we found significant differences between catkin and leaf expression in both sexes, although to a lesser extent in females (Figure S1). On the other hand, for catkin female-biased genes, we also observed large differences in male expression between catkin and leaf samples; however, we found little to no female expression changes between the two tissues (Figure S1).

We further divided catkin sex-biased genes into autosomal (including the pseudoautosomal region of the sex chromosomes) and Z-linked genes. On the autosomes, we found 3,536 sex-biased genes (1,728 male-biased and 1,808 female-biased genes). On the nonrecombining region of the Z chromosome, we found only 31 sex-biased genes (19 male-biased and 12 female-biased genes); however, considering the narrow region of recombination suppression between the sex chromosomes (Pucholt et al., 2017, 3.5–8.8 Mbp), these sex-biased genes represented 44% of the total identified gene content in the nonrecombining sex-chromosome region.

3.3 | Rates of evolution

We compared the overall ratios of nonsynonymous-to-synonymous nucleotide substitutions (d_N/d_S) between catkin and leaf and found
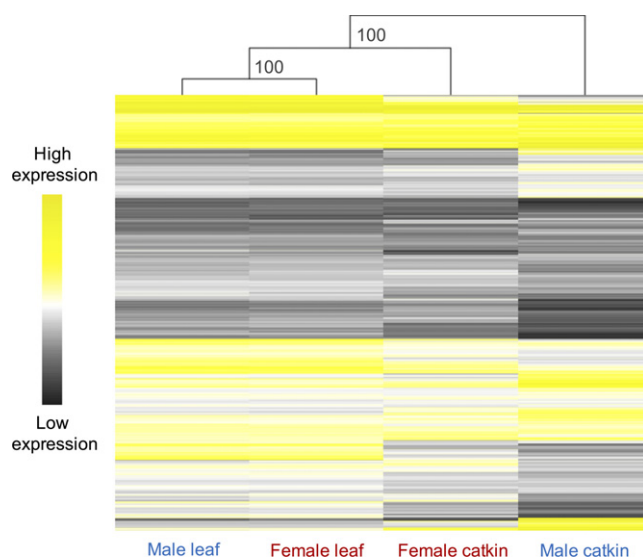
FIGURE 3   Sex-biased gene expression in *Salix viminalis*. (a) Proportion and range of differentially expressed and unbiased genes in catkin and leaf. (b) Comparison between male and female average expression for sex-biased and unbiased genes in catkin. Numbers in brackets represent the number of genes in each category. Significant differences between male and female expression based on Wilcoxon rank sum tests are denoted (ns = nonsignificant, ***$p < .001$)
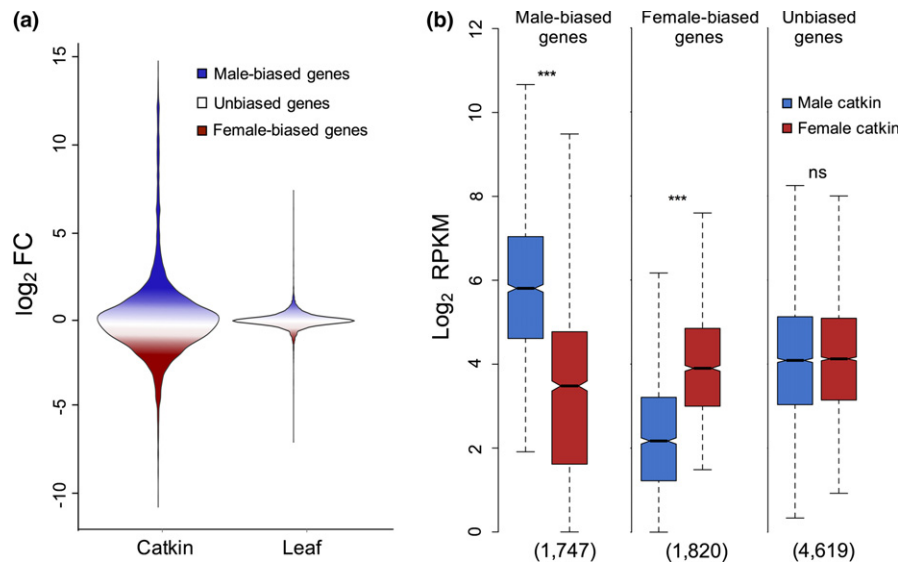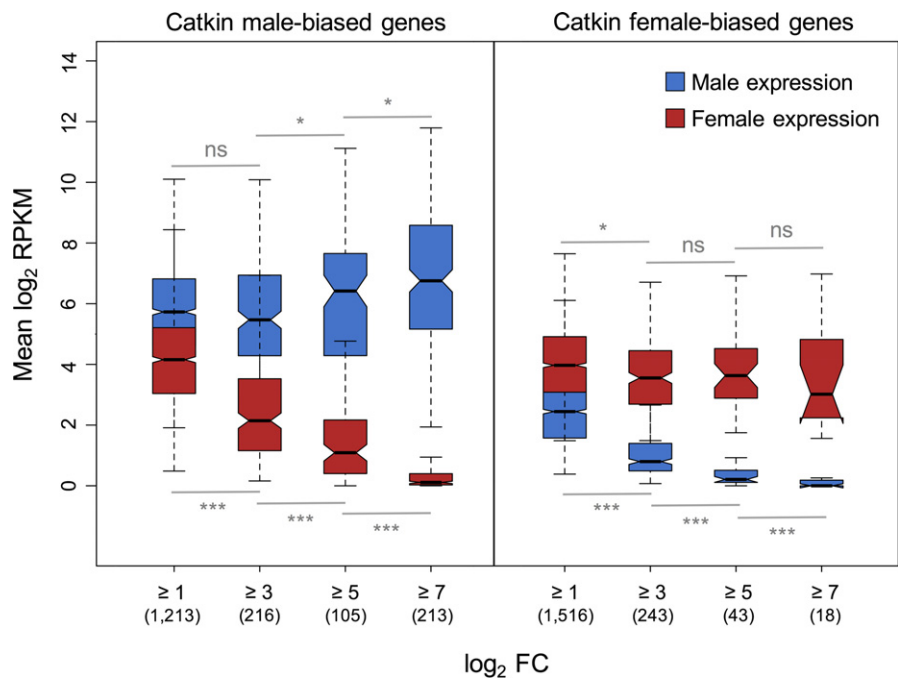

FIGURE 4   Average male and female catkin gene expression at different sex-bias fold change thresholds for all assessed catkin male-biased and female-biased genes. Numbers in brackets represent the number of genes in each fold change category. Significance level is based on Wilcoxon rank sum tests (ns = nonsignificant, *$p < .05$, ***$p < .001$)

no significant differences between the two ($p = .476$, significance based on permutation tests with 1,000 replicates). We also did not find a significant difference in the evolution of unbiased genes between the two tissues ($p = .056$ from permutation tests with 1,000 replicates), likely influenced by the large overlap of genes between them (97% of catkin unbiased genes represent 58% of the unbiased genes expressed in leaf). We found too few significantly sex-biased genes in the leaf tissue to make any statistical comparisons of rates of sequence evolution between catkin and leaf sex-biased genes.

We also compared the ratio of d_N/d_S between sex-biased and unbiased genes in catkin to test for differences in the rate of evolutionary divergence. Interestingly, we found that on autosomes, although male-biased genes have more amino acid substitutions than both unbiased and female-biased genes, as shown by significantly higher d_N values, d_N/d_S for male-biased genes was significantly lower, indicating slower rates of functional evolution relative to unbiased (Table 1; Table S2) and female-biased genes ($p < .001$, significance based on permutation tests with 1,000 replicates). Similar results were obtained when we estimated d_N/d_S from a data set of 1:1 orthologs that excluded cases of co-orthology (Table S2), as well as from omega values resulting from running CODEML branch model 2 in PAML on concatenated sequences of genes in each sex-bias gene category (Table S3). This lower d_N/d_S ratio is caused in part by a disproportionate increase in synonymous substitutions compared to nonsynonymous substitutions, causing the relationship between d_N and d_S in male-biased genes to lie further away from direct proportionality than in the case of unbiased genes (Figure S2).

TABLE 1 Divergence and polymorphism estimates for catkin gene categories on autosomes and the nonrecombining Z region

Tissue	Location	Category[a]	n Genes[b]	d_N (95% CI) sig.[c]	d_S (95% CI) sig.[c]	d_N/d_S (95% CI) sig.[c]	p_N (95% CI) sig.[c]	p_S (95% CI) sig.[c]	p_N/p_S (95% CI) sig.[c]	DoS sig.[d]
Catkin	Autosomes and recombining Z	UB	1,754	0.0030 (0.0028–0.0031)	0.0135 (0.0130–0.0141)	0.2204 (0.2101–0.2311)	0.0027 (0.0025–0.0028)	0.0109 (0.0103–0.0114)	0.2456 (0.2328–0.2584)	−0.0495
		MB	674	**0.0032 (0.0030–0.0035)** p = .012	**0.0162 (0.0147–0.0187)** p < .001	**0.1951 (0.1769–0.2157)** p < .001	**0.0029 (0.0026–0.0032)** p = .022	**0.0116 (0.0108–0.0124)** p = .042	0.2491 (0.2260–0.2727) p = .682	−0.0346 p = .627
		FB	732	0.0031 (0.0029–0.0033) p = .094	**0.0149 (0.0141–0.0158)** p < .001	0.2095 (0.1938–0.2256) p = .082	**0.0030 (0.0028–0.0033)** p = .002	**0.0121 (0.0113–0.0131)** p < .001	0.2477 (0.2293–0.2666) p = .774	−0.0375 p = .916
	Nonrecombining Z	UB	12	0.0032 (0.0024–0.0043)	0.0102 (0.0056–0.0145)	0.3130 (0.2141–0.5498)	0.0015 (0.0007–0.0032)	0.0045 (0.0022–0.0078)	0.3407 (0.1778–0.5588)	0.0800
		MB	3	0.0029 (0.0–0.0140) p = .378	0.0143 (0.0091–0.0396) p = .730	0.2019 (0.0–0.3533) p = .244	0.0029 (0.0–0.0210) p = .396	0.0104 (0.0039–0.0505) p = .084	0.2781 (0.0–0.4151) p = .082	0.0088 p = .563
		FB	4	0.0032 (0.0021–0.0037) p = .964	0.0100 (0.0061–0.0207) p = .926	0.3172 (0.1649–0.4996) p = .948	**0.0045 (0.0008–0.0082)** p = .026	**0.0138 (0.0055–0.0229)** p < .001	0.3243 (0.0845–0.4657) p = .882	0.0770 p = .761

[a]Unbiased (UB), male-biased (MB) and female-biased (FB) genes.
[b]Number of genes with both divergence and polymorphism data.
[c]p values based on 1,000 replicates permutation tests comparing male-biased and female-biased genes with unbiased genes. Significant p values (< .05) are shown in bold.
[d]p values from Wilcoxon nonparametric tests comparing male-biased and female-biased genes with unbiased genes. Significant p values (< .05) are shown in bold.

Female-biased autosomal loci also showed the same pattern as male-biased genes relative to unbiased genes; however, this result was not significant (Table 1; Table S2). On the nonrecombining Z, male-biased genes also show lower rates of evolution compared to unbiased genes; however, this finding was not significant, likely due to the small sample size of male-biased genes ($n = 3$). In contrast, female-biased Z-linked loci showed accelerated rates of evolution in comparison with male-biased Z-linked genes ($p < .001$, significance based on permutation tests with 1,000 replicates).

Highly expressed genes are often observed to exhibit lower d_N/d_S values (Cherry, 2010; Drummond, Bloom, Adami, Wilke, & Arnold, 2005; Pál, Papp, & Hurst, 2001; Slotte et al., 2011); therefore, to determine whether expression level might explain our results, we divided sex-biased and unbiased genes into quartiles based on overall expression. As expected, we found that as gene expression level increases, the rate of sequence divergence decreases and this holds true for both sex-biased and unbiased genes (Figure S3). To further investigate the effect of expression level on the variation in rates of sequence divergence between sex-bias categories, we used a multiple regression analysis to predict d_N/d_S results based on expression level and degree of sex-bias. For defining the degree of sex-bias, genes were classed into five groups, highly female-biased genes (FC ≤ -3), lowly female-biased genes ($-3 <$ FC ≤ -1), unbiased genes ($-1 <$ FC < 1), lowly male-biased genes ($1 \leq$ FC < 3) and highly male-biased genes (FC ≥ 3). We found a significant negative relationship between d_N/d_S values and both average $\log_2$ RPKM expression level ($\beta = -.03$, $p < .001$) and degree of sex-bias ($\beta = -.04$, $p = .014$). There was no significant effect of the interaction between expression level and degree of sex-bias on d_N/d_S results, suggesting that any differences in the rates of sequence evolution due to sex-bias are independent of the gene expression level for each sex-bias category. Despite these results, the adjusted r^2 was very low ($r^2 = .01$), indicating that other factors, such as purifying or haploid selection, largely explain the vast majority of sequence divergence results.

We also estimated average levels of synonymous codon usage bias for sex-biased and unbiased genes to determine whether this could explain the differences in the rates of synonymous substitutions between the gene categories. Stronger codon usage bias has been associated with higher gene expression as selective forces act to increase translational efficiency (Duret, 2002; Ingvarsson, 2010). Codon bias has also been shown to differ between differentially expressed genes, with male-biased genes undergoing weaker codon usage bias than female-biased (Mank et al., 2008; Magnusson et al., 2011; however, this varied across different developmental stages; Whittle, Malik, & Krochko, 2007) and unbiased genes (Hambuch & Parsch, 2005). Additionally, greater codon bias has been estimated for genes with lower rates of synonymous substitutions (Urrutia & Hurst, 2001).

We estimated codon usage bias for genes in each category through the effective number of codons (ENC), where stronger codon bias was indicated by lower ENC values. The differences in codon bias between the different gene categories were subtle, and the gene frequency spectra for all categories were distributed towards the higher end of the effective number of codons (ENC), hence lower codon usage bias (Figure S4). However, male-biased genes had significantly lower codon usage bias than both unbiased (Table 2) and female-biased genes ($p < .001$, significance based on permutation tests with 1,000 replicates). These findings, together with the higher rates of synonymous substitutions in male-biased genes compared to unbiased and female-biased genes, indicate weaker purifying selection on silent mutations in male-biased genes (Sharp & Li, 1987).

We used polymorphism data to calculate the ratio of nonsynonymous-to-synonymous polymorphisms (p_N/p_S). Sex-biased genes on both autosomes and the nonrecombining Z region have significantly higher nonsynonymous and synonymous polymorphism levels compared to unbiased genes; however, the p_N/p_S ratio was not significantly different in either of the comparisons (Table 1). To distinguish between the selective pressures acting on sequence evolution, we used the McDonald–Kreitman test of selection, comparing the ratios of d_N/d_S to p_N/p_S for each gene category. Following filtering, we recovered six unbiased, one male-biased and two female-biased genes showing signatures of positive selection (Table 3). However, there was no significant difference in the proportion of genes evolving under positive selection between either of the gene categories (Table 3, significance denoted in table). Because the McDonald–Kreitman test is extremely conservative, we also assessed selection pressures on sex-biased genes using the direction of selection test

TABLE 2 Codon usage bias for catkin sex-bias gene categories

Tissue	Location	Category	n Genes[a]	ENC[b] sig.[c]
Catkin	Autosomes and recombining Z	Unbiased	1,754	52.15
		Male biased	674	52.71 **p < .001**
		Female biased	732	52.20 p = .588

[a]Number of genes with both divergence and polymorphism data.
[b]Average effective number of codons for each gene category.
[c]p values based on 1,000 replicates permutation test comparing male-biased and female-biased genes relative to unbiased genes. Significant p values (< .05) are shown in bold.

TABLE 3 McDonald–Kreitman test of selection

Tissue	Location	Category	n Genes[a]	Positive selection[b] sig.[c]
Catkin	Autosomes and recombining Z	Unbiased	1,766	6
		Male biased	677	1 ns
		Female biased	736	2 ns

[a]Number of genes with both divergence and polymorphism data.
[b]Number of genes with significant positive selection indicated by significant deviations in D_N, D_S, P_N and P_S and $d_N/d_S > p_N/p_S$.
[c]Significance based on Fisher's exact test comparing sex-biased to unbiased genes (ns = nonsignificant).

(Stoletzki & Eyre-Walker, 2011). Through the DoS statistic, we recovered 681 unbiased, 262 male-biased and 282 female-biased genes under putative positive selection (DoS > 0), yet, consistent with the McDonald–Kreitman test, we found no significant differences in the proportion of genes evolving under positive selection (Fisher's exact test $p > .9$ for both female-biased and male-biased genes in comparison with unbiased genes). Taken together, the divergence and polymorphism analyses, through tests of positive selection, suggest that the lower rates of sequence evolution seen in male-biased genes could be due to purifying selection acting to remove deleterious recessive mutations.

4 | DISCUSSION

The evolution of sex-biased gene sequence has been extensively analysed in animal systems. In contrast, far less is known about the evolution of sex-biased genes in plants in general and in dioecious angiosperms in particular. Previous work in *A. thaliana*, an annual and largely selfing hermaphroditic species, found low rates of evolution in pollen-expressed genes, although with evidence of a higher proportion of sites under positive selection (Gossmann et al., 2014). This could be the result of the greater haploid selection in plants; however, it could also be, at least partially, the result of the selfing mating system in this species, which leads to the purging of recessive deleterious variation. Similarly, in the self-incompatible close relative of *A. thaliana*, *C. grandiflora*, a larger fraction of pollen-specific genes was found to evolve under strong purifying selection and to also exhibit faster protein evolution rates compared to sporophytic genes (Arunkumar et al., 2013). This is suggested to be the result of both higher pollen competition and the haploid nature of the pollen-specific tissue.

Here, we investigate the evolution of sex-biased genes in *S. viminalis*, a perennial dioecious (obligate outcrossing) species with partial wind pollination. Similarly to *C. grandiflora* (Kao & McCubbin, 1996), *S. viminalis* theoretically experiences far higher levels of pollen competition than *A. thaliana*, particularly intermale competition. Although we might expect the high levels of sperm competition in *S. viminalis* to produce higher rates of protein evolution for male-biased genes, we observed the opposite. Moreover, in contrast to work in *C. grandiflora* (Arunkumar et al., 2013), we did not find evidence of a high proportion of male-biased genes under positive selection.

The observed dynamics of sex-biased gene expression in *S. viminalis* is consistent with previous reports in a wide range of species. Equivalent to studies on somatic and reproductive tissues in animal systems (Mank, 2017; Pointer et al., 2013; Yang et al., 2016), we found that the reproductive tissue was far more transcriptionally dimorphic than the vegetative tissue (Figures 2 and 3). Additionally, in plant species in particular, very few studies have been able to identify any significant sex-biased genes in nonreproductive tissues (Robinson et al., 2014; Zemp, Minder, & Widmer, 2014; Zluvova et al., 2010). We also found that, in catkin, male-biased genes were expressed at significantly higher levels and had a higher magnitude

of sex-bias than female-biased genes (Figure 3). The level of sex-biased gene expression found in the *S. viminalis* reproductive tissue is markedly lower than that in animal species (Jiang & Machado, 2009; Pointer et al., 2013), consistent with the significantly higher degree of sexual dimorphism in animal systems. On the other hand, we found a larger percentage of sex-biased genes compared to several plant and algae species with low levels of sexual dimorphism (Harkess et al., 2015; Lipinska et al., 2015; Zemp et al., 2016). This is indicative of higher intersexual morphological differences in the *S. viminalis* reproductive tissue, which is consistent with previous descriptions of the structural differences between male and female catkins (Cronk et al., 2015).

Contrary to findings from the dioecious *Silene latifolia* (Zemp et al., 2016), however similarly to reports from animal and algae systems (Lipinska et al., 2015; Perry et al., 2014), our results indicate that sex-biased gene expression has likely evolved as an outcome of expression changes in males (Figure S1). This would also explain why catkin male samples are more transcriptionally different than catkin female samples with respect to leaf samples (Figure 2). These results suggest that ancestral intralocus sexual conflict may have been more detrimental to males, leading to the evolution of sex-biased gene expression in order to resolve such conflicts.

Additionally, although not statistically significant, we found that male-biased genes had higher p_N/p_S values compared to both unbiased and female-biased genes, which is in stark contrast to divergence results where we found male-biased genes to have significantly lower d_N/d_S values. Given that perturbations in population size can alter estimates of polymorphism (Pool & Nielsen, 2007; Tajima, 1989), it is difficult to assess the causes of the contrasting results between d_N/d_S and p_N/p_S estimates for sex-biased genes. Nevertheless, divergence estimates are less sensitive to demographic fluctuations and we more strongly rely on this measurement in our analyses of evolutionary rates of sex-biased genes.

Sex-biased genes in willow exhibit higher expression levels than unbiased genes, and highly expressed male-biased and female-biased genes had significantly lower rates of evolution than unbiased and lowly expressed sex-biased genes (Figure S3). The fact that highly expressed genes evolve more slowly could be due to a range of different reasons, which are still highly debated (Drummond et al., 2005). The structural or functional features of the proteins they encode (Drummond et al., 2005), high pleiotropic constraints acting on the genes (Pál et al., 2001) as well as gene conversion events (Petes & Hill, 1988) have all been suggested as potential mechanisms through which highly expressed genes could have lower rates of sequence evolution. Although the high expression of many sex-biased genes in *S. viminalis* may partially explain their slower rates of evolution, our analysis revealed a very weak correlation between expression level and rate of evolution, indicating that, in this case, expression level does not largely explain the low rates of sex-biased gene evolution.

It is interesting that the lower d_N/d_S values of male-biased genes are associated with an overall increase in synonymous mutations

relative to nonsynonymous mutations (Figure S2). This, plus our observation that male-biased genes experience lower levels of codon usage bias (Table 2), could suggest that our d_N/d_S results have been influenced by different levels of codon usage across gene expression categories. Different selection forces are thought to lead to codon usage bias, such as positive selection for preferred synonymous mutations (mutations that lead to preferred codons) and purifying selection acting on unfavourable mutations, preventing a decrease in the frequency of preferred codons (Hershberg & Petrov, 2008). Despite previous expectations that selection acting at synonymous sites is weak (Akashi, 1995; Hershberg & Petrov, 2008), several studies suggest that a range of selection strengths, spanning from weak to strong selection, influence the evolution of synonymous mutations, and hence codon usage bias measures (Hershberg & Petrov, 2008; Lawrie, Messer, Hershberg, & Petrov, 2013). However, although differential codon bias across expression categories has the potential to influence our d_N/d_S estimates, our additional PAML analysis (Table S3) indicates that this is not likely to be the case.

Similar to the findings from *A. thaliana* and *C. grandiflora*, the unusual rates of evolution of sex-biased genes in *S. viminalis* could also be explained by the differential selection pressures acting on diploid versus haploid life stages. Haploid selection (Joseph & Kirkpatrick, 2004) is more effective at removing recessive deleterious mutations than selection in the diploid life stages, where dominant alleles can mask the effects of deleterious recessive alleles (Kondrashov & Crow, 1991). Although all predominantly diploid organisms pass through both haploid and diploid phases, animal species employ different mechanisms through which selection on the haploid stage is minimized (Otto, Scott, & Immler, 2015). Not only can aneuploid spermatids still be potentially viable (Lindsley & Grell, 1969), indicating limited haploid expression, but studies in mice have shown that genetically haploid spermatids evade haploid selection by sharing gene products through cytoplasmic bridges (Erickson, 1973), becoming thus phenotypically diploid (Braun, Behringer, Peschon, Brinster, & Palmiter, 1989).

Haploid selection is far more extensive in plants due to both the larger proportion of the life cycle spent in the haploid phase and active gene transcription, which has been observed in gametes, particularly in pollen (Otto et al., 2015). In addition to haploid selection, male gametophytes in angiosperm species are under strong sexual selection pressures (Erbar, 2003; Snow & Spira, 1996), particularly in outcrossing species. Mechanisms of sexual selection in angiosperms include pollen tube and pistil interactions and pollen competition over ovules, which is exacerbated in outcrossing species (Bernasconi et al., 2004).

It is important to note that the reduced floral structure and microscopic nature of the catkin (Cronk et al., 2015) makes it nearly impossible to separate haploid from diploid reproductive tissue in this species. However, our catkin preparations are highly enriched for haploid cells (Figure 1) when compared to the vegetative samples. We expect that rates of evolution for purely haploid sex-biased tissue would be even lower than what we observe if haploid selection is indeed the primary cause of the slower rates of evolution.

Apart from insect pollen dispersal, willows also have wind-dispersed pollination (Peeters & Totland, 1999) and experience high levels of pollen competition. The observed patterns of gene sequence evolution in *S. viminalis* support the notion that pollen competition in conjunction with haploid selection produces greater levels of purifying selection on male-biased genes. This would remove deleterious variation and lead to significantly slower rates of functional gene sequence evolution. Interestingly, the algae *Ectocarpus*, a species where sex-biased genes are subject almost entirely to haploid selection, shows accelerated rates of evolution for both male- and female-biased genes (Lipinska et al., 2015). This suggests that haploid selection may not be the only force that influences the rate of evolution of sex-biased genes in haploid cells. Indeed, data from haploid-specific genes (pollen-specific genes in *S. viminalis*) would help to more precisely determine the degree to which the currently observed lower rates of evolution of male-biased genes can be explained by haploid selection or other factors such as expression breath (Arunkumar et al., 2013; Gossmann et al., 2014; Szövényi et al., 2013).

In summary, our findings are generally consistent with previous reports on the patterns of sex-bias gene expression in plant and animal species. However, different forces may differentiate patterns of evolution between animal and plant systems. The reduction in haploid selection in animals may limit the power of purifying selection to remove mildly deleterious variation, particularly when it is largely recessive. In *S. viminalis*, we observe reduced rates of evolution for male-biased genes, consistent with increased purifying selection from the extended haploid phase. Even though male-biased genes show relaxed levels of codon bias, this does not seem to be a major driver of the reduced rate of evolution. Future work should focus on investigating the differences in the relative strength of haploid versus diploid selection in dioecious angiosperm species in shaping the evolution of sex-biased genes.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

DATA ACCESSIBILITY

ORCID

Iulia Darolti [iD] http://orcid.org/0000-0002-5865-4969

Pascal Pucholt [iD] http://orcid.org/0000-0003-3342-1373

REFERENCES

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. *Genetics*, *139*, 1067–1076.

Ainsworth, C. (2000). Boys and girls come out to play: The molecular biology of dioecious plants. *Annals of Botany*, *86*, 211–221. https://doi.org/10.1006/anbo.2000.1201

Altenhoff, A. M., & Dessimoz, C. (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Computational Biology*, *5*, e1000262. https://doi.org/10.1371/journal.pcbi.1000262

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

Andolfatto, P. (2008). Controlling type-I error of the McDonald–Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics*, *180*, 1767–1771. https://doi.org/10.1534/genetics.108.091850

Arunkumar, R., Josephs, E. B., Williamson, R. J., & Wright, S. I. (2013). Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Molecular Biology and Evolution*, *30*, 2475–2486. https://doi.org/10.1093/molbev/mst149

Assis, R., Zhou, Q., & Bachtrog, D. (2012). Sex-biased transcriptome evolution in Drosophila. *Genome Biology and Evolution*, *4*, 1189–1200. https://doi.org/10.1093/gbe/evs093

Axelsson, E., Hultin-Rosenberg, L., Brandström, M., Zwahlén, M., Clayton, D. F., & Ellegren, H. (2008). Natural selection in avian protein-coding genes expressed in brain. *Molecular Ecology*, *17*, 3008–3017. https://doi.org/10.1111/j.1365-294X.2008.03795.x

Barrett, S. C. H., & Hough, J. (2013). Sexual dimorphism in flowering plants. *Journal of Experimental Botany*, *64*, 67–82. https://doi.org/10.1093/jxb/ers308

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y. P., Hahn, M. W., . . . Langley, C. H. (2007). Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, *5*, e310. https://doi.org/10.1371/journal.pbio.0050310

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, *57*, 289–300.

Berlin, S., Trybush, S. O., Fogelqvist, J., Gyllenstrand, N., Hallingbäck, H. R., Åhman, I., . . . Hanley, S. J. (2014). Genetic diversity, population structure and phenotypic variation in European *Salix viminalis* L. (Salicaceae). *Tree Genetics and Genomes*, *10*, 1595–1610. https://doi.org/10.1007/s11295-014-0782-5

Bernasconi, G., Ashman, T. L., Birkhead, T. R., Bishop, J. D. D., Gross-niklaus, U., Kubli, E., . . . Hellriegel, B. (2004). Evolutionary ecology of the prezygotic stage. *Science*, *303*, 971–975. https://doi.org/10.1126/science.1092180

Bierne, N., & Eyre-Walker, A. (2003). The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: Implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics*, *165*, 1587–1597.

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., & Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature*, *337*, 373–376. https://doi.org/10.1038/337373a0

Charlesworth, D. (2002). Plant sex determination and sex chromosomes. *Heredity*, *88*, 94–101. https://doi.org/10.1038/sj/hdy/6800016

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist*, *130*, 113–146. https://doi.org/10.1086/284701

Cherry, J. L. (2010). Highly expressed and slowly evolving proteins share compositional properties with thermophilic proteins. *Molecular Biology and Evolution*, *27*, 735–741. https://doi.org/10.1093/molbev/msp270

Comeron, J. M., & Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. *Journal of Molecular Evolution*, *47*, 268–274. https://doi.org/10.1007/PL00006384

Connallon, T., & Knowles, L. L. (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics*, *21*, 495–499. https://doi.org/10.1016/j.tig.2005.07.006

Cronk, Q. C., Needham, I., & Rudall, P. J. (2015). Evolution of catkins: Inflorescence morphology of selected Salicaceae in an evolutionary and developmental context. *Frontiers in Plant Science*, *6*, 1030. https://doi.org/10.3389/fpls.2015.01030

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., & Zhang, D. (2015). Young genes out of the male: An insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant*, *8*, 935–945. https://doi.org/10.1016/j.molp.2014.12.008

Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G. A., . . . Yin, T. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Research*, *24*, 1274–1277. https://doi.org/10.1038/cr.2014.83

Dean, R., Wright, A. E., Marsh-Rollo, S. E., Nugent, B. M., Alonzo, S. H., & Mank, J. E. (2017). Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. *Molecular Ecology*, *26*, 505–518. https://doi.org/10.1111/mec.13919

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., & Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the USA*, *102*, 14338–14343. https://doi.org/10.1073/pnas.0504070102

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development*, *12*, 640–649. https://doi.org/10.1016/S0959-437X(02)00353-2

Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*, 689–698. https://doi.org/10.1038/nrg2167

Erbar, C. (2003). Pollen tube transmitting tissue: Place of competition of male gametophytes. *International Journal of Plant Sciences*, *164*, S265–S277. https://doi.org/10.1086/377061

Erickson, R. P. (1973). Haploid gene expression versus meiotic drive: The relevance of intercellular bridges during spermatogenesis. *Nature: New biology*, *243*, 210–212.

Fisher, M. J. (1928). The morphology and anatomy of flowers of the Salicaceae I. *American Journal of Botany*, *15*, 307–326. https://doi.org/10.2307/2435733

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., & Searle, S. M. J. (2014). Ensembl 2014. *Nucleic Acids Research*, *42*, 749–755. https://doi.org/10.1093/nar/gkt1196

Gershoni, M., & Pietrokovski, S. (2014). Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. *Nature Communications*, *5*, 4438. https://doi.org/10.1038/ncomms5438

Ghelardini, L., Berlin, S., Weih, M., Lagercrantz, U., Gyllenstrand, N., & Rönnberg-Wästljung, A. C. (2014). Genetic architecture of spring and autumn phenology in *Salix*. *BMC Plant Biology*, *14*, 31. https://doi.org/10.1186/1471-2229-14-31

Gossmann, T. I., Saleh, D., Schmid, M. W., Spence, M. A., & Schmid, K. J. (2016). Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. *Molecular Biology and Evolution*, *33*, 1669–1678. https://doi.org/10.1093/molbev/msw044

Gossmann, T. I., Schmid, M. W., Grossniklaus, U., & Schmid, K. J. (2014). Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *31*, 574–583. https://doi.org/10.1093/molbev/mst226

Grath, S., & Parsch, J. (2016). Sex-biased gene expression. *Annual Review of Genetics*, *50*, 29–44. https://doi.org/10.1146/annurev-genet-120215-035429

Hallingbäck, H. R., Fogelqvist, J., Powers, S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., ... Rönnberg-Wästljung, A. C. (2016). Association mapping in *Salix viminalis* L. (Salicaceae) - identification of candidate genes associated with growth and phenology. *GCB Bioenergy*, *8*, 670–685. https://doi.org/10.1111/gcbb.12280

Hambuch, T. M., & Parsch, J. (2005). Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. *Genetics*, *170*, 1691–1700. https://doi.org/10.1534/genetics.104.038109

Harkess, A., Mercati, F., Shan, H. Y., Sunseri, F., Falavigna, A., & Leebens-Mack, J. (2015). Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytologist*, *207*, 883–892. https://doi.org/10.1111/nph.13389

Harrison, P. W., Jordan, G. E., & Montgomery, S. H. (2014). SWAMP: Sliding window alignment masker for PAML. *Evolutionary Bioinformatics Online*, *10*, 197–204. https://doi.org/10.4137/EBO.S18193

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the USA*, *112*, 4393–4398. https://doi.org/10.1073/pnas.1501339112

Heilbuth, J. C. (2000). Lower species richness in dioecious clades. *American Naturalist*, *156*, 221–241. https://doi.org/10.1086/303389

Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. *Annual Review of Genetics*, *42*, 287–299. https://doi.org/10.1146/annurev.genet.42.110807.091442

Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Yi, S. V., Keller, L., & Goodisman, M. A. D. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proceedings of the National Academy of Sciences of the USA*, *108*, 15936–15941. https://doi.org/10.1073/pnas.1104825108

Ingvarsson, P. K. (2010). Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in Populus tremula. *Molecular Biology and Evolution*, *27*, 650–660. https://doi.org/10.1093/molbev/msp255

Jiang, Z. F., & Machado, C. A. (2009). Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics*, *183*, 1175–1185. https://doi.org/10.1534/genetics.109.105775

Jiang, H., Peng, S., Zhang, S., Li, X., Korpelainen, H., & Li, C. (2012). Transcriptional profiling analysis in *Populus yunnanensis* provides insights into molecular mechanisms of sexual differences in salinity tolerance. *Journal of Experimental Botany*, *63*, 3709–3726. https://doi.org/10.1093/jxb/ers064

Joseph, S. B., & Kirkpatrick, M. (2004). Haploid selection in animals. *Trends Ecology and Evolution*, *19*, 592–597. https://doi.org/10.1016/j.tree.2004.08.004

Käfer, J., Marais, G. A. B., & Pannell, J. R. (2017). On the rarity of dioecy in flowering plants. *Molecular Ecology*, *26*, 1225–1241. https://doi.org/10.1111/mec.14020

Kao, T. H., & McCubbin, A. G. (1996). How flowering plants discriminate between self and non-self pollen to prevent inbreeding. *Proceedings of the National Academy of Sciences of the USA*, *93*, 12059–12065. https://doi.org/10.1073/pnas.93.22.12059

Karrenberg, S., Kollmann, J., & Edwards, P. J. (2002). Pollen vectors and inflorescence morphology in four species of *Salix*. *Plant Systematics and Evolution*, *235*, 181–188. https://doi.org/10.1007/s00606-002-0231-z

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., ... Pääbo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, *309*, 1850–1854. https://doi.org/10.1126/science.1108296

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*, 357–360. https://doi.org/10.1038/nmeth.3317

Koboldt, D. C., Larson, D. E., & Wilson, R. K. (2014). Using VARSCAN 2 for germline variant calling and somatic mutation detection. *Current Protocols in Bioinformatics*, *44*, 15.14.11–15.14.17. https://doi.org/10.1002/0471250953.bi1504s44

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., ... Wilson, R. K. (2012). VARSCAN 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, *22*, 568–576. https://doi.org/10.1101/gr.129684.111

Kolde, R. (2012). PHEATMAP: Pretty heatmap. *R package version 61*.

Kondrashov, A. S., & Crow, J. F. (1991). Haploidy or diploidy: Which is better? *Nature*, *351*, 314–315. https://doi.org/10.1038/351314a0

Lawrie, D. S., Messer, P. W., Hershberg, R., & Petrov, D. A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics*, *9*, e1003527. https://doi.org/10.1371/journal.pgen.1003527

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMTOOLS. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). ORTHOMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*, 2178–2189. https://doi.org/10.1101/gr.1224503

Lindsley, D. L., & Grell, E. H. (1969). Spermatogenesis without chromosomes in Drosophila melanogaster. *Genetics*, *61*, 69–78.

Lipinska, A., Cormier, A., Luthringer, R., Peters, A. F., Corre, E., Gachon, C. M. M., ... Coelho, S. M. (2015). Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Molecular Biology and Evolution*, *32*, 1581–1597. https://doi.org/10.1093/molbev/msv049

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., & Usadel, B. (2012). ROBINA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, *40*, 622–627. https://doi.org/10.1093/nar/gks540

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap replacement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*, 1632–1635. https://doi.org/10.1126/science.1158395

Magnusson, K., Mendes, A. M., Windbichler, N., Papathanos, P. A., Nolan, T., Dottorini, T., ... Crisanti, A. (2011). Transcription regulation of sex-biased genes during ontogeny in the malarial vector *Anopheles gambiae*. *PLoS ONE*, *6*, e21572. https://doi.org/10.1371/journal.pone.0021572

Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nature Ecology & Evolution*, *1*, 0006. https://doi.org/10.1038/s41559-016-0006

Mank, J. E., Hultin-Rosenberg, L., Axelsson, E., & Ellegren, H. (2007). Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Molecular Biology and Evolution*, 24, 2698–2706. https://doi.org/10.1093/molbev/msm208

Mank, J. E., Hultin-Rosenberg, L., Webster, M. T., & Ellegren, H. (2008). The unique genomic properties of sex-biased genes: Insights from avian microarray data. *BMC Genomics*, 9, 148. https://doi.org/10.1186/1471-2164-9-148

Mank, J. E., Nam, K., Brunström, B., & Ellegren, H. (2010a). Ontogenic complexity of sexual dimorphism and sex-specific selection. *Molecular Biology and Evolution*, 27, 1570–1578. https://doi.org/10.1093/molbev/msq042

Mank, J. E., Vicoso, B., Berlin, S., & Charlesworth, B. (2010b). Effective population size and the Faster-X effect: Empirical results and their interpretation. *Evolution*, 64, 663–674. https://doi.org/10.1111/j.1558-5646.2009.00853.x

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature*, 351, 652–654. https://doi.org/10.1038/351652a0

Moore, J. C., & Pannell, J. R. (2011). Sexual selection in plants. *Current Biology*, 21, R176–R182. https://doi.org/10.1016/j.cub.2010.12.035

Muyle, A., Shearn, R., & Marais, G. A. B. (2017). The evolution of sex chromosomes and dosage compensation in plants. *Genome Biology and Evolution*, 9, 627–645. https://doi.org/10.1093/gbe/evw282

Otto, S. P., Scott, M. F., & Immler, S. (2015). Evolution of haploid selection in predominantly diploid organisms. *Proceedings of the National Academy of Sciences of the USA*, 112, 15952–15957. https://doi.org/10.1073/pnas.1512004112

Pál, C., Papp, B., & Hurst, L. D. (2001). Highly expressed genes in yeast evolve slowly. *Genetics*, 158, 927–931.

Parsch, J., & Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. *Nature Reviews Genetics*, 14, 83–87. https://doi.org/10.1038/nrg3376

Peeters, L., & Totland, Ø. (1999). Wind to insect pollination ratios and floral traits in five alpine *Salix* species. *Canadian Journal of Botany*, 77, 556–563. https://doi.org/10.1139/b99-003

Perry, J. C., Harrison, P. W., & Mank, J. E. (2014). The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 31, 1206–1219. https://doi.org/10.1093/molbev/msu072

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). STRINGTIE enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33, 290–295. https://doi.org/10.1038/nbt.3122

Petes, T. D., & Hill, C. W. (1988). Recombination between repeated genes in microorganisms. *Annual Review of Genetics*, 22, 147–168. https://doi.org/10.1146/annurev.ge.22.120188.001051

Pointer, M. A., Harrison, P. W., Wright, A. E., & Mank, J. E. (2013). Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genetics*, 9, e1003697. https://doi.org/10.1371/journal.pgen.1003697

Pool, J. E., & Nielsen, R. (2007). Population size changes reshape genomic patterns of diversity. *Evolution*, 63, 3001–3006. https://doi.org/10.1111/j.1558-5646.2007.00238.x

Pucholt, P., Wright, A. E., Conze, L. L., Mank, J. E., & Berlin, S. (2017). Recent sex chromosomes divergence despite ancient dioecy in the willow *Salix viminalis*. *Molecular Biology and Evolution*, 34, 1991–2001. https://doi.org/10.1093/molbev/msx144

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

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D., & Hartl, D. L. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*, 300, 1742–1745. https://doi.org/10.1126/science.1085881

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*, 101, 1588–1596. https://doi.org/10.3732/ajb.1400196

Robinson, K. M., Delhomme, N., Mähler, N., Schiffthaler, B., Önskog, J., Albrectsen, B. R., ... Street, N. R. (2014). *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. *BMC Plant Biology*, 14, 276. https://doi.org/10.1186/s12870-014-0276-5

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). EDGER: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Sharma, E., Künstner, A., Fraser, B. A., Zipprich, G., Kottler, V. A., Henz, S. R., ... Dreyer, C. (2014). Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genomics*, 15, 400. https://doi.org/10.1186/1471-2164-15-400

Sharp, P. M., & Li, W. H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15, 1281–1295. https://doi.org/10.1093/nar/15.3.1281

Slotte, T., Bataillon, T., Hansen, T. T., St Onge, K., Wright, S. I., & Schierup, M. H. (2011). Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biology and Evolution*, 3, 1210–1219. https://doi.org/10.1093/gbe/evr094

Snow, A. A., & Spira, T. P. (1996). Pollen-tube competition and male fitness in Hibiscus moscheutos. *Evolution*, 50, 1866–1870. https://doi.org/10.1111/j.1558-5646.1996.tb03573.x

Stoletzki, N., & Eyre-Walker, A. (2011). Estimation of the neutrality index. *Molecular Biology and Evolution*, 28, 63–70. https://doi.org/10.1093/molbev/msq249

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjödin, A., ... Street, N. R. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist*, 208, 1149–1156. https://doi.org/10.1111/nph.13557

Suzuki, R., & Shimodaira, H. (2006). PVCLUST: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22, 1540–1542. https://doi.org/10.1093/bioinformatics/btl117

Szövényi, P., Devos, N., Weston, D. J., Yang, X., Hock, Z., Shaw, J. A., ... Wagner, A. (2014). Efficient purging of deleterious mutations in plants with haploid selfing. *Genome Biology and Evolution*, 6, 1238–1252. https://doi.org/10.1093/gbe/evu099

Szövényi, P., Ricca, M., Hock, Z., Shaw, J. A., Shimizu, K. K., & Wagner, A. (2013). Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Molecular Biology and Evolution*, 30, 1929–1939. https://doi.org/10.1093/molbev/mst095

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123, 597–601.

Urrutia, A. O., & Hurst, L. D. (2001). Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics*, 159, 1191–1199.

Whittle, C. A., & Johannesson, H. (2013). Evolutionary dynamics of sex-biased genes in a hermaphrodite fungus. *Molecular Biology and Evolution*, 30, 2435–2446. https://doi.org/10.1093/molbev/mst143

Whittle, C. A., & Johnston, M. O. (2003). Male-biased transmission of deleterious mutations to the progeny in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA*, 100, 4055–4059. https://doi.org/10.1073/pnas.0730639100

Whittle, C. A., Malik, M. R., & Krochko, J. E. (2007). Gender-specific selection on codon usage in plant genomes. *BMC Genomics*, 8, 169. https://doi.org/10.1186/1471-2164-8-169

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, *87*, 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Wright, A. E., Harrison, P. W., Zimmer, F., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Molecular Ecology*, *24*, 1218–1235. https://doi.org/10.1111/mec.13113

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, L., Zhang, Z., & He, S. (2016). Both male-biased and female-biased genes evolve faster in fish genomes. *Genome Biology and Evolution*, *8*, 1–22. https://doi.org/10.1093/gbe/evw239

Zemp, N., Minder, A., & Widmer, A. (2014). Identification of internal reference genes for gene expression normalization between the two sexes in dioecious white Campion. *PLoS ONE*, *9*, e92893. https://doi.org/10.1371/journal.pone.0092893

Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G. A., & Widmer, A. (2016). Evolution of sex-biased gene expression in a dioecious plant. *Nature Plants*, *2*, 16168. https://doi.org/10.1038/nplants.2016.168

Zemp, N., Tavares, R., & Widmer, A. (2015). Fungal infection induces sex-specific transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene latifolia*. *PLoS Genetics*, *11*, e1005536. https://doi.org/10.1371/journal.pgen.1005536

Zluvova, J., Zak, J., Janousek, B., & Vyskot, B. (2010). Dioecious *Silene latifolia* plants show sexual dimorphism in the vegetative stage. *BMC Plant Biology*, *10*, 208. https://doi.org/10.1186/1471-2229-10-208

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

References

Ahmed, S., Cock, J. M., Pessia, E., Luthringer, R., Cormier, A., Robuchon, M., Sterck, L., Peters, A. F., Dittami, S. M., Corre, E., Valero, M., Aury, J.-M., Roze, D., Van De Peer, Y., Bothwell, J., Marais, G. A. B., and Coelho, S. M. (2014). A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Current Biology, 24*(17), 1945-1957.

Ainsworth, C. (2000). Boys and girls come out to play: The molecular biology of dioecious plants. *Annals of botany, 86*(2), 211-221.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics, 139*(2), 1067-1076.

Albritton, S. E., Kranz, A.-L., Rao, P., Kramer, M., Dieterich, C., and Ercan, S. (2014). Sex-biased gene expression and evolution of the X chromosome in nematodes. *Genetics, 197*(3), 865-883.

Almeida, P., Proux-Wera, E., Churcher, A., Soler, L., Dainat, J., Pucholt, P., Nordlund, J., Martin, T., Rönnberg-Wästljung, A. C., and Nystedt, B. (2019). Single-molecule genome assembly of the Basket Willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *bioRxiv*, 589804.

Altenhoff, A. M., and Dessimoz, C. (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Computational Biology, 5*(1), e1000262.

Altschul, S. F., Gish, W., Miller, W., Myers, E., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-169.

Andersson, M. (1994). *Sexual selection*. Princeton, New Jersey: Princeton University Press.

Andolfatto, P. (2008). Controlling type-I error of the McDonald-Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics, 180*(3), 1767-1771.

Arnqvist, G., and Rowe, L. (2005). *Sexual conflict*. Princeton, New Jersey: Princeton University Press.

Arunkumar, K., Mita, K., and Nagaraju, J. (2009). The silkworm Z chromosome is enriched in testis-specific genes. *Genetics, 182*(2), 493-501.

Arunkumar, R., Josephs, E. B., Williamson, R. J., and Wright, S. I. (2013). Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Molecular Biology and Evolution, 30*(11), 2475-2486.

Assis, R., Zhou, Q., and Bachtrog, D. (2012). Sex-biased transcriptome evolution in *Drosophila*. *Genome Biology and Evolution, 4*(11), 1189-1200.

Axelsson, E., Hultin-Rosenberg, L., Brandström, M., Zwahlén, M., Clayton, D. F., and Ellegren, H. (2008). Natural selection in avian protein-coding genes expressed in brain. *Molecular Ecology, 17*(12), 3008-3017.

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

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., and Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics, 27*(9), 350-357.

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., and The Tree of Sex Consortium (2014). Sex determination: why so many ways of doing it? *PLoS Biology, 12*(7), e1001899.

Bachtrog, D., Toda, N. R. T., and Lockton, S. (2010). Dosage compensation and demasculinization of X chromosomes in *Drosophila*. *Current Biology, 20*(16), 1476-1481.

Badyaev, A. V. (2002). Growing apart: An ontogenic perspective on the evolution of sexual size dimorphism. *Trends in Ecology and Evolution, 17*(8), 369-378.

Barrett, S. C. H., and Hough, J. (2013). Sexual dimorphism in flowering plants. *Journal of Experimental Botany, 64*(1), 67-82.

Bateman, A. J. (1948). Intra-sexual selection in *Drosophila*. *Heredity, 2*(3), 349-368.

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y. P., Hahn, M. W., Nista, P. M., Jones, C. D., Kern, A. D., Dewey, C. N., Pachter, L., Myers, E., and Langley, C. H. (2007). Population genomics: whole-genome

analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology, 5*(11), e310.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Sociecty: series B (Methodological), 57*(1), 289-300.

Bergero, R., Forrest, A., Kamau, E., and Charlesworth, D. (2007). Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics, 175*(4), 1945-1954.

Bergero, R., Gardner, J., Bader, B., Yong, L., and 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.

Bergero, R., Qiu, S., Forrest, A., Borthwick, H., and Charlesworth, D. (2013). Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. *Genetics, 194*(3), 673-686.

Berlin, S., Trybush, S. O., Fogelqvist, J., Gyllenstrand, N., Hallingbäck, H. R., Åhman, I., Nordh, N. E., Shield, I., Powers, S. J., Weih, M., Lagercrantz, U., Ronnberg-Wastljung, A.-C., Karp, A., and Hanley, S. J. (2014). Genetic diversity, population structure and phenotypic variation in European *Salix viminalis* L. (Salicaceae). *Tree Genetics & Genomes, 10*(6), 1595-1610.

Bernasconi, G., Ashman, T.-L., Birkhead, T. R., Bishop, J. D. D., Grossniklaus, U., Kubli, E., Marshall, D. L., Schmid, B., Skogsmyr, I., Snook, R. R., Taylor, D., Till-Bottraud, I., Ward, P. I., Zeh, D. W., and Hellriegel, B. (2004). Evolutionary ecology of the prezygotic stage. *Science, 303*(5660), 971-975.

Beukeboom, L. W., and Perrin, N. (2014). *The evolution of sex determination*: Oxford University Press, USA.

Bick, Y. A., and Jackson, W. D. (1967). DNA content of monotremes. *Nature, 215*(5097), 192-193.

Bierne, N., and Eyre-Walker, A. (2003). The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. *Genetics, 165*(3), 1587-1597.

Blaser, O., Grossen, C., Neuenschwander, S., and Perrin, N. (2013). Sex‐chromosome turnovers induced by deleterious mutation load. *Evolution: International Journal of Organic Evolution, 67*(3), 635-645.

Bloch, N. I., Corral-López, A., Buechel, S. D., Kotrschal, A., Kolm, N., and Mank, J. E. (2018). Early neurogenomic response associated with variation in guppy female mate preference. *Nature Ecology & Evolution, 2*(11), 1772.

Böhne, A., Sengstag, T., and Salzburger, W. (2014). Comparative transcriptomics in East African cichlids reveals sex-and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biology and Evolution, 6*(9), 2567-2585.

Bonduriansky, R., and Chenoweth, S. F. (2009). Intralocus sexual conflict. *Trends in Ecology & Evolution, 24*(5), 280-288.

Boyes, D. C., Nasrallah, M. E., Vrebalov, J., and Nasrallah, J. B. (1997). The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. *Plant Cell, 9*(2), 237-247.

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., and Palmiter, R. D. (1989). Genetically haploid spermatids are phenotipically diploid. *Nature, 337*, 373-376.

Breden, F., and Stoner, G. (1987). Male predation risk determines female preference in the Trinidad guppy. *Nature, 329*(6142), 831.

Bull, J. J. (1983). *Evolution of sex determining mechanisms*: The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Bull, J. J., and Charnov, E. L. (1977). Changes in the heterogametic mechanism of sex determination. *Heredity, 39*(1), 1.

Campos, J., Qiu, S., Guirao-Rico, S., Bergero, R., and Charlesworth, D. (2017). Recombination changes at the boundaries of fully and partially sex-linked regions between closely related *Silene* species pairs. *Heredity, 118*(4), 395.

Catalán, A., Hutter, S., and Parsch, J. (2012). Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics, 13*(1), 654.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America, 75*(11), 5618-5622.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics, 10*(3), 195.

Charlesworth, B., and Charlesworth, D. (2000). The degenration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 355*(1403), 1563-1572.

Charlesworth, B., Coyne, J. A., and Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist, 130*(1), 113-146.

Charlesworth, D. (2002). Plant sex determination and sex chromosomes. *Heredity, 88*(2), 94-101.

Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity, 95*(2), 118-128.

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C., and Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics, 8*(10), 762.

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J. N., Hong, Y., Li, Q., Sha, Z., Zhou, H., Xie, M., Yu, Q., Liu, Y., Xiang, H., Wang, N., Wu, K., Yang, C., Zhou, Q., Liao, X., Yang, L., Hu, Q., Zhang, J., Meng, L., Jin, L., Tian, Y., Lian, J., Yang, J., Miao, G., Liu, S., Liang, Z., Yan, F., Li, Y., Sun, B., Zhang, H., Zhang, J., Zhu, Y., Du, M., Zhao, Y., Schartl, M., Tang, Q., and Wang, J. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics, 46*(3), 253-260.

Cherry, J. L. (2010). Highly expressed and slowly evolving proteins share compositional properties with thermophilic proteins. *Molecular Biology and Evolution, 27*(3), 735-741.

Chibalina, M. V., and Filatov, D. A. (2011). Plant Y chromosome degeneration is retarded by haploid purifying selection. *Current Biology, 21*(17), 1475-1479.

Chikhi, R., and Medvedev, P. (2014). Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics, 30*(1), 31-37.

Chippindale, A. K., Gibson, J. R., and Rice, W. R. (2001). Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America, 98*(4), 1671-1675.

Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D. J., Tomasino, E., Coutanceau, J.-P., Pepey, E., Shirak, A., and Kocher, T. D. (2008). Genetics of sex determination in tilapiine species. *Sexual Development, 2*(1), 43-54.

Comeron, J. M., and Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol., 47*(3), 268-274.

Comeron, J. M., Williford, A., and Kliman, R. (2008). The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity, 100*(1), 19.

Connallon, T., and Clark, A. G. (2010). Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution, 64*(12), 3417-3442.

Connallon, T., and Knowles, L. L. (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics, 21*(9), 495-499.

Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., and Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature, 508*(7497), 488.

Cossard, G. G., Toups, M. A., and Pannell, J. R. (2018). Sexual dimorphism and rapid turnover in gene expression in pre-reproductive seedlings of a dioecious herb. *Annals of botany, 123*(7), 1119-1131.

Cox, R. M., and Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist, 173*(2), 176-187.

Cronk, Q. C., Needham, I., and Rudall, P. J. (2015). Evolution of catkins: Inflorescence morphology of selected Salicaceae in an evolutionary and developmental context. *Frontiers in Plant Science, 6*, 1030.

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D. (2015). Young genes out of the male: An insight from evolutionary age analysis of the pollen transcriptome. *Molecular Plant, 8*(6), 935-945.

Cutter, A. D., and Ward, S. (2004). Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Molecular Biology and Evolution, 22*(1), 178-188.

Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G. A., Milne, R., Chen, Y., Wan, Z., Wang, Z., Luo, W., Wang, K., Wan, D., Wang, M., Wang, J., Liu, J., and Yin, T. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Research, 24*(10), 1274-1277.

Darolti, I., Wright, A. E., Pucholt, P., Berlin, S., and Mank, J. E. (2018). Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*. *Molecular Ecology, 27*(3), 694-708.

Darolti, I., Wright, A. E., Sandkam, B. A., Morris, J., Bloch, N. I., Farré, M., Fuller, R. C., Bourne, G. R., Larkin, D. M., Breden, F., and Mank, J. E. (2019). Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *bioRxiv*, 589788.

228

Darwin, C. (1871). *The descent of man, and selecton in relation to sex*. London: Murray.

Dean, R., Wright, A. E., Marsh-Rollo, S. E., Nugent, B. M., Alonzo, S. H., and Mank, J. E. (2017). Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. *Molecular Ecology, 26*(2), 505-518.

Devlin, R. H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture, 208*(3-4), 191-364.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Dobzhansky, T., and Dobzhansky, T. G. (1970). *Genetics of the evolutionary process*. New York: Columbia University Press.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the United States of America, 102*(40), 14338-14343

Dufresnes, C., Borzee, A., Horn, A., Stöck, M., Ostini, M., Sermier, R., Wassef, J., Livinchuck, S. N., Kosch, T. A., Waldman, B., Jang, Y., Brelsford, A., and Perrin, N. (2015). Sex-chromosome homomorphy in Palearctic tree frogs results from both turnovers and X–Y recombination. *Molecular Biology and Evolution, 32*(9), 2328-2337.

Dufresnes, C., Majtyka, T., Baird, S. J., Gerchen, J. F., Borzée, A., Savary, R., Ogielska, M., Perrin, N., and Stöck, M. (2016). Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. *Scientific reports, 6*, 21029.

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Current Opinions in Genetics and Development, 12*(6), 640-649.

Eads, B. D., Colbourne, J. K., Bohuski, E., and Andrews, J. (2007). Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genomics, 8*(1), 464.

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology, 3*(3), e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics, 10*(1), 48.

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics, 8*(9), 689-698.

Emerson, J., Kaessmann, H., Betrán, E., and Long, M. (2004). Extensive gene traffic on the mammalian X chromosome. *Science, 303*(5657), 537-540.

Endler, J. A. (1980). Natural selection on color patterns in *Poecilia reticulata. Evolution, 34*(1), 76-91.

Endler, J. A. (1984). Natural and sexual selection on color patterns in poeciliid fishes. In *Evolutionary ecology of neotropical freshwater fishes* (pp. 95-111): Springer.

Endler, J. A. (1995). Multiple-trait coevolution and environmental gradients in guppies. *Trends in Ecology & Evolution, 10*(1), 22-29.

Erbar, C. (2003). Pollen tube transmitting tissue: Place of competition of male gametophytes. *International Journal of Plant Sciences, 164*(S5), S265-S277.

Ercan, S., Giresi, P. G., Whittle, C. M., Zhang, X., Green, R. D., and Lieb, J. D. (2007). X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nature Genetics, 39*(3), 403-408.

Erickson, R. P. (1973). Haploid gene expression versus meiotic drive: The relevance of intercellular bridges during spermatogenesis. *Nature New Biology, 243*(128), 210-212.

Ezaz, T., Sarre, S. D., O'Meally, D., Graves, J. A., and Georges, A. (2009). Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenetic and Genome Research, 127*(2-4), 249-260.

Fay, J. C., Wyckoff, G. J., and Wu, C.-I. (2001). Positive and negative selection on the human genome. *Genetics, 158*(3), 1227-1234.

Fisher, M. J. (1928). The morphology and anatomy of the flowers of the Salicaceae I. *American Journal of Botany, 15*(5), 307-326.

Fisher, R. A. (1931). The evolution of dominance. *Biological reviews, 6*(4), 345-368.

Fitzpatrick, M. J. (2004). Pleiotropy and the genomic location of sexually selected genes. *The American Naturalist, 163*(6), 800-808.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C. G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kahari, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy,

D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T. J. P., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., and Searle, S. M. J. (2014). Ensembl 2014. *Nucleic Acids Research, 42*(D1), D749-D755.

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G. J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthelemy, R.-M., Bast, J., Bazykin, G. A., Charbol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., Gouret, P., Hallatschek, O., Hecox-Lea, B., Labadie, K., Lejeune, B., Piskurek, O., Poulain, J., Rodriguez, F., Ryan, J. F., Vakhrusheva, O. A., Wajnberg, E., Wirth, B., Yushenova, I., Kellis, M., Kondrashov, A. S., Welch, D. B. M., Pontarotti, P., Weissenbach, J., Wincker, P., Jaillon, O., and Van Doninck, K. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature, 500*(7463), 453.

Franchini, P., Jones, J. C., Xiong, P., Kneitz, S., Gompert, Z., Warren, W. C., Walter, R. B., Meyer, A., and Schartl, M. (2018). Long-term experimental hybridisation results in the evolution of a new sex chromosome in swordtail fish. *Nature Communications, 9*(1), 5136.

Fraser, B. A., Künstner, A., Reznick, D. N., Dreyer, C., and Weigel, D. (2015). Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). *Molecular Ecology, 24*(2), 389-408.

Fridolfsson, A. K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H. C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J., and Ellegren, H. (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences of the United States of America, 95*(14), 8147-8152.

Fujito, S., Takahata, S., Suzuki, R., Hoshino, Y., Ohmido, N., and Onodera, Y. (2015). Evidence for a common origin of homomorphic and Heteromorphic sex chromosomes in distinct *Spinacia* species. *G3: Genes, Genomes, Genetics, 5*(8), 1663-1673.

Gamble, T., Geneva, A. J., Glor, R. E., and Zarkower, D. (2014). Anolis sex chromosomes are derived from a single ancestral pair. *Evolution, 68*(4), 1027-1041.

Gershoni, M., and Pietrokovski, S. (2014). Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. *Nature Communications, 5*, 4438.

Ghelardini, L., Berlin, S., Weih, M., Lagercrantz, U., Gyllenstrand, N., and Ronnberg-Wastljung, A. C. (2014). Genetic architecture of spring and autumn phenology in *Salix*. *BMC Plant Biology, 14*(1), 31.

Gibson, J. R., Chippindale, A. K., and Rice, W. R. (2002). The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 269*(1490), 499-505.

Gordon, S. P., Lopez-Sepulcre, A., and Reznick, D. N. (2012). Predation-associated differences in sex linkage of wild guppy coloration. *Evolution, 66*(3), 912-918.

Gordon, S. P., Reznick, D., Arendt, J. D., Roughton, A., Ontiveros Hernandez, M. N., Bentzen, P., and López-Sepulcre, A. (2015). Selection analysis on the rapid evolution of a secondary sexual trait. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 282*(1813), 20151244.

Gossmann, T. I., Saleh, D., Schmid, M. W., Spence, M. A., and Schmid, K. J. (2016). Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. *Molecular Biology and Evolution, 33*(7), 1669-1678.

Gossmann, T. I., Schmid, M. W., Grossniklaus, U., and Schmid, K. J. (2014). Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Molecular Biology and Evolution, 31*(3), 574-583.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-toh, K., Friedman, N., and Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology, 29*(7), 644-652.

Grath, S., and Parsch, J. (2016). Sex-biased gene expression. *Annual Review of Genetics, 50*, 29-44.

Guo, M., Davis, D., and Birchler, J. A. (1996). Dosage effects on gene expression in a maize ploidy series. *Genetics, 142*(4), 1349-1355.

Gupta, V., Parisi, M., Sturgill, D., Nuttall, R., Doctolero, M., Dudko, O. K., Malley, J. D., Eastman, P. S., and Oliver, B. (2006). Global analysis of X-chromosome dosage compensation. *Journal of Biology, 5*(1), 3.

Haldane, J., and Haldane, J. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics, 12*(02).

Hale, M. C., McKinney, G. J., Thrower, F. P., and Nichols, K. M. (2018). Evidence of sex-bias in gene expression in the brain transcriptome of two populations of rainbow trout (*Oncorhynchus mykiss*) with divergent life histories. *PLoS ONE, 13*(2), e0193009

Hale, M. C., Xu, P., Scardina, J., Wheeler, P. A., Thorgaard, G. H., and Nichols, K. M. (2011). Differential gene expression in male and female rainbow trout embryos prior to the onset of gross morphological differentiation of the gonads. *BMC Genomics, 12*(1), 404.

Hallingback, H. R., Fogelqvist, J., Powers, S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., Martin, T., Weih, M., Gyllenstrand, N., Karp, A., Lagercrantz, U., Hanley, S. J., Berlin, S., and Ronnberg-Wastljung, A. C. (2016). Association mapping in *Salix viminalis* L. (Salicaceae) - identification of candidate genes associated with growth and phenology. *Global Change Biology Bioenergy, 8*(3), 670-685.

Hambuch, T. M., and Parsch, J. (2005). Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. *Genetics, 170*(4), 1691-1700.

Handley, L.-J. L., Ceplitis, H., and Ellegren, H. (2004). Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics, 167*(1), 367-376.

Harkess, A., Mercati, F., Shan, H. Y., Sunseri, F., Falavigna, A., and Leebens-Mack, J. (2015). Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytologist, 207*(3), 883-892.

Harrison, P. W., Jordan, G. E., and Montgomery, S. H. (2014). SWAMP: sliding window alignment masker for PAML. *Evolutionary Bioinformatics, 10*, EBO. S18193.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., and Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America, 112*(14), 4393-4398.

Helibuth, J. C. (2000). Lower species richness in dioecious clades. *The American Naturalist, 156*(3), 221-241.

Hense, W., Baines, J. F., and Parsch, J. (2007). X chromosome inactivation during *Drosophila spermatogenesis*. *PLoS Biology, 5*(10), e273.

Hershberg, R., and Petrov, D. A. (2008). Selection on codon bias. *Annual Review of Genetics, 42*, 287-299.

Hill, W. G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research, 8*(3), 269-294.

Hillis, D. M., and Green, D. M. (1990). Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. *Journal of Evolutionary Biology, 3*(1-2), 49-64.

Houde, A. E., and Endler, J. A. (1990). Correlated evolution of female mating preferences and male color patterns in the guppy *Poecilia reticulata*. *Science, 248*(4961), 1405-1408.

Hough, J., Hollister, J. D., Wang, W., Barrett, S. C., and Wright, S. I. (2014). Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proceedings of the National Academy of Sciences of the United States of America, 111*(21), 7713-7718.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research, 9*(9), 868-877.

Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Yi, S. V., Keller, L., and Goodisman, M. A. D. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticty. *Proceedings of the National Academy of Sciences of the United States of America, 108*(38), 15936-15941.

Husby, A., Schielzeth, H., Forstmeier, W., Gustafsson, L., and Qvarnström, A. (2013). Sex chromosome linked genetic variance and the evolution of sexual dimorphism of quantitative traits. *Evolution: International Journal of Organic Evolution, 67*(3), 609-619.

Ingleby, F. C., Flis, I., and Morrow, E. H. (2015). Sex-biased gene expression and sexual conflict throughout development. *Cold Spring Harbor Perspectives in Biology, 7*(1), a017632.

Ingvarsson, P. K. (2010). Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. *Molecular Biology and Evolution, 27*(3), 650-660.

Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., Barrow, L. N., Canestrelli, D., Crochet, P.-A., and Dufresnes, C. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications, 9*(1), 4088.

Jiang, H., Peng, S., Zhang, S., Li, X., Korpelainen, H., and Li, C. (2012). Transcriptional profiling analysis in *Populus yunnanensis* provides insights into molecular mechanisms of sexual differences in salinity tolerance. *Jorunal of Experimental Botany, 63*(10), 3709-3726.

Jiang, Z. F., and Machado, C. A. (2009). Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics, 183*(3), 1175-1185.

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics, 27*(21), 3070-3071.

Joseph, S. B., and Kirkpatrick, M. (2004). Haploid selection in animals. *Trends in Ecology & Evolution, 19*(11), 592-597.

Käfer, J., Marais, G. A. B., and Pannell, J. R. (2017). On the rarity of dioecy in flowering plants. *Molecular Ecology, 26*(5), 1225-1241.

Kaiser, V. B., and Ellegren, H. (2006). Nonrandom distribution of genes with sex-biased expression in the chicken genome. *Evolution, 60*(9), 1945-1951.

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., and Kikuchi, K. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genetics, 8*(7), e1002798.

Kao, T. H., and McCubbin, A. G. (1996). How flowering plants discriminate between self and non-self pollen to prevent inbreeding. *Proceedings of the National Academy of Sciences of the United States of America, 93*(22), 12059-12065.

Karrenberg, S., Kollmann, J., and Edwards, P. J. (2002). Pollen vectors and inflorescence morphology in four species of *Salix. Plant Systematics and Evolution, 235*(1), 181-188.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin, I. T., Takeda, H., Morishita, S., and Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature, 447*(7145), 714-719.

Kelley, D. R., Schatz, M. C., and Salzberg, S. L. (2010). Quake: quality-aware detection and correction of sequencing errors. *Genome Biology, 11*(11), R116.

Kemp, D. J., Reznick, D. N., Grether, G. F., and Endler, J. A. (2009). Predicting the direction of ornament evolution in Trinidadian guppies (Poecilia reticulata). *Proceedings of the Royal Society B: Biological Sciences, 276*(1677), 4335-4343.

Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America, 100*(20), 11484-11489.

Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., and Pääbo, S. (2005). Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science, 309*(5742), 1850-1854.

Khil, P. P., Smirnova, N. A., Romanienko, P. J., and 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.

Khila, A., Abouheif, E., and Rowe, L. (2012). Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. *Science, 336*(6081), 585-589.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods, 12*(4), 357-360.

Kim, J., Larkin, D. M., Cai, Q., Asan, Zhang, Y., Ge, R. L., Auvil, L., Capitanu, B., Zhang, G., Lewin, H. A., and Ma, J. (2013). Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 1785-1790.

Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., Absher, D. M., Grimwood, J., Schmutz, J., and Myers, R. M. (2009). A role for a neo-sex chromosome in stickleback speciation. *Nature, 461*(7267), 1079.

Koboldt, D. C., Larson, D. E., and Wilson, R. K. (2014). Using VarScan 2 for germline variant calling and somatic mutation detection. *Current Protocols in Bioinformatics, 44*(1), 15.14.11-15.14.17.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., and Wilson, R. K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research, 22*(3), 568-576.

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

Kolde, R. (2012). Pheatmap: Pretty heatmap. *R package version 61*.

Kondo, M., Nagao, E., Mitani, H., and Shima, A. (2001). Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet Research, 78*(1), 23-30.

Kondo, M., Nanda, I., Schmid, M., and Schartl, M. (2009). Sex cetermination and sex chromosome evolution: insights from medaka. *Sexual Development, 3*(2-3), 88-98.

Kondrashov, A. S., and Crow, J. F. (1991). Haploidy or diploidy: Which is better? *Nature, 351*(6324), 314-315.

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brannstrom, I., Immler, S., Maklakov, A., and Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology, 23*(2), 168-171.

Kozielska, M., Pen, I., Beukeboom, L. W., and Weissing, F. J. (2006). Sex ratio selection and multi-factorial sex determination in the housefly: a dynamic model. *Jorunal of Evolutionary Biology, 19*(3), 879-888.

Lahn, B. T., and Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science, 286*(5441), 964-967.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3), R25.

Larschan, E., Bishop, E. P., Kharchenko, P. V., Core, L. J., Lis, J. T., Park, P. J., and Kuroda, M. I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature, 471*(7336), 115-118.

Lawrie, D. S., Messer, P. W., Hershberg, R., and Petrov, D. A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics, 9*(5), e1003527.

Leder, E. H., Cano, J. M., Leinonen, T., O'Hara, R. B., Nikinmaa, M., Primmer, C. R., and Merila, J. (2010). Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. *Molecular Biology and Evolution, 27*(7), 1495-1503.

Lenormand, T. (2003). The evolution of sex dimorphism in recombination. *Genetics, 163*(2), 811-822.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment / map format and SAMtools. *Bioinformatics, 25*, 2078-2079.

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research, 13*(9), 2178-2189.

Lindholm, A., and Breden, F. (2002). Sex chromosomes and sexual selection in poeciliid fishes. *The American Naturalist, 160*(S6), S214-224.

Lindholm, A. K., Brooks, R., and Breden, F. (2004). Extreme polymorphism in a Y-linked sexually selected trait. *Heredity, 92*(3), 156-162.

Lindholm, A. K., Sandkam, B., Pohl, K., and Breden, F. (2015). *Poecilia picta*, a close relative to the guppy, exhibits red male coloration polymorphism: a system for phylogenetic comparisons. *PLoS ONE, 10*(11), e0142089.

Lindsley, D. L., and Grell, E. H. (1969). Spermatogenesis without chromosomes in *Drosophila melanogaster*. *Genetics, 61*, 69-78.

Lipinska, A., Cormier, A., Luthringer, R., Peters, A. F., Corre, E., Gachon, C. M. M., Cock, J. M., and Coelho, S. M. (2015). Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Molecular Biology and Evolution, 32*(6), 1581-1597.

Lisachov, A. P., Zadesenets, K. S., Rubtsov, N. B., and Borodin, P. M. (2015). Sex chromosome synapsis and recombination in male guppies. *Zebrafish, 12*(2), 174-180.

Liu, Z., Moore, P. H., Ma, H., Ackerman, C. M., Ragiba, M., Yu, Q., Pearl, H. M., Kim, M. S., Charlton, J. W., and Stiles, J. I. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature, 427*(6972), 348.

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., and Usadel, B. (2012). RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research, 40*, 622-627.

Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap replacement prevents errors in sequence alignment and evolutionary analysis. *Science, 320*, 1632-1635.

Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research, 21*(6), 936-939.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S. M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T. W., and Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience, 1*(1), 18.

Magnusson, K., Mendes, A. M., Windbichler, N., Papathanos, P. A., Nolan, T., Dottorini, T., Rizzi, E., Christophides, G. K., and Crisanti, A. (2011). Transcription regulation of sex-biased genes during ontogeny in the malaria vector *Anopheles gambiae*. *PLoS ONE, 6*(6), e21572.

Malone, J. H., Cho, D. Y., Mattiuzzo, N. R., Artieri, C. G., Jiang, L., Dale, R. K., Smith, H. E., McDaniel, J., Munro, S., Salit, M., Andrews, J., Przytycka, T. M., and Oliver, B. (2012). Mediation of *Drosophila* autosomal dosage effects and compensation by network interactions. *Genome Biology, 13*(4), R28.

Mank, J. E. (2009). The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in Genetics, 25*(5), 226-233.

Mank, J. E. (2013). Sex chromosome dosage compensation: definitely not for everyone. *Trends in Genetics, 29*(12), 677-683.

Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nature Ecology and Evolution, 1*(1), 6.

Mank, J. E., and Avise, J. C. (2009). Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sexual Development, 3*(2-3), 60-67.

Mank, J. E., Hall, D. W., Kirkpatrick, M., and Avise, J. C. (2005). Sex chromosomes and male ornaments: a comparative evaluation in ray-finned fishes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 273*(1583), 233-236.

Mank, J. E., Hosken, D. J., and Wedell, N. (2011). Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. *Evolution, 65*(8), 2133-2144.

Mank, J. E., Hosken, D. J., and Wedell, N. (2014). Conflict on the sex chromosomes: cause, effect, and complexity. *Cold Spring Harbor Perspectives in Biology, 6*(12), a017715.

Mank, J. E., Hultin-Rosenberg, L., Axelsson, E., and Ellegren, H. (2007). Fast-X on the Z: rapid evolution of sex-linked genes in birds. *Genome Research, 17*(5), 618-624.

Mank, J. E., Hultin-Rosenberg, L., Webster, M. T., and Ellegren, H. (2008). The unique genomic properties of sex-biased genes: insights from avian microarray data. *BMC Genomics, 9*(1), 148.

Mank, J. E., Nam, K., Brunstrom, B., and Ellegren, H. (2010a). Ontogenetic complexity of sexual dimorphism and sex-specific selection. *Molecular Biology and Evolution, 27*(7), 1570-1578.

Mank, J. E., Nam, K., and Ellegren, H. (2009). Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution, 27*(3), 661-670.

Mank, J. E., Promislow, D. E., and Avise, J. C. (2006). Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society, 87*(1), 83-93.

Mank, J. E., Vicoso, B., Berlin, S., and Charlesworth, B. (2010b). Effective population size and the Faster-X effect: empirical results and their interpretation. *Evolution, 64*(3), 663-674.

Masly, J. P., and Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in *Drosophila. PLoS Biology, 5*(9), e243.

Matsubara, K., Tarui, H., Toriba, M., Yamada, K., Nishida-Umehara, C., Agata, K., and Matsuda, Y. (2006). Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proceedings of the National Academy of Sciences of the United States of America, 103*(48), 18190-18195.

Matsuda, M., Sotoyama, S., Hamaguchi, S., and Sakaizumi, M. (1999). Male-specific restriction of recombination frequency in the sex chromosomes of the medaka, *Oryzias latipes*. *Genetics Research, 73*(3), 225-231.

McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature, 351*(6328), 652-654.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research, 20*(9), 1297-1303.

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

Meisel, R. P., Malone, J. H., and Clark, A. G. (2012). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research, 22*(7), 1255-1265.

Meredith, R. W., Pires, M. N., Reznick, D. N., and Springer, M. S. (2011). Molecular phylogenetic relationships and the coevolution of placentotrophy and superfetation in *Poecilia* (Poeciliidae: Cyprinodontiformes). *Molecular Phylogenetics and Evolution, 59*(1), 148-157.

Metta, M., Gudavalli, R., Gibert, J.-M., and Schlötterer, C. (2006). No accelerated rate of protein evolution in male-biased *Drosophila pseudoobscura* genes. *Genetics, 174*(1), 411-420.

Miura, I. (2007). An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. *Sexual Development, 1*(6), 323-331.

Moore, J. C., and Pannell, J. R. (2011). Sexual selection in plants. *Current Biology, 21*(5), R176-182.

Mořkovský, L., Storchová, R., Plachý, J., Ivánek, R., Divina, P., and Hejnar, J. (2010). The chicken Z chromosome is enriched for genes with preferential expression in ovarian somatic cells. *Journal of Molecular Evolution, 70*(2), 129-136.

Morris, J., Darolti, I., Bloch, N. I., Wright, A. E., and Mank, J. E. (2018). Shared and species-specific patterns of nascent Y chromosome evolution in two guppy species. *Genes, 9*(5), 238.

Müller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist, 66*(703), 118-138.

Mullon, C., Wright, A. E., Reuter, M., Pomiankowski, A., and Mank, J. E. (2015). Evolution of dosage compensation under sexual selection differs between X and Z chromosomes. *Nature Communications, 6*, 7720.

Muyle, A., Kafer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G. A. (2016). SEX-DETector: A probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution, 8*(8), 2530-2543.

Muyle, A., Shearn, R., and Marais, G. A. B. (2017). The evolution of sex chromosomes and dosage compensation in plants. *Genome Biology and Evolution, 9*(3), 627-645.

Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., and Marais, G. A. (2012). Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biology, 10*(4), e1001308.

Nanda, I., Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., and Schartl, M. (2014). Sex chromosome polymorphism in guppies. *Chromosoma, 123*(4), 373-383.

Natri, H. M., Shikano, T., and Merila, J. (2013). Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. *Molecular Biology and Evolution, 30*(5), 1131-1144.

Naurin, S., Hansson, B., Hasselquist, D., Kim, Y.-H., and Bensch, S. (2011). The sex-biased brain: sexual dimorphism in gene expression in two species of songbirds. *BMC Genomics, 12*(1), 37.

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., and Monéger, F. (2004). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biology, 3*(1), e4.

Nielsen, S. V., Banks, J. L., Diaz, R. E., Trainor, P. A., and Gamble, T. (2018). Dynamic sex chromosomes in Old World chameleons (Squamata: Chamaeleonidae). *Journal of Evolutionary Biology, 31*(4), 484-490.

Ohno, S. (1967). *Sex chromosomes and sex-linked genes*. New York: Springer.

Oliver, B., and Parisi, M. (2004). Battle of the Xs. *Bioessays, 26*(5), 543-548.

Ometto, L., Shoemaker, D., Ross, K. G., and Keller, L. (2010). Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Molecular Biology and Evolution, 28*(4), 1381-1392.

Orr, H. A., and Kim, Y. (1998). An adaptive hypothesis for the evolution of the Y chromosome. *Genetics, 150*(4), 1693-1698.

Orzack, S. H., Sohn, J. J., Kallman, K. D., Levin, S. A., and Johnston, R. (1980). Maintenance of the three sex chromosome polymorphism in the platyfish, *Xiphophorus maculatus*. *Evolution, 34*(4), 663-672.

Otto, S. P., Scott, M. F., and Immler, S. (2015). Evolution of haploid selection in predominantly diploid organisms. *Proceedings of the National Academy of Sciences of the United States of America, 112*(52), 15952-15957.

Pál, C., Papp, B., and Hurst, L. D. (2001). Highly expressed genes in yeast evolve slowly. *Genetics, 158*(2), 927-931.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*(2), 289-290.

Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S., and Oliver, B. (2003). Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science, 299*(5607), 697-700.

Parker, G. A., and Partridge, L. (1998). Sexual conflict and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1366), 261-274.

Payer, B., and Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics, 42*, 733-772.

Peeters, L., and Totland, Ø. (1999). Wind to insect pollination ratios and floral traits in five alpine *Salix* species. *Canadian Journal of Botany, 77*(4), 556-563.

Pennell, M. W., Kirkpatrick, M., Otto, S. P., Vamosi, J. C., Peichel, C. L., Valenzuela, N., and Kitano, J. (2015). Y fuse? Sex chromosome fusions in fishes and reptiles. *PLoS Genetics, 11*(5), e1005237.

Pennell, M. W., Mank, J. E., and Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology, 27*(19), 3950-3963.

Perrin, N. (2009). Sex reversal: a fountain of youth for sex chromosomes? *Evolution, 63*(12), 3043-3049.

Perry, J. C., Harrison, P. W., and Mank, J. E. (2014). The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Molecular Biology and Evolution, 31*(5), 1206-1219.

Perry, J. C., and Rowe, L. (2015). The evolution of sexually antagonistic phenotypes. *Cold Spring Harbor Perspectives in Biology, 7*(6), a017558.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology, 33*(3), 290-295.

Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., and Marais, G. A. (2012). Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proceedings of the National Academy of Sciences of the United States of America, 109*(14), 5346-5351.

Petes, T. D., and Hill, C. W. (1998). Recombination between repeated genes in microorganisms. *Annual Review of Genetics, 22*(1), 147-168.

Pointer, M. A., Harrison, P. W., Wright, A. E., and Mank, J. E. (2013). Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genetics, 9*(8), e1003697.

Pollux, B., Meredith, R., Springer, M., Garland, T., and Reznick, D. (2014). The evolution of the placenta drives a shift in sexual selection in livebearing fish. *Nature, 513*(7517), 233.

Pool, J. E., and Nielsen, R. (2007). Population size changes reshape genomic patterns of diversity. *Evolution: International Journal of Organic Evolution, 61*(12), 3001-3006.

Pucholt, P., Wright, A. E., Liu Conze, L., and Mank, J. E. (2017). Recent sex chromosomes divergence despite ancient dioecy in the willow *Salix viminalis*. *Molecular Biology and Evolution, 34*(8), 1991-2001.

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

Quinn, A., Juneja, P., and Jiggins, F. M. (2014). Estimates of allele-specific expression in Drosophila with a single genome sequence and RNA-seq data. *Bioinformatics, 30*(18), 2603-2610.

R Core Team. (2015). R: A language and environment for statistical computing.

Rahman, A., Hallgrimsdottir, I., Eisen, M., and Pachter, L. (2018). Association mapping from sequencing reads using *k*-mers. *Elife, 7*, e32920.

Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D., and Hartl, D. L. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science, 300*(5626), 1742-1745.

Reichwald, K., Petzold, A., Koch, P., Downie, B. R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., and Bens, M. (2015). Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell, 163*(6), 1527-1538.

Reinius, B., Saetre, P., Leonard, J. A., Blekhman, R., Merino-Martinez, R., Gilad, Y., and Jazin, E. (2008). An evolutionarily conserved sexual signature in the primate brain. *PLoS Genetics, 4*(6), e1000100.

Reinke, V., San Gil, I., Ward, S., and Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development, 131*(2), 311-323.

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Jorunal of Botany, 101*(10), 1588-1596.

Rice, W., and Chippindale, A. (2001). Intersexual ontogenetic conflict. *Journal of Evolutionary Biology, 14*(5), 685-693.

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution, 38*(4), 735-742.

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution, 41*(4), 911-914.

Rice, W. R. (1992). Sexually antagonistic genes: experimental evidence. *Science, 256*(5062), 1436-1439.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., WGS500 Consortium, Wilkie, A. O., McVean, G., and Lunter, G. (2014). Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics, 46*(8), 912-918.

Roberts, N. B., Juntti, S. A., Coyle, K. P., Dumont, B. L., Stanley, M. K., Ryan, A. Q., Fernald, R. D., and Roberts, R. B. (2016). Polygenic sex determination in the cichlid fish Astatotilapia burtoni. *BMC Genomics, 17*(1), 835.

Roberts, R. B., Ser, J. R., and Kocher, T. D. (2009). Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science, 326*(5955), 998-1001.

Robinson, K. M., Delhomme, N., Mahler, N., Schiffthaler, B., Onskog, J., Albrectsen, B. R., Ingvarsson, P. K., Hvidsten, T. R., Jansson, S., and Street, N. R. (2014). *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. *BMC Plant Biology, 14*, 276.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Roesti, M., Moser, D., and Berner, D. (2013). Recombination in the threespine stickleback genome--patterns and consequences. *Molecular Ecology, 22*(11), 3014-3027.

Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., and Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). *PLoS Genetics, 5*(2), e1000391.

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., and Bird, C. P. (2005). The DNA sequence of the human X chromosome. *Nature, 434*(7031), 325.

Rosser, Z. H., Balaresque, P., and Jobling, M. A. (2009). Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. *The American Journal of Human Genetics, 85*(1), 130-134.

Russell, J., and Pannell, J. (2015). Sex determination in dioecious *Mercurialis annua* and its close diploid and polyploid relatives. *Heredity, 114*(3), 262-271.

Russo, C., Rocco, L., Stingo, V., Aprea, G., and Odierna, G. (2009). A cytogenetic analysis of *Gambusia holbrooki* (Cyprinodontiformes, Poecilidae) from the River Sarno. *Italian Journal of Zoology, 66*(3), 291-296.

Sætre, G. P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C., and Syvänen, A. C. (2003). Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1510), 53-59.

Sanderson, B. J., Wang, L., Tiffin, P., Wu, Z., and Olson, M. S. (2019). Sex‐biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*. *New Phytologist, 221*(1), 527-539.

Sandkam, B., Young, C. M., and Breden, F. (2015). Beauty in the eyes of the beholders: colour vision is tuned to mate preference in the Trinidadian guppy (*Poecilia reticulata*). *Molecular Ecology, 24*(3), 596-609.

Santos, E. M., Kille, P., Workman, V. L., Paull, G. C., and Tyler, C. R. (2008). Sexually dimorphic gene expression in the brains of mature zebrafish. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 149*(3), 314-324.

Schultheis, C., Bohne, A., Schartl, M., Volff, J. N., and Galiana-Arnoux, D. (2009). Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sexual Development, 3*(2-3), 68-77.

Ser, J. R., Roberts, R. B., and Kocher, T. D. (2010). Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution, 64*(2), 486-501.

Sessions, S. K., Bizjak, M. L., Green, D. M., Trifonov, V., and Ferguson-Smith, M. (2016). Evidence for sex chromosome turnover in protein salamanders. *Cytogenetic Genome Research, 148*(4), 305-313.

Sessions, S. K., and Kezer, J. (2012). Evolutionary cytogenetics of Bolitoglossine salamanders. *Amphibian cytogenetics and evolution*, 89-130.

Sharma, E., Kunstner, A., Fraser, B. A., Zipprich, G., Kottler, V. A., Henz, S. R., Weigel, D., and Dreyer, C. (2014). Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genomics, 15*(1), 400.

Sharp, P. M., and Li, W. H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research, 15*(3), 1281-1295.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Hillier, H. S. C., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., and Chinwalla, A. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature, 423*(6942), 825-837.

Slattery, J. P., Murphy, W. J., and O'Brien, S. J. (2000). Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. *Molecular Biology and Evolution, 17*(5), 825-829.

Slotte, T., Bataillon, T., Hansen, T. T., St Onge, K., Wright, S. I., and Schierup, M. H. (2011). Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome Biology and Evolution, 3*, 1210-1219.

Small, C. M., Carney, G. E., Mo, Q., Vannucci, M., and Jones, A. G. (2009). A microarray analysis of sex- and gonad-biased gene expression in the

zebrafish: evidence for masculinization of the transcriptome. *BMC Genomics, 10*(1), 579.

Smith, J. M. (1991). Theories of sexual selection. *Trends in Ecology & Evolution, 6*(5), 146-151.

Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research, 23*(1), 23-35.

Snow, A. A., and Spira, T. P. (1996). Pollen-tube competition and male fitness in *Hibiscus mosceutos*. *Evolution, 50*, 1866-1870.

Sola, J. L., Rossi, A. R., Iaselli, V., Rasch, E. M., and Monaco, P. J. (1992). Cytogenetics of bisexual/unisexual species of *Poecilia*. *Cytogenetic Genome Research, 60*(1), 229-235.

Sola, L., Bressanello, S., Rasch, E. M., and Monaco, P. J. (1993). Cytogenetics of bisexual/unisexual species of *Poecilia*. IV. Sex chromosomes, sex chromatin composition and Ag-NOR polymorphisms in *Poecilia latipinna*: a population from Mexico. *Heredity, 70*(1), 67.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313.

Stevenson, K. R., Coolon, J. D., and Wittkopp, P. J. (2013). Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC Genomics, 14*(1), 536.

Stock, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., and Maddalena, T. (2011). Ever-young sex chromosomes in European tree frogs. *PLoS Biology, 9*, e1001062. doi:0.1371/journal.pbio.1001062

Stock, M., Savary, R., Betto-Colliard, C., Biollay, S., Jourdan-Pineau, H., and Perrin, N. (2013). Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). *Journal of Evolutionary Biology, 26*(3), 674-682.

Stoletzki, N., and Eyre-Walker, A. (2011). Estimation of the neutrality index. *Molecular Biology and Evolution, 28*(1), 63-70.

Storchova, R., and Divina, P. (2006). Nonrandom representation of sex-biased genes on chicken Z chromosome. *Journal of Molecular Evolution, 63*(5), 676-681.

Sturgill, D., Zhang, Y., Parisi, M., and Oliver, B. (2007). Demasculinization of X chromosomes in the *Drosophila* genus. *Nature, 450*(7167), 238.

Sun, L., Johnson, A. F., Li, J., Lambdin, A. S., Cheng, J., and Birchler, J. A. (2013). Differential effect of aneuploidy on the X chromosome and genes with sex-biased expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America, 110*(41), 16514-16519.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjodin, A., Van De Peer, Y., Jansson, S., Hvidsten, T. R., and Street, N. R. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytologist, 208*, 1149-1156.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics, 22*(12), 1540-1542.

Szovenyi, P., Devos, N., Weston, D. J., Yang, X., Hock, Z., Shaw, J. A., Shimizu, K. K., McDaniel, S. F., and Wagner, A. (2014). Efficient purging of deleterious mutations in plants with haploid selfing. *Genome Biology and Evolution, 6*(5), 1238-1252.

Szovenyi, P., Ricca, M., Hock, Z., Shaw, J. A., Shimizu, K. K., and Wagner, A. (2013). Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. *Molecular Biology and Evolution, 30*(8), 1929-1939.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics, 123*(3), 597-601.

Traut, W., and Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Research, 9*(8), 659-672.

Tripathi, N., Hoffmann, M., Weigel, D., and Dreyer, C. (2009a). Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (*Poecilia reticulata*). *Genetics, 182*(1), 365-374.

Tripathi, N., Hoffmann, M., Willing, E. M., Lanz, C., Weigel, D., and Dreyer, C. (2009b). Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 276*(1665), 2195-2208.

Trombetta, B., Cruciani, F., Underhill, P. A., Sellitto, D., and Scozzari, R. (2009). Footprints of X-to-Y gene conversion in recent human evolution. *Molecular Biology and Evolution, 27*(3), 714-725.

Uebbing, S., Künstner, A., Mäkinen, H., and Ellegren, H. (2013). Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biology and Evolution, 5*(8), 1555-1566.

Urrutia, A. O., and Hurst, L. D. (2001). Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics, 159*(3), 1191-1199.

Van Doorn, G. S., and Kirkpatrick, M. (2010). Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics, 186*(2), 629-645.

Veller, C., Muralidhar, P., Constable, G. W., and Nowak, M. A. (2017). Drift-induced selection between male and female heterogamety. *Genetics, 207*(2), 711-727.

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

Vicoso, B., and Bachtrog, D. (2015). Numerous transitions of sex chromosomes in Diptera. *PLoS Biology, 13*(4), e1002078.

Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics, 7*(8), 645-653.

Vicoso, B., and Charlesworth, B. (2009). The deficit of male-biased genes on the *D. melanogaster* X chromosome is expression-dependent: A consequence of dosage compensation? *Journal of Molecular Evolution, 68*(5), 576-583.

Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S., and Bachtrog, D. (2013a). Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Bioogy, 11*(8), e1001643.

Vicoso, B., Kaiser, V. B., and Bachtrog, D. (2013b). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedins of the National Academy of Sciences of the United States of America, 110*(16), 6453-6458.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. *Proceedins of the National Academy of Sciences of the United States of America, 103*(9), 3220-3225.

Volff, J. N., and Schartl, M. (2001). Variability of genetic sex determination in poeciliid fishes. *Genetica, 111*(1-3), 101-110.

Wang, J., Na, J.-K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J. E., and Zhang, W. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedins of the National Academy of Sciences of the United States of America, 109*(34), 13710-13715.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature, 493*(7434), 664-668.

White, M. A., Kitano, J., and 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.

Whittle, C. A., and Extavour, C. G. (2017). Rapid evolution of ovarian-biased genes in the yellow fever mosquito (*Aedes aegypti*). *Genetics, 206*(4), 2119-2137.

Whittle, C. A., and Johannesson, H. (2013). Evolutionary dynamics of sex-biased genes in a hermaphrodite fungus. *Molecular Biology and Evolution, 30*(11), 2435-2446.

Whittle, C. A., and Johnston, M. O. (2003). Male-biased transmission of deleterious mutations to the progeny in *Arabidopsis thaliana*. *Proceedings of the National Acadey of Science of the United States of America, 100*(7), 4055-4059.

Whittle, C. A., Malik, M. R., and Krochko, J. E. (2007). Gender-specific selection on codon usage in plant genomes. *BMC Genomics, 8*(1), 169.

Winge, Ö. (1927). The location of eighteen genes in *Lebistes reticulatus*. *Journal of Genetics, 18*, 1-43.

Wright, A. E., Darolti, I., Bloch, N. I., Oostra, V., Sandkam, B., Buechel, S. D., Kolm, N., Breden, F., Vicoso, B., and Mank, J. E. (2017). Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nature Communications, 8*, 14251.

Wright, A. E., Darolti, I., Bloch, N. I., Oostra, V., Sandkam, B. A., Buechel, S. D., Kolm, N., Breden, F., Vicoso, B., and Mank, J. E. (2019a). On the power to detect rare recombination events. *Proceedings of the National Academy of Sciences of the United States of America, 116*(26), 12607-12608.

Wright, A. E., Dean, R., Zimmer, F., and Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications, 7*, 12087.

Wright, A. E., Fumagalli, M., Cooney, C. R., Bloch, N. I., Vieira, F. G., Buechel, S. D., Kolm, N., and Mank, J. E. (2018). Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters, 2*(2), 52-61.

Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A., and Mank, J. E. (2014). Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution, 68*(11), 3281-3295.

Wright, A. E., Harrison, P. W., Zimmer, F., Montgomery, S. H., Pointer, M. A., and Mank, J. E. (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Molecular Ecology, 24*(6), 1218-1235.

Wright, A. E., Moghadam, H. K., and Mank, J. E. (2012). Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics, 192*(4), 1433-1445.

Wright, A. E., Rogers, T. F., Fumagalli, M., Cooney, C. R., and Mank, J. E. (2019b). Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict. *Molecular Ecology, 28*(11), 2860-2871.

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene, 87*(1), 23-29.

Xu, L., Sin, S. Y. W., Grayson, P., Janes, D. E., Edwards, S. V., and Sackton, T. B. (2018). Evolutionary dynamics of sex chromosomes of palaeognathous birds. *bioRxiv,* 295089.

Yang, L. (2016). Both male-biased and female-biased genes evolve faster in fish genomes. *Genome Biology and Evolution, 8*(11), 1-22.

Yang, X., Schadt, E. E., Wang, S., Wang, H., Arnold, A. P., Ingram-Drake, L., Drake, T. A., and Lusis, A. J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Research, 16*(8), 995-1004.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*(8), 1586-1591.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution, 17*(1), 32-43.

Zemp, N., Minder, A., and Widmer, A. (2014). Identification of internal reference genes for gene expression normalization between the two sexes in dioecious white Campion. *PLoS ONE, 9*(3), e92893.

Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G. A., and Widmer, A. (2016). Evolution of sex-biased gene expression in a dioecious plant. *Nature Plants, 2*(11), 16168.

Zemp, N., Tavares, R., and Widmer, A. (2015). Fungal infection induces sex-specific transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene latifolia*. *PLoS Genetics, 11*(10), e1005536.

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

Zhao, M., Zha, X.-F., Liu, J., Zhang, W.-J., He, N.-J., Cheng, D.-J., Dai, Y., Xiang, Z.-H., and Q-Y, X. (2011). Global expression profile of silkworm genes from larval to pupal stages: toward a comprehensive understanding of sexual differences. *Insect Science, 18*(6), 607-618.

Zhou, Q., and Bachtrog, D. (2012). Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science, 337*(6092), 341-345.

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E. D., Gilbert, M. T. P., and Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science, 346*(6215), 1246338.

Zimmer, F., Harrison, P. W., Dessimoz, C., and Mank, J. E. (2016). Compensation of dosage-sensitive genes on the chicken Z chromosome. *Genome Biology and Evolution, 8*(4), 1233-1242.

Zluvova, J., Zak, J., Janousek, B., and Vyskot, B. (2010). Dioecious *Silene latifolia* plants show sexual dimorphism in the vegetative stage. *BMC Plant Biology, 10*, 208.