EARLY STAGES OF SEX CHROMOSOME EVOLUTION

by

Andrea Mrnjavac

November, 2024

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

Committee in charge:
Krzysztof Pietrzak, Chair
Beatriz Vicoso
Nick Barton
Andrea Betancourt

The thesis of Andrea Mrnjavac, titled *Early stages of sex chromosome evolution*, is approved by:

Supervisor: Beatriz Vicoso, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Nick Barton, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Andrea Betancourt,University of Liverpool, Liverpool, UK

Signature: _____

Defense Chair: Krzysztof Pietrzak, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

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

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

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

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

Signature: _____

Andrea Mrnjavac

November, 2024

Signed page is on file

Abstract

Sex chromosomes and autosomes exhibit very different evolutionary dynamics. The Y chromosome usually degenerates, leaving many X-linked loci hemizygous in males. Since recessive X-linked mutations are always exposed to selection in males, selection is more efficient on the X chromosome than on autosomes on recessive mutations, leading to faster adaptation on the X chromosome than other genomic regions, if beneficial mutations are on average recessive (known as the Faster-X effect). In the presence of the functional, but non-recombining gametolog on the Y (as is often the case in young non-recombining regions), recessive mutations are sheltered from selection on the X chromosome. We model this scenario and show that the efficiency of selection is reduced on diploid X loci due to sheltering by the Y chromosome. Reduced efficiency of selection leads to slower adaptation and increased accumulation of deleterious mutations (Slower-X effect). We extended this model to explore the effect of sex-specific selection on degeneration of sex chromosomes, showing theoretically that male-limited genes degenerate on the X chromosome and female-biased genes degenerate on the Y chromosome. This prediction depends on the effective population size and the mutation rate, explaining the variety of sex chromosome degeneration patterns observed in nature.

To test for direct evidence of a Slower-X (or Slower-Z) effect, we analyzed the ZW sex chromosomes of the flatworm *Schistosoma japonicum*, which have a very young non-recombining region with non-degenerated W. Diploid Z-linked genes have higher ratios of non-synonymous to synonymous polymorphisms than autosomal genes, supporting reduced efficiency of selection on the diploid Z region. These results provide evidence of sheltering by the W chromosome, a mechanism that could contribute to Z (X) chromosome degeneration, and illustrate contrasting evolutionary patterns in old and young sex chromosome regions. In addition, genes with sex-specific patterns of expression show opposite patterns of selection in the young (diploid) and old (hemizygous) Z, showing the complex manner in which sex-specific selection shapes the evolutionary patterns of sex chromosomes.

Acknowledgments

I would like to thank Beatriz Vicoso, Ksenia Khudiakova, Nick Barton, Tim Connallon, Vicoso group, Marwan Elkrewi, Filip Ruzicka, Christelle Fraisse, Daniel Jeffries, Melissa Toups, Andrea Betancourt, ISTA Scientific Computing Unit and IT Support, Grants Office and Christine Ostermann.

About the Author

Andrea Mrnjavac completed her Bsc and Msc in Molecular Biology at the University of Zagreb and did an internship at the University of Kiel, before joining ISTA in September 2019. Her main research interests include molecular evolution and evolutionary genomics. During her PhD she published chapters of her thesis in journals such as Evolution Letters and Genetics, presented her work in several conferences, such as ESEB, Evolution and PopGroup, where she was awarded the Best Student Talk award, presented her work in a seminar at UCL and went for a research stay with Tim Connallon at Monash University.

List of Collaborators and Publications

Collaborators

- Beatriz Vicoso (ISTA) (Supervised Chapter 2 and Chapter 4, and contributed to the analysis in Chapter 3)
- Ksenia A Khudiakova (ISTA) (contributed to the analysis in Chapter 2)
- Nicholas H Barton (ISTA) (contributed to the analysis in Chapter 2)
- Tim Connallon (Monash University, Australia) (Supervised Chapter 3 and contributed to the analysis in Chapter 3)

Publications

Mrnjavac, A., K. A. Khudiakova, N. H. Barton, and B. Vicoso. 2023. Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. Evolution Letters 7:4–12.

Mrnjavac, A., and B. Vicoso. 2024. Evidence of a Slower-Z effect in Schistosoma japonicum. bioRxiv.

Mrnjavac, A., B. Vicoso and T. Connallon. Effects of sheltering on sex chromosome degeneration and sex-biased gene content evolution (in revision)

Table of Contents

Chapter 1: Introduction

Sex chromosome evolution: the canonical model

Sexual reproduction is widespread in eukaryotes as it allows more efficient adaptation (Gillespie, 2004, Chapter 7). Sex can be determined environmentally or genetically (Bachtrog et al., 2014). Genetic determination of sex by sex chromosomes was discovered by Nettie Stevens in 1905, following the observation that male mealworms have one chromosome that is much smaller than others, while females carried equally sized chromosomes (Stevens, 1905, reviewed in Furman et al., 2020). Today we know that sex chromosomes have evolved independently in many different taxa but share many common features. The canonical model of sex chromosome evolution includes 4 steps (reviewed recently in Furman et al., 2020, Zhu et al., 2024). 1. Sex chromosomes originate from a standard pair of autosomes when one of the autosomes acquires a sex-determining (SD) locus (Bull 1983, Bachtrog et al., 2014, Vicoso 2019). 2. Recombination suppression between the X and the Y usually evolves in the surrounding region and extends along the chromosome (reviewed in Ponnikas et al., 2018, Charlesworth, 2023, Jay et al., 2024). 3. In the absence of recombination, selection is less efficient and the Y chromosome degenerates (Bachtrog 2013, Charlesworth, 2021). 4. Dosage compensation mechanisms are selectively favoured as dosage sensitive genes in the Y degenerate. In some cases, dosage compensation can quickly spread along the X chromosome, accelerating degeneration of Y the (reviewed in Wright et al.,2016).

The usual steps in sex chromosome evolution are summarised in Figure 1. Our example illustrates the male heterogametic system, but female heterogametic sex

chromosomes (ZW) follow the same trajectory. In the following sections, we summarise some key steps in this classical model and the evidence supporting them. In the Discussion (chapter 5), we further highlight several recent extensions to this model that, along with our results, are re-shaping our broad understanding of sex chromosome evolution.

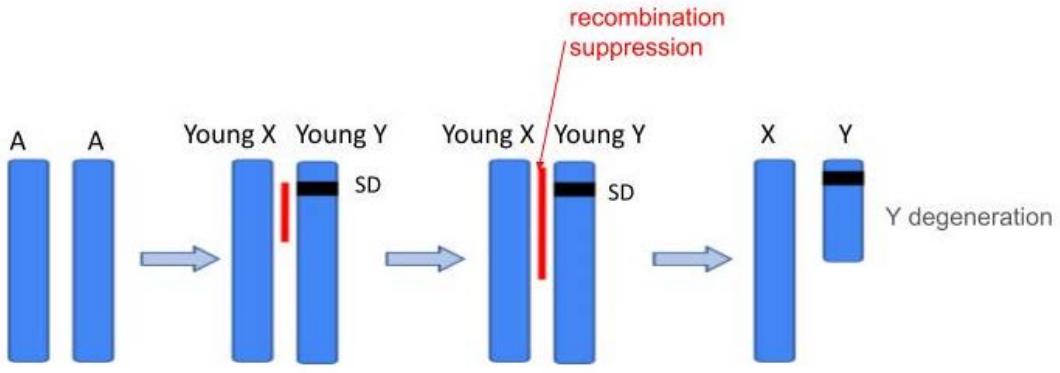


Figure 1. The usual steps in sex chromosome evolution: one of the autosomes (A) from a pair acquires a sex determining locus (SD). This is usually followed by recombination suppression between the X and the Y, which progressively evolves along the chromosome. The Y chromosome does not recombine and consequently degenerates.

Processes of Y chromosome degeneration

Once recombination is suppressed between X and Y chromosomes, X chromosomes can still recombine in females, while Y chromosomes do not recombine at all, and their effective population size is reduced to one quarter of the autosomal effective population size. Selection is less efficient in the absence of recombination (Felsenstein, 1974), which makes adaptation on the Y slower and accumulation of deleterious mutations by Muller's ratchet and genetic hitchhiking faster (reviewed in Bachtrog, 2013), leading to exponential gene loss on the Y (Bachtrog, 2008). Muller

first proposed a scenario of inevitable genome degeneration in the absence of recombination (Muller, 1964). Since without recombination selection cannot purge deleterious mutations fast enough, with a sufficiently high mutation rate, after some time, all existing haplotypes will have at least one deleterious mutation, i.e. there will have been a click of the “ratchet”. Under this mechanism, all the haplotypes will over time accumulate more and more deleterious mutations, leading to progressive degeneration and eventually, extinction (Gillespie, 2004, Chapter 7). Furthermore, deleterious mutations can be fixed in the population if they are linked to a beneficial mutation that provides a net advantage for the whole haplotype (hitchhiking, Rice, 1987). In a similar way, adaptive evolution will be counteracted by purifying selection on linked deleterious mutations (the so-called “Ruby in the rubbish” effect, Peck, 1994, Orr and Kim, 1998). Maladaptation and accumulation of deleterious mutations and transposable elements lead to gene silencing and eventual gene loss on the Y chromosome. Old heteromorphic Y chromosomes in mammals and *Drosophila melanogaster* have lost the majority of their ancestral gene content, apart from a subset of genes involved in male fertility, and some key dosage-sensitive genes in the case of mammals (Bachtrog, 2013). On the other hand, there are examples of younger Y chromosomes that are only partially degenerated (i.e. they still retain many of their ancestral genes), such as the young Y chromosome in *Drosophila miranda* (reviewed in Bachtrog, 2013) and several plant Y chromosomes (reviewed in Charlesworth, 2021). These intermediate stages of degeneration have yielded important insights into this process, showing for instance that gene loss is not random, and that genes with male-biased expression are preferentially retained (Kaiser et al., 2011).

Dosage compensation of the X chromosome

The degeneration of homologous genes on the Y chromosome results in X-linked genes being present in only one copy in males. Since gene copy number and gene expression dosage are correlated, this should lead to a deficit of X-linked gene products in males. Being aware of the Y chromosome degeneration in *Drosophila* (Muller, 1914), over 100 years ago *Drosophila* geneticists wondered how X-linked genes could have the same phenotype in males and females (reviewed in Gartler, 2014). Muller (1932) showed evidence of hyperexpression of the *Drosophila* X chromosome in males, and termed the phenomenon as “dosage compensation”. Ohno (1967) studied dosage compensation in mammals, where one of the X chromosomes is inactivated in females, and proposed a classic two-step hypothesis for the evolution of global dosage compensation. In the first step, degeneration of the Y chromosome poses a selective pressure for the evolution of upregulation of X-linked genes (assuming the ancestral gene expression is under stabilising selection). However, this upregulation causes the X chromosome to be overexpressed in females, which leads to the evolution of the X chromosome inactivation in females (reviewed in Gu and Walters, 2017, Zhu et al., 2024). Studies on non-model species have shown that global dosage compensation is not the rule. In some species, such as nematodes, and Lepidoptera, the X (or Z) chromosome has balanced expression in males and females, but is not dosage compensated to the ancestral expression level (Gu and Walters, 2017). Dosage compensation can also be partial, incomplete, and restricted to a few dosage-sensitive genes, such as in birds, snakes and schistosomes (reviewed in Gu and Walters, 2017, Furman et al., 2020). Interestingly, incomplete dosage compensation is more often found in female heterogametic systems. These different

responses to Y degeneration highlight the complexity of sex chromosome evolution, and the need to understand what forces are shaping both the early gene loss and the regulatory responses on the nascent XY pair.

Different evolutionary dynamics of the X chromosome and autosomes

Since the Y chromosome degenerates, many X-linked loci are present in only one copy in males, i.e., they are hemizygous. The effect of each mutation on X will be expressed in males, regardless of its dominance coefficient, so patterns of selection on the X differ from those in autosomes. Hemizygosity in males is thought to cause many peculiarities of the X chromosome. For instance, the X chromosome is expected to be enriched for loci underlying sexual dimorphism. Rice (1984) showed that sexually antagonistic mutations increase in frequency more readily on the X chromosome than on the autosomes, if they are recessive and beneficial for males, and if they are dominant and beneficial for females. Sexual conflict is likely often resolved by the evolution of expression modifiers, which results in sex-limited expression (Rice, 1984). A prediction of this model is therefore an excess of either genes primarily expressed in males ("male-biased genes", if the mutations shaping these genes are overall recessive) or female-biased genes (if mutations are dominant) on the X. While it is unclear if sexually antagonistic mutations are dominant or recessive, sexual antagonism has been invoked to account for the differential accumulation of both male- and female-biased genes on the X and autosomes (Rice, 1984). For example, male-biased genes are underrepresented on the *Drosophila* X chromosome, but overrepresented on the mammalian X (Gurbich and Bachtrog, 2008).

Furthermore, because of hemizygosity of X-linked genes in males, new recessive mutations on the X are under more efficient selection than on the autosomes, as they are always exposed to selection in males. With equal sex ratios, the effective population size of the X chromosome is ¾ the autosomal effective population size. However, despite reduced effective population size, strong haploid selection in males should lead to faster adaptive evolution on the X chromosome than autosomes for (partially) recessive mutations (Charlesworth et al., 1987, Vicoso and Charlesworth, 2006), especially for male-biased genes (Figure 2). Faster evolutionary rates on the X chromosome, compared to the autosomes, is known as the Faster-X effect. These models assume that the effective population size (N_e) of the X chromosome is ¾ the autosomal effective population size. As selection is more efficient in larger populations, higher N_{eX}/N_{eA} ratio relaxes the conditions for the Faster-X evolution, while lower Ne_X/Ne_A ratio decreases the probability of faster adaptive evolution on the X chromosome (Vicoso and Charlesworth, 2009). These previously mentioned models assume that a new beneficial mutation is necessary for adaptation. However, adaptation can occur from standing variation, for example, if the environment changes and previously neutral or deleterious alleles become beneficial. Since autosomes usually have larger effective population size than the X chromosome, and therefore, higher neutral diversity, adaptation occurring from standing variation should be faster on autosomes than the X chromosome regardless of the dominance (Orr and Betancourt, 2001). Finally, Connallon et al. (2012) developed a model that unifies adaptation from the standing variation and adaptation from new mutations, and showed that the Faster-X effect might be smaller in larger populations, where there is more opportunity for the adaptation from standing

variation. This might explain some of the differences in Faster-X patterns in different taxa (Connallon et al., 2012).

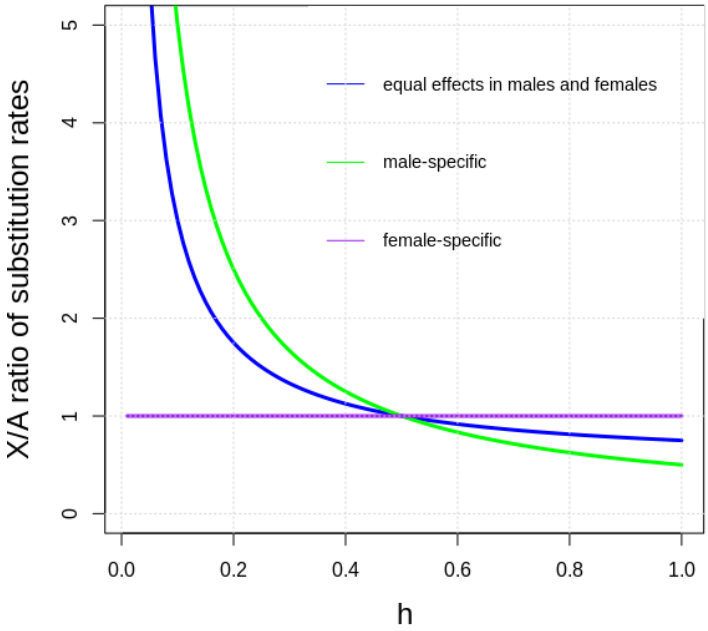


Figure 2. Faster-X effect, adapted from the equations from Charlesworth et al., 1987 and Vicoso and Charlesworth, 2006. X/A ratio of substitution rates for beneficial mutations as a function of dominance (h). The model assumes $N_{eX}/N_{eA}=3/4$, equal mutation rates in males and females, complete dosage compensation and hemizygosity of the X chromosome in males.

Young sex chromosomes and sex-linked regions

There is substantial empirical evidence for sexualization of the X chromosome gene content and the Faster-X effect (Ellegren, 2011, Gurbich and Bachtrog, 2008, Parsch and Ellegren, 2013, Meisel and Connallon, 2013, Charlesworth et al., 2018). However, empirical and theoretical studies have been biased so far towards well differentiated sex chromosomes with degenerated Y, which can be easily detected through cytogenetic or genomic coverage analyses. Homomorphic sex chromosomes

are harder to identify: population data is typically needed to identify very young non-recombining regions, with non-degenerated Y (reviewed in Vicoso, 2019). Very young non-differentiated sex-linked regions have the same gene content on the X and the Y, and homologous regions in X and Y differ only in independently accumulated alleles since recombination suppression. Genetic differentiation between X and Y can be estimated from F_{st} between male and female populations: F_{st} is expected to be elevated in the non-recombining but non-degenerated region. Recently, such data has become available, and more young sex-linked regions are being identified (Elkrewi et al., 2021, Elkrewi et al., 2022). Surprisingly, young X-linked regions sometimes appeared to have unusual evolutionary dynamics, including an excess of gene loss, despite not being hemizygous (Nozawa et al., 2016). While some hypotheses were verbally put forward for why this might be, the evolutionary dynamics of young sex chromosomes and young sex-linked regions had been largely theoretically and empirically unexplored, and this motivated the work described. In particular, during my PhD, I explored the early stages of sex chromosomes evolution both theoretically, thereby providing a novel framework for understanding some of these earlier results, and empirically.

Summary of research

In young sex chromosomes, or young sex-linked regions, many X-linked loci have a functional, but non-recombining gametolog on the Y, i.e., such X-linked loci are diploid in males and not hemizygous. We first modelled this scenario and showed that the efficiency of selection is reduced on such diploid X loci due to sheltering by the Y chromosome. As sheltering occurs in males, male-important genes are mostly affected

by reduced efficiency of selection, which leads to slower adaptation and increased accumulation of deleterious mutations on the X chromosome (Slower-X effect) (Chapter 2).

We extended this model to explore the effect of sheltering in the presence of sex-specific selection on the degeneration of sex chromosomes. We showed theoretically that sheltering can cause male-limited genes to degenerate on the X chromosome while female-biased genes degenerate on the Y chromosome (Chapter 3). While sheltering was the first hypothesis put forward to account for Y degeneration, follow up theory that did not consider sex-specific selection brought its role into question. Our extended model revives this classic theory and shows that it can in fact shape the gene content of both the Y and of the X.

To empirically test for direct evidence of Slower-X (or Slower-Z) effect, we analysed the ZW sex chromosomes of the flatworm *Schistosoma japonicum*, which have a very young non-recombining region with non-degenerated W (Elkrewi et al., 2021, Xu et al., 2023). Female-biased genes in the diploid Z region show signatures of relaxed purifying selection, compared to genes in the hemizygous Z region and autosomes. These results are consistent with theoretical predictions and provide empirical evidence of sheltering by the W chromosome, a mechanism that could contribute to Z (X) chromosome degeneration, and illustrate contrasting evolutionary patterns in old and young sex chromosome regions (Chapter 4).

References

Bachtrog, D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. Genetics 179:1513–1525.

Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14:113–124.

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, and T. T. of S. Consortium. 2014. Sex determination: why so many ways of doing it? PLOS Biology 12:e1001899.

Bull, J. J. 1983. Evolution of sex determining mechanisms. Benjamin/Cummings Publishing Company, Advanced Book Program.

Charlesworth, B., J. L. Campos, and B. C. Jackson. 2018. Faster-X evolution: Theory and evidence from *Drosophila*. Molecular Ecology 27:3753–3771.

Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. The American Naturalist 130:113–146.

Charlesworth, D. 2021. The timing of genetic degeneration of sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 376:20200093.

Charlesworth, D. 2023. Why and how do Y chromosome stop recombining? Journal of Evolutionary Biology 36:632–636.

Connallon, T., N. D. Singh, and A. G. Clark. 2012. Impact of genetic architecture on the relative rates of X versus autosomal adaptive substitution. Molecular Biology and Evolution 29:1933–1942.

Elkrewi, M., U. Khauratovich, M. A. Toups, V. K. Bett, A. Mrnjavac, A. Macon, C. Fraisse, L. Sax, A. K. Huylmans, F. Hontoria, and B. Vicoso. 2022. ZW sex-chromosome evolution and contagious parthenogenesis in *Artemia* brine shrimp. Genetics 222:iyac123.

Elkrewi, M., M. A. Moldovan, M. A. L. Picard, and B. Vicoso. 2021. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Molecular Biology and Evolution 38:5345–5358.

Ellegren, H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet 12:157–166.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.

Furman, B. L. S., D. C. H. Metzger, I. Darolti, A. E. Wright, B. A. Sandkam, P. Almeida, J. J. Shu, and J. E. Mank. 2020. Sex chromosome evolution: So many exceptions to the rules. Genome Biology and Evolution 12:750–763.

Gartler, S. M. 2014. A brief history of dosage compensation. J Genet 93:591–595.

Gillespie, J. H. 2004. Population Genetics. Johns Hopkins University Press.

Gurbich, T. A., and D. Bachtrog. 2008. Gene content evolution on the X chromosome. Current Opinion in Genetics & Development 18:493–498.

Jay, P., D. Jeffries, F. E. Hartmann, A. Véber, and T. Giraud. 2024. Why do sex chromosomes progressively lose recombination? Trends in Genetics 40:564-579.

Kaiser, V. B., Q. Zhou, and D. Bachtrog. 2011. Nonrandom gene loss from the *Drosophila miranda* Neo-Y chromosome. Genome Biology and Evolution 3:1329–1337.

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

Muller, H. 1932. Further studies on the nature and causes of gene mutations. Proceedings of the 6th International Congress of Genetics, 1932 213–255.

Muller, H. J. 1914. A gene for the fourth chromosome of *Drosophila*. Journal of Experimental Zoology 17:325–336.

Muller, H. J. 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1:2–9.

Nozawa, M., K. Onizuka, M. Fujimi, K. Ikeo, and T. Gojobori. 2016. Accelerated pseudogenization on the neo-X chromosome in *Drosophila miranda*. Nat Commun 7:13659.

Ohno, S. 1967. Sex chromosomes and sex-linked genes. Springer, Berlin, Heidelberg.

Orr, H. A., and A. J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics 157:875–884.

Orr, H. A., and Y. Kim. 1998. An adaptive hypothesis for the evolution of the Y chromosome. Genetics 150:1693–1698.

Parsch, J., and H. Ellegren. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet 14:83–87.

Peck, J. R. 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics 137:597–606.

Ponnikas, S., H. Sigeman, J. K. Abbott, and B. Hansson. 2018. Why do sex chromosomes stop recombining? Trends in Genetics 34:492–503.

Rice, W. R. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116:161–167.

Rice, W. R. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742.

Stevens, N. M. 1905. Studies in Spermatogenesis ... Carnegie Institution of Washington.

Vicoso, B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol 3:1632–1641.

Vicoso, B., and B. Charlesworth. 2009. Effective population size and the faster-X effect: an extended model. Evolution 63:2413–2426.

Vicoso, B., and B. Charlesworth. 2006. Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet 7:645–653.

Wright, A. E., R. Dean, F. Zimmer, and J. E. Mank. 2016. How to make a sex chromosome. Nat Commun 7:12087.

Xu, X., Y. Wang, C. Wang, G. Guo, X. Yu, Y. Dai, Y. Liu, G. Wei, X. He, G. Jin, Z. Zhang, Q. Guan, A. Pain, S. Wang, W. Zhang, N. D. Young, R. B. Gasser, D. P. McManus, J. Cao, Q. Zhou, and Q. Zhang. 2023. Chromosome-level genome assembly defines female-biased genes associated with sex determination and differentiation in the human blood fluke *Schistosoma japonicum*. Molecular Ecology Resources 23:205–221.

Zhu, Z., L. Younas, and Q. Zhou. 2024. Evolution and regulation of animal sex chromosomes. Nat Rev Genet 1–16.

Chapter 2: Slower-X: Reduced efficiency of selection in the early stages of X chromosome evolution

Andrea Mrnjavac, Ksenia A Khudiakova, Nicholas H Barton, Beatriz Vicoso

Evolution Letters, Volume 7, Issue 1, 1 February 2023, Pages 4–12,

https://doi.org/10.1093/evlett/qrac004

Author contributions: Beatriz Vicoso sparked the initial idea. BV and Andrea Mrnjavac designed the study. AM performed the analysis. Ksenia Khudiakova provided guidance on the methods. Nick Barton contributed to the analysis (equations on pages 22 and 23). AM wrote the first draft of the manuscript, BV and AM the current version. All the authors gave comments on the manuscript and approved the final version.

Abstract

Differentiated X chromosomes are expected to have higher rates of adaptive divergence than autosomes, if new beneficial mutations are recessive (the “faster-X effect”), largely because these mutations are immediately exposed to selection in males. The evolution of X chromosomes after they stop recombining in males, but before they become hemizygous, has not been well explored theoretically. We use the diffusion approximation to infer substitution rates of beneficial and deleterious mutations under such a scenario. Our results show that selection is less efficient on diploid X loci than autosomal and hemizygous X loci under a wide range of parameters. This “Slower-X” effect is stronger for genes affecting primarily (or only) male fitness, and for sexually antagonistic genes. These unusual dynamics suggest

that some of the peculiar features of X chromosomes, such as the differential accumulation of genes with sex-specific functions, may start arising earlier than previously appreciated.

Introduction

In many species, the sex of an individual is determined by a pair of sex chromosomes, such as the X and Y of mammals, or Z and W in the case of the female heterogametic system (Bachtrog et al., 2014). Although we focus on the more commonly studied XY case, the models discussed here also apply to ZW systems (by switching the sexes). Sex chromosomes arise from a pair of autosomes when one of them acquires a sex-determining gene (Wright et al., 2016, Vicoso, 2019). Sex chromosome evolution is often coupled with the suppression of recombination between X and Y chromosomes in the region surrounding the sex-determining gene. Recombination suppression can spread progressively along the chromosomes in a stepwise manner, creating distinct “evolutionary strata”, i.e. regions that stopped recombining at the same time, and which consequently have different levels of XY divergence (Lahn and Page, 1999, Ponnikas et al., 2018, Jefferies et al., 2021). Recombination suppression between X and Y reduces the efficiency of selection on the Y chromosome, leading to progressive gene loss on the Y (Engelstädter, 2008, Bachtrog, 2013). The resulting imbalance in gene copy number in males drives the evolution of dosage compensation mechanisms, which results in equal expression levels in males and females in somatic tissues. In male gonads, on the other hand, X chromosomes are often down-regulated or completely inactivated (Larson et al.,

2018, Mahadevaraju et al., 2021). In addition to this unusual regulatory architecture, X chromosomes have been found to differ from autosomes in various ways. One consistent feature is the over- and under-representation of genes with sex-specific patterns of expression (sex-biased genes), although the specific direction of the enrichment varies across species. For instance, the Drosophila X chromosome has a deficit of male-biased genes, whereas the mammalian X is enriched for genes with male-specific functions (Gurbich and Bachtrog, 2008). X chromosomes also often have more transposable elements and repeats, and have different gene densities, than autosomes. Finally, genes move out of X chromosomes more often than expected in both mammals and Drosophila (Gurbich and Bachtrog, 2008). Understanding what evolutionary processes drive these patterns has been the goal of extensive theoretical and empirical research (Rice, 1984, Vicoso and Charlesworth, 2006, Gurbich and Bachtrog, 2008, Meisel & Connallon, 2013, Charlesworth et al., 2018).

Since males have only one X chromosome, whereas females have two, X chromosomes differ from autosomes in key population parameters. In a population with equal sex ratio and N individuals, there are $1.5N$ X chromosomes and $2N$ sets of autosomes, so the population size of an X chromosome is ¾ the population size of an autosome. Furthermore, X chromosomes are transmitted ⅔ of the time through females and ⅓ of the time through males, whereas autosomes spend an equal amount of time in males and females. Finally, once Y-linked genes have been lost, recessive mutations arising on an X chromosome are immediately selected in hemizygous males. How these peculiarities affect the evolutionary dynamics of X-linked loci has been previously modelled (Avery, 1984, Rice, 1984, Charlesworth et al., 1987, Vicoso and Charlesworth, 2009, Meisel and Connallon, 2013, Patten,

2018, Hitchcock and Gardner, 2020). Rice (1984) found that new recessive sexually antagonistic mutations (i.e. mutations with opposite fitness effects in males and females) can invade a population more easily if they are X-linked than autosomal, and suggested that this may lead to an excess of X-linked genes underlying sexual dimorphism. However, this prediction depends on the dominance coefficient of sexually antagonistic mutations (Rice, 1984), for which we have little empirical evidence, so that it is hard to make clear predictions as to whether the X or autosomes are more favorable to the invasion of sexually antagonistic mutations (Ruzicka and Connallon, 2020). Charlesworth et al. (1987) further showed that selection on recessive mutations is stronger on the X chromosome than on the autosomes, resulting in faster substitution rates of recessive beneficial mutations on the X chromosome than on the autosomes, a pattern known as the “Faster-X effect” (Vicoso and Charlesworth, 2006, 2009). Conversely, they found that slower substitution rates are expected on X-linked loci for deleterious recessive mutations. Mutations with stronger effects on male than female fitness are particularly prone to faster-X evolution (whereas mutations with female-limited effects are exempt). Extensions of this theory have further shown that male-biased mutation rates (Kirkpatrick and Hall, 2004) and increased variance in male relative to female reproductive success (through its effect on X and autosome effective population size) can increase the faster-X effect (Vicoso and Charlesworth, 2009). On the other hand, if positive selection acts on standing variation rather than new mutations, faster-X evolution is not expected (Orr and Betancourt, 2001). Several studies have attempted to detect a faster-X effect empirically, for instance by testing for a higher proportion of adaptive substitutions on the X chromosome compared to autosomes. Faster-X divergence and faster-X adaptation have been found in various vertebrate

and invertebrate clades (Meisel & Connallon, 2013, Charlesworth et al., 2018, Bechsgaard et al., 2019, Mongue et al., 2022, Sackton et al., 2014, Llopart, 2018, Rupp et al., 2017), but not all (Pinharanda et al., 2018, Rousselle et al., 2016, Radhakrishnan and Valenzuela, 2018). Similarly, an excess of sexually antagonistic effects has been suggested in some studies (Gibson et al, 2002, Foerster et al., 2007, Innocenti and Morrow, 2010, Abbott et al., 2020) but not others (Fry, 2010, Ruzicka et al., 2019, Ruzicka and Connallon, 2022). This is further complicated by the fact that the quantitative genetic measures typically used to detect sexual antagonism are biased towards the detection of X-linked effects, due to X hemizygosity causing larger variance in fitness on the X chromosome (Ruzicka and Connallon, 2020). Because these theories predict different adaptive trajectories for the X and autosomes, they have been invoked to account for various unusual patterns observed on X chromosomes, such as the differential representation of genes with sex-biased expression or the excess movement out of the X (reviewed in Vicoso and Charlesworth, 2006, Gurbich and Bachtrog, 2008).

Much less is known about the evolution of X-linked genes during the early stages of sex chromosome evolution, when the majority of genes on the Y chromosome are functional, but recombination between X and Y is suppressed. The evolution of X-linked genes under such a scenario (in which there is no recombination between homologous loci on the X and Y, but both X and Y homologs are functional and affect fitness), has not been theoretically explored (but see Engelstaedter, 2008, who modelled how the accumulation of deleterious mutations on the X affects Y-chromosome evolution). By contrast, several empirical studies recognized that young, diploid X-linked loci can have unusual evolutionary dynamics (Nozawa et al., 2016, Nozawa et al., 2021, Wright et al, 2017). Faster rates of

nonsynonymous substitutions and enrichment of sex-biased genes have been detected on young and undifferentiated X/Z chromosomes (Wright et al., 2017, Pucholt et al., 2017). Furthermore, Nozawa et al. (2016, 2021) found evidence of accelerated pseudogenization rates on the young X chromosomes of several Drosophila lineages, compared with both the ancient X chromosome and the autosomes. They hypothesized that mechanisms similar to the ones causing degeneration of the Y chromosome could be driving degeneration of the X chromosome: first, the X chromosome has a smaller population size compared to the autosomes, which makes selection less efficient. Second, since X chromosomes do not recombine in males, their effective population size can be further reduced. However, the opposite is expected in *Drosophila*, as in this clade recombination is restricted to females, where X chromosomes are found ⅔ of the time. Third, female-biased transmission could be driving the loss of genes that are unimportant for females. Another effect that could contribute to the accumulation of deleterious mutations on young X-linked genes is sheltering by the functional gene copy on the Y chromosome. New mutations arising on a diploid X-locus are always heterozygous in males because there is no recombination between X and Y, and their phenotypic effect is masked by the ancestral allele on the Y. While the role of sheltering has been appreciated in other contexts, such as the degeneration of Y chromosomes (Muller, 1914, Nei, 1970, Bachtrog, 2013) and the evolution of recombination suppression (Charlesworth and Wall, 1999, Antonovics and Abrams, 2004, Jay et al., 2022, Olito et al., 2022), it is unclear to what extent it affects early X-chromosome evolution. Here, we model evolutionary rates of X-linked loci with functional Y copies under various selective regimes, to formally explore how these different processes shape the early stages of X chromosome evolution.

Methods

The diffusion approximation

Substitution rates can be calculated as the average number of mutations entering a population in one generation times the fixation probability of those mutations (Kimura and Ohta, 1971). To derive probabilities of fixation for autosomal and hemizygous X-linked loci, Vicoso and Charlesworth (2009) used the diffusion approximation. Here, we extend their model to diploid X-linked loci.

Let A_1 and A_2 be alleles for some locus, with frequencies *(1-p)* and *p*, respectively, and fitness effects as noted in Table 1.

Table 1. Relative fitnesses in females and males for autosomal, hemizygous X-linked and diploid X-linked loci

		Females			**Males**		
Autosomal	Genotypes	A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
	Fitness	*1*	$1+hs_f$	$1+s_f$	*1*	$1+hs_m$	$1+s_m$
Hemizygous X-linked	Genotypes	$A_{x1}A_{x1}$	$A_{x1}A_{x2}$	$A_{x2}A_{x2}$	A_{x1}	A_{x2}	
	Fitness	*1*	$1+hs_f$	$1+s_f$	*1*	$1+s_m$	
Diploid X-linked	Genotypes	$A_{x1}A_{x1}$	$A_{x1}A_{x2}$	$A_{x2}A_{x2}$	$A_{y1}A_{x1}$	$A_{y1}A_{x2}$	
	Fitness	*1*	$1+hs_f$	$1+s_f$	*1*	$1+hs_m$	

We can estimate the fixation probability of allele A_2 using the diffusion approximation (Ewens, 2004). The fixation probability of an allele with the initial frequency *p* is given by the function *U(p)*:

$$U(p) = \frac{\int_0^p G(y)dy}{\int_0^1 G(y)dy}, \quad (1)$$

with $$G(y) = exp(-2\int_0^y \frac{M(x)}{V(x)}dx),$$

where *M(x)* and *V(x)* are respectively the expectation and the variance of the change of allele frequency.

Assuming a weak effect of selection in each sex, so that second-order terms are small enough to be neglected, fitness of a genotype can be approximated as the average of fitness effects in males and females (Nagylaki, 1979, Charlesworth and Charlesworth, 2010, ch 3.1). For a diploid X-linked locus, the expected change of an allele A_2 frequency due to selection is given by:

$$M_{dX}(x) \approx x(1-x)(\frac{2}{3}(w_{2f} - w_{1f}) + \frac{1}{3}(w_{2m} - w_{1m})),$$

where *x* is the frequency of the allele A_2, w_{1f} and w_{1m} are marginal fitnesses of allele A_1 in females and males respectively, and w_{2f} and w_{2m} are marginal fitnesses of allele A_2 in females and males respectively. Calculating these marginal fitnesses from Table 1, we have:

$$M_{dX}(x) \approx x(1-x)(\frac{2}{3}s_f(h + x(1-2h)) + \frac{1}{3}hs_m),$$

where *h* is the dominance coefficient.

We now divide $M_{dX}(x)$ by the variance in the change of allele A_2 frequency due to sampling drift,

$$V_{dX}(x) \approx \frac{x(1-x)}{2N_{eX}},$$

where N_{eX} is the effective population size of the X chromosome, and integrate to find *G(y)*. This can be written in terms of mean selection averaged across sexes, $\bar{s} = \frac{2}{3}\frac{s_f}{2} + \frac{1}{3}hs_m = \frac{s_f + hs_m}{3}$, and $\sigma = \frac{s_f(1-2h)}{3}$, which is zero if there is no dominance (*h=1/2*). Then:

$$G_{dX}(y) = exp(-2\int_0^y \frac{M(x)}{V(x)}dx)$$

$$= exp(-\frac{4}{3}N_{eX}(2s_f + s_m)h\int_0^y dx - \frac{8}{3}N_{eX}sf(1-2h)\int_0^y xdx$$

$$= exp(-\frac{4}{3}N_{eX}(2s_f + s_m)hy - \frac{4}{3}N_{eX}s_f(1-2h)y^2)$$

$$= exp(-4N_{eX}(\frac{1}{3}(s_f + hs_m)y - \frac{1}{3}s_f(1-2h)y(1-y)))$$

$$= exp(-4N_{eX}(\bar{s}y - \sigma y(1-y))) \quad (2).$$

We are interested in the probability of fixation of single new mutation, which is initially at $p = \frac{2}{3N}$; the substitution rate is $1.5N\mu U(p) = \mu U(p)/p$, where μ is the mutation rate. Following the Eq.1, the fixation probability *U(p)* becomes:

$$U(p) = \frac{\int_0^p G(y)dy}{\int_0^1 G(y)dy}$$

$$= \frac{erf(\frac{N_{eX}(\bar{s}+\sigma(2p-1))}{\sqrt{N_{eX}\sigma}}) - erf(\frac{N_{eX}(\bar{s}-\sigma)}{\sqrt{N_{eX}\sigma}})}{erf(\frac{N_{eX}(s-\sigma)}{\sqrt{N_{eX}\sigma}}) - erf(\frac{N_{eX}(\bar{s}-\sigma)}{\sqrt{N_{eX}\sigma}})}$$

$$= \frac{\frac{4p\sqrt{N_{eX}\sigma}}{\sqrt{\pi}}exp(-\frac{N_{eX}(\bar{s}-\sigma)^2}{\sigma})}{erf(\frac{N_{eX}(\bar{s}-\sigma)}{\sqrt{N_{eX}\sigma}}) - erf(\frac{N_{eX}(\bar{s}-\sigma)}{\sqrt{N_{eX}\sigma}})} + O(p^2),$$

where $O(p^2)$ are higher order terms in Taylor series when *p* is close to zero. Assuming weak selection and sufficiently large effective population size, $O(p^2)$ can

be neglected for new mutations. Then, the first order approximation in *p* for substitution rate is:

$$\mu(U(p)/p) = 4\mu\sqrt{\frac{N_{eX}\sigma}{\pi}}\,\frac{exp(-N_{eX}(\bar{s}-\sigma)^2/\sigma)}{(erf(N_{eX}(\bar{s}+\sigma)/\sqrt{N_{eX}\sigma})-erf(N_{eX}(\bar{s}-\sigma)/\sqrt{N_{eX}\sigma}))}$$

Note that the substitution rate relative to mutation depends only on $N_{eX}\bar{s}$ and $N_{eX}\sigma$; in the additive case (*h*=1/2), it simplifies to Kimura's formula, $4N_{eX}\bar{s}/(1 - exp(-4N_{eX}\bar{s}))$.

Also, note from Table 1 that the only difference between the hemizygous and the diploid X cases is that s_m is replaced by hs_m in the latter case; thus, the results of Vicoso and Charlesworth (2009) for the hemizygous case can be found from the results given here simply by replacing s_m by s_m/h.

We implemented the numerical integration in R (*version 4.1.1*, code available at https://git.ist.ac.at/amrnjava/x-chromosome-theory/-/blob/main/fixation_probaility_functions.R), which allowed us to derive substitution rates for autosomal, hemizygous X-linked and diploid X-linked loci over the range of dominance coefficients and selective effects in males and females. A GUI application with implemented fixation probability functions for autosomal, hemizygous X-linked and diploid X-linked loci is available at https://degenerate-x.science.ista.ac.at/, and allows the user to explore the difference among substitution rates of autosomal, hemizygous X-linked and diploid X-linked loci over the range of parameter values (selective effects in males and females and dominance coefficient), assuming effective population size of an X chromosome is $\frac{3}{4}$ of autosomal effective population size. In addition, we modelled evolutionary rates on X chromosomes and autosomes assuming equal population

sizes for autosomes and X chromosomes, in order to sequester the effects of effective population size and sheltering.

Branching process approximation

To model the substitution rates of strongly beneficial mutations at diploid X-linked loci with functional but non-recombining Y gametolog, and allow comparisons with the classic Charlesworth et al. (1987) result, we used Haldane's branching process approximation (Haldane, 1927, Charlesworth et al., 1987, Vicoso and Charlesworth, 2006, Meisel & Connallon, 2013, Charlesworth et al., 2018).

If $N_e sh$ is sufficiently large ($1/N_e << sh << 1$), the fixation probability of a new beneficial mutation can be approximated as twice the advantage of a heterozygous genotype (Haldane, 1927).

Taking into account the effects of a mutation in males and females separately (Vicoso and Charlesworth, 2006), the fixation probability of a single new beneficial mutation at an autosomal locus can be approximated as:

$$P_A(\frac{1}{2N}) \approx \frac{1}{2}2hs_f + \frac{1}{2}2hs_m \approx h(s_f + s_m),$$

where s_f and s_m are selection coefficients in females and males, respectively, *h* the dominance coefficient, and *N* the number of diploid individuals, as before.

For diploid X-linked loci, fixation probability of a new mutation is then:

$$P_X(\frac{1}{1.5}) \approx \frac{2}{3}2hs_f + \frac{1}{3}2hs_m \approx \frac{2}{3}h(2s_f + s_m).$$

Following substitution rates are then:

$K_A \approx 2N\mu h(s_f + s_m)$ for autosomal loci, and

$K_X \approx \frac{3}{2}N\mu\frac{2}{3}h(2s_f + s_m) \approx N\mu h(2s_f + s_m)$ for diploid X loci.

The ratio of diploid X to autosomal adaptive substitution rates is then:

$$R \approx \frac{2N\mu h(s_f+s_m)}{N\mu h(2s_f+s_m)} \approx \frac{2s_f+s_m}{2s_f+2s_m}, \quad (4)$$

if we assume the number of X chromosomes in a population is ¾ the number of autosomes, and dominance is the same in males and females, as well as the mutation rate.

Results

Diploid X loci adapt slower and accumulate deleterious mutations faster

In addition to previously derived substitution rates on hemizygous X loci and autosomal loci (Charlesworth et al., 1987, Vicoso and Charlesworth, 2009), we used the diffusion approximation to estimate fixation probabilities of new mutations arising at diploid X-linked loci (which have a functional homolog on the Y), and their substitution rates. We verified these estimates with individual-based forward genetic simulations (Fig. S1, details of simulations are given in Supplementary Material). The X/A ratios of substitution rates (R) for beneficial and deleterious mutations and for diploid and hemizygous X-linked loci are visualised in Figure 1. We recover the previously described faster-X effect for hemizygous X-linked loci (Charlesworth et al., 1987, Vicoso and Charlesworth, 2006, 2009), where recessive beneficial mutations accumulate faster on the hemizygous X loci, while dominant beneficial mutations accumulate faster on the autosomes compared to the hemizygous X loci. On the contrary, diploid X loci exhibit a Slower-X effect regardless of the dominance coefficient: the substitution rate of beneficial mutations is lower at diploid X loci compared to autosomal and hemizygous X loci, in accordance with reduced efficiency of selection on diploid X loci due to sheltering and reduced effective population size (Table 1). Substitution rates of deleterious mutations follow the opposite pattern, with diploid X loci accumulating deleterious mutations faster than autosomal and hemizygous X loci. These results, therefore, show that young diploid X-linked genes have reduced adaptive potential, and increased accumulation of

deleterious mutations, relative to both hemizygous X-linked genes and autosomal genes.

To disentangle the effects of reduced effective population size and sheltering on substitution rates of diploid X loci, we calculated the ratio of substitution rates in the case where $N_{eX}=N_{eA}$, as is approximately the case in *Drosophila melanogaster* due to the effect of selection at linked sites (although differences in male and female fitness variance may also play a role, (Charlesworth, 2001)): since there is no recombination in males X chromosome have higher effective recombination rates than autosomes, which increases the N_{eX}/N_{eA} ratio (Charlesworth et al., 2018.). The results show that in this case there is still a “Slower-X” effect for diploid X loci, but only for recessive mutations (S.Figure 4).

As an example, Figure 1 shows the ratios of substitution rates (R) for beneficial mutations with $N_{eA}s$= 3, and for deleterious mutations with $N_{eA}s$= -1, where equal effects in males and females are assumed ($s=s_f=s_m$), but a wide range of positive and negative $N_{eA}s$ values (we modelled $N_{eA}s$ values from -3 to 5) yield the same qualitative pattern. The effect of different $N_{eA}s$ values of a mutation on the X/A ratio of substitution rates can be explored in a GUI web application provided at https://degenerate-x.science.ista.ac.at/. However, it is worth noting that the ratio of substitution rates at diploid X loci compared to autosomal loci (R) increases exponentially with the strength of the deleterious effect of a mutation (S.Figure 2). This means that mutations with stronger deleterious effects will be fixed much more often in diploid X loci relative to an autosome.

While the original faster-X publication (Charlesworth et al., 1987) focused on R (in that case the autosome:X substitution rate ratio), evolutionary rates are usually measured as the rate of nonsynonymous substitutions (dN), normalized by the

synonymous rate of substitutions (dN/dS, with dS acting as a proxy for neutral divergence) (Meisel & Connallon, 2013). To facilitate comparisons with such data, we also plot the substitution rate of beneficial and deleterious mutations after normalizing them by the neutral substitution rate (qualitative patterns remain the same, S.Figure 3).

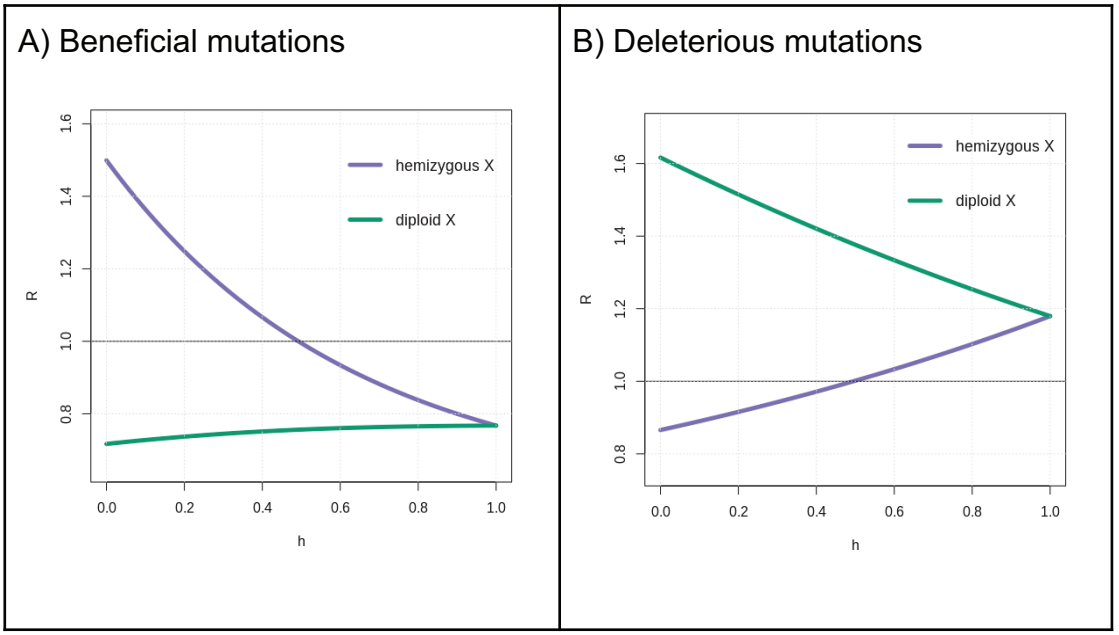


Figure 1. Ratios of substitution rates (R) on hemizygous X to autosomal loci and diploid X to autosomal loci as functions of dominance coefficient h, for beneficial, $N_{eA}s$=3 (A) and deleterious, $N_{eA}s$=-1 mutations (B).

Similar results are recovered with the branching process approximation (see methods). Equation 4 shows that for strongly beneficial mutations (where $N_e sh$ is sufficiently large), the X/A ratio of substitution rates does not depend on the dominance coefficient at diploid X loci, in contrast with the previously described X/A ratio of adaptive evolutionary rates at hemizygous X loci, which exhibit Faster-X effect for recessive mutations. More precisely, our results show that for mutations with female-limited selective effects, substitution rates are the same at autosomal

and diploid X loci, that is, R=1; for mutations with male-limited selective effects, R=½, that is, adaptation rate on diploid X is half the adaptation rate on autosomes, and for mutations with equal effects in males and females, R=¾. These results indicate Slower-X effect for all strongly beneficial mutations arising on diploid X loci and having an effect in males.

The "Slower X" effect is strongest for male-biased mutations

We also aimed to disentangle how selective effects in males and females separately affect the substitution rate at diploid X loci. Mutations can have different effects in males and females: they can be sex-limited, affecting only the fitness of one sex, sex-biased, if they have a stronger effect on the fitness of one sex than the other, or sexually antagonistic, if they have fitness effects of opposite signs in the two sexes. We can intuitively see from the fitness table (Table 1) that the differences in evolutionary rates of diploid X, hemizygous X and autosomal loci result from the differences in the male part of the fitness table. Indeed, similar to Charlesworth et al. (1987), we find that evolutionary rates are the same at autosomal, hemizygous X-linked and diploid X-linked loci for female-limited mutations (mutations affecting only female fitness, e.g. $N_e s_m=0$, $N_e s_f=3$ in Figure 2A, or $N_e s_m=0$, $N_e s_f=-1$ in Fig. 2B). On the other hand, the X/A ratio of substitution rates (R) for beneficial male-limited mutations ($N_e s_m=3$, $N_e s_f=0$), and to a smaller extent male-biased mutations, is lower than R for mutations with an equal effect in both males and females ($N_e s_m=3$, $N_e s_f=3$) (Fig 2A). It is important to note in Figure 2. that in the case with equal strength of selection in males and females, the average selection coefficient is larger than in the

case of sex-limited fitness effects. Male-limited and male-biased mutations are primarily under selection in males, where their effect is masked by the ancestral allele on the Y, resulting in a stronger “Slower-X effect”. Analogously, deleterious male-limited and male-biased mutations accumulate faster at diploid X loci than at autosomal loci, and faster than deleterious female-limited and female-biased loci with the corresponding fitness effects (e.g. comparing male-biased mutations with $N_e s_m$=*-1*, $N_e s_f$=*-0.5* to female-biased mutations with $N_e s_m$=*-0.5*, $N_e s_f$=*-1*). Counterintuitively, Fig 2B shows that R for a diploid X is larger for mutations with equal fitness effects in males and females than for male-limited and male-biased mutations. This is because mutations affecting both sexes ($N_e s_m$=*-1*, $N_e s_f$=*-1*) are overall more deleterious than mutations affecting males ($N_e s_m$=*-1*, $N_e s_f$=*0*) or females only, and R for the diploid X increases exponentially with the strength of the deleterious effect of a mutation (S.Figure 2). To summarize, the X/A ratio of substitution rates for deleterious mutations at diploid X will be greater than 1 as long as the mutation has an effect in males, and mutations with a stronger deleterious effect in males than in females will accumulate in excess on the young X compared to autosomes.

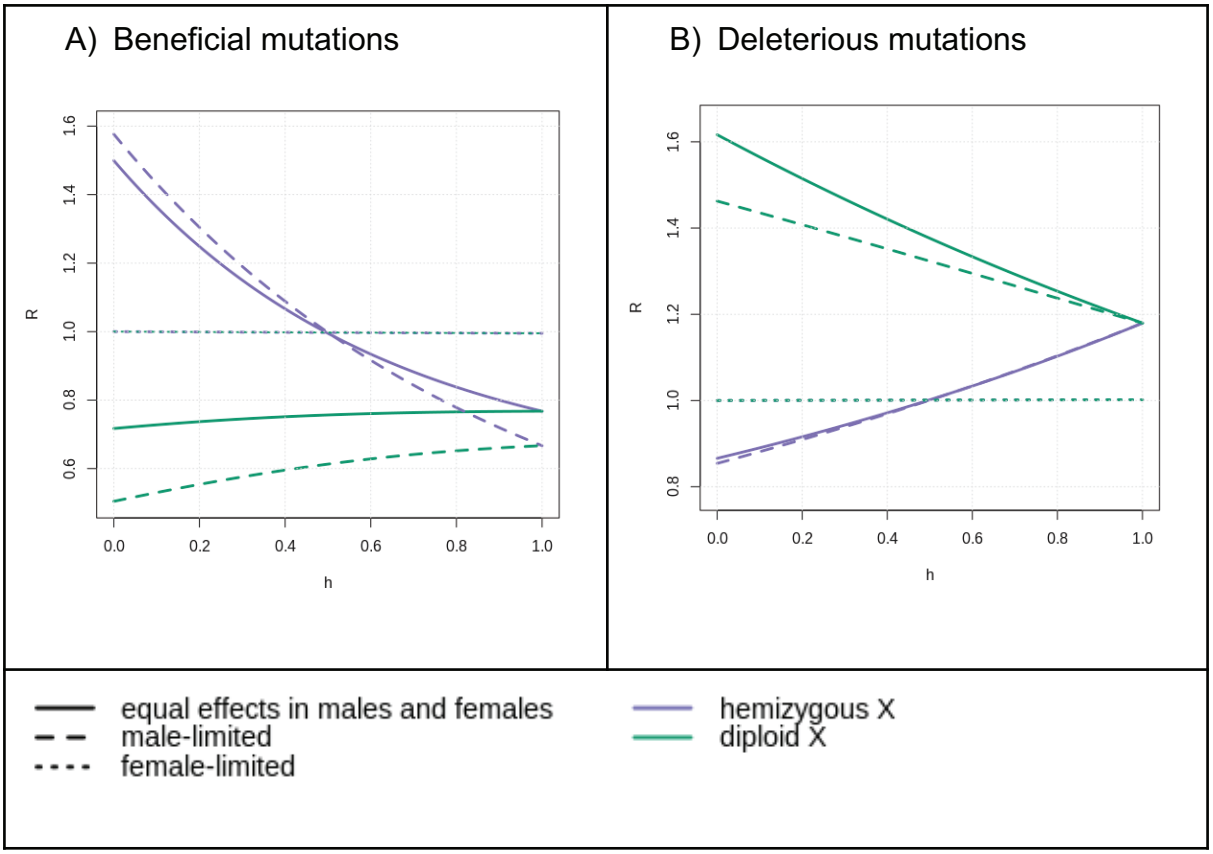


Figure 2 Ratios of X/A substitution rates (R) of male-limited mutations ($N_e s_m$=*3*, $N_e s_f$=*0* for and $N_e s_m$=*-1*, $N_e s_f$=*0* for beneficial and deleterious mutations respectively) are plotted alongside ratios of substitution rates of mutations with equal effects ($N_e s_m$=*3*, $N_e s_f$=*3* for and $N_e s_m$=*-1*, $N_e s_f$=*-1* for beneficial and deleterious mutations respectively), and female-limited ($N_e s_m$=*0*, $N_e s_f$=*3* for and $N_e s_m$=*0*, $N_e s_f$=*-1* for beneficial and deleterious mutations respectively) effects for beneficial (A) and deleterious (B) mutations, as a function of dominance coefficient, *h.* R for female-biased mutations is between R for female-limited mutations and R for mutations with equal effects in males and females, while R for male-biased mutations is between R for male-limited mutations and R for mutations with equal effects in both sexes.

Sexually antagonistic mutations are expected to accumulate faster on an ancient hemizygous X than on the autosomes if they are recessive and male-beneficial, or dominant and female-beneficial (Rice, 1984, Vicoso and Charlesworth, 2009, and Fig. 3). One assumption of these studies, that we also make here, is that the dominance coefficient of antagonistic mutations is the same in males and females (see Fry, 2010, for results when this does not hold). The resulting differential accumulation of sexually antagonistic mutations has been proposed to influence the evolution of sex chromosomes and their role in encoding sexual dimorphism (Rice, 1984, Fry, 2010). Figure 3 shows that mutations carrying male-advantage but female-disadvantage ($N_e s_m=3$, $N_e s_f=-3$) accumulate much more slowly at diploid X loci than at autosomal or hemizygous X loci, independent of their dominance coefficient. Mutations carrying female-advantage and male-disadvantage ($N_e s_m=-3$, $N_e s_f=3$), on the other hand, accumulate faster at diploid X loci. Taken together, these results show that the reduced efficacy of selection on a diploid X is highly influenced by sex-specific fitness effects, with mutations that benefit primarily males tending to accumulate slower, and mutations that are detrimental to males accumulating faster whether they benefit females or not.

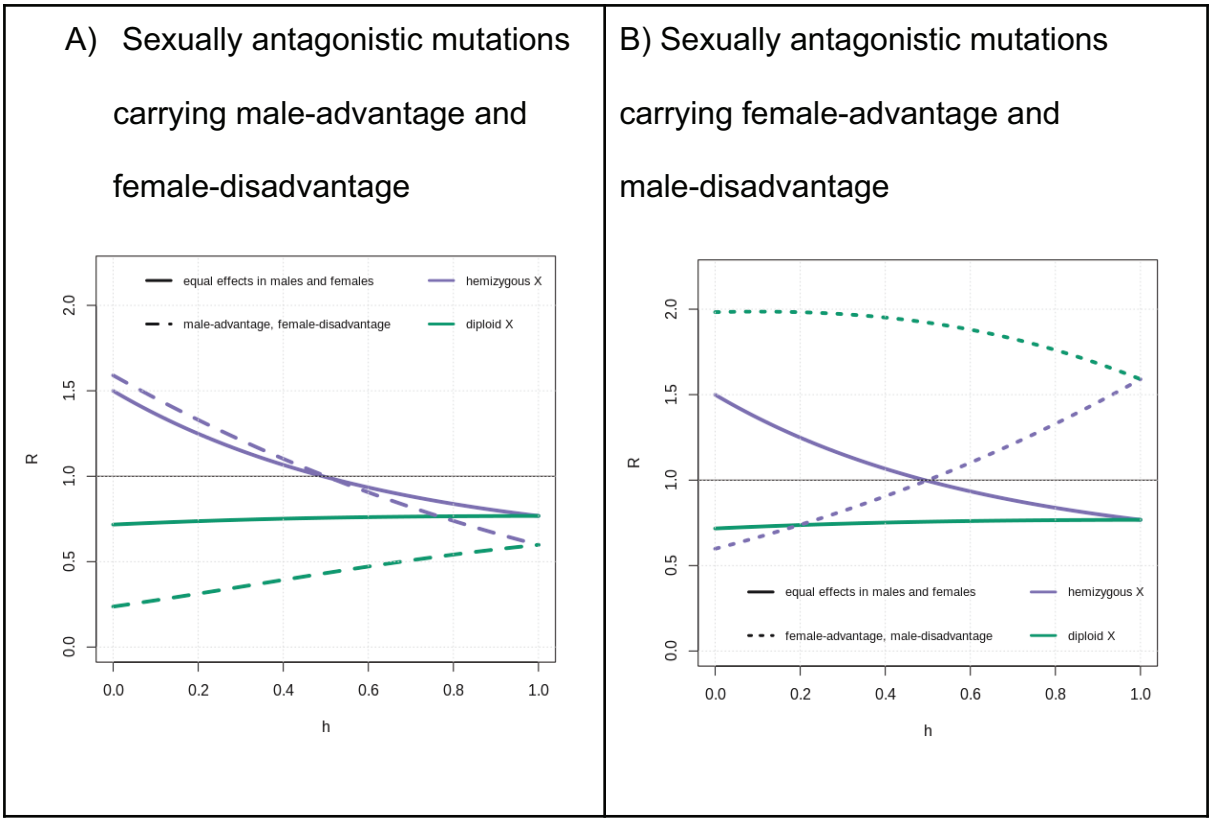


Figure 3 Ratios of X/A substitution rates (R) for sexually antagonistic mutations carrying male-advantage and female-disadvantage ($N_e s_m$=*3*, $N_e s_f$=*-3*) (**A**) and mutations carrying female-advantage and male-disadvantage $N_e s_m$=*-3*, $N_e s_f$=*3* (**B**) plotted along the R for mutations with equal effects in males and females, as a function of dominance coefficient, *h*.

Discussion

There has been extensive theoretical and empirical work on the “faster-X effect” expected on X chromosomes with a degenerated Y counterpart (reviewed in Meisel & Connallon, 2013, Charlesworth et al., 2018). Here, we modelled evolutionary rates of diploid X-linked loci, with a functional, but non-recombining, gametolog on the Y counterpart. Our results show very different evolutionary dynamics for diploid and hemizygous X loci. We make two key predictions for the evolution of young X-chromosomes: 1. selection efficacy is reduced, such that fewer beneficial mutations and more deleterious mutations will fix there compared to an autosomal locus 2. The relaxation of selection is stronger for mutations that primarily affect male fitness and/or benefit males at the expense of females. Over time, this may lead to the “demasculinization” of the young X chromosome, i.e. the degeneration of genes with male-important functions, and/or the failure to acquire new genes with male functions. This is in contrast to hemizygous X-linked loci on differentiated X chromosomes, which may exhibit faster adaptation and masculinization (as long as beneficial mutations are generally recessive).

These peculiar evolutionary dynamics of diploid X loci are caused by: 1. a smaller effective population size compared to autosomes, 2. female-biased transmission and 3. sheltering of partly recessive X-linked mutations in males by an ancestral allele on the Y. By exploring a range of parameters, we could to some extent quantify the individual contribution of these effects. Sheltering does not affect fully dominant mutations, and the difference in R when h=0 and h=1 shows that it can have a substantial effect on rates of adaptive and maladaptive divergence (Fig. 1-3). When the reduction in N_e is removed (N_{eA}=N_{eX}), reduced efficacy of selection is

only detected when mutations are at least partly recessive (due to sheltering, Sup. Fig. 4A,B) and/or have male-biased or male-limited effects (due to female-biased transmission and a stronger effect of sheltering, Sup. Fig. 4C,D). The relative importance of these effects depends on the strength of fitness effects in males, as sheltering only affects mutations expressed in this sex. Finally, both female-biased transmission and sheltering contribute to the proposed demasculinization, as male-important mutations are affected disproportionately (S.Figure 4).

Only a few empirical studies have explicitly compared the evolution of diploid X-linked genes to autosomal control. Several of them investigated the young neo-X chromosome found in the *Drosophila miranda* lineage, where a pair of autosomes fused with the ancestral Y chromosome and became a neo-X and neo-Y a little over 1 million years ago. Around 40 % of genes on the neo-Y are still functional (Zhou and Bachtrog, 2012, Nozawa et al. 2016). Zhou and Bachtrog (2012) found that hemizygous X loci adapt faster than diploid X loci on the neo-X chromosome in *D.miranda*. Nozawa et al. (2016) further detected evidence of an accelerated pseudogenization rate on the neo-X chromosome in *D.miranda* after it became X-linked. They compared the neo-X chromosome in *D.miranda* to the corresponding autosome in *D.pseudoobscura*, and found that genes that were ancestrally under reduced selective constraints (have higher dN/dS values in *D.pseudoobscura*) and genes with ancestrally male-biased function (approximated from *D.pseudoobscura* F/M expression ratios) were more likely to be pseudogenized on the neo-X than we would expect if it was an autosome. Their analyses suggest that the reduction in efficiency of selection, especially for male-biased genes, is causing the accelerated pseudogenization rate on the young X chromosome. Recently, Nozawa et al. (2021) found degeneration of neo-X chromosomes in two other *Drosophila* species with

independently acquired neo-X. These results are generally in line with our theoretical predictions of maladaptive evolution of young X chromosomes and accompanying demasculinization, but work in various other systems is needed to understand how universal this pattern is and how much of a contribution it makes to X-chromosome evolution.

In particular, one question that we did not address here is the time frame over which maladaptive evolution occurs. In the early stages of sex chromosome differentiation, the majority of the ancestral gene content on the Y chromosome will still be functional, and corresponding X-linked loci will be diploid and evolve under reduced selective efficacy. The accelerated rate of pseudogenization on the X will slow down as the Y chromosome degenerates and more X-linked loci become hemizygous, causing a shift in evolutionary dynamics to standard “faster-X”. Theory predicts that Y chromosomes first degenerate quickly after the recombination suppression but after they have lost about half of the gene content the process slows down, since the rates of degeneration by Muller’s ratchet, background selection and genetic hitchhiking correlate with the number of active genes (Bachtrog, 2008). For instance, more than half of the genes on the *Drosophila miranda* neo-Y chromosome have been lost in a little over 1 million years, and the 15 million year old neo-Y of *D.pseudoobscura* is already highly degenerated (reviewed in Charlesworth, 2021). However, it is unclear if the neo-sex chromosomes of Drosophila, which have quickly co-opted a preexisting mechanism of dosage compensation, are representative of typical dynamics of Y degeneration, and much slower Y degeneration has been described in other systems (Li et al., 2021, Charlesworth, 2021). Nonrecombining regions with intermediate or low levels of Y/W degeneration have been described in various taxa (Charlesworth, 2021), e.g.: schistosomes (Elkrewi et al., 2021), frogs

(Furman and Evans, 2018), crustaceans (Elkrewi et al., 2022), birds (Liu et al., 2021), fish (Sardell et al., 2021), and plants (Veltsos et al., 2019). It is therefore clear that many young X-linked genes can remain diploid for substantial periods of time.

The existence of a period of maladaptive evolution has implications for young homomorphic X chromosomes, but may also contribute to patterns observed on older sex chromosomes. The fact that X-linked genes with male-specific functions should be more prone to early maladaptive evolution suggests that the "demasculinization" that is observed on differentiated X-chromosomes of various species (Gurbich and Bachtrog, 2008) may begin before the degeneration of the Y. The subsequent shift in evolutionary dynamics could contribute to the differences in sex-biased gene contents of X chromosomes, which are for instance masculinized in mammals and demasculinized in *Drosophila* lineage (Gurbich and Bachtrog, 2008). The importance of temporal dynamics to the process of demasculinization is well appreciated in Drosophila, where only ancient male-biased genes are depleted from the X chromosome, whereas newly evolved ones are enriched on it (Zhang et al., 2010). While this potentially supports a role of diploid X evolution, the explanations brought forward to explain it typically assume a hemizygous X. Furthermore, Engelstaeder (2008) used modelling to show that the accumulation of deleterious mutations on diploid X-linked loci can slow down the degeneration of their Y-linked loci. An illustration of this is *PRSSLY*, a gene that was lost from the mammalian X chromosome in eutherians but retained on the Y (Hughes et al., 2022). If genes that function primarily in males are the ones that tend to accumulate deleterious mutations on a young X, this may lead to their preferential maintenance on degenerating Y chromosomes. The preservation of genes with male-biased

expression on the Y has been observed (Kaiser et al., 2011, Zhou and Bachtrog, 2012, Mahajan and Bachtrog, 2017, Crowson et al., 2017) and is usually assumed to be driven by male-specific selection on Y-linked genes. In fact, degeneration of male-important genes on the X may drive their conservation on the Y, as well as the other way around. If the accumulation of deleterious mutations on X-linked genes promotes the maintenance of at least partial activity of their Y homeolog (and *vice versa*), this may promote the maintenance of both copies and partly account for long-lived homomorphic sex chromosomes.

Other peculiarities of X chromosomes may first arise early in their evolution. Repetitive sequences and transposable elements are overrepresented on X chromosomes as well as Y chromosomes (Bellott et al., 2010). The reduction in effective population size of young X chromosomes (due to their lower population size, and potentially further exacerbated because they do not recombine in males) may already contribute to the accumulation of repeats. Finally, X-chromosomes are central to the two "rules of speciation": Haldane's rule (hybrid sterility or inviability tends to affect the heterogametic sex more than the homogametic sex), and the "large X effect" (an excessive proportion of hybrid sterility loci maps to X chromosomes). Most of the clades that obey these rules have differentiated sex chromosomes, and explanations have typically invoked the faster-X hypothesis (along with other models) (Dufresnes and Crochet, 2022). However, hybridization patterns in *Aedes* mosquitoes, which have undifferentiated sex chromosomes, follow Haldane's rule (Presgraves and Orr, 1998). Similarly, Dufresne et al. (2016) detected an excessive role of an undifferentiated X-chromosome in the reproductive isolation of tree frogs, and these rules have been suggested to apply to some clades with homomorphic sex chromosomes (Filatov, 2017). An excessive accumulation of

deleterious mutations on homomorphic X chromosomes may provide an explanation for a large X effect in this context, if compensatory mutations arise elsewhere in the genome. Similarly, if Y-degradation and X-sheltering affect different genes in close species, a mismatch between them may contribute to male sterility in hybrids. Studies of hybrids over a wide range of sex chromosome differentiation will in the future allow us to quantify the temporal dynamics of the large-X effect and Haldane's rule, and the contribution of diploid X evolution.

In short, X-linked loci are expected to undergo a period of maladaptive evolution and demasculinization in the early stages of their differentiation. Our results show that contrary to what is often assumed, the peculiar evolutionary patterns on the X chromosome may arise before substantial degeneration of the Y has occurred, and provide a novel framework for interpreting the increasing amount of data available for clades with young sex chromosomes.

Author Contributions

A.M. and B.V. designed the study and wrote the first draft. A.M., K.K. and N.B. performed the analysis. All the authors contributed to and approved the final version of the manuscript. All authors declare no conflict of interest.

Acknowledgments

We thank the Vicoso and Barton groups and ISTA Scientific Computing Unit.

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreements no. 715257 and no. 716117).

References

Abbott, J.K., Chippindale, A.K. & Morrow, E.H. (2020) The microevolutionary response to male-limited X-chromosome evolution in *Drosophila melanogaster* reflects macroevolutionary patterns. Journal of Evolutionary Biology, 33, 738–750.

Antonovics, J. & Abrams, J.Y. (2004) Intratetrad mating and the evolution of linkage relationships. Evolution, 58, 702–709.

Avery, P. J. 1984. The population genetics of haplo-diploids and X-linked genes. Genetics Research 44:321–341.

Bachtrog, D. (2008) The temporal dynamics of processes underlying Y chromosome degeneration. Genetics, 179, 1513–1525.

Bachtrog, D. (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet, 14, 113–124.

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., et al. (2014) Sex determination: why so many ways of doing it? PLOS Biology, 12, e1001899.

Bechsgaard, J., Schou, M.F., Vanthournout, B., Hendrickx, F., Knudsen, B., Settepani, V., et al. (2019) Evidence for Faster X chromosome evolution in spiders. Molecular Biology and Evolution, 36, 1281–1293.

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., et al. (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature, 466, 612–616.

Charlesworth, B. (2001) The effect of life-history and mode of inheritance on neutral genetic variability. Genetics Research, 77, 153–166.

Charlesworth, B., Campos, J.L. & Jackson, B.C. (2018) Faster-X evolution: Theory and evidence from Drosophila. Molecular Ecology, 27, 3753–3771.

Charlesworth, B. & Charlesworth, D. (2010) Elements of Evolutionary Genetics. W. H. Freeman.

Charlesworth, B., Coyne, J.A. & Barton, N.H. (1987) The relative rates of evolution of sex chromosomes and autosomes. The American Naturalist, 130, 113–146.

Charlesworth, B. & Wall, J.D. (1999) Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 266, 51–56.

Charlesworth, D. (2021) The timing of genetic degeneration of sex chromosomes. Philos Trans R Soc Lond B Biol Sci, 376, 20200093.

Crowson, D., Barrett, S.C.H. & Wright, S.I. (2017) Purifying and positive selection influence patterns of gene loss and gene expression in the evolution of a plant sex chromosome system. Molecular Biology and Evolution, 34, 1140–1154.

Dufresnes, C. & Crochet, P.-A. (2022) Sex chromosomes as supergenes of speciation: why amphibians defy the rules? Philosophical Transactions of the Royal Society B: Biological Sciences, 377, 20210202.

Dufresnes, C., Majtyka, T., Baird, S.J.E., Gerchen, J.F., Borzée, A., Savary, R., et al. (2016) Empirical evidence for large X-effects in animals with undifferentiated sex chromosomes. Sci Rep, 6, 21029.

Elkrewi, M., Khauratovich, U., Toups, M.A., Bett, V.K., Mrnjavac, A., Macon, A., et al. (2022) ZW sex-chromosome evolution and contagious parthenogenesis in Artemia brine shrimp. Genetics, 222, iyac123.

Elkrewi, M., Moldovan, M.A., Picard, M.A.L. & Vicoso, B. (2021) Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Molecular Biology and Evolution, 38, 5345–5358.

Engelstädter, J. (2008) Muller's ratchet and the degeneration of Y chromosomes: a simulation study. Genetics, 180, 957–967.

Ewens, W.J. (2004) Mathematical Population Genetics. Interdisciplinary Applied Mathematics. Springer, New York, NY.

Filatov, D.A. (2018) The two "rules of speciation" in species with young sex chromosomes. Molecular Ecology, 27, 3799–3810.

Foerster, K., Coulson, T., Sheldon, B.C., Pemberton, J.M., Clutton-Brock, T.H. & Kruuk, L.E.B. (2007) Sexually antagonistic genetic variation for fitness in red deer. Nature, 447, 1107–1110.

Fry, J.D. (2010) The genomic location of sexually antagonistic variation: some cautionary comments. Evolution, 64, 1510–1516.

Furman, B.L.S. & Evans, B.J. (2018) Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. Genome Biology and Evolution, 10, 742–755.

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

Gurbich, T.A. & Bachtrog, D. (2008) Gene content evolution on the X chromosome. Curr Opin Genet Dev, 18, 493–498.

Haldane, J.B.S. (1927) A Mathematical theory of natural and artificial selection, Part V: Selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society, 23, 838–844.

Haller, B.C. & Messer, P.W. (2019) SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Molecular Biology and Evolution, 36, 632–637.

Hitchcock, T.J. & Gardner, A. (2020) A gene's-eye view of sexual antagonism. Proceedings of the Royal Society B: Biological Sciences, 287, 20201633.

Hughes, J.F., Skaletsky, H., Nicholls, P.K., Drake, A., Pyntikova, T., Cho, T.-J., et al. (2022) A gene deriving from the ancestral sex chromosomes was lost from the X and retained on the Y chromosome in eutherian mammals. BMC Biology, 20, 133.

Innocenti, P. & Morrow, E.H. (2010) The sexually antagonistic genes of *Drosophila melanogaster*. PLOS Biology, 8, e1000335.

Jay, P., Tezenas, E., Véber, A. & Giraud, T. (2022) Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLOS Biology, 20, e3001698.

Jeffries, D.L., Gerchen, J.F., Scharmann, M. & Pannell, J.R. (2021) A neutral model for the loss of recombination on sex chromosomes. Phil. Trans. R. Soc. B, 376, 20200096.

Kaiser, V.B., Zhou, Q. & Bachtrog, D. (2011) Nonrandom gene loss from the *Drosophila miranda* Neo-Y Chromosome. Genome Biology and Evolution, 3, 1329–1337.

Kimura, M. & Ohta, T. (1971) On the rate of molecular evolution. J Mol Evol, 1, 1–17.

Kirkpatrick, M. & Hall, D.W. (2004) Male-biased mutation, sex linkage, and the rate of adaptive evolution. Evolution, 58, 437–440.

Lahn, B. T., and D. C. Page. 1999. Four evolutionary strata on the human X chromosome. Science 286:964–967.

Larson, E.L., Kopania, E.E.K. & Good, J.M. (2018) Spermatogenesis and the Evolution of Mammalian Sex Chromosomes. Trends in Genetics, 34, 722–732.

Li, M., Zhang, R., Fan, G., Xu, W., Zhou, Q., Wang, L., et al. (2021) Reconstruction of the origin of a Neo-Y sex chromosome and its evolution in the spotted knifejaw, *Oplegnathus punctatus*. Molecular Biology and Evolution, 38, 2615–2626.

Liu, J., Wang, Z., Li, J., Xu, L., Liu, J., Feng, S., et al. (2021) A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. Genome Res., 31, 497–511.

Llopart, A. (2018) Faster-X evolution of gene expression is driven by recessive adaptive cis-regulatory variation in Drosophila. Molecular Ecology, 27, 3811–3821.

Mahadevaraju, S., Fear, J.M., Akeju, M., Galletta, B.J., Pinheiro, M.M.L.S., Avelino, C.C., et al. (2021) Dynamic sex chromosome expression in Drosophila male germ cells. Nat Commun, 12, 892.

Mahajan, S. & Bachtrog, D. (2017) Convergent evolution of Y chromosome gene content in flies. Nat Commun, 8, 785.

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

Mongue, A.J., Hansen, M.E. & Walters, J.R. (2022) Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages*. Evolution, 76, 332–345.

Muller, H.J. (1914) A gene for the fourth chromosome of Drosophila. Journal of Experimental Zoology, 17, 325–336.

Nagylaki, T. (1979), Selection in dioecious populations. Annals of Human Genetics, 43: 143-150.

Nei, M. (1970) Accumulation of nonfunctional genes on sheltered chromosomes. The American Naturalist, 104, 311–322.

Nozawa, M., Minakuchi, Y., Satomura, K., Kondo, S., Toyoda, A. & Tamura, K. (2021) Shared evolutionary trajectories of three independent neo-sex chromosomes in Drosophila. Genome Res, 31, 2069–2079.

Nozawa, M., Onizuka, K., Fujimi, M., Ikeo, K. & Gojobori, T. (2016) Accelerated pseudogenization on the neo-X chromosome in Drosophila miranda. Nat Commun, 7, 13659.

Olito, C., Ponnikas, S., Hansson, B. & Abbott, J.K. (2022) Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes. Evolution, 76, 1320–1330.

Orr, H.A. & Betancourt, A.J. (2001) Haldane's sieve and adaptation from the standing genetic variation. Genetics, 157, 875–884.

Patten, M.M. (2019) The X chromosome favors males under sexually antagonistic selection. Evolution, 73, 84–91.

Pinharanda, A., Rousselle, M., Martin, S.H., Hanly, J.J., Davey, J.W., Kumar, S., et al. (2019) Sexually dimorphic gene expression and transcriptome evolution provide mixed evidence for a fast-Z effect in Heliconius. Journal of Evolutionary Biology, 32, 194–204.

Ponnikas, S., Sigeman, H., Abbott, J.K. & Hansson, B. (2018) Why do sex chromosomes stop recombining? Trends Genet, 34, 492–503.

Presgraves, D.C. & Orr, H.A. (1998) Haldane's rule in taxa lacking a hemizygous X. Science, 282, 952–954.

Pucholt, P., Wright, A.E., Conze, L.L., Mank, J.E. & Berlin, S. (2017) Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. Molecular Biology and Evolution, 34, 1991–2001.

Radhakrishnan, S. & Valenzuela, N. (2017) Chromosomal context affects the molecular evolution of sex-linked genes and their autosomal counterparts in turtles and other vertebrates. Journal of Heredity, 108, 720–730.

Rice, W.R. (1984) Sex chromosomes and the evolution of sexual dimorphism. Evolution, 38, 735–742.

Rousselle, M., Faivre, N., Ballenghien, M., Galtier, N. & Nabholz, B. (2016) Hemizygosity enhances purifying selection: lack of Fast-Z evolution in two Satyrine butterflies. Genome Biology and Evolution, 8, 3108–3119.

Rupp, S.M., Webster, T.H., Olney, K.C., Hutchins, E.D., Kusumi, K. & Wilson Sayres, M.A. (2017) Evolution of dosage compensation in *Anolis carolinensis*, a reptile with XX/XY chromosomal sex determination. Genome Biology and Evolution, 9, 231–240.

Ruzicka, F. & Connallon, T. (2020) Is the X chromosome a hot spot for sexually antagonistic polymorphisms? Biases in current empirical tests of classical theory. Proceedings of the Royal Society B: Biological Sciences, 287, 20201869.

Ruzicka, F. & Connallon, T. (2022) An unbiased test reveals no enrichment of sexually antagonistic polymorphisms on the human X chromosome. Proceedings of the Royal Society B: Biological Sciences, 289, 20212314.

Ruzicka, F., Hill, M.S., Pennell, T.M., Flis, I., Ingleby, F.C., Mott, R., et al. (2019) Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. PLOS Biology, 17, e3000244.

Sackton, T.B., Corbett-Detig, R.B., Nagaraju, J., Vaishna, L., Arunkumar, K.P. & Hartl, D.L. (2014) Positive selection drives Faster-Z evolution in silkmoths. Evolution, 68, 2331–2342.

Sardell, J.M., Josephson, M.P., Dalziel, A.C., Peichel, C.L. & Kirkpatrick, M. (2021) Heterogeneous histories of recombination suppression on stickleback sex chromosomes. Molecular Biology and Evolution, 38, 4403–4418.

Veltsos, P., Ridout, K.E., Toups, M.A., González-Martínez, S.C., Muyle, A., Emery, O., et al. (2019) Early sex-chromosome evolution in the diploid dioecious plant *Mercurialis annua*. Genetics, 212, 815–835.

Vicoso, B. (2019) Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol, 3, 1632–1641.

Vicoso, B. & Charlesworth, B. (2006) Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet, 7, 645–653.

Vicoso, B. & Charlesworth, B. (2009) Effective population size and the Faster-X effect: An extended model. Evolution, 63, 2413–2426.

Wright, A.E., Darolti, I., Bloch, N.I., Oostra, V., Sandkam, B., Buechel, S.D., et al. (2017) Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nat Commun, 8, 14251.

Wright, A.E., Dean, R., Zimmer, F. & Mank, J.E. (2016) How to make a sex chromosome. Nat Commun, 7, 12087.

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

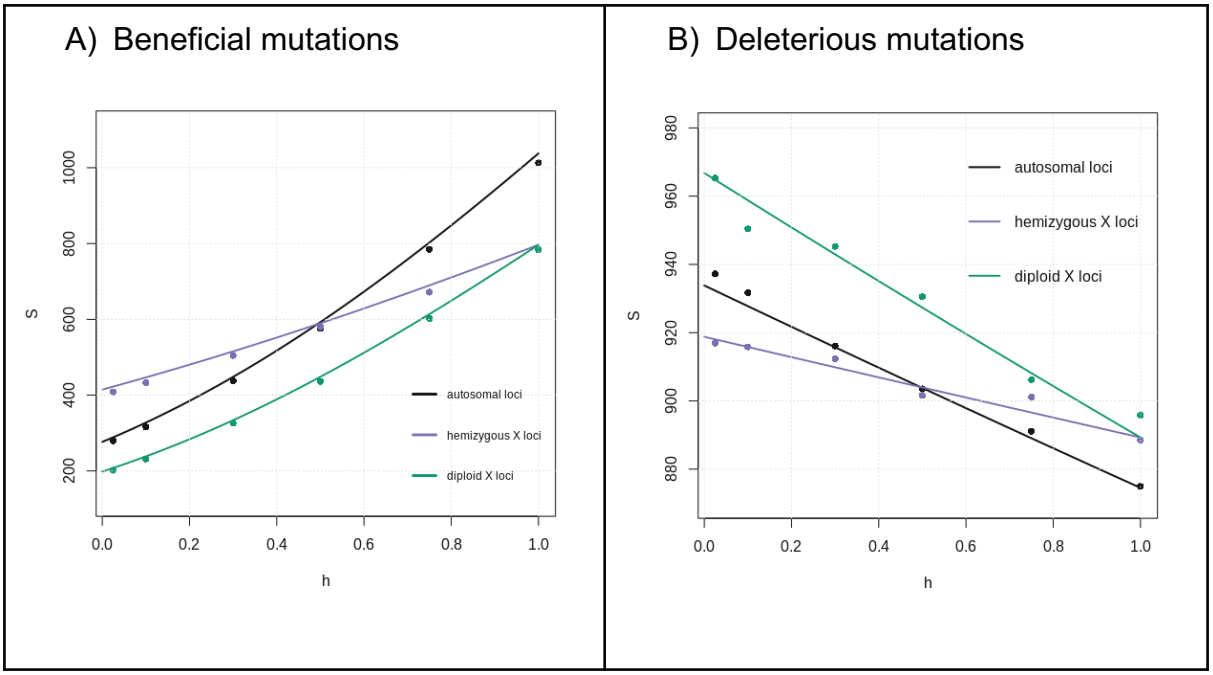
Zhou, Q. & Bachtrog, D. (2012) Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science, 337, 341–345.

Supplementary material and methods

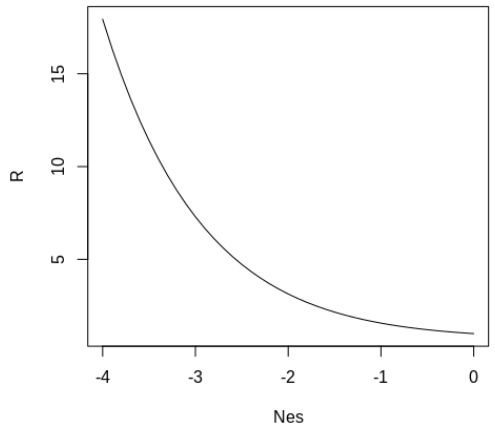
Individual-based simulations

To confirm our inference of fixation probabilities and resulting substitution rates using diffusion approximation, we ran individual-based forward Wright-Fisher simulations, using the simulation package SLiM, version 3.6 (Haller and Messer, 2019). We simulated three cases: autosomal loci, hemizygous X-linked loci and diploid X-linked loci. We simulated 100 diploid individuals with separate sexes and an equal sex ratio. Each simulated genome had 100 loci. In generation zero, all the individuals are homozygous (or hemizygous) at all the loci for ancestral alleles and have the same fitness. Beneficial mutations with selective effects given in Table 1, occur at a rate of 10^{-6} per generation per locus and deleterious mutations occur at rate 10^{-5} per generation per locus. When simulating sex chromosomes, mutations were set to occur only on the X chromosome to represent the diploid X-linked locus case from Table 1., where the corresponding locus on the Y is fixed for an ancestral allele, while when simulating hemizygous X-linked loci, the Y chromosome was assumed to be empty/degenerated. Every individual contributes to the gamete pool of the following generation with the probability proportional to individual fitness. Individual fitness can be calculated as the product of the fitness effects of all the loci. The recombination rate was set to 0.5 to simulate free recombination between the loci. There was no recombination between the X and Y chromosomes. Each simulation was run for 1005000 generations, and the number of substitutions in the last 1 million generations was tracked. We only counted the substitutions occurring in the last 1000000 generations as it takes approximately $4N_e$ for a mutation to fix, so it

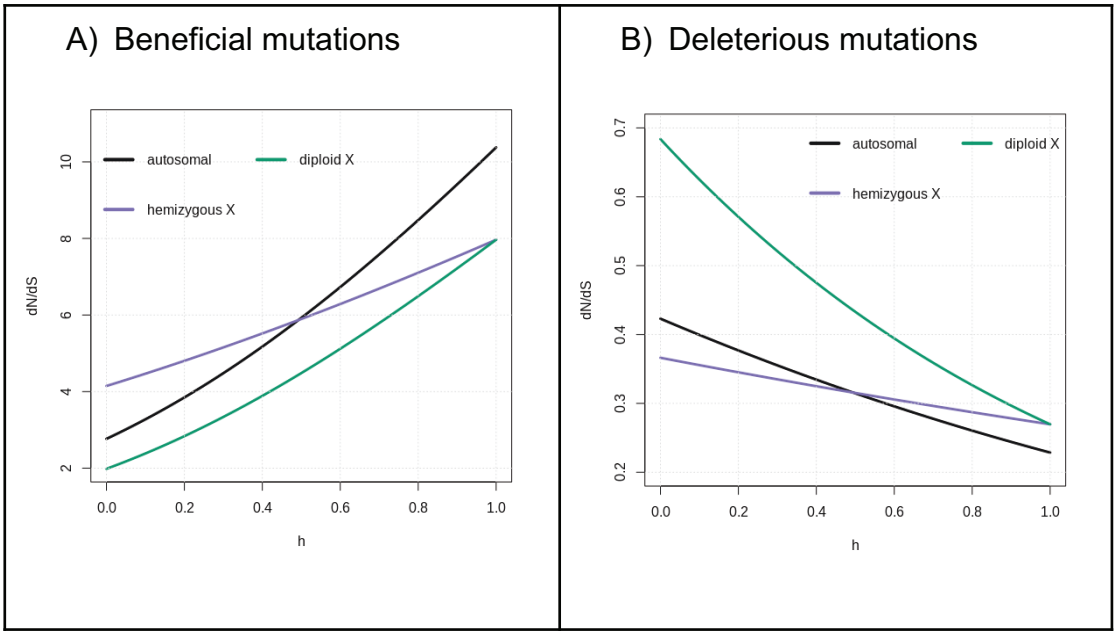
takes approximately 400 generations to reach a steady substitution rate in the simulation, and we filtered the first 5000 to be on the safe side. We ran separate simulations for different values of selection coefficient, *s*, corresponding to beneficial and deleterious mutations, and a range of dominance coefficients, *h*. For each of the cases, we ran 100 simulation replicates. The simulation code is available at https://git.ist.ac.at/amrnjava/x-chromosome-theory/-/blob/main/README.md.



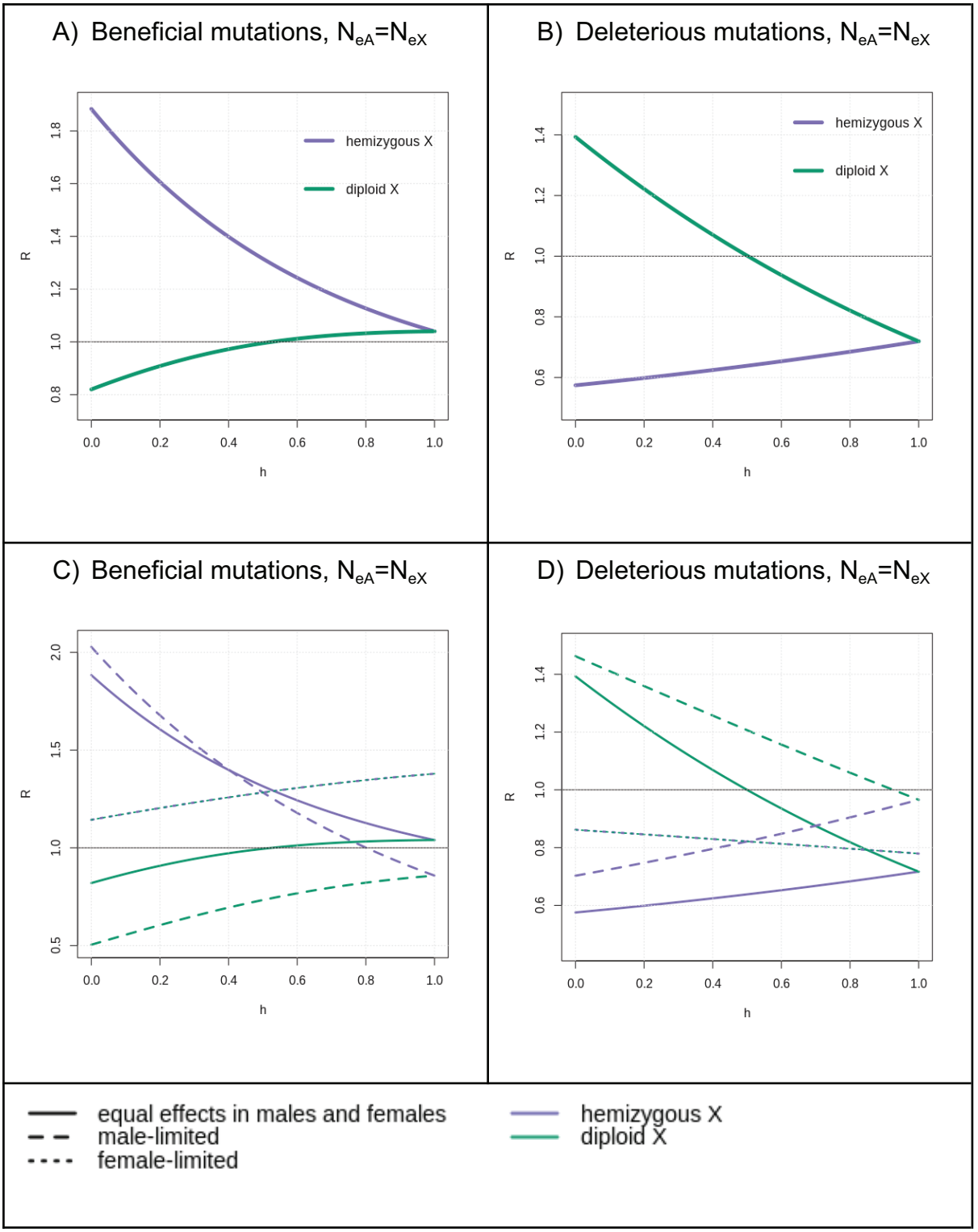
S.Figure 1 The mean number of beneficial (A) and deleterious (B) substitutions in simulations, for 100 loci in 1000000 generations, as a function of dominance coefficient *h*, for autosomal, hemizygous X-linked loci and diploid X-linked loci (points), plotted along the theoretical predictions (lines) for the number of substitutions with corresponding parameters. For beneficial mutations $N_{eA}s=3$, and for deleterious $N_{eA}s=-0.1$. The maximum percentage error between the theoretical predictions and simulation replicate means is 3.5% for autosomal loci, 2.9% for diploid X loci and 3.4% for hemizygous X loci, for beneficial mutations; and 0.5% for autosomal loci, 0.9% for diploid X loci and 0.5% for hemizygous X loci, for deleterious mutations.



S.Figure 2. Ratio of substitution rates (R) at diploid X loci compared to autosomal loci, for recessive (h=0.1), deleterious mutations, as a function of selective effect of a mutation $N_e s$. The ratio exponentially increases as deleterious effect increases.



S.Figure 3. Expected dN/dS values for beneficial ($N_{eA}s=3$) and deleterious ($N_{eA}s=-1$) mutations for diploid X, hemizygous X and autosomal loci.



S.Figure 4. Ratios of substitution rates (*R*) on hemizygous X to autosomal loci and diploid X to autosomal loci as functions of dominance coefficient *h*, for beneficial, $N_{eA}s$= 3 (A and C) and deleterious, $N_{eA}s$= -1 mutations (B and D), where $N_{eX}=N_{eA}$. Mutations with equal effects in males and females are plotted in A and B, and in addition, in C and D, male-specific and female-specific mutations were plotted as

well. Here we can disentangle the effect of sheltering and the effect of reduced N_e, and fitness effects in males and females separately on the X/A ratio of substitution rates.

Chapter 3: Effects of sheltering on sex chromosome degeneration and sex-biased gene content evolution

Andrea Mrnjavac[1], Beatriz Vicoso[1], Tim Connallon[2*]

[1] Institute of Science and Technology Austria, Klosterneuburg, Austria

[2] School of Biological Sciences, Monash University, Clayton, Victoria, Australia

Author contributions: Tim Connallon and Andrea Mrnjavac designed the study. TC and AM performed the analysis together for the single-locus dynamics. TC performed the analysis on dominance (Appendix 1) and multi-locus dynamics (Selective interference and fixation of Y-linked LOF alleles section, Figure 4. and Appendix 4). Beatriz Vicoso and AM performed the sex-bias survey (Figure 5. and Supplementary methods). TC and AM wrote the initial version of the manuscript. TC wrote the revised version of the manuscript, with comments from BV and AM.

Abstract

Sex chromosome systems have arisen many times, independently, yet they exhibit many common features. Y chromosomes often evolve extensive gene losses and become enriched for genes important for male fertility, whereas X chromosomes evolve modest gene losses and, in some cases, become enriched for genes primarily expressed by females. These features are thought to reflect an array of evolutionary mechanisms that include selective interference between nonrecombining Y-linked variants, gene expression coevolution between sex chromosomes, and inter-chromosomal gene traffic. Here, we argue that these features are also consistent with a simple model of 'sheltering'. Sheltering, as originally proposed by Muller (1914), refers to the evolutionary accumulation of recessive loss-of-function (LOF) mutations

in Y-linked regions that are permanently heterozygous with the X. This hypothesis was once popular but fell out of favour in the 1970s because early mathematical models failed to support Muller's intuition and available data on dominance and dosage compensation were viewed as incompatible with the hypothesis. We reconsider the major arguments against Muller's model, given modern data on dominance and dosage compensation, and develop an extended model of sheltering that relaxes restrictive assumptions of earlier models. Our results suggest that sheltering might contribute substantially to sex chromosome evolution, including commonly observed patterns of sex-linked gene losses and enrichments for female- and male-biased genes on the X and Y. Sex chromosomes are undoubtedly influenced by a wide range of evolutionary factors. We argue that sheltering should be seriously considered as a potentially important one.

Introduction

Each new X and Y chromosome pair is undifferentiated and carries an identical set of genes, yet this initial state is typically transient (Charlesworth 1996; Bachtrog 2013; Bachtrog et al. 2014; for exceptions, see: Stöck et al. 2011; Kamiya et al. 2012; Vicoso et al. 2013). The suppression of recombination between the X and Y—which either evolves or occurs naturally in cases where males do not recombine—initiates a cascade of evolutionary changes that ultimately yield predictable patterns of differentiation between the sex chromosomes. Much of this differentiation involves gene losses from the Y (Bachtrog 2013; Abbott et al. 2017; Furman et al. 2020), though recent studies on *Drosophila* and mammals also report gene losses from the X (Nozawa et al. 2016, 2021; Hughes et al. 2022). Beyond their conspicuous disparities in gene numbers, the X and Y also evolve differences in the *types* of genes that they carry, with the Y typically becoming enriched for genes involved in male fertility (Skaletsky et al. 2003; Bellot et al. 2014; Mahajan and Bachtrog 2017; Nozawa et al. 2021; Shaw and White 2022; Wei et al. 2024), and the X becoming enriched for genes preferentially expressed by females ('female-biased genes'; Ellegren 2011; Meisel et al. 2012; Albritton et al. 2014; Papa et al. 2017; Foster et al. 2020; Hu et al. 2022; Lasne et al. 2023; Mora et al. 2024).

Several processes are thought to explain these patterns of sex chromosome gene content. Y-linked gene losses are widely attributed to selective interference (i.e., Hill-Robertson effects), in which natural selection becomes overwhelmed by genetic drift in genomic regions that lack recombination (Felsenstein 1974; Charlesworth 1978; Bachtrog 2013). Under this view, the lack of crossing over between the X and Y hinders the evolutionary removal of deleterious mutations and the fixation of beneficial

mutations, resulting in a gradual decay of functional Y-linked genes, and a retention of their homologs on the X. Divergence between the X and Y in the proportions of sex-biased genes that they carry is potentially influenced by several factors (Vicoso and Charlesworth 2006), including chromosomal differences in the rates at which male- and female-biased genes evolutionarily accumulate through gene duplication (Connallon and Clark 2011), translocations or centric fusions between sex chromosomes and autosomes (Charlesworth and Charlesworth 1980; Pennell et al. 2015), *de novo* gene formation (Begun et al. 2007), and gene regulatory divergence leading to sex-biased gene expression (Rice 1984; Connallon and Clark 2010). Regulatory constraints associated with dosage compensation and meiotic X inactivation—both of which occur in mature sex chromosome systems with a highly differentiated X and Y—might further contribute by making the X a transcriptionally inhospitable environment for male-biased genes (Vibranovski et al. 2009; Bachtrog et al. 2010; Meisel et al. 2022). Notably, while each of the above scenarios can influence specific features of sex chromosome evolution, several must act concurrently to broadly explain the patterns of gene content observed on the X and Y.

Here, we argue that an extended version of Muller's classic 'sheltering' hypothesis (Muller 1914, 1918; Muller and Painter 1932) can contribute to both the gene loss and sex-biased gene content evolutionary patterns of many sex chromosome systems. Muller's hypothesis—the first influential explanation for Y chromosome gene decay—fell out of favour during the 1970s and is now disregarded. We therefore begin by presenting a brief history of the hypothesis along with the major arguments that led to its rejection. We then re-evaluate the arguments that were raised against the sheltering hypothesis and find that they are not conclusive, particularly in view of contemporary data on gene expression, dosage compensation, and the fitness

effects of loss-of-function mutations. Finally, we present an extended mathematical model that relaxes restrictive assumptions of earlier sheltering models. Our extension shows—in contrast to earlier models—that sheltering provides a viable and potentially important mechanism for Y- and X-linked gene losses and concurrent enrichment for different categories of sex-biased genes. We clarify conditions under which sheltering should be an important factor in sex chromosome evolution and highlight several empirical findings that are consistent with the model.

The rise and fall of Muller's sheltering hypothesis

In 1914, Herman Muller (co-crediting fly-room colleague Alfred Sturtevant) proposed the "sheltering hypothesis" for Y chromosome degeneration (Muller 1914; Muller 1918; Muller and Painter 1932; Fisher 1935; Nei 1970). Muller argued that an ancestral Y chromosome—which contains the same set of genes but lacks recombination with the ancestral X—should degenerate over time through the accumulation of recessive Y-linked mutations. If, as he predicted, Y-linked loss-of-function mutations are invariably heterozygous with functional X-linked copies of the same genes, then the fitness effects of the Y-linked variants will be sheltered from selection. Such mutations can, therefore, accumulate and become fixed on the Y chromosome, which decays in the number of functional genes it carries.

Muller's hypothesis for Y chromosome degeneration nevertheless fell out of favour during the mid-to-latter half of the 20th Century, and when mentioned today (which is rare), is usually summarily dismissed as little more than a historical note (*e.g.*, Charlesworth 1978, 1991; Rice 1996; Orr and Kim 1998; Abbott et al. 2017; Vicoso

2019). The eventual fall in popularity of the sheltering hypothesis was due to three main arguments. The first, and the most convincing, was that formal population genetic models undermined Muller's intuition about the evolutionary consequences of sheltering. Muller correctly noted that Y-linked mutations can be masked by functional copies of the same genes on the X, yet he erred in assuming that such sheltering effects would necessarily be strong enough to cause Y-linked gene degeneration. Using a deterministic model, Fisher (1935) showed that recessive lethal mutations cannot reach high frequencies on the Y chromosome because similar mutations also arise on the X and render Y-linked sheltering effects incomplete. In fact, when X- and Y-linked mutation rates are equal, lethal alleles evolve to identical equilibrium frequencies on the X and Y, remaining rare on both chromosomes. Extensions of the model to include male-biased mutation rates (Fisher 1935) and inbreeding (Frota-Pessoa and Aratangy 1968) permit lethal alleles to reach higher frequencies on the Y than the X, though these conditions are still not sufficient for lethals to become fixed. Finally, Nei (1970) explored the effect of genetic drift on the fixation rates of Y-linked lethals and found that the population size must be exceptionally small for fixation to be likely. Collectively, these early attempts to formally model sheltering convinced many of the inadequacy of the mechanism (Charlesworth 1978, 1991; Rice 1996; Orr and Kim 1998).

Charlesworth (1978) raised two further arguments against the sheltering hypothesis, which revolved around the empirical evidence (at the time) for dosage compensation of the X chromosome and the dominance of deleterious mutations. Sheltering effects are strongest when mutations are completely recessive (*i.e.*, when they have dominance coefficients of $h = 0$). Incompletely recessivity (dominance coefficients within the range: $0 < h < 0.5$) amplifies the efficiency of purifying selection

and reduces the capacity of sheltering effects to influence Y chromosome evolution (Nei 1970). Mutation accumulation data available during the 1970s clearly showed that the *average* dominance coefficient for mildly deleterious mutations was within the range $0.1 < h < 0.5$ (Simmons and Crow 1977; Crow 1993), indicating some expression in heterozygotes. Although estimates of the dominance of loss-of-function mutations were unavailable at that time, estimates of the mean dominance of lethal mutations (as possible proxies for loss-of-function alleles) suggested strong yet incomplete recessivity. Most of the mutation data available at the time were insufficient for estimating other aspects of the distribution of dominance (Halligan and Keightley 2009). Nevertheless, the evidence that mutations were only partially recessive, on average, reinforced the view that sheltering was unlikely to be important in sex chromosome evolution (Charlesworth 1978, 1991; Rice 1996; Orr and Kim 1998).

The final argument is based on the observation that species with highly degenerate Y chromosomes often evolve dosage compensation, which implies that there must be a fitness cost to males of losing genes on the Y (Charlesworth 1978, 1991). This argument rules out sheltering as a universal explanation for Y chromosome degeneration in species that have evolved any degree of dosage compensation, which requires that at least some Y-linked gene losses reduce male fitness. Yet such observations do not rule out significant contributions of sheltering to sex chromosome evolution alongside other processes (*e.g.*, selective interference; Charlesworth 1978; Bachtrog 2013) that do favour the evolution of dosage compensation.

A reassessment of the major arguments against the sheltering hypothesis

A contemporary reassessment of data on dosage compensation, dominance, sex-specific transcription, and the fitness effects of loss-of-function (LOF) mutations, suggests that the sheltering hypothesis might have been prematurely dismissed. The three main arguments against sheltering—based on dosage compensation, dominance, and early theoretical models—are suggestive but far from damning. Below, we address each in turn.

Arguments based on observations of dosage compensation. The argument against sheltering from observations of dosage compensation has not fared well. While older model systems like *Drosophila*, mammals and nematodes suggested that dosage compensation was the norm, we now know that many species have highly degenerate Y (or W) chromosomes, yet have not evolved complete dosage compensation, which can instead be partial, gene-by-gene, and far from complete (Gu and Walters 2017; Zhu et al. 2024). Incomplete dosage compensation implies that dosage-related fitness costs of Y- or W-linked gene losses might often be negligible. In cases where full dosage compensation does evolve, it merely implies that some, though not necessarily all, gene losses are costly in heterozygous state. This of course tells us nothing about the proportion of loss-of-function mutations that might incur such costs.

Arguments based on estimates of dominance. The average deleterious mutation is not completely recessive (the mean dominance coefficient for mildly deleterious mutations is roughly $\bar{h} \approx 0.25$; Manna et al. 2011), yet this fact tells us little about the broader distribution of dominance for all mutations or specific mutational categories (Halligan and Keightley 2009). For example, classical studies of *Drosophila*

indicate that the average dominance coefficient for lethal mutations is within the range $0.01 < \bar{h} < 0.05$ (Simmons and Crow 1977; Crow 1993), though this still leaves open the possibility that a meaningful fraction could be completely recessive or overdominant with respect to fitness. As Nei (1970) aptly points out:

> *This, of course, does not mean that there are no completely recessive or overdominant lethals. On the contrary, it seems that there are lethals with varying degrees of dominance from slight overdominance to a rather high degree of partial dominance. If this is the case, those genes which are overdominant or completely recessive would be fixed in the population rather quickly but the others would be fixed only slowly, depending on the degree of dominance and population size.*

In a recent review of the evolution of homozygous lethal mutations, Marion and Noor (2023) similarly caution against conflating mean dominance—which has been estimated in *Drosophila* and other model organisms—with the distribution of dominance, which has not. For example, the classical *Drosophila* data provide estimates of the fitness effects of lethal-bearing chromosomes, each carrying one or more lethal alleles whose individual dominance effects cannot be assessed (Marion and Noor 2023).

A few studies have reported estimates of higher moments of the distribution of dominance for lethal mutations (*e.g.*, the variance), and these provide information about the types of distributions that might be compatible with data on dominance. Yoshikawa and Mukai (1970) reported point estimates of the mean and variance of dominance for lethal mutations accumulated on 2nd chromosomes of *D. melanogaster* ($\bar{h} = 0.027$ and $\sigma_h^2 = 0.0027$, respectively). If we were to assume a gamma or a beta distribution for *h* (and thus constrain dominance coefficients to be positive: $h > 0$), these estimates of $\bar{h}$ and σ_h^2 imply a right-skewed distribution with a high proportion of nearly recessive lethals (~60% have $h < 0.01$; see Appendix 1). This interpretation aligns with other *Drosophila* data showing that the mean dominance of *segregating*

lethal alleles is significantly lower than the mean dominance of new lethal mutations (*e.g.*, Crow 1991, 1993), which requires substantial variation in the dominance coefficients of new lethal mutations.

Given that most genes are not essential (Rancati et al. 2018), and lethal mutations do not necessarily confer loss-of-function (Marion and Noor 2023), it is questionable whether studies of the dominance of lethal mutations are representative of loss-of-function alleles in general. More recent, targeted deletion datasets permit direct estimates of the fitness effects of individual loss-of-function alleles. Using yeast data for non-essential genes, Agrawal and Whitlock (2011) estimated the mean, variance, and skew of the distribution of dominance for whole-gene deletions. Among deletions with relatively large homozygous effects (whose fitness effects can be reliably estimated; see Manna et al. 2012), the distribution of dominance is right skewed with mean dominance near zero (see Fig. 5 of Manna et al. 2012, which reanalyses the top ~20% largest-effect deletions from Agrawal and Whitlock's dataset), which is consistent with a high proportion of deletions exhibiting complete recessivity. Furthermore, a recent study in *Arabidopsis* shows that more deleterious mutations tend to be more recessive than less deleterious mutations (Huber et al., 2018).

Modern population genomics has provided new opportunities for evaluating the fitness effects of deleterious mutations (Eyre-Walker and Keightley 2007), including protein-truncating variants (a type of loss-of-function allele). While assessing the dominance of LOF mutations and Mendelian disease alleles remains a challenge (see Fuller et al. 2019; Balick et al. 2022), complete recessivity appears to be plausible for a large fraction of genes. For example, Mendelian disease mutations are often close enough to recessive to be categorized as such (Blekhman et al. 2008).

Haplosufficiency is widespread across a range of species (e.g., yeast to humans; Balick et al. 2022; Zschocke et al. 2023), consistent with LOF mutations often having recessive effects. Finally, the frequencies of homozygous lethal disease alleles in human populations are higher than expected if they were expressed in heterozygotes (Amorim et al. 2017), suggesting that such alleles could be completely recessive or mildly overdominant, though ascertainment bias might also contribute to inflated frequencies of these alleles (Amorim et al. 2017; Marion and Noor 2023). Finally, while the average number of derived nucleotide variants is remarkably stable among human populations, frequencies of LOF alleles exhibit strong sensitivity to founder events and population bottlenecks, as expected if a substantial fraction is completely recessive (Simons and Sella 2016).

Arguments based on mathematical models of sheltering. While previous models of sheltering suggest that the conditions for Y-linked gene loss are narrow, this conclusion could reflect two restrictive assumptions of these models. Firstly, the models largely focus on the potential for fixation of homozygous lethal mutations, yet most genes in multicellular Eukaryotes appear to be non-essential (i.e., their LOF alleles are not lethal; Rancati et al .2018) and conditions for their loss through sheltering should be more permissive. Secondly, the models invariably assume that LOF mutations incur equal fitness costs when expressed by each sex. While such an assumption is usually reasonable for lethal mutations (Ashburner et al. 2005) it is unlikely to hold for LOF mutations in general. For example, we know that mutations conferring sterility in one sex do not typically cause sterility in the other (Lindsley and Lifschytz 1972), and that the fitness effects of spontaneous mutations are often sexually dimorphic, particularly so with respect to their effects on adult fitness components (Mallet et al. 2011; Sharp and Agrawal 2013). Two decades of

transcriptomics research clearly shows that large fractions of the genome exhibit sex-biased gene expression (Ellegren and Parsch 2007; Parch and Ellegren 2013; Grath and Parsch 2016). This sexual dimorphism in gene expression is at least somewhat indicative of sex differences a gene's functional importance. Genes with sex-limited expression appear to be common within animal genomes (see our Discussion) and, by definition, must have sex-limited fitness effects. Likewise, mutations in sex-biased genes often exhibit sexually dimorphic phenotypic effects (Connallon and Clark 2011; van der Bijl W, Mank JE. 2021). It seems likely that LOF alleles of sex-biased genes will exhibit asymmetric fitness costs in each sex, though this has not yet been systematically tested.

Recent theory has shown that X-linked deleterious mutations—particularly mutations with male-biased fitness costs—can be strongly sheltered from selection when the Y is fixed for functional (wild-type) copies of the same genes (Mrnjavac et al. 2023). What remains to be shown is how segregating LOF mutations on the X and Y will interact to determine the evolutionary fates of X- and Y-linked genes in cases where LOF mutations differentially affect the fitness of each sex. We turn to this issue, below.

An extended model of X and Y chromosome sheltering

We consider the evolutionary accumulation of loss-of-function (LOF) mutations within homologous regions of an X and Y chromosome pair that does not recombine, and which initially carries the same set of functional genes (for full details of the model, see Appendix 2). Our focus on LOF mutations reflects our interest in modelling gene losses, along with the empirical observation that LOF alleles are the most likely

variants to be recessive. Because there are many ways a gene can lose function, from single-nucleotide mutations that introduce premature stop codons to deletions or insertions that disrupt the protein or its regulatory sequences, LOF mutation rates per gene are expected to be relatively high (*e.g.*, with rates on the order of μ ~10^{-4} to 10^{-6}; Drake et al. 1998; Monroe et al. 2021; Balick et al. 2022). Our most important departure from earlier models of sheltering (Fisher 1935; Frota-Pessoa and Aratangy 1968; Nei 1970) is that we allow mutations to differentially affect female and male fitness. We also study the effects of hitchhiking (via selective sweeps or background selection) on the fixation of sheltered LOF alleles, which previous theory has not considered.

Deterministic evolutionary dynamics of LOF mutations on the X and Y chromosome

We begin by considering a deterministic model of recurrent mutation and selection at single genes, and later consider the effects of drift and hitchhiking on the fixation of LOF alleles. We assume that functional copies of each gene mutate to LOF alleles at rates μ_f and μ_m in females and males, respectively. Back-mutation is ignored and likely to be negligible for LOF alleles. LOF alleles of a given gene reduce fitness by s_f and s_m in female and male homozygotes ($0 \leq s_f, s_m \leq 1$) and by $s_f h$ and $s_m h$ in heterozygotes, where h is the dominance coefficient ($0 \leq h \leq 1$, with $h = 0$ corresponding to complete recessivity). We assume throughout that the sex-averaged strength of selection against LOF mutations is much stronger than sex-averaged mutation rates ($\bar{s} \gg \bar{\mu}$, where $\bar{s} = \left(s_f + s_m\right)/2$ and $\bar{\mu} = \left(\mu_f + \mu_m\right)/2$). These conditions prevent gene losses on autosomes, though as we shall see, they are permissive of sex chromosome gene losses.

When mutations are recessive and fitness effects are equally strong in each sex (h = 0 and $s_f = s_m$), we recapture the results of earlier sheltering models. In this case, LOF mutations remain rare on both the X and the Y. As previously noted by Fisher (1935), the mutation-selection equilibria are identical between the X and Y when mutation rates are also equal between the sexes (*i.e.*, $\hat{p}_X = \hat{p}_Y$ when $\mu_f = \mu_m$ and $s_f = s_m$, where $\hat{p}_X$ and $\hat{p}_Y$ represent the equilibrium LOF allele frequencies on the X and Y, respectively; see Fig. 1). Male-biased mutation rates within the range that is typically observed in animals (*i.e.*, $\mu_f \leq \mu_m \leq 4\mu_f$; Connallon et al. 2022) elevate LOF frequencies on the Y relative to the X, though LOF alleles remain rare on both chromosomes.

Sexually dimorphic fitness costs allow LOF mutations to differentially accumulate on the X or Y (Fig. 1). LOF mutations become enriched on the X chromosome in genes that are disproportionately important for male fitness ($\hat{p}_X > \hat{p}_Y$ for genes where $s_m/s_f > \mu_m/\mu_f$), while LOF mutations preferentially accumulate on the Y in genes that are disproportionately important for female fitness ($\hat{p}_X < \hat{p}_Y$ for genes where $s_m/s_f < \mu_m/\mu_f$). Complete fixation of a LOF allele on the X or the Y (*i.e.*, $\hat{p}_X = 1$ or $\hat{p}_Y = 1$) occurs when sheltering effects are strong and LOF alleles have sufficiently pronounced sexual dimorphism in their homozygous fitness effects. Completely recessive LOF mutations become fixed in X-linked genes with male-limited functions (*i.e.*, $\hat{p}_X = 1$ when h = 0 and $s_f \leq \mu_f$). A modest amount of expression in females ($s_f > \mu_f$) or heterozygous expression (*e.g.*, h = 0.01 in Fig. 1) is sufficient to prevent fixation of X-linked LOF mutations. Conditions for fixation of Y-linked LOF mutations are much more permissive. Genes expressed by both sexes are lost from the Y if they are substantially more important for females than for males, but male-limited expression is not required for fixation (*i.e.*, LOF allele fixation occurs when $s_m <$

$\mu_m\sqrt{s_f/\mu_f}$, which corresponds to the grey shaded regions in Fig. 1). Incomplete recessivity (h > 0) reduces the scope for Y-linked LOF allele fixation (though gene losses remain possible), whereas male-biased mutation rates expand the scope (Fig. 1).

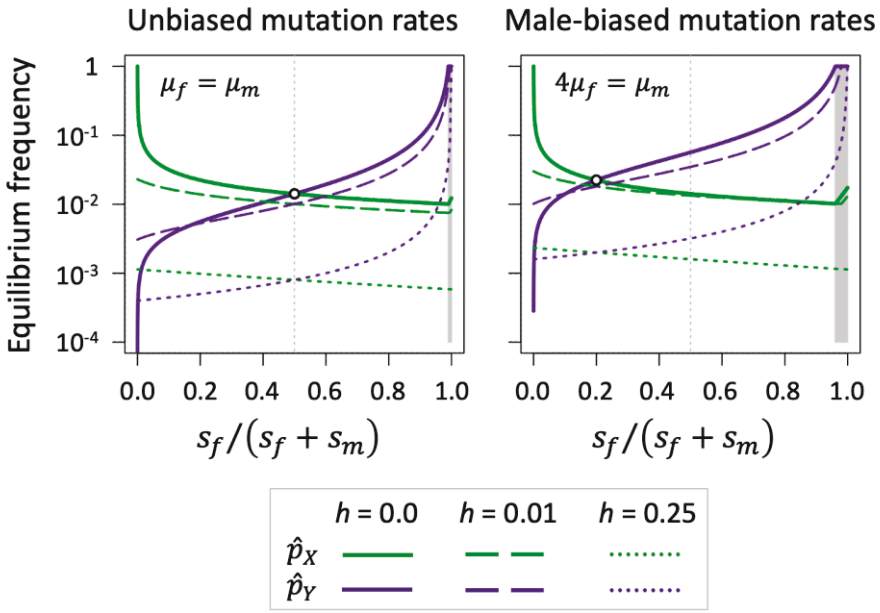


Figure 1. Sexually dimorphic fitness effects of LOF mutations promote sex chromosome gene losses. Each panel shows mutation-selection equilibrium frequencies (in $\log_{10}$ scale) for X- and Y-linked genes under different scenarios of dominance for LOF mutations (completely recessive: h = 0; partially recessive: h = 0.01 and h = 0.25). The x-axis spans the gradient between male-limited genes ($s_f/(s_f + s_m)$ = 0 when s_m > 0 and s_f = 0), genes that with equal fitness effects in each sex ($s_f/(s_f + s_m)$ = 0.5 when $s_m = s_f$) and female-limited genes ($s_f/(s_f + s_m)$ = 1 when s_m = 0 and s_f > 0). The shaded grey regions show parameter space favouring deterministic fixation of recessive LOF mutations on the Y chromosome; the parameter region leading to X-linked gene losses is much narrower and includes genes with male-limited functions. Results show cases where $\bar{s} = 0.5$ and $\mu_f = 10^{-4}$. Equilibria were determined analytically in the cases where LOF mutations were completely recessive (h = 0), and they were otherwise determined numerically.

Effects of genetic drift and sheltering on X and Y chromosome gene losses

We next evaluated effects of genetic drift on LOF allele fixation using a combination of analytical and simulation approaches (see Appendix 3). We first used a diffusion approximation to calculate the fixation probabilities for new Y-linked recessive LOF mutations that enter a population in which X-linked genetic variation segregates at mutation-selection-drift balance. Following Nei (1970), we assumed that X-linked variation evolved in the absence of ancestral Y-linked variants, which allows X-linked mutations to reach higher frequencies than they would if the Y were polymorphic. This assumption reduces sheltering effects for Y-linked variants and the following results should therefore be viewed as a conservative lower bound for the fixation probabilities of Y-linked mutations. Secondly, we carried out full stochastic forward simulations in which recurrent mutation, selection, and genetic drift affect the evolutionary dynamics of LOF alleles at both X- and Y-linked genes.

Genetic drift alters the predictions of our deterministic model in two important ways. First, drift substantially expands the scope for LOF allele fixation on the Y chromosome, owing to its relatively small effective population size (effective population sizes are taken to be $2N_e$ for autosomes, $1.5N_e$ for the X, and $0.5N_e$ for the Y; Hartl and Clark 2007). Y-linked LOF mutations not only fix with high probabilities under parameter conditions leading to deterministic degeneration on the Y, but they also fix under parameter conditions which do not favour deterministic fixation (Fig. 2). This expanded scope for degeneration is pronounced in contexts where the population-scaled LOF mutation rate is small (*e.g.*, $N_e\bar{\mu} \ll 1$). Overall, genes with strongly female-biased functions are consistently lost from the Y. However, LOF mutations that substantially reduce fitness of both sexes are also expected to fix in cases where the population-scaled LOF mutation rate per gene is small (*i.e.*, $N_e\bar{\mu}$ is

small; Fig. 2), which applies to small genes (where $\bar{\mu}$ is relatively small) and lineages with small population sizes.

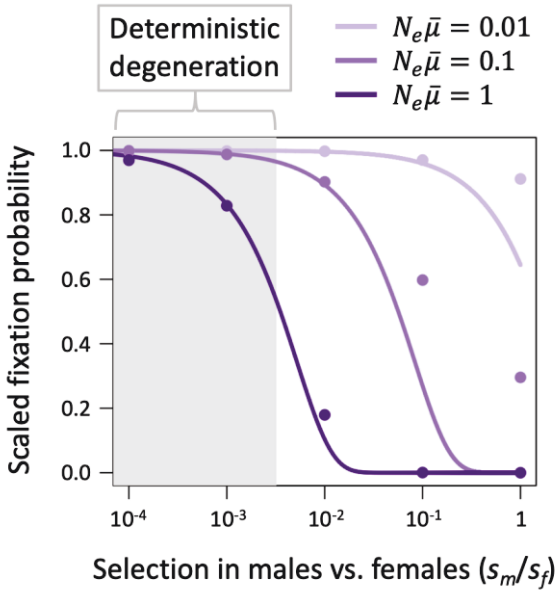


Figure 2. Fixation probabilities of Y-linked LOF mutations depend on interactions between drift and sex differences in selection. The curves show analytical predictions for unique Y-linked mutations entering a population at mutation-selection-drift equilibrium for X, with no ancestral genetic variation on the Y (based on eq. (2) with f_0 = 1). The circles show proportions of Y-linked LOF mutations that fix in replicate computer simulations that incorporate selection and drift on X and Y chromosomes, and recurrent mutation on the X. Each datapoint is based on 10^7-10^9 replicate simulations ending in fixation or loss of the Y-linked variant. All fixation probabilities are scaled relative to those of neutral mutations. The shaded region represents the parameter space leading to deterministic fixation (which corresponds to the condition $s_m < \mu_m\sqrt{s_f/\mu_f}$). Results are shown for a LOF mutation rate of $\bar{\mu} = \mu_f = \mu_m = 10^{-5}$ and three population sizes ($N = N_e = 10^5$, 10^4, and 10^3). LOF alleles are completely recessive (h = 0) and genes are essential for females (s_f = 1).

Second, genetic drift affects the predictability with which male-limited genes are lost from the X versus the Y chromosome (Fig. 3). Male-limited genes with intermediate-to-large population-scaled mutation rates (e.g., $N_e\bar{\mu} > 0.2$) are reliably lost from the X and retained on the Y, consistent with our deterministic predictions. By contrast, small population-scaled mutation rates lead to a mix of male-limited

gene losses from the X and Y, which reflects the strong sheltering effects that arise on both chromosomes in cases where deleterious mutations rarely segregate simultaneously on the X and Y (Fig. 3A). As population-scaled LOF mutation rates approach zero, the proportion of male-limited genes lost from the X versus the Y approaches the limit $f_X = \left(2 + \mu_m/\mu_f\right)/\left(2 + 4\mu_m/\mu_f\right)$, in which at least half of all gene losses are Y-linked ($f_X = 0.5$ when $\mu_m = \mu_f$; $f_X < 0.5$ when $\mu_m > \mu_f$; Fig. 3A).

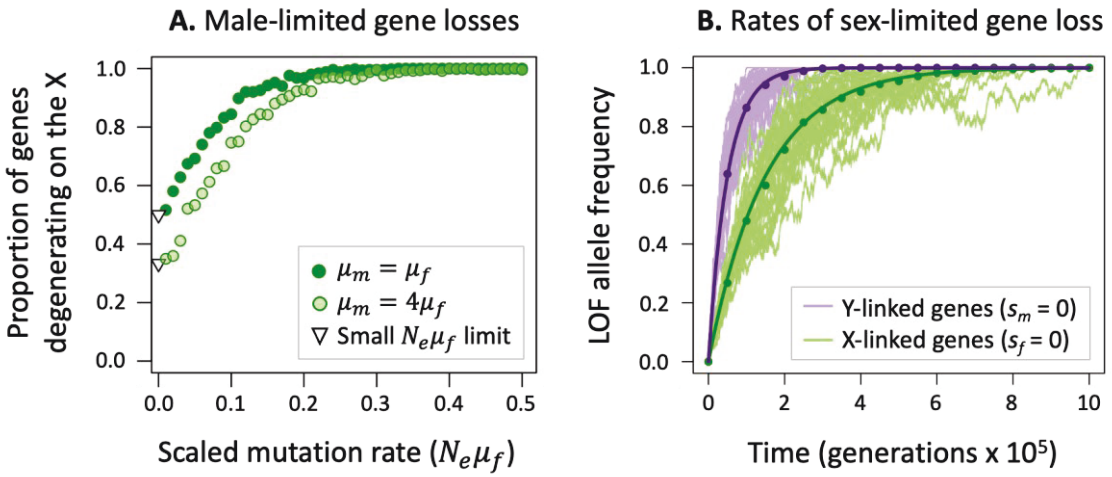


Figure 3. Genetic drift affects the predictability of X- and Y-linked gene loss. Panel A: Genes with male-limited functions eventually degenerate from either the Y or the X chromosome, owing to fixation of recessive LOF alleles. In each of 500 replicate simulations, LOF mutations were initially absent from the X and Y; both chromosomes were permitted to evolve under recurrent mutation, selection and drift, until a LOF allele was fixed on the X or the Y (with $h = 0$, $\mu_f = 10^{-4}$, $s_m = 0.3$, $s_f = 0$, N_e ranging from 10^2 to 10^4). Circles shows the proportion of male-limited genes degenerating from the X but retained on the Y; the remaining proportion degenerate from the Y and retained on the X. Triangles show the limit in which the population-scaled mutation rate approaches zero. **Panel B:** Recessive LOF mutations eventually fix in sex-limited genes, with drift causing variation in evolutionary trajectories of LOF alleles over time. Results show examples of degeneration in a large population ($N_e = 10^5$) with parameters $\mu_f = 10^{-5}$, $\mu_m = 2\mu_f$, $s_f = 1$ and $s_m = 0$ for Y-linked genes, and $s_m = 1$ and $s_f = 0$ for X-linked genes. 30 simulation runs were carried out for each chromosome, with individual trajectories (thin, pale lines) scattered about the analytical predictions (bold curves) and circles denoting mean LOF frequencies across the set of simulated trajectories.

Finally, the evolutionary rates at which sex-limited genes become lost from X and Y chromosomes are both predictable and rapid relative to the age of many sex chromosome systems (Fig. 3B). For male-limited genes on the X, the average frequency of LOF mutations after *t* generations of selection and mutation is approximately $p_{X,t} = 1 - e^{-2\mu_f t/3}$, and for female-limited genes on the Y, the average frequency of LOF mutations is $p_{Y,t} = 1 - e^{-\mu_m t}$ (see Appendix 3). These simple predictions closely match to the averages of simulated LOF allele trajectories under selection, recurrent mutation and drift (Fig. 3B; Fig. S1). These rates of gene loss are rapid enough to cause gene losses over empirically relevant timescales of evolutionary divergence for plant and animal sex chromosome systems—from very old systems that emerged many millions of years ago to neo-sex chromosome systems (*e.g.*, in flies) that are less than a million years old (we elaborate on this point in the Discussion).

Selective interference and fixation of Y-linked LOF alleles

In cases where sheltering, by itself, does not lead to the decay of Y-linked genes, hitchhiking effects can readily fix Y-linked LOF alleles. As in previous sheltering models, our results thus far consider the evolutionary dynamics of LOF alleles at single genes and ignore effects of interference caused by selection at physically linked Y-linked loci. Yet the decline of the sheltering hypothesis coincided with the rise of models of Y chromosome decay due to the hitchhiking effects of selective sweeps and background selection (see Bachtrog 2006, 2013)—processes that should co-occur with the accumulation of LOF alleles by sheltering. We therefore sought to quantify how these forms of selective interference influence the fixation of recessive LOF alleles on the Y. For clarity, we treat each mechanism separately (with derivations

provided in Appendix 4), though both should occur simultaneously in nature (Bachtrog 2008).

In cases where a beneficial mutation arises on a Y chromosome that carries one or more LOF alleles, a selective sweep will fix the beneficial variant and the linked LOF alleles, resulting in Y-linked gene loss. To determine the probability of such an event, we assume that LOF alleles are recessive, the fitness effects of new beneficial mutations are exponentially distributed with a mean of $\bar{s}_b$, and the population size is sufficiently large that the LOF alleles of a gene initially segregate at mutation-selection equilibrium ($\hat{p}_Y = \mu_m/(s_m\hat{p}_X)$ and $\hat{p}_X = \sqrt{\mu_f/s_f}$ on the Y and X, respectively). In the case where Y chromosomes segregate for LOF alleles at a single gene, the probability that the beneficial mutation arises in association with the LOF allele and then carries it fixation is:

$$\Pr(\text{LOF fixes}) \approx 2\bar{s}_b\hat{p}_Y \exp(-s_m\hat{p}_X/\bar{s}_b) \quad (1)$$

This prediction, which closely matches results from stochastic simulations, shows that hitchhiking substantially elevates fixation probabilities of Y-linked LOF alleles relative to fixation probabilities of neutral mutations arising on the Y (the latter being 2/*N*; Fig. 4A). The effect of hitchhiking is pronounced for genes that are more important for female than male fitness (*i.e.*, genes where $s_m/s_f < 1$), though fixation probabilities for LOF alleles remain relatively high in cases where the gene is essential for both sexes ($s_m = s_f = 1$). These results are extended to cases where LOF segregate at multiple loci (see Appendix 4).

Under background selection, weakly deleterious alleles can become fixed by drift when they arise on Y chromosomes that are otherwise free of deleterious mutations. As in the genetic drift model presented above (Fig. 2), we suppose for a given gene that LOF alleles are initially absent from the Y, and they segregate at

mutation-selection-drift balance on the X. Letting f_0 represent the proportion of Y chromosomes that is free of background deleterious genetic variation (*i.e.*, mildly deleterious variants maintained at mutation-selection balance), the probability that a recessive LOF allele becomes fixed on the Y is:

$$\Pr(\text{LOF fixes}) \approx f_0 \frac{\exp\left(2\frac{N_e}{N}s_m\bar{p}_X\right) - 1}{\exp(f_0 N_e s_m \bar{p}_X) - 1} \tag{2}$$

where $\bar{p}_X = \Gamma\left(1.5N_e\mu + \frac{1}{2}\right)\left(\sqrt{N_e s_f}\,\Gamma(1.5N_e\mu)\right)^{-1}$ is the expected frequency of the LOF allele on the X, $\Gamma()$ refers to the gamma function, N_e is the effective population size, and N is the census size (see Appendix 4). Numerical evaluation of eq. (2) shows that background selection, by reducing f_0, expands the conditions under which LOF alleles can fix on the Y (Fig. 4A).

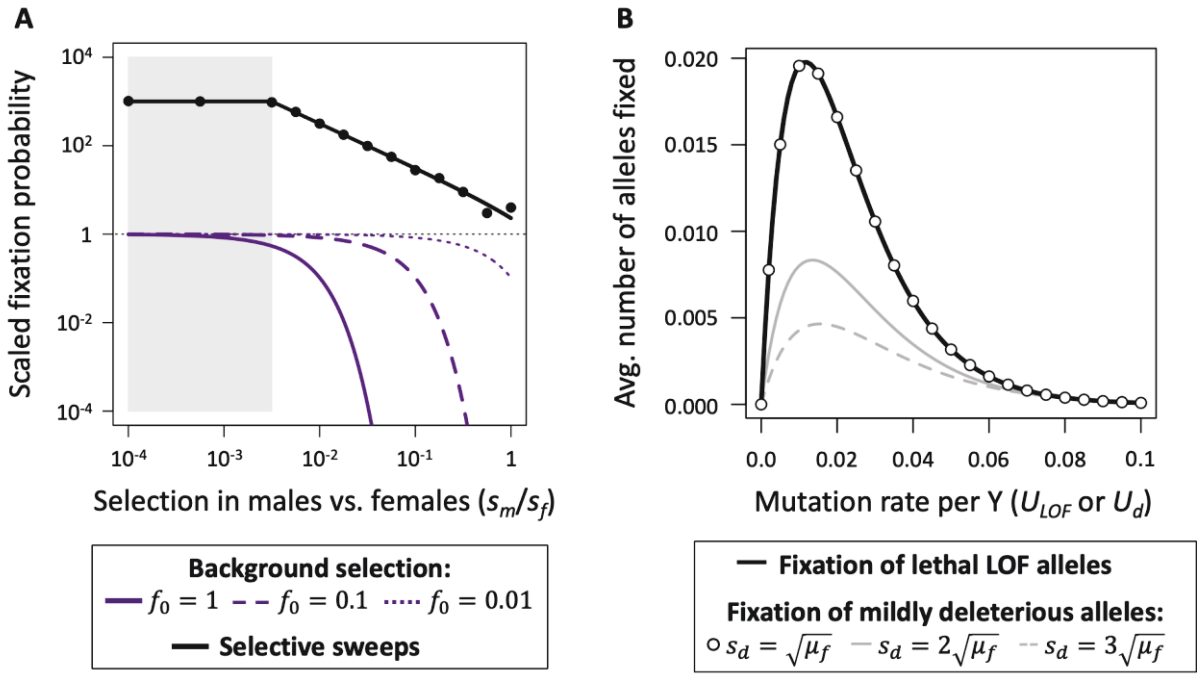


Figure 4. Conditions for Y-linked LOF allele fixation through hitchhiking and background selection. Panel A: The grey shaded region shows the parameter space for deterministic fixation of recessive LOF alleles ($s_m < \mu_m\sqrt{s_f/\mu_f}$). The solid black line (eq. (1)) shows the probability that a LOF allele at a single Y-linked gene hitchhikes to fixation with a beneficial mutation. Circles show the proportion of 10^5 simulated

beneficial mutations that sweep to fixation with a LOF allele. Purple curves (eq. (2)) show the fixation probabilities a new Y-linked LOF mutations under three levels of background selection f_0, which represents proportion of Y chromosomes that are free of deleterious mutations. Fixation probabilities in panel A are scaled relative to those of neutral mutations (*i.e.*, $2/N$ for Y-linked mutations, shown by the horizontal broken black line). **Panel B:** effects of hitchhiking when many deleterious variants segregate simultaneously on the Y chromosome. The black curve shows cases where recessive lethal alleles segregate at mutation-selection balance and potentially hitchhike with a beneficial mutation. The remaining results show cases of hitchhiking of mildly deleterious mutations with heterozygous effects of s_d. Curves are based on equations presented in Appendix 4 of the Supplementary Material. Results use the parameters $\bar{\mu} = \mu_f = \mu_m = 10^{-5}$, $N = N_e = 10^5$, $s_f = 1$, and $\bar{s}_b = 0.01$.

How do the above fixation probabilities of sheltered LOF alleles compare to those of mildly deleterious Y-linked mutations? Each can be understood in relation their effective fitness costs. The effective cost of a mildly deleterious Y-linked variant depends on its heterozygous fitness effect on males (which we define as s_d), whereas the effective cost of a recessive LOF mutation is $s_m p_X$, which depends on its homozygous cost (s_m in males) and the frequency of LOF alleles for the X-linked copy of the gene (p_X). Hitchhiking disproportionately promotes the fixation of sheltered LOF alleles relative to mildly deleterious mutations whenever $s_m p_X < s_d$. For the extreme where the gene is essential for both sexes ($s_f = s_m = 1$) and the population is at mutation-selection equilibrium (at $\hat{p}_X$ and $\hat{p}_Y$), selective sweeps fix LOF alleles more readily than mildly deleterious mutations provided $s_d > \sqrt{\mu_f}$, where μ_f is the genic LOF mutation rate (Fig. 4B; Appendix 4). LOF alleles fix even more permissively when they are more harmful for females than males ($s_f > s_m$), they are non-essential for either sex, and/or drift causes the frequency of X-linked LOF alleles to drop below their mutation-selection equilibria. Comparable results apply under background selection,

where recessive LOF mutations fix more readily than mildly deleterious mutations if $s_d > s_m \bar{p}_X$.

Discussion

We have shown that the sheltering of LOF mutations can promote the evolution of gene losses from X and Y chromosomes, resulting in an enrichment of female-biased genes on the X and male-biased genes on the Y. These consequences of sheltering should occur when three specific conditions are met. First, recombination must be suppressed between homologous regions of the X and Y, which is a common feature of sex chromosomes. Second, loss-of-function (LOF) mutations must often be recessive or nearly so—a scenario that, as we argue above, is compatible with current data on dominance, though the question requires further study. Third, LOF mutations must often exhibit strong sex-biased fitness effects. LOF mutations must be male-limited and completely recessive to reliably fix on the X chromosome. Conditions for fixation are much more permissive for Y-linked genes and include scenarios where LOF alleles are not completely recessive (Fig 1) or where selective interference promotes the fixation of alleles with severe fitness effects in both sexes (*e.g.*, lethals; Fig. 4).

The evolution of separate sexes and sexual dimorphism predates the origin of animal sex chromosomes systems, and genes with sex-specific functions were probably prevalent on ancestral autosomes that became sex chromosomes. Multiple lines of evidence suggest that large-effect mutations, including LOF alleles, often have strongly sex-biased fitness effects. In addition to observations that sterility alleles

typically have sex-specific effects (Lindsley and Lifschytz 1972; Ashburner et al. 2005) and whole-gene knockouts often have sex-biased or sex-limited phenotypic effects (van der Bijl W, Mank JE. 2021), the pervasiveness of sex-biased gene expression (Grath and Parsch 2016) implies that many genes are sexually dimorphic in their functional importance. While relatively few studies report estimates of the proportion of the genome that is sex-limited in expression, those that have indicate substantial proportions of sex-limited genes (*e.g.*, Perry et al. 2014; Mongue et al. 2021; Bain et al. 2021; Yu et al. 2023).

To gain a clearer picture of the prevalence of strongly sex-biased and sex-limited genes, we searched for high-quality gene expression studies, prioritizing those using whole-body samples (which should be more conclusive, given our purpose, than studies focusing on a small number of tissues) and when possible multiple life stages. Although our search was not exhaustive, it includes species across several animal phyla, including Arthropoda, Chordata, Platyhelminthes, Nematoda, and Tardigrada (details of the search and analysis can be found in the Supplementary Material). We estimated proportions of genes expressed at different degrees of sex-bias, including sex-limited expression, which are summarized in Fig. 5. The analysis shows extensive variation among taxa in the degree of sex-biased expression throughout the genome, with several species showing large fractions of genes with 5- to 10-fold or higher expression in one sex relative to the other, and others showing minimal sex-biased expression. There was a clear elevation in the number of strongly male-biased and male-limited genes relative to the number of strongly female-biased and female-limited genes, which hints at substantial opportunity for the decay of X-linked genes through sheltering effects. Our survey suggests that the processes that we have outlined here may vary in importance among species, though there is little question that gene

knockout and sex-specific phenotyping experiments are needed to more firmly establish the fitness consequences of LOF mutations in each sex.

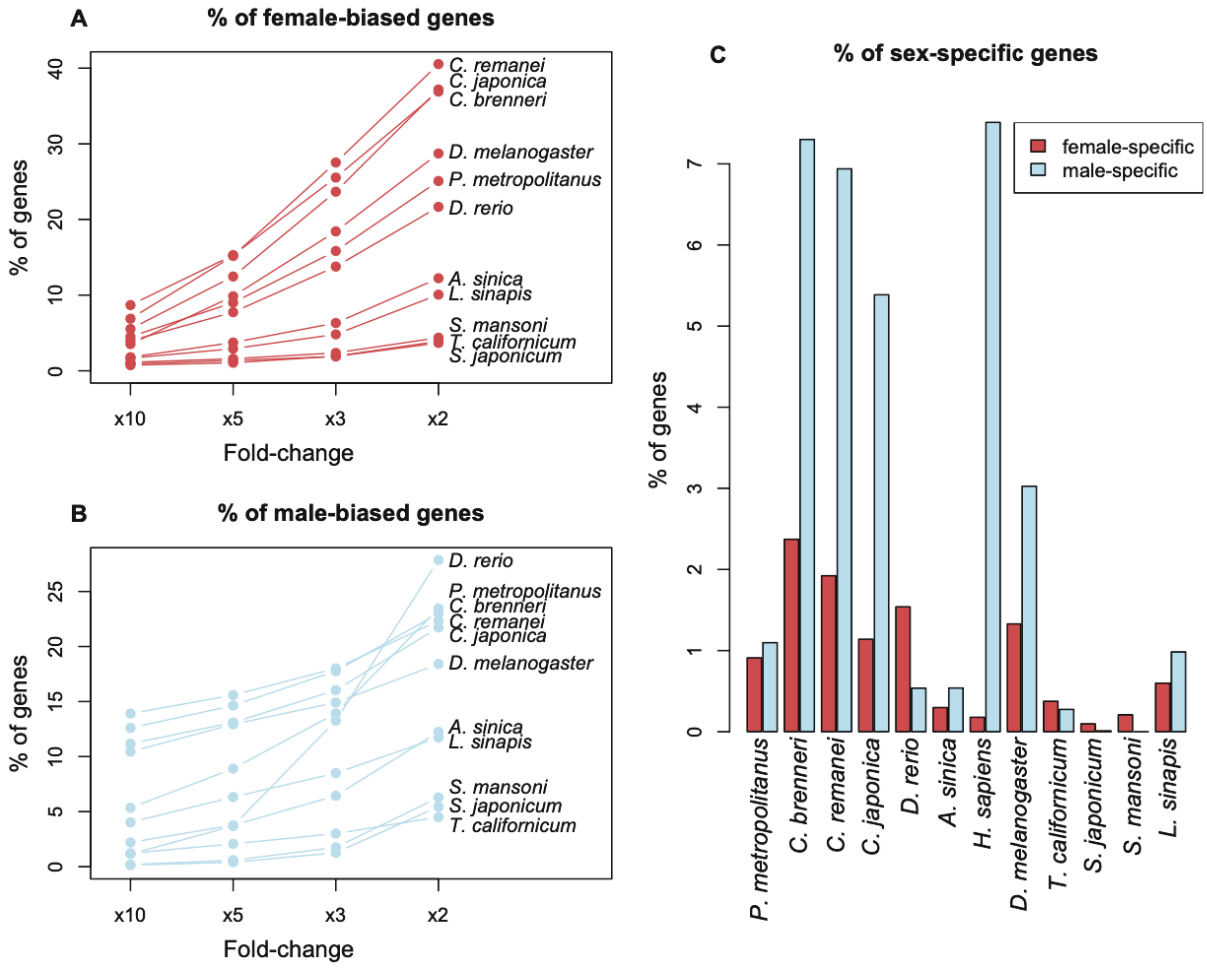


Figure 5. A survey of sex-biased and sex-limited gene expression in animals. Panel A: Percentage of female-biased genes in different species, using various thresholds. **Panel B:** Percentage of male-biased genes at various thresholds. **Panel C:** Percentage of genes with female- and male-specific expression (i.e., >99% of expression is in one of the sexes). Expression data were derived from Thomas et al. 2012; Lonsdale et al. 2013; Leal et al. 2018; Djordjevic et al. 2022; King and Zenker 2023; Elkrewi et al. 2021, 2023; Sugiura et al. 2024.

Our model is broadly consistent with several features of sex chromosome gene content evolution. Firstly, sheltering can contribute to the evolution of gene losses and masculinization of the Y chromosome. Deterministic fixation of recessive LOF alleles

is predicted in genes that are much more important for females than males (Fig. 1), and this process is facilitated by male-biased mutation rates. Sheltering further promotes the decay of Y-linked genes, including those with important functions in males, in species with small population size (Figs. 2-3) or in cases where selective interference co-occurs with the sheltering effects of LOF alleles (Fig. 4). It should be noted that selective interference and other processes are expected to drive Y chromosome degeneration in the absence of sheltering effects (Bachtrog 2006, 2013; Lenormand and Roze 2022). Nevertheless, sheltering could contribute substantially to gene losses, including genes that are more important for female than male fitness.

Secondly, our model predicts the loss of X-linked genes with male-limited functions, particularly within lineages with large population sizes and weak or non sex-biased mutation rates, and for genes where the population-scaled LOF mutation rate is large ($N_e\bar{\mu} \gg 0$; genes with small LOF mutation rates are more likely to become lost from the Y). These predictions validate Neuhaus's (1939) intuition that male-limited genes might be prone to loss from the X, and they build upon the model of Mrnjavac et al. (2023), which showed that male-limited deleterious mutations were prone to accumulation on X chromosomes paired with an undifferentiated Y. Predictions of our model align remarkably well with observations from animal genomes. X chromosome demasculinization has been widely reported in insects such as *Drosophila*, which happen to have historically large effective population sizes and similar mutation rates per sex, yet such patterns are less pronounced in mammals, where N_e is smaller and mutation rates tend to be male-biased (Vicoso and Charlesworth 2006). Recent studies of neo-X chromosomes in multiple *Drosophila* species show accelerated pseudogenization rates for genes that are important for males (and these genes remain functional on the neo-Y; Nozawa et al. 2016, 2021). While evidence of X-linked

gene losses in vertebrates are sparse, the rare exceptions are consistent with our model. The gene *PRSSLY* was present on the ancestral X and Y chromosomes of mammals and is currently the only gene known to have been lost from the mammalian X and retained on the Y (Hughes et al. 2022). *PRSSLY* has testis-specific expression, and its exceptional length could mean that it represents a large target for LOF mutations (*i.e.*, $N_e\bar{\mu}$ may be large for *PRSSLY*, though $N_e\bar{\mu}$ would be small for most other mammalian genes).

Thirdly, our model predicts that gene loss can be rapid, including in lineages with very large population sizes. Indeed, the average rate of degeneration for sex-limited genes is predicted to be roughly independent of the population size. With LOF mutation rates in the range of 10^{-4} to 10^{-6} (Monroe et al. 2021; Balick et al. 2022), rates of gene loss predicted by our model are easily compatible with those observed in nature (reviewed in Charlesworth 2021). For example, with a LOF mutation rate of $\mu = 10^{-5}$ per sex, our model predicts that X-linked male-limited LOF alleles will reach frequencies of 90% within 350,000 generations, while Y-linked female-limited LOF alleles will reach 90% within 230,000 generations. With 15 generations per year in *Drosophila* (Turelli and Hoffmann 1995; Pool 2015), these timescales equate to ~23,000 and ~15,000 years, respectively. These rates can account for the rapid gene losses observed in young neo-sex chromosomes from multiple *Drosophila* species (*e.g.*, autosome to sex chromosome translocations occurred at ~1.1 mya, ~0.5 mya, and 0.25 mya in *D. miranda*, *D. americana*, and *D. albomicans*, respectively; Nozawa 2021).

One interesting feature of our model is that gene loss can, in some cases, prove to be beneficial. For example, fixation of recessive LOF alleles in male-limited genes on the X or female-limited genes on the Y will generally *decrease* genetic loads at the

new mutation-selection equilibrium (see the Supplementary Information). In these instances, the remaining functional gene copies reside on the chromosome in which selection is most efficient, thereby alleviating load relative to the ancestral, diploid state. However, as in most models for Y chromosome degeneration, gene losses will often impose a fitness cost to males, including when LOF alleles are incompletely recessive (Fig. 1) or fixed by hitchhiking. As previous studies have shown, instances in which Y-linked fixations are damaging to males can lead to downregulation of Y-linked genes and upregulation of their functional X-linked homologs, leading to dosage compensation (Charlesworth 1978; Lenormand and Roze, 2020).

There is no single process that explains all features of sex chromosome evolution. Rather, the key questions are who the main players are and how each contributes to empirical patterns of X- and Y-linked gene losses, gains, and regulation. We argue that an extended version of Muller's (1914) sheltering hypothesis might contribute substantially to sex chromosome differentiation, following the cessation of recombination between the X and Y. An appeal of this model is its simplicity and compatibility with several features of sex chromosome evolution. Evaluating the broader importance of this sheltering scenario will require renewed effort in defining the distribution of dominance among deleterious mutations, and broadly evaluating the sex-specific fitness costs of gene losses. While both empirical aims are challenging, modern genetic engineering approaches coupled with high throughput phenotyping place us in a good position to accomplish these goals.

Acknowledgements

We thank Filip Ruzicka, Colin Olito, Akane Uesugi, Deborah Charlesworth, Brian Charlesworth, Melissa Toups, Daniel Jeffries, and two anonymous reviewers for comments on an earlier (and very different) version of the paper. Technical support was provided by ISTA Scientific Computing Unit.

References

Abbott JK, Nordén AK, Hansson B. 2017 Sex chromosome evolution: historical insights and future perspectives. *Proc R Soc B* 284:20162806.

Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187:553-566.

Albritton SE, Kranz AL, Rao P, Kramer M, Dietrich C, Ercan S. 2014. Sex-biased gene expression and evolution of the X chromosome in nematodes. *Genetics* 197:865-883.

Amorim CEG, Gao Z, Baker Z, Diesel JF, Simons YB, Haque IS, Pickrell J, Przeworski M. 2017. The population genetics of human disease: the case of recessive, lethal mutations. *PLoS Genet.* 13:e1006915.

Ashburner M, Golic KG, Hawley RS. 2005. *Drosophila: a laboratory handbook*, 2nd ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr Op Genet Dev*. 16:578-585.

Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* 179:1513-1525.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 14:113-124.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayors I, Ming R, Perrin N, Ross L, Valenzuela N, Vamosi JC. 2014. Sex determination: why so many ways of doing it? *PLoS Biol*. 12:e1001899.

Bachtrog D, Toda NRT, Lockton S. 2010. Dosage compensation and demasculinization of X chromosomes in *Drosophila*. Curr Biol. 20:1476-1481.

Bain SA, Marshall H, de la Filia AG, Laetsch DR, Husnik F, Ross L. 2021. Sex-specific expression and DNA methylation in a species with extreme sexual dimorphism and paternal genome elimination. *Mol Ecol.* 30:5687-5703.

Balick DJ, Jordan DM, Sunyaev S, Do R. 2022. Overcoming constraints on the detection of recessive selection in human genes from population frequency data. *Am J Hum Genet*. 109:33-49

Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* clade. *Genetics* 176:1131-1137.

Bellot DW, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508:494-499.

Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, Bustamante CD, Teshima KM, Przeworski M. 2008. Natural Selection on Genes that Underlie Human Disease Susceptibility. *Curr Biol.* 18:883-889.

Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci USA* 75:5618-5622.

Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 25:1030-1033.

Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol*. 6:142-162.

Charlesworth D. 2021 The timing of genetic degeneration of sex chromosomes. *Phil Trans R Soc B* 376:20200093.

Charlesworth D, Charlesworth B. 1980. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet Res*. 35:205-214.

Connallon T, Beasley IJ, McDonough Y, Ruzicka F. 2022. How much does the unguarded X contribute to sex differences in life span? *Evol Letters* 6:319-329.

Connallon T, Clark AG. 2010. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* 64:3417-3442.

Connallon T, Clark AG. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* 187:919-937.

Connallon T, Clark AG. 2011. Association between Sex-Biased Gene Expression and Mutations with Sex-Specific Phenotypic Consequences in Drosophila. *Genome Biol Evol*. 3:151-155.

Crow JF. 1993. Mutation, mean fitness, and genetic load. In Futuyma D, Antonovics J (eds.) *Oxford Surveys in Evolutionary Biology*, Vol. 9. Oxford University press: Oxford, pp. 3-42.

Crow JF. 1991. Professor Mukai: The man and his work. *Jpn J Genet*. 66:669-682.

Djordjevic J, Dumas Z, Robinson-Rechavi M, Schwander T, Parker DJ. 2022. Dynamics of sex-biased gene expression during development in the stick insect *Timema californicum*. *Heredity* 129:113-122.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 148:1667-1686.

Elkrewi M, Moldovan MA, Picard MAL, Vicoso B. 2021. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. *Mol Biol Evol*. 38:5345-5358.

Elkrewi M, Khauratovich U, Toups MA, Bett VK, Mrnjavac A, Macon A, Fraisse C, Sax L, Huylmans AK, Hontoria F, Vicoso B. 2023. ZW sex-chromosome evolution and contagious parthenogenesis in Artemia brine shrimp. *Genetics* 222:iyac123.

Ellegren H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet*. 12:157-166.

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

Eyre-Walker A, Keightly PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet*. 8:610-618.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* 78:737-756.

Fisher RA. 1935. Sheltering of lethals. *Am Nat.* 699:446-455.

Foster JM, et al. 2020. Sex chromosome evolution in parasitic nematodes of humans. *Nature Communications* 11:1964

Frota-Pessoa O, Aratangy LR. 1968. The degeneration of the Y chromosome. *Rev Bras de Pesquisas Med Biol*. 1:241-244.

Fuller ZL, Berg JJ, Mostafavi H, Sella G, Przeworski M. 2019. Measuring intolerance to mutation in human genetics. *Nat Genet*. 51:772-776.

Furman BLS, Metzger DCH, Darolti I, Wright AE, Sandkam BA, Almeida P, Shu JJ, Mank JE. 2020. Sex chromosome evolution: so many exceptions to the rules. *Genome Biol Evol.* 12:750-763.

Gu L, Walters JR. 2017. Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. *Genome Biol Evol*. 9:2461-2476.

Grath S, Parsch J. 2016. Sex-biased gene expression. *Annu Rev Genet*. 50:29-44.

Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Evol Syst*. 40:151-172.

Hartl DL, Clark AG. 2007. *Principles of Population Genetics*, 4th Ed. Sinauer Assoc.: Sunderland, MA.

Hu QL, Ye YX, Zhuo JC, Huang HJ, Li JM, Zhang CX. 2022. Chromosome-level assembly, dosage compensation and sex-biased gene expression in the small brown plant hopper, Laodelphax striatellus. *Genome Biol Evol*. 14:evac160

Huber, C.D., Durvasula, A., Hancock, A.M., Lohmueller K.E. 2018. Gene expression drives the evolution of dominance. *Nat Commun* **9**, 2750

Hughes JF, Skaletsky H, Nicholls PK, Drake A, Pyntikova T, Cho TJ, Bellott DW, Page DC. 2022. A gene deriving from the ancestral sex chromosomes was lost from the X and retained on the Y chromosome in eutherian mammals. *BMC Biology* 20:133

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, Kikuchi K. 2012. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (Fugu). *PLoS Genet* 8: e1002798.

King AC, Zenker AK. 2023. Sex blind: bridging the gap between drug exposure and sex-related gene expression in Danio rerio using next-generation sequencing (NGS) data and a literature review to find the missing links in pharmaceutical and environmental toxicology. *Frontiers in Toxicology* 5:1187302

Lasne C, Elkrewi M, Toups MA, Layana L, Macon A, Vicoso B. 2023. The scorpionfly (Panorama cognate) genome highlights conserved and derived features of the peculiar dipteran X chromosome. *Mol Biol. Evol*. 40:msad245

Leal L, Talla V, Källman T, Friberg M, Wiklund C, Dincă V, Vila R, Backström N. 2018. Gene expression profiling across ontogenetic stages in the wood white (Leptidea sinapis) reveals pathways linked to butterfly diapause regulation. *Molecular Ecology* 27:935-948.

Lenormand T, Roze D. 2022. Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science* 375:663-666.

Lindsley DL, Lifschytz E. Beatty RA, Gluecksohn-Waelsch S. 1972. The genetic control of spermiogenesis in *Drosophila*. *Edinburgh Symposium on the Genetics of the Spermatozoon*. Bogtrykkeriet Forum: Copenhagen, Denmark, pp. 203-222.

Lonsdale J, et al. 2013. The genotype-tissue expression (GTEx) project. *Nature Genetics* 45:580-585.

Mahajan S, Bachtrog D. 2017. Convergent evolution of Y chromosome gene content in flies. *Nature Communications* 8:785.

Mallet MA, Bouchard JM, Kimber CM, Chippindale AK. 2011. Experimental mutation-accumulation on the X chromosome of Drosophila melanogaster reveals stronger selection on males than females. *BMC Evol Biol*. 11:156.

Manna F, Gallet R, Martin G, Lenormand T. 2012. The high-throughput yeast deletion fitness data and the theories of dominance. *J Evol Biol*. 25:892-903.

Manna F, Martin G, Lenormand T. 2011. Fitness landscapes: an alternative theory for the dominance of mutation. *Genetics* 189:923-937.

Marion SB, Noor MAF. 2023. Interrogating the roles of mutation-selection balance, heterozygote advantage, and linked selection in maintaining recessive lethal variation in natural populations. *Annu Rev Anim Biosci*. 11:77-91.

Meisel RP, Asgari D, Schlamp F, Uncles RL. 2022. Induction and inhibition of Drosophila X chromosome gene expression are both impeded by the dosage compensation complex. *G3* 12:jkac165.

Meisel RP, Malone JH, Clark AG. 2012. Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Research* 22:1255-1265.

Mongue AJ, Hansen ME, Walters JR. 2021. Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages. *Evolution* 76:332-345.

Monroe JG, McKay JK, Weigel D, Flood PJ. 2021. The population genomics of adaptive loss of function. Heredity 126:383-395.

Mora P, Hospodářská M, Voleníková AC, Koutecký P, Štundlová J, Dalíková M, Walters JR, Nguyen P. 2024. Sex-biased gene content is associated with sex chromosome turnover in Danaini butterflies. *Molecular Ecology*, in press.

Muller HJ. 1914. A gene for the fourth chromosome of *Drosophila*. *J Exp Zoology* 17:325-336.

Muller HJ. 1918. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* 3:422-499.

Muller HJ, Painter TS. 1932. The differentiation of the sex chromosomes of Drosophila into genetically active and inert regions. *Z Indukt Abstamm Vererblehre*. 62:316-365.

Mrnjavac A, Khudiakova KA, Barton NH, Vicoso B. 2023. Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. *Evol Letters* 7:4–12.

Nei M. 1970. Accumulation of nonfunctional genes on sheltered chromosomes. *Am Nat.* 104:311-322.

Neuhaus MJ. 1939. A cytogenetic study of the Y-chromosome of *Drosophila melanogaster*. *J Genetics* 37:229-254.

Nozawa M, Minakuchi Y, Satomura K, Kondo S, Toyoda A, Tamura K. 2021. Shared evolutionary trajectories of three independent neo-sex chromosomes in Drosophila. Genome Res. 31:2069–2079.

Nozawa M, Onizuka K, Fujimi M, Ikeo K, Gojobori T. 2016. Accelerated pseudogenization on the neo-X chromosome in Drosophila miranda. *Nat Comm*. 7:13659.

Orr HA. 2000. The rate of adaptation in asexuals. *Genetics* 155:961-968.

Orr HA. Kim Y. 1998. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* 150:1693-1698.

Papa F, Windbichler N, Waterhouse RM, Canetti A, D'Amato R, Persampieri T, Lawniczak MKN, Nolan T, Papathanos PA. 2017. Rapid evolution of female-biased genes among four species of Anopheles malaria mosquitoes. *Genome Research* 27:1536-1548

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet*. 14:83-87.

Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, Kitano J. 2015. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. *PLoS Genet.* 11:e1005237.

Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster. Mol Biol Evol*. 31:1206-1219.

Pool JE. 2015. The Mosaic Ancestry of the Drosophila genetic reference panel and the D. melanogaster reference genome reveals a network of epistatic fitness interactions. *Mol Biol Evol*. 32:3236-3251.

Rancati G, Moffat J, Typas A, Pavelka N. 2018. Emerging and evolving concepts in gene essentiality. *Nat Rev Genet*. 19:34-49.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735-742.

Rice WR. 1996. Evolution of the Y sex chromosomes in animals. *BioScience* 46:331-343.

Sharp NP, Agrawal AF. 2013. Male-biased fitness effects of spontaneous mutations in Drosophila melanogaster. *Evolution* 67:1189-1195.

Shaw DE, White MA. 2022. The evolution of gene regulation on sex chromosomes. *Trends in Genetics* 38:P844-855.

Simmons MJ, Crow JF. 1977. Mutations affecting fitness in Drosophila populations. *Annu Rev Genet*. 11:49-78.

Simons YB, Sella G. 2016. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr Opin Genet Dev*. 41:150-158.

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

Stöck M, Horn A, Grossen C, Lindtke D, Sermier R, Betto-Colliard C, Dufresnes C, Bonjour E, Dumas Z, Luquet E, Maddalena T, Sousa HC, Martinez-Solano I, Perrin

N. 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biol* 9: e1001062.

Sugiura K, Yoshida Y, Hayashi K, Arakawa K, Kunieda T, Matsumoto M. 2024. Sexual dimorphism in the tardigrade *Paramacrobiotus metropolitanus*. *Zoological Letters* 10:11.

Thomas CG, Li R, Smith HE, Woodruff GC, Oliver B, Haag ES. 2012. Simplification and desexualization of gene expression in self-fertile nematodes. *Current Biology* 22:2167-2172

Turelli M, Hoffmann AA. 1995. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics* 140:1319-1338.

van der Bijl W, Mank JE. 2021. Widespread cryptic variation in genetic architecture between the sexes. *Evol Letters* 5:359-369.

Vibranovski MD, Lopes HF, Karr TL, Long M. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed Genes. *PLoS Genet.* 5: e1000731.

Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat Ecol Evol*. 3:1632–1641.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 7:645-653.

Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol* 11:e1001643.

Wei KHC, Chatla K, Bachtrog D. 2024. Single-cell RNA-seq of Drosophila miranda testis reveals the evolution and trajectory of germline sex chromosome regulation. *PLoS Biology* 22:e3002605.

Yoshikawa I, Mukai T. 1970. Heterozygous effects on viability of spontaneous lethal genes in Drosophila melanogaster. *Jpn J Genet*. 45:443-455.

Yu X, Marshall H, Liu Y, Xiong Y, Zeng X, Yu H, Chen W, Zhou G, Zhu B, Ross L, Lu Z. 2023. Sex-specific transcription and DNA methylationlandscapes of the Asian citrus psyllid, a vector of huanglongbing pathogens. *Evolution* 77:1203-1215.

Zhu Z, Younas L, Zhou Q. 2024. Evolution and regulation of animal sex chromosomes. *Nat Rev Genet*. doi: 10.1038/s41576-024-00757-3

Zschocke J, Byers PH, Wilkie AOM. 2023. Mendelian inheritance revisited: dominance and recessiveness in medical genetics. *Nat Rev Genet*. 24:442-463.

Supplementary Materials

Appendix 1. Possible distributions of dominance for lethal mutations

The beta and gamma distributions provide two flexible distributions that might apply to the distribution of *h* for specific mutation classes, such as lethals or LOF mutations. For both of distributions, dominance is constrained to be within the range $h > 0$, so that a substantial fraction of mutations will be recessive, or nearly so, in models where the highest probability density occurs at *h* = 0. The parameters of each model can be inferred from empirical estimates of the mean and variance of the distribution of *h*.

Under a beta distribution for *h*, the mean and the variance for *h* are defined as:

$$\bar{h} = \frac{\alpha}{\alpha + \beta}$$

$$\sigma_h^2 = \frac{\bar{h}(1 - \bar{h})}{1 + \alpha + \beta}$$

where α and β are the shape parameters of the beta distribution. The parameters of the distribution can then be inferred from information about the mean and variance of *h*, by rearranging:

$$\alpha = \bar{h}\left(\frac{\bar{h}(1 - \bar{h})}{\sigma_h^2} - 1\right)$$

$$\beta = (1 - \bar{h})\left(\frac{\bar{h}(1 - \bar{h})}{\sigma_h^2} - 1\right)$$

Under a gamma distribution, the mean and the variance for *h* are given by as:

$$\bar{h} = k\theta$$

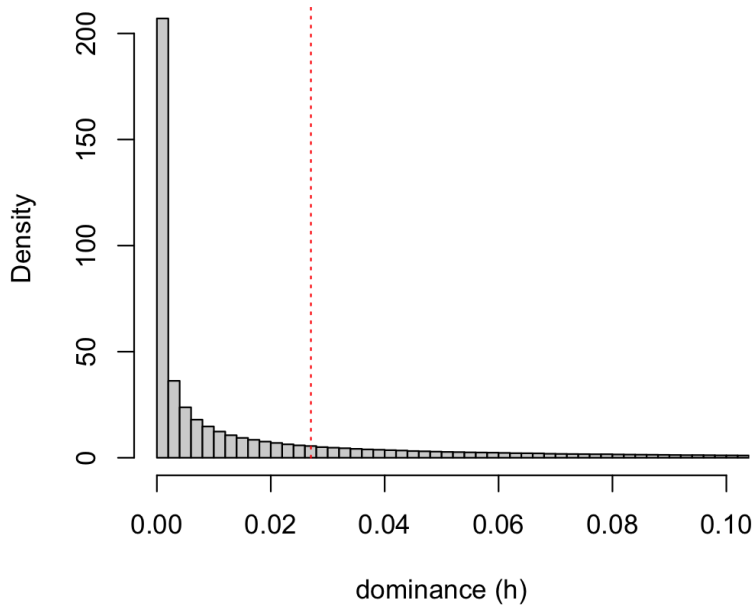
$$\sigma_h^2 = k\theta^2$$

where *k* is the shape and θ is the scale parameter of the gamma distribution. The parameters of the gamma distribution can be inferred from information on the mean and variance of h using the rearrangements:

$$k = \frac{\bar{h}^2}{\sigma_h^2}$$

$$\theta = \frac{\sigma_h^2}{\bar{h}}$$

Yoshikawa and Mukai (1970) present estimates of the mean and variance of *h* for lethal mutations accumulated on *Drosophila melanogaster* 2nd chromosomes: $\bar{h} = 0.027$ and $\sigma_h^2 = 0.0027$. When these estimates are inserted into the above equations, we obtain a beta distribution with parameters $\alpha \approx 0.236$ and $\beta \approx 8.49$, which predicts a strongly right-skewed distribution with roughly 60% of lethals have dominance coefficients below *h* = 0.01 (*i.e.*, the following plot shows the distribution of dominance for the model with the mean in red).



Similar predictions emerge under a gamma distribution of *h*, in which the point estimates ($\bar{h} = 0.027$ and $\sigma_h^2 = 0.0027$) imply a gamma distribution with $k = 0.27$ and $\theta = 0.1$. This distribution predicts that ~58% of lethals will have dominance coefficients below *h* = 0.01.

In their analysis of yeast gene deletion data, Agrawal and Whitlock (2011) found that the distribution of *h* for gene deletions provided a reasonably fit to a shifted gamma

distribution, with the average *h* changing across the spectrum of homozygous selection coefficients of deletions (*s*), but no change in the variance. The relevant gamma distribution for each value of *s* includes shape (*k*), scale (θ), and displacement parameters. The shape and scale parameters can be inferred from the parameters δ and σ_h^2 reported in Table 3 of Agrawal and Whitlock (2011). Manna et al. (2012 *JEB*) have shown that the displacement parameter is negligible for deletions with large values of *s* (note as well that the large-*s* range is most suitable for accurately inferring dominance; see Manna et al. 2012). Using $\delta = 0.037$ and $\sigma_h^2 = 0.011$, the shape and scale parameters of the gamma distribution for *h* are $k = 0.124$ and $\theta = 0.297$, in which case the distribution is, again, strongly right skewed with roughly 69% of deletions having dominance coefficients below the threshold *h* = 0.01.

Appendix 2. Deterministic model of sheltering

We focus on genes present in functional and non-functional states on a pair of X and Y chromosomes that do not recombine with one another. Each gene has two major alleles: a functional allele that is favoured by natural selection and a deleterious loss-of-function (LOF) allele. Mutations in each gene are unidirectional, with functional alleles mutating to LOF alleles at rates of μ_f in female gametes and μ_m in male gametes. Gain of function mutations are assumed to be sufficiently rare that they can be ignored.

At a given gene, A will be the functional allele and A_0 will be the LOF allele. Let p_m and p_f represent the frequencies (respectively) of the X-linked LOF allele in male and female gametes contributing to fertilization; p_Y is the frequency of the same allele on Y-bearing male gametes contributing to fertilization. The corresponding functional allele frequencies are $q_m = 1 - p_m$, $q_f = 1 - p_f$, and $q_Y = 1 - p_Y$. Female and male gametes combine randomly to form zygotes of the next generation. Table S1 presents the zygotic frequencies and fitness per genotype. Selection and mutation complete a full generation cycle, at which point the frequency of the *A* allele in X-bearing male gametes will be:

$$1 - p'_m = \frac{\big(1 - p_f\big)(1 - p_Y s_m h)(1 - \mu_m)}{p_f\big(1 - s_m h - p_Y s_m(1 - h)\big) + \big(1 - p_f\big)(1 - p_Y s_m h)} \qquad [1a]$$

The frequency of the *A* allele in female gametes will be:

$$1 - p'_f = \frac{\frac{1}{2}\Big(p_f(1 - p_m) + p_m\big(1 - p_f\big)\Big)\big(1 - s_f h\big) + \big(1 - p_f\big)(1 - p_m)}{1 - s_f p_f p_m - \Big(p_f(1 - p_m) + p_m\big(1 - p_f\big)\Big)s_f h}\big(1 - \mu_f\big) \qquad [1b]$$

And the frequency of *A* on *Y*-bearing gametes will be:

$$1 - p'_Y = \frac{(1 - p_Y)\big(1 - p_f s_m h\big)(1 - \mu_m)}{p_Y\Big(1 - s_m h - p_f s_m(1 - h)\Big) + (1 - p_Y)\big(1 - p_f s_m h\big)} \qquad [1c]$$

Eqs. [1a-c] are exact and serve as the basis for all of our deterministic predictions.

Table S1. Frequency and fitness for each genotype and sex[1]

	Genotype		
	A_0A_0	A_0A	AA
Frequency in female zygotes	$p_f p_m$	$p_f q_m + p_m q_f$	$q_f q_m$
Frequency in male zygotes	$p_f p_Y$	$p_f q_Y + p_Y q_f$	$q_f q_Y$
Fitness in females	$1 - s_f$	$1 - s_f h$	1
Fitness in males	$1 - s_m$	$1 - s_m h$	1

[1] Note that males have two possible orientations of the heterozygous genotype: (*i*) heterozygous with an X-linked LOF allele, which occurs with frequency $p_f q_Y$, or (*ii*) heterozygous with a Y-linked LOF allele, which occurs with frequency $p_Y q_f$.

Evolutionary dynamics and equilibria for recessive LOF alleles

For cases where deleterious mutations are recessive ($h = 0$), and mutation rates are positive, the recursions simplify to:

$$p'_m = 1 - \frac{(1 - p_f)(1 - \mu_m)}{1 - s_m p_Y p_f}$$

$$p'_f = 1 - \frac{1 - \frac{1}{2}(p_f + p_m)}{1 - s_f p_f p_m}(1 - \mu_f)$$

$$p'_Y = 1 - \frac{(1 - p_Y)(1 - \mu_m)}{1 - s_m p_Y p_f}$$

In this case, there are four possible equilibria:

1. Fixation on the X and Y ($\hat{p}_Y = \hat{p}_f = \hat{p}_m = 1$)
2. Fixation on the Y and polymorphism on the X ($\hat{p}_Y = 1; \hat{p}_f < 1; \hat{p}_m < 1$)
3. Fixation on the X and polymorphism on the Y ($\hat{p}_Y < 1; \hat{p}_f = \hat{p}_m = 1$)
4. Polymorphism on both the X and Y ($\hat{p}_Y < 1; \hat{p}_f < 1; \hat{p}_m < 1$)

The fourth equilibrium state, when valid, is:

$$\hat{p}_X = \hat{p}_m = \hat{p}_f = \sqrt{\frac{\mu_f}{s_f}} \qquad [2a]$$

$$\hat{p}_Y = \frac{\mu_m}{s_m \hat{p}_X} = \frac{\mu_m}{s_m}\sqrt{\frac{s_f}{\mu_f}} \qquad [2b]$$

The X-linked equilibrium will be higher than the Y-linked equilibrium when:

$$\hat{p}_X > \hat{p}_Y \Leftrightarrow \frac{s_m}{s_f} > \frac{\mu_m}{\mu_f}$$

The X-linked equilibrium will be higher than the Y-linked equilibrium when:

$$\hat{p}_X < \hat{p}_Y \Leftrightarrow \frac{s_m}{s_f} < \frac{\mu_m}{\mu_f}$$

The equilibrium with polymorphism on the X and Y will be valid under the condition: $\mu_m/s_m < \sqrt{\mu_f/s_f} < 1$, which defines the minimum requirements of selection each sex to prevent degeneration on either chromosome. With equal mutation rates and selection in each sex ($s_f = s_m$; $\mu_f = \mu_m$), this equilibrium further simplifies to:

$$\hat{p}_f = \hat{p}_m = \hat{p}_Y = \sqrt{\frac{\mu}{s}}$$

which matches Fisher's (1935) result (*22*), in which harmful mutations (lethal alleles in his model: $s = 1$) evolve to equal frequencies on the X and Y chromosome.

Deterministic fixation of LOF alleles on the X

The X-linked LOF allele will be driven to fixation when $s_f \leq \mu_f$, which should apply to male-limited genes (genes with no effect on female fitness: $s_f = 0$). This result is implied by the equilibrium in eq. [2a], and it can be formally evaluated by way of a linear stability analysis for the equilibrium in which the LOF is fixed on the X and the Y

evolves to a polymorphic equilibrium at mutation-selection balance. This equilibrium corresponds to: $\hat{p}_f = \hat{p}_m = 1$ and $\hat{p}_Y = \mu_m(1 - hs_m)/s_m(1 - h)$. The characteristic polynomial for this equilibrium (by way of the Jacobian matrix for the set of exact recursion equations; see chapter 8 of Otto and Day 2007) is:

$$\det\begin{pmatrix} -\lambda & 1 & 0 \\ \frac{(1-\mu_f)}{2(1-s_f)} & \frac{(1-\mu_f)}{2(1-s_f)} - \lambda & 0 \\ 0 & 1 - \frac{\mu_m}{s_m} & \frac{1-s_m}{1-\mu_m} - \lambda \end{pmatrix}$$

$$= \left(\lambda^2 - \lambda\frac{(1-\mu_f)}{2(1-s_f)} - \frac{(1-\mu_f)}{2(1-s_f)}\right)\left(\frac{1-s_m}{1-\mu_m} - \lambda\right) = 0$$

which yields the eigenvalues:

$$\lambda = \frac{1-s_m}{1-\mu_m}$$

$$\lambda = \frac{\frac{(1-\mu_f)}{2(1-s_f)} - \sqrt{\left(\frac{(1-\mu_f)}{2(1-s_f)}\right)^2 + 4\frac{(1-\mu_f)}{2(1-s_f)}}}{2}$$

$$\lambda = \frac{\frac{(1-\mu_f)}{2(1-s_f)} + \sqrt{\left(\frac{(1-\mu_f)}{2(1-s_f)}\right)^2 + 4\frac{(1-\mu_f)}{2(1-s_f)}}}{2}$$

Provided $\mu_m < s_m$ (which must be true for $\hat{p}_Y$ < 1 given *h* = 0), then the first eigenvalue is will always be less than one. Because $\frac{(1-\mu_f)}{2(1-s_f)} > 0$, the third eigenvalue is the one we need to focus on. The equilibrium will be stable when that eigenvalue is less than one, *i.e.*:

$$1 > \frac{\frac{(1-\mu_f)}{2(1-s_f)} + \sqrt{\left(\frac{(1-\mu_f)}{2(1-s_f)}\right)^2 + 4\frac{(1-\mu_f)}{2(1-s_f)}}}{2}$$

which simplifies to $s_f < \mu_f$. The same eigenvalue will be greater than one—and LOF allele fixation will therefore be unstable—when $s_f > \mu_f$.

Deterministic fixation of LOF alleles on the Y

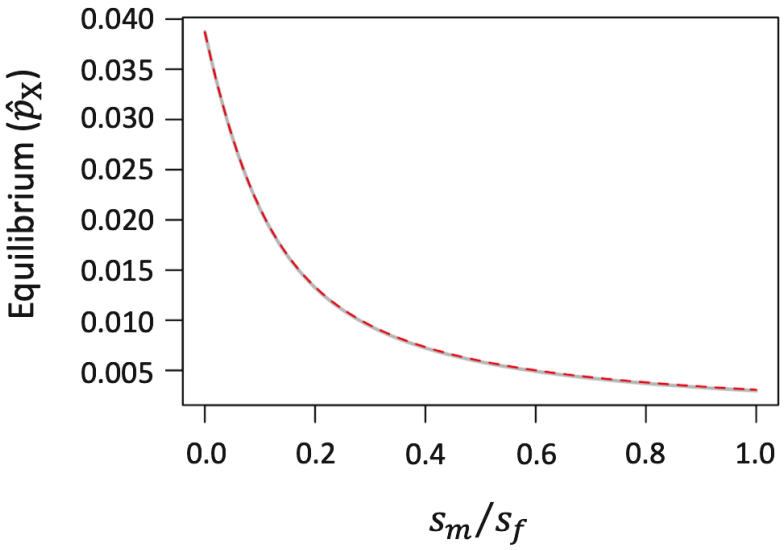
The Y-linked LOF allele will be driven to fixation when $s_m \leq \alpha\sqrt{\mu_f s_f}$, where $\alpha = \mu_m/\mu_f$. The condition for stable fixation is implied by eq. [2b], and a stability analysis of the equilibrium with $p_Y = 1$ confirms the prediction. Specifically, when the LOF allele is near fixation on the Y and the *A* allele is near fixation on the X (the latter is expected when selection in females is strong relative to the mutation rate), then the X-linked dynamics are approximately:

$$p'_X = p_X + \frac{-2s_f p_X^2 - s_m p_Y p_X + 2\mu_f + \mu_m}{3}$$

which yields the following X-linked equilibrium when the LOF allele is fixed on the Y:

$$\hat{p}_X = \frac{\sqrt{s_m^2 + 8s_f(2\mu_f + \mu_m)} - s_m}{4s_f}$$

Deterministic simulations using the exact recursions show that the equilibrium is accurate in cases where selection in females is strong relative to the mutation rate. The following figure compares the approximate equilibrium (grey) against exact forward simulations to equilibrium using the full and exact recursion equations (red) (results use $\mu_f = \mu_m = 10^{-4}$ and $s_f = 0.1$).



Using the approximate recursions, we obtain the following Jacobian matrix for the equilibrium with Y-linked LOF alleles fixed:

$$J = \begin{pmatrix} 1 - \frac{4s_f\hat{p}_X + s_m}{3} & -\frac{s_m\hat{p}_X}{3} \\ 0 & \frac{1-\mu_m}{1-s_m\hat{p}_X} \end{pmatrix}$$

which has the leading eigenvalue:

$$\lambda_L = \frac{1-\mu_m}{1-s_m\hat{p}_X}$$

The equilibrium will be stable if $\lambda_L < 1$, which requires:

$$s_m < \frac{\mu_m}{\hat{p}_X}$$

Substituting the approximation for $\hat{p}_X$, we have:

$$s_m < \alpha\sqrt{\mu_f s_f}$$

where $\alpha = \mu_m/\mu_f$, which confirms the same result stated above.

It is clear from this last result that conditions for fixation of LOF alleles are more permissive on the Y chromosome than they are on the X. Recall that on the X chromosome, fixation was only likely for genes that have negligible importance in females. In contrast, deterministic gene degeneration can occur on the Y chromosome

in cases where the gene has a meaningful fitness effect in males. For example, for genes that are essential for females ($s_f = 1$) and that have relatively large LOF mutation rates ($\mu_f = \mu_m = 10^{-4}$), LOF mutations on the Y will deterministically spread to fixation when $s_m < 0.01$ (*i.e.*, mutations altering fitness by up to 1%). Male-biased mutation rates will further expand the criteria for deterministic degeneration. This effect arises because strong selection in females keeps LOF alleles on the X at low frequencies, which generates strong sheltering of Y-linked genes in males.

General dynamics of sex-limited genes

For genes with female-limited functions ($s_f > 0 = s_m$), the Y-linked recursion simplifies to:

$$1 - p_Y' = (1 - p_Y)(1 - \mu_m)$$

which is independent of selection in females and allows for a general solution of the allele frequency dynamics on the Y. The expected frequency of Y-linked LOF mutations after *t* generation of evolution will be:

$$p_{Y,t} = 1 - \left(1 - p_{Y,0}\right)(1 - \mu_m)^t \approx 1 - \left(1 - p_{Y,0}\right)e^{-\mu_m t}$$

where $p_{Y,0}$ is the initial frequency; the final approximation applies because μ_m is small. If we assume that the LOF allele was initially absent from the population ($p_{Y,0} = 0$), the result further simplifies to $p_{Y,t} \approx 1 - e^{-\mu_m t}$, as presented in the main text. For the mirror image case involving LOF mutations in male-limited genes, we can approximate the evolutionary dynamics on the X and Y chromosome as:

$$p_X' \approx p_X + \frac{-s_m p_Y p_X (1 - p_X) + \left(\mu_m + 2\mu_f\right)(1 - p_X)}{3}$$

$$p_Y' \approx p_Y - s_m p_X p_Y (1 - p_Y) + \mu_m (1 - p_Y)$$

These dynamics suggest that the Y chromosome should respond more quickly to selection (by a factor of three) than the X. We may therefore use a separation of timescales approximation, in which we assume that the Y-linked allele frequencies will quickly approach the following a quasi-equilibrium state given the frequency of the X-linked mutation:

$$\tilde{p}_Y \approx \frac{\mu_m}{s_m p_X}$$

Plugging this quasi-equilibrium into the expression for change p_X gives us:

$$1 - p_X' \approx (1 - p_X)\left(1 - \frac{2\mu_f}{3}\right)$$

which yields the general solution:

$$p_{X,t} \approx 1 - \left(1 - p_{X,0}\right)\left(1 - \frac{2\mu_f}{3}\right)^t \approx 1 - \left(1 - p_{X,0}\right)\exp\left(-\frac{2}{3}\mu_f t\right)$$

where $p_{X,0}$ is the initial X-linked frequency of the LOF allele, and $p_{X,t}$ is its frequency after *t* generations. With the X-linked LOF initially absent from the population, the last result further simplifies to $p_{X,t} \approx 1 - \exp\left(-\frac{2}{3}\mu_f t\right)$, as presented in the main text.

Deterministic changes in the genetic load

The equilibrium genetic load contributed by a locus mutating to recessive LOF alleles is define as follows. For an autosomal locus—which includes the ancestral state preceding the origin of a new sex chromosome system or a translocation event that generates neo-X and neo-Y chromosomes—the female and male genetic loads will be:

$$L_f = \hat{p}_m \hat{p}_f s_f$$

and

$$L_m = \hat{p}_m \hat{p}_f s_m$$

where $\hat{p}_f$ and $\hat{p}_m$ are the equilibrium frequencies of LOF alleles in female and male gametes contributing to fertilization in each generation.

For an autosomal locus that mutates to recessive LOF alleles with homozygous fitness effects of s_m in males and s_f in females, the evolutionary dynamics under mutation and selection will be:

$$1 - p'_m = \left(1 - \frac{p_m p_f (1 - s_m) + \frac{1}{2}\left(p_f + p_m - 2p_f p_m\right)}{1 - p_m p_f s_m}\right)(1 - \mu_m)$$

$$1 - p'_f = \left(1 - \frac{p_m p_f \left(1 - s_f\right) + \frac{1}{2}\left(p_f + p_m - 2p_f p_m\right)}{1 - p_m p_f s_f}\right)\left(1 - \mu_f\right)$$

Note that these frequency dynamics should also be roughly equivalent to those of pseudo-autosomal regions (PAR) provided the locus in question is not closely linked to the male-limited region of the Y. At equilibrium for this model, we have the identity:

$$2 = \frac{1 - \mu_f}{1 - \hat{p}_m \hat{p}_f s_f} + \frac{1 - \mu_m}{1 - \hat{p}_m \hat{p}_f s_m}$$

For female-limited loci, the equilibrium load (in females) simplifies to:

$$L_f = \hat{p}_m \hat{p}_f s_f = \frac{\mu_f + \mu_m}{1 + \mu_m} = \mu_f + \frac{\mu_m \left(1 - \mu_f\right)}{1 + \mu_m}$$

And for male-limited loci, the equilibrium load (in males) becomes:

$$L_m = \hat{p}_m \hat{p}_f s_m = \frac{\mu_m + \mu_f}{1 + \mu_f} = \mu_m + \frac{\mu_f (1 - \mu_m)}{1 + \mu_f}$$

What happens to the load at equilibrium for genes on sex chromosomes, particularly genes that have become degenerate on either the X or Y? We focus on the bookend cases of our deterministic model, which include male-limited gene degeneration is expected on the X chromosome, and female-limited gene degeneration is expected on the Y.

For a female-limited gene, the Y-linked copy is expected to degenerate while the X-linked copy will evolve to the equilibrium at mutation-selection balance. In this case, the equilibrium for the X is approximately:

$$\hat{p}_X = \sqrt{\frac{2\mu_f + \mu_m}{2s_f}}$$

and the equilibrium load for females becomes:

$$L_f = s_f \hat{p}_X^2 \approx \mu_f + \frac{\mu_m}{2}$$

which is less than the ancestral autosomal load as long as $1 > \mu_m + 2\mu_f$. This condition will always be true given biologically realistic mutation rates.

For a male-limited gene, the X-linked copy is expected to degenerate while the Y-linked copy will evolve to the equilibrium at mutation-selection balance. In this case, the equilibrium for the Y is:

$$\hat{p}_Y = \frac{\mu_m}{s_m}$$

and the male load will be:

$$L_m = s_m \hat{p}_Y = \mu_m$$

which is again an improvement over the ancestral autosomal state provided $\mu_f > 0$.

Finally, consider the intermediate case of a gene in which LOF mutations equally affect the sexes. Although, sheltering in the absence of drift is not expect to cause degeneration in this case, the transition from an ancestral autosomal state to a sex-linked state does have an effect on the equilibrium genetic load. For an autosome with $s = s_f = s_m$, the equilibrium load at mutation-selection balance is:

$$L_f = L_m = \hat{p}_m \hat{p}_f s = \frac{\mu_f + \mu_m}{2}$$

In cases where both the X and Y chromosome equilibria are polymorphic, which includes the specific case of equal selection in each sex (see above), the female and male genetic loads become:

$$L_f = s_f \hat{p}_f \hat{p}_m = \mu_f$$

$$L_m = s_m \hat{p}_f \hat{p}_Y = \mu_m$$

The autosomal and sex-linked loads will all be the same in the special case where mutation rates show no sex bias. If, as in many species, the mutation rate is male-biased ($\mu_m > \mu_f$), evolutionary transition from an autosomal to a sex-linked state will result in a reduction of the female load and elevation of the male load.

Appendix 3: Sheltering under mutation, selection and genetic drift

We incorporate drift into our models using a combination of exact computer simulations of X- and Y-linked evolutionary dynamics in a Wright-Fisher population where multinomial sampling of genotypes mimics the process of genetic drift (see pp. 229-230 of Charlesworth and Charlesworth 2010). Specifically, we assume in each generation that there are a fixed number of adult females and males (N_f and N_m, respectively, though we focus on the simplest case where $N_f = N_m$). The deterministic model presented in Appendix 1 yields predictions for the expected frequencies of each female and male genotype in the post-selection pool of individuals contributing to reproduction. We then carry out independent multinomial sampling in each sex (using a pseudo-random number generator) to determine the actual genotype frequencies in the pool of reproducing adults. Mutations occur after the multinomial sampling step and cause some deviation between LOF allele frequencies of adults and the allele frequencies in gametes contributing to fertilization in the next generation. Given an equal sex ratio, the effective population size for autosomes will be $2N_e$, where $N_e = N_f + N_m$; Effective sizes for the X and Y will be $1.5N_e$ and $0.5N_e$, respectively.

Y-linked fixation probabilities. Each of our simulations of Y-linked fixation probabilities (as in Fig. 2) begin with a single initial copy of a Y-linked LOF allele. The initial frequency of the X-linked LOF allele corresponds to the mean of the stationary distribution predicted for the X chromosome in an ancestral population in which all Y-linked copies were functional (the initial condition is derived immediately below). The system was then allowed to evolve under recurrent mutation on the X (but not the Y), and selection and drift on both the X and Y, until the Y-linked variant is either lost from the population or fixed. The proportion of fixations among the set of simulation runs was used to calculate the fixation probability of a Y-linked mutation.

The starting conditions for these fixation probability simulations require some further explanation. While the initial conditions allow us to make clear conclusions about the evolutionary fates of unique Y-linked variants entering the population, the predictions they yield should be conservative (they should underestimate actual fixation probabilities on the Y) and therefore interpreted as such. The initial state is of a population where X-linked LOF alleles will, if anything, be artificially high because we leave no opportunity for selection on males to influence the initial X-linked variability. Thus, once we introduce a new Y-linked variant, the sheltering effect it experiences will be dampened relative to a population in which both X and Y segregate at mutation-selection balance. This is why fixation probabilities can be less than that of a neutral mutation in scenarios in which our deterministic model predicts their fixation.

Simulations of X versus Y chromosome degeneration. Our models exploring the rates and relative probabilities of gene degeneration on the X and Y (as in Fig. 3) are based on the full model that includes selection, drift, and recurrent mutation at both X-linked and Y-linked loci. For these simulations, our initial population is fixed at both chromosomes for the functional allele. From this initial state, we carried out stochastic simulations until a LOF allele is fixed on either the X or the Y and we recorded the outcome and the dynamical trajectory of the allele that reaches fixation.

All simulations were carried out in R (R Core Team. 2021).

Fixation probabilities of new Y-linked mutations: analytical approximations

Our analytical approach follows that of Nei (*23*), who modelled the fixation probabilities for Y-linked mutations entering a population that was polymorphic for the X. We will assume here that purifying selection on the X is strong relative to the

mutation rate, which ensures that X-linked LOF mutations will be rare. The requirement is easily met as long as selection in females is strong relative to the mutation rate, and the population scaled selection coefficient ($N_e s_f$) is large. We will also ignore X-linked allele frequency differences between sexes, which will be negligible under the preceding assumption that X-linked purifying selection is strong. We focus on recessive LOF alleles.

In ancestral population with no Y-linked variation, males cannot be homozygous, whereas females will sometimes be homozygous for X-linked alleles. Thus, all ancestral purifying selection that governs the initial diversity on the X is due to selection on females. As in Nei (*23*), we assume that the initial X-linked diversity is at equilibrium between mutation, purifying selection, and genetic drift.

In the ancestral population, the expected change in X-linked LOF alleles, per generation, is:

$$M_X \approx -\frac{2}{3} s_f p_X^2 (1 - p_X) + \frac{2\mu_f + \mu_m}{3} (1 - p_X)$$

which is a function of selection in females and mutation in both sexes. Assuming that X-linked genes have three-quarters effective population size of autosomes ($2N_e$ for autosomes; $1.5N_e$ for the X), then the variance in X-linked allele frequency change, per generation, is:

$$V_X = \frac{p_X(1 - p_X)}{1.5N_e}$$

Given these approximations, and applying the standard diffusion approximation for mutation-selection-drift balance (***e.g.***, Wright 1945; Crow and Kimura 1970), the stationary distribution for X-linked LOF mutations will be:

$$f(p_X) = \frac{C}{V} \exp\left(2 \int \frac{M_X}{V_X} dp\right) = C p_X^{N_e(2\mu_f + \mu_m) - 1} (1 - p_X)^{-1} e^{-N_e s_f p_X^2}$$

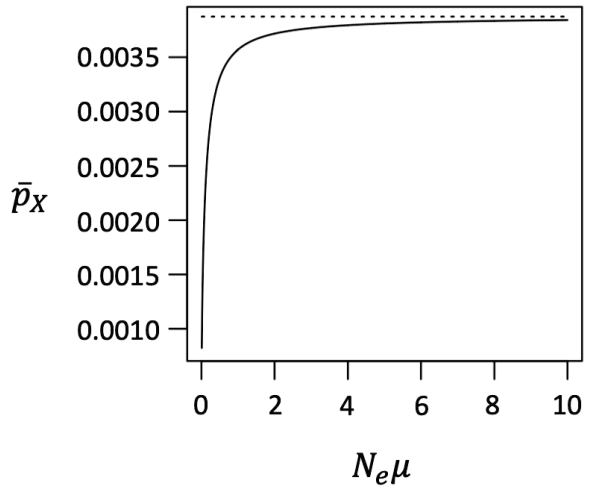
where *C* is a constant that ensures that the distribution integrates to one. Following Nei (1968), and assuming that $N_e s_f$ is large so that the terms $(1 - p_X)^{-1}$ can be ignored, then the mean of the stationary distribution can be approximated as:

$$\bar{p}_X = \frac{\Gamma\left(\frac{1}{2}N_e(2\mu_f + \mu_m) + \frac{1}{2}\right)}{\sqrt{N_e s_f}\,\Gamma\left(\frac{1}{2}N_e(2\mu_f + \mu_m)\right)}$$

where $\Gamma(x)$ is the gamma function. In the special case where $\mu = \mu_f = \mu_m$, the last result simplifies to

$$\bar{p}_X = \frac{\Gamma\left(1.5N_e\mu + \frac{1}{2}\right)}{\sqrt{N_e s_f}\,\Gamma(1.5N_e\mu)}$$

which mirrors eq. (4) of Nei (*23*), but with one exception. In our model we use the term $\sqrt{N_e s_f}$ rather than $\sqrt{1.5N_e s}$ (as in (*23*)) because purifying selection is female-limited owing to the absence of segregating LOF alleles on the Y (Nei's model implies that purifying selection can occur in males as well, but this is not possible under the assumption that the ancestral Y is monomorphic for functional alleles). The following figure plots our expression for $\bar{p}_X$ as a function of $N_e\mu$ (solid curve) and compares it to the deterministic mutation-selection equilibrium, $\hat{p}_X = \sqrt{(2\mu_f + \mu_m)/2s_f}$ (broken line). The latter is a reasonable approximation of the former when $N_e\mu \gg 1$. Smaller population-scaled mutation rates result in mean LOF allele frequencies that are lower than deterministic.



Given an X-linked LOF allele frequency of $\bar{p}_X$, the expected allele frequency change on the Y chromosome (ignoring additional mutations) is:

$$M_Y \approx -s_m \bar{p}_X p_Y (1 - p_Y)$$

The variance in Y-linked allele frequency change per generation is $V_Y = 2p_Y(1 - p_Y)/N_e$. Following Kimura (1962) and Nei (1970), the fixation probability for a Y-linked mutation with initial frequency p_0 will be:

$$U_Y(p_0) = \frac{\int_0^{p_0} G(x)dx}{\int_0^1 G(x)dx} = \frac{e^{N_e s_m \bar{p}_X p_0} - 1}{e^{N_e s_m \bar{p}_X} - 1}$$

where:

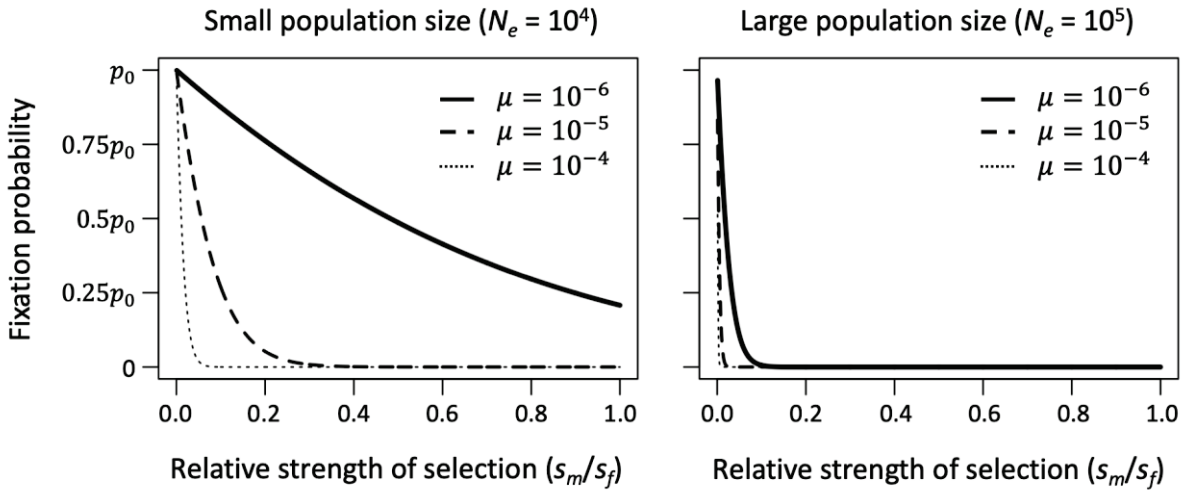
$$G(x) = \exp\left(-2\int_0^x \frac{M_Y}{V_Y} dp_Y\right) = \exp(N_e s_m \bar{p}_X x)$$

With $0.5N_e\mu_m$ new Y-linked mutations per generation, each with initial frequency $p_0 = 2/N_e$, then the rate of fixation for Y-linked LOF alleles should be:

$$R_Y = \frac{1}{2} N_e \mu_m U_Y(p_0) = \frac{1}{2} N_e \mu_m \frac{e^{2s_m \bar{p}_X} - 1}{e^{N_e s_m \bar{p}_X} - 1}$$

The following figure plots fixation probabilities of Y-linked LOF mutations in genes that are essential for females (s_f = 1). Two different population sizes and three LOF mutation rates are considered ($\mu = (2\mu_f + \mu_m)/3$). The results show that sheltering

promotes Y-linked degeneration in cases where the gene is much less important in males than females (s_m << s_f), effective population size is small, and LOF mutation rates are low (*e.g.,* small genes). The latter effect is attributable to the fact that $\bar{p}_X$ declines relative to deterministic predictions with declining $N_e\mu$ (see the preceding figure), which enhances the masking effect the X chromosome has on Y-linked LOF alleles.



Our diffusion approximations, like those of Nei (*23*), predict fixation probabilities at Y-linked genes while holding the X-linked LOF allele frequency constant and equal to the mean predicted from the stationary distribution for the X. In other words, the following approximations neglect evolution on the X chromosome following the introduction of new genetic variants on the Y. This is obviously an unrealistic assumption, but a necessity for any analytical progress with the model. We therefore present the approximations in combination with full stochastic forward simulations that relax this critical assumption by allowing the X and Y to dynamically co-evolve with one another. The simulations suggest that the analytical approximation is most accurate when $N_e\mu$ is large, and it otherwise underestimates the true fixation probability (see Fig. 2 in the main text), presumably because drift often causes

deleterious allele on the X to drop sharply enough that the that sheltering of Y-linked effects become amplified and increase the likelihood that the Y-linked allele then becomes fixed.

Evolutionary fates of male-limited genes in small populations. In large populations, male-limited genes are expected to degenerate on the X chromosome while being retained on the Y. However, as the population declines and the population-scaled mutation rate becomes small, sheltering effects become stronger on both types of chromosomes because LOF alleles become so rare. New X-linked LOF variants become strongly sheltered because the Y tends to be fixed for the functional copy. Likewise, Y-linked variants become strongly sheltered because the X also tends to be fixed for the functional copy. In the limit of mutation-limited evolution, where $N_e\mu \to 0$, the X and Y chromosomes essentially compete to contribute the first LOF allele fixation. Whichever chromosome degenerates first for the male-limited gene, will cause effectively strong selection on the other chromosome to maintain the function of the same gene. And because the population is largely homomorphic prior to the fixation event, the evolutionary dynamics of each male-limited LOF allele that enters the population will be dominated by drift. Such alleles will tend to be either lost or fixed before the next LOF mutation enters the population.

In this mutation-limited environment, the probability that a new X-linked variant arises and is ultimately fixed will be $R_X = \left(2\mu_f + \mu_m\right)/3$ per generation. The corresponding probability on the Y will be $R_Y = \mu_m$. The probability that the first fixation event is on the X will be:

$$\frac{R_X}{R_X + R_Y} = \frac{2 + \mu_m/\mu_f}{2\left(1 + 2\,\mu_m/\mu_f\right)}$$

which corresponds to the limit presented in the main text (e.g., see Fig. 3a).

Appendix 4: Hitchhiking and background selection models

Fixation of LOF alleles via hitchhiking: results for a single Y-linked gene

Exact deterministic dynamics of the hitchhiking model. Assume that are two alleles at the X-linked locus (the functional allele *A* and a LOF allele *a*) and *effectively* three alleles on the Y (owing to complete linkage): the functional allele, a LOF allele in an otherwise ancestral Y chromosome background, and a LOF allele on a background with a beneficial mutation. We assume that the beneficial allele increases male fitness by a factor of 1 + s_b, and fitness effects across loci are multiplicative. The LOF allele frequencies on the X are represented by p_m and p_f (as before), the frequency of the LOF allele on the ancestral background is p_Y, and the frequency of the LOF/beneficial combination is p_b. The evolutionary dynamics are described by the following recursions.

The frequency of the *A* allele in X-bearing male gametes will be:

$$1 - p'_m = \frac{\left(1 - p_f\right)\left((1 - p_Y - p_b) + p_Y(1 - s_m h) + p_b(1 - s_m h)(1 + s_b)\right)(1 - \mu_m)}{\bar{w}_m}$$

where:

$$\bar{w}_m = \left(1 - p_f\right)(1 - p_Y - p_b) + p_Y\left(1 - p_f\right)(1 - s_m h) + p_b\left(1 - p_f\right)(1 - s_m h)(1 + s_b) + p_f(1 - p_Y - p_b)(1 - s_m h) + p_Y p_f(1 - s_m) + p_b p_f(1 - s_m)(1 + s_b)$$

The frequency of the *A* allele in female gametes will be:

$$1 - p'_f = \frac{\frac{1}{2}\left(p_f(1 - p_m) + p_m\left(1 - p_f\right)\right)\left(1 - s_f h\right) + \left(1 - p_f\right)(1 - p_m)}{1 - s_f p_f p_m - \left(p_f(1 - p_m) + p_m\left(1 - p_f\right)\right)s_f h}\left(1 - \mu_f\right)$$

The frequency of *A* on *Y*-bearing gametes will be:

$$1 - p'_Y - p'_b = \frac{(1 - p_Y)\left(1 - p_f s_m h\right)(1 - \mu_m)}{\bar{w}_m}$$

The frequency of the LOF allele on Y-bearing gametes ***without*** the beneficial mutation is:

$$p'_Y = \frac{p_Y\left(1 - s_m h - p_f s_m(1-h)\right)}{\bar{w}_m} + \frac{\mu_m(1-p_Y)\left(1 - p_f s_m h\right)}{\bar{w}_m}$$

The frequency of the LOF allele on Y-bearing gametes ***with*** the beneficial mutation is:

$$p'_b = \frac{p_b\left(1 - s_m h - p_f s_m(1-h)\right)(1+s_b)}{\bar{w}_m}$$

When LOF alleles are completely recessive (h = 0), the system simplifies to:

$$p'_m = 1 - \frac{\left(1-p_f\right)(1+p_b s_b)(1-\mu_m)}{\bar{w}_m}$$

$$p'_f = 1 - \frac{\frac{1}{2}\left(p_f(1-p_m) + p_m\left(1-p_f\right)\right) + \left(1-p_f\right)(1-p_m)}{1 - s_f p_f p_m}\left(1-\mu_f\right)$$

$$p'_Y = \frac{p_Y\left(1 - p_f s_m\right)}{\bar{w}_m} + \frac{\mu_m(1-p_Y)}{\bar{w}_m}$$

$$p'_b = \frac{p_b\left(1 - p_f s_m\right)(1+s_b)}{\bar{w}_m}$$

$$\bar{w}_m = 1 - p_Y p_f s_m + p_b\left(s_b - p_f s_m(1+s_b)\right)$$

Approximations for the fixation probability of a LOF allele. Whether or not a beneficial mutation becomes established on the Y is affected by the marginal fitness of Y chromosomes that carry the beneficial variant. Suppose a beneficial mutation arises on a Y chromosome that carries a LOF allele with homozygous fitness effects s_f and s_m in females and males, respectively. The marginal fitness of a Y chromosome with the pair of alleles is:

$$w_b = (1+s_b)(1-p_X) + (1+s_b)(1-s_m)p_X = (1+s_b)(1-s_m p_X)$$

The marginal fitness of a Y chromosome that carries neither allele is unity. Assume that the LOF mutation is initially at the deterministic equilibrium ($\hat{p}_X$ and $\hat{p}_Y$ on the X

and Y, respectively). Provided there is a net beneficial effect of a Y haplotype carrying a new beneficial mutation ($w_b > 1$), then the probability of establishment for a double-mutant haplotype will be:

$$\sim 2(w_b - 1) = 2(1 + s_b)(1 - s_m\hat{p}_X) - 2 \approx 2(s_b - s_m\hat{p}_X)$$

The approximation is valid when $s_b > s_m\hat{p}_X$ and the establishment probability is zero otherwise.

Supposing that s_b follows an exponential distribution with mean $\bar{s}_b$, then the probability that a new beneficial mutation both arises in association with the LOF allele and establishes is:

$$\Pr(\text{LOF fixes}) \approx \hat{p}_Y \int_{s_m\hat{p}_X}^{\infty} 2(s_b - s_m\hat{p}_X)\frac{1}{\bar{s}_b}e^{-s_b/\bar{s}_b}ds_b = 2\bar{s}_b\hat{p}_Y \exp\left(-\frac{s_m\hat{p}_X}{\bar{s}_b}\right)$$

where $\bar{s}_b^{-1}e^{-s_b/\bar{s}_b}$ is the probability density function for beneficial mutation effects.

Simulations. Assume that the beneficial mutation is initially absent and both the X and Y are at polymorphic equilibrium for the LOF allele. Assuming that LOF mutations are recessive, the equilibria are: $\hat{p}_X = \hat{p}_m = \hat{p}_f = \sqrt{\frac{\mu_f}{s_f}}$ and $\hat{p}_Y = \min\left(1, \frac{\mu_m}{s_m\hat{p}_X}\right)$. To calculate the probability that a single beneficial mutation arising on the Y results in a hitchhiking event that fixes the LOF allele, we first calculated the probability that the mutation arises in association with the LOF allele by sampling from a Bernoulli distribution where $\hat{p}_Y$ is probability of success. In cases the mutation was associated with the LOF allele, we carried out simulations with selection, mutation and drift until the beneficial mutation was fixed or lost from the population. For each simulation run, we sampled a selection coefficient for the beneficial mutation (s_b) from an exponential distribution with mean $\bar{s}_b$. Simulations used the initial allele frequencies: $p_f = p_m = \hat{p}_X$, $p_Y = \hat{p}_Y - N_m^{-1}$ and $p_b = N_m^{-1}$, where N_m is the number of

males in the population (we assume an equal sex ratio so that $N_m = N_f = N/2$, where N is the effective population size).

Each generation of the simulation used exact deterministic selection equations to predict the expected frequencies of females and males after selection. The actual number of breeding adults with each genotype was based on multinomial sampling of genotypes whose sampling probabilities were based on deterministic predictions. The breeding population was comprised of N_m and N_f adult males and females. The fixation probability was estimated as the proportion of beneficial mutations that landed on a Y carrying a LOF allele and was then fixed.

Fixation of LOF alleles via hitchhiking: multiple Y-linked genes

With multiple functional genes on the Y, single hitchhiking events can potentially result in fixation of LOF alleles at multiple genes. For simplicity, suppose that there are n Y-linked genes, each with the same mutation rate to LOF alleles and the same homozygous selection coefficients in each sex. The number of LOF alleles per Y chromosome will then be Poisson distributed with mean and variance $n\hat{p}_Y$. The fixation probability of a beneficial mutation with fitness effect s_b that is associated with k LOF alleles is $\sim 2(1+s_b)(1-s_m\hat{p}_X)^k - 2 \approx 2(s_b - s_m\hat{p}_X k)$, which is valid when $s_b > s_m\hat{p}_X k$ and the probability is zero otherwise. Assuming that beneficial effects are drawn from an exponential distribution, then the fixation probability of a random beneficial mutation that is initially associated with k LOF alleles is:

$$\Pr(\text{fix}|k) \approx \int_{s_m\hat{p}_X k}^{\infty} 2(s_b - s_m\hat{p}_X k)\frac{1}{\bar{s}_b}e^{-s_b/\bar{s}_b}ds_b = 2\bar{s}_b \exp\left(-\frac{s_m\hat{p}_X k}{\bar{s}_b}\right)$$

The expected number of LOF mutations fixed for each new beneficial mutation entering the population is:

$$\mathrm{E}(k_{LOF}) = \sum_{k=0}^{n} k \Pr(\text{fix}|k) \frac{e^{-n\hat{p}_Y}(n\hat{p}_Y)^k}{k!} \approx 2\bar{s}_b \sum_{k=0}^{n} k \exp\left(-\frac{s_m \hat{p}_X k}{\bar{s}_b}\right) \frac{e^{-n\hat{p}_Y}(n\hat{p}_Y)^k}{k!}$$

$$= 2\bar{s}_b n\hat{p}_Y \exp\left(-n\hat{p}_Y\left(1 - \exp\left(-\frac{s_m \hat{p}_X}{\bar{s}_b}\right)\right) - \frac{s_m \hat{p}_X}{\bar{s}_b}\right)$$

where $\hat{p}_X = \hat{p}_m = \hat{p}_f = \sqrt{\frac{\mu_f}{s_f}}$ and $\hat{p}_Y = \min\left(1, \frac{\mu_m}{s_m \hat{p}_X}\right)$

Fixation of mildly deleterious mutations via hitchhiking

It is worth comparing the rates at which sheltered alleles fix by hitchhiking relative to the fixation rates of mildly deleterious mutations. Mildly deleterious mutations are expressed in heterozygotes, with typical dominance coefficients of *h* ~ 0.25 (Manna et al. 2011; Charlesworth 2015). For loci mutating to mildly deleterious alleles, we assume that the heterozygous fitness effects are strong relative to the mutation rate ($s_m h, s_f h \gg \mu_m, \mu_f$), in which case the evolutionary dynamics under mutation and selection are well-approximated by:

$$p'_m = p_f(1 - s_m h) + \mu_m$$

$$p'_f = \frac{1}{2}(p_f + p_m)(1 - s_f h) + \mu_f$$

$$p'_Y = p_Y(1 - s_m h) + \mu_m$$

yielding the following equilibria:

$$\hat{p}_f = \frac{\mu_m(1 - s_f h) + 2\mu_f}{2s_f h + s_m h(1 - s_f h)}$$

$$\hat{p}_m = \hat{p}_f(1 - s_m h) + \mu_m$$

$$\hat{p}_Y = \frac{\mu_m}{s_m h}$$

To distinguish between the mutation rate and fitness effects of mildly deleterious alleles and those of LOF alleles in our sheltering model, let s_d represent

the *heterozygous* effect of mildly deleterious mutations in males, and μ_d represent the mildly deleterious mutation rate per Y-linked locus, so that $\hat{p}_{Y,d} = \mu_m/s_d$ is equilibrium frequency for the locus. Following a similar model by Orr and Kim (1998), we assume that the fitness effects of mildly deleterious alleles are constant across loci, and fitness effects are multiplicative across loci. The fitness associated with a Y chromosome that carries a beneficial mutation and *k* mildly deleterious alleles is $(1 - s_d)^k(1 + s_b)$, and the probability of establishment for such a haplotype is:

$$\sim 2((1 - s_d)^k(1 + s_b) - 1) \approx 2(s_b - s_d k)$$

which is valid $s_b > s_d k$, and the fixation probability is otherwise zero. Supposing that s_b follows an exponential distribution with mean $\bar{s}_b$, then the fixation probability of a new beneficial mutation that arises in association with *k* deleterious alleles is:

$$\Pr(\text{fix}|k) \approx \int_{s_d k}^{\infty} 2(s_b - s_d k)\frac{1}{\bar{s}_b} e^{-s_b/\bar{s}_b} ds_b = 2\bar{s}_b \exp\left(-\frac{s_d k}{\bar{s}_b}\right)$$

The number of mildly deleterious mutations segregating on the Y chromosome is Poisson distributed with mean of $L\mu_m/s_d$, where *L* is the number of Y-linked loci mutating to mildly deleterious alleles. The expected number of mildly deleterious mutations that become fixed for each new beneficial mutation entering the population is:

$$\mathrm{E}(k_d) = \sum_{k=0}^{n} k \Pr(\text{fix}|k) \frac{e^{-L\mu_m/s_d}(L\mu_m/s_d)^k}{k!} \approx 2\bar{s}_b \sum_{k=0}^{n} k \exp\left(-\frac{s_d k}{\bar{s}_b}\right) \frac{e^{-L\mu_m/s_d}(L\mu_m/s_d)^k}{k!}$$

$$= 2\bar{s}_b \frac{U_d}{s_d} \exp\left(-\frac{U_d}{s_d}\left(1 - \exp\left(-\frac{s_d}{\bar{s}_b}\right)\right) - \frac{s_d}{\bar{s}_b}\right)$$

where $U_d = L\mu_m$ is the total Y chromosome mutation rate to mildly deleterious alleles.

For purposes of comparison, let us consider the case where LOF alleles are homozygous lethal ($s_f = s_m = 1$), in which case

$$\mathrm{E}(k_{LOF}) \approx 2\bar{s}_b \frac{U_{LOF}}{\sqrt{\mu_f}} \exp\left(-\frac{U_{LOF}}{\sqrt{\mu_f}}\left(1-\exp\left(-\frac{\sqrt{\mu_f}}{\bar{s}_b}\right)\right)-\frac{\sqrt{\mu_f}}{\bar{s}_b}\right)$$

where $U_{LOF} = n\mu_m$ is the total Y chromosome mutation rate to LOF alleles. Here, we see that the expressions for mildly deleterious and LOF mutations become identical when $s_d = \sqrt{\mu_f}$. Fixation of homozygous lethal LOF alleles becomes more permissible than mildly deleterious alleles when $s_d > \sqrt{\mu_f}$. Fixation of LOF alleles becomes even more permissible if genes are not essential for males ($s_m < 1$).

Fixation of LOF alleles via background selection

We will assume that LOF mutations are recessive, and that a set of background loci are primary selected in heterozygous state. We will assume that background loci are at deterministic mutations-selection balance, and that LOF mutations are initially absent from the Y chromosome. The X-linked LOF allele frequencies are at the mutation-selection-drift equilibrium conditioned on the absence of LOF mutations on the Y.

Following the general approach of Charlesworth (1994) for modelling background selection, the fixation probability of a new LOF mutation arising on the Y will depend on:

- The homozygous fitness cost of the LOF allele in males (s_m)
- The census (N) effective population size (N_e); we assume an equal sex ratio and equal variance in reproductive success
- The initial frequency of new mutations on the Y chromosome ($p = 2/N$)
- The fraction of the Y chromosome that is free of deleterious mutations at background loci (f_0, further defined below)

Following the standard theory of mutation-selection balance (Orr 2000), and assuming no LD among the background loci, the frequency of the least loaded Y chromosome class will be:

$$f_0 = e^{-U_Y/s_H}$$

where U_Y is the total Y-linked mutation rate at the background loci and s_H is the harmonic mean heterozygous fitness cost of background mutations.

Because a new LOF will be eliminated if it arises in a loaded background, we focus on the subset that lands on an unloaded background. For those mutations, the initial frequency will be $p_0 = p/f_0$ (Charlesworth 1994). Conditioned on the LOF mutation residing in the least loaded class, the mean and variance in allele frequency change for a LOF mutation at frequency p_Y will be (respectively):

$$M_Y \approx -s_m \bar{p}_X p_Y (1 - p_Y)$$

And

$$V_Y = \frac{2p_Y(1 - p_Y)}{f_0 N_e}$$

where $\bar{p}_X$ is the expected X-linked frequency of the LOF allele (Nei 1970; see Appendix 3). Conditioned on the Y-linked LOF mutation landing in the least loaded class, the probability of its fixation will be:

$$u_Y(p_0) = \frac{\exp(f_0 N_e s_m \bar{p}_X p_0) - 1}{\exp(f_0 N_e s_m \bar{p}_X) - 1} = \frac{\exp\left(\frac{2N_e}{N} s_m \bar{p}_X\right) - 1}{\exp(f_0 N_e s_m \bar{p}_X) - 1}$$

Since the probability of a random LOF mutation landing on an unloaded background is f_0, the overall fixation probability of a LOF mutation will be:

$$\Pr(\text{LOF fixes}) = f_0 u_Y(p_0) = f_0 \frac{\exp\left(\frac{2N_e}{N} s_m \bar{p}_X\right) - 1}{\exp(f_0 N_e s_m \bar{p}_X) - 1}$$

For point of contrast, we can also calculate the fixation rate of mildly deleterious mutations through background selection. For mutation with heterozygous fitness cost of s_d, the probability of fixation under background selection is:

$$\Pr(\text{del. mutation fixes}) = f_0 \frac{\exp\left(\frac{2N_e}{N} s_d\right) - 1}{\exp(f_0 N_e s_d) - 1}$$

The fixation probability for LOF alleles will be higher whenever $s_m \bar{p}_X < s_d$. In the case of lethal mutations in a very large population, this condition simplifies to $\sqrt{\mu_f} < s_d$. The condition becomes more permissive for LOF mutations in non-essential genes.

Supplementary Figures

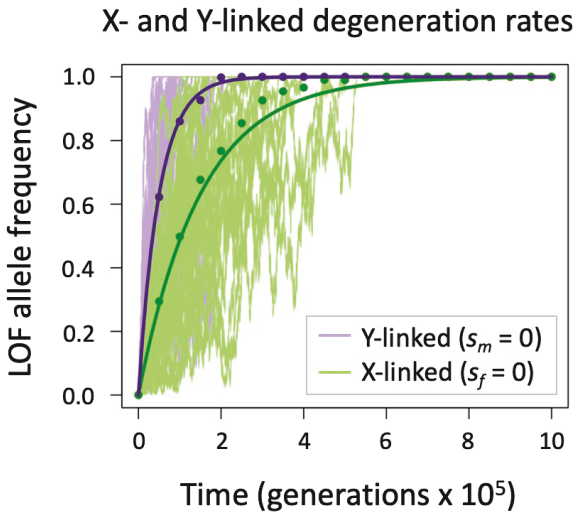


Figure S1. Evolutionary dynamics of LOF mutations that eventually become fixed in sex-limited genes: Further results. Degeneration is shown in a smaller population than is presented in Fig. 3b of the main text. Here, the population size is an order of magnitude smaller (N_e = 10^4 rather than N_e = 10^5), but with all other parameters remaining the same (*i.e.*, $\mu_f = 10^{-5}$, $\mu_m = 2\mu_f$, s_f = 1 for Y-linked genes, and s_m = 1 for X-linked genes). 30 simulation runs are shown for each chromosome, with individual trajectories (thin, pale lines) scattered about the analytical predictions (bold curves) and circles denoting mean LOF frequencies across the set of simulated trajectories. The average trajectories LOF allele frequency trajectories are highly predictable (i.e., we see good alignment between the bold curves and the circles). However, individual allele frequency trajectories exhibit greater variability in small compared to large populations (se Fig. 3b in the main text for contrast), as indicated by the broad range of frequency states denoted in pale purple (for the Y) and pale green (for the X).

Supplementary Methods

The following databases were mined for RNA-seq datasets containing whole-body data from both males and females: the NCBI Short Reads Archive (https://www.ncbi.nlm.nih.gov/sra), the NCBI GEO Datasets (https://www.ncbi.nlm.nih.gov/gds), and Google Scholar (https://scholar.google.com). Priority was given to datasets for which both adult and juvenile stages were sampled, or which allowed for broader phylogenetic sampling (brine shrimp, tardigrade, and zebrafish datasets were included despite only containing adult data as their (sub)phylum was otherwise not represented). The list of datasets with information on replicate number, developmental stages that were sampled, and data type, are provided in Supplementary Table S1.

When gene expression values were provided, either as GEO datasets or supplementary tables of the associated manuscripts, these were downloaded and used directly. If expression values were provided as raw read counts, CPM values (counts per million) were estimated before proceeding. FPKM, RKPKM and TPM values were used directly. Only RNA-seq reads were available for zebrafish (Danio rerio) and wood white butterflies (Leptidea sinapis). For these two species, CDS sequences were obtained from Ensembl (Danio_rerio.GRCz11.cds.all.fa) and from the NCBI genome assembly page (GCF_905404315.1_ilLepSina1.1_cds_from_genomic.fna) respectively. Only the longest CDS was kept for each gene, and expression values were obtained using Kallisto (version 0.50.1; Bray et al. 2016).

Gene expression values were quantile-normalized across each dataset with NormalyzerDE (Willforss et al. 2019). For each gene, expression values were averaged by sex. In order to avoid including very low expression genes, which may not be biologically relevant, we removed genes that had summed male and female expression below the 25th percentile. A second round of quantile normalization was performed on the averaged male and female values (as filtering after averaging replicates can lead to uneven distributions despite the initial normalization). In order to avoid removing sex-specific genes, 0.1 was added to all expression values before the log2 ratio of male to female expression was calculated; this ratio was then used to select genes with varying fold-change levels of sex-bias. To infer the number of sex-specific genes, we also calculated the metric male_expression/ (male_expression+female_expression), and selected genes for which this metric was below 0.01 (female-specific) or above 0.99 (male-specific).

In the case of humans, no whole-body data was obviously available, and they were therefore not included in the estimation of sex-bias. Instead, median gene expression values for specific tissues (obtained from 4 to many hundreds of individuals depending on the tissue) were downloaded from GTEX:

GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz from: https://www.gtexportal.org/home/downloads/adult-gtex#bulk_tissue_expression

While for most tissues the median expression is derived from male and female individuals, several tissues are sex specific: Uterus, Vagina, Ovary, Ectocervix, Endocervix, Fallopian Tube, and Breast Mammary are female-specific, and Prostate

and Testis are male-specific. For each gene, expression was summed across all tissues. Genes with total expression below the 25th percentile were removed, to avoid including very low expression genes. Genes were then classified as male-specific, if over 99% of their expression came from male-specific tissues, and female-specific, if over 99% of their expression came from female-specific tissues.

Supplementary References

Bray NL, Pimentel H, Melsten P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34:525-527.

Charlesworth B, Charlesworth D. 2010. *Elements of Evolutionary Genetics*. Roberts and Company Publishers, Greenwood Village, CO.

Crow JF, Kimura M. 1970. *An Introduction to population genetics theory*. Harper and Row: New York, NY.

Kimura M. 1962. On the Probability of Fixation of Mutant Genes in a Population. *Genetics* 47:713–719.

Nei M. 1968. The frequency distribution of lethal chromosomes in finite populations. *Proc Natl Acad Sci USA* 60:517-524.

Otto SP, Day T. 2007. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press: Princeton, NJ.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Willforss J, Charade A, Levander F. 2019. NormalyzerDE: Online tool for improved normalisation of comics expression data and high-sensitivity differential expression analysis. *J Proteome Research* 18:732-740.

Wright S. 1945. The differential equation of the distribution of gene frequencies. *Proc Natl Acad Sci USA*. 31:382-389.

Chapter 4: Evidence of a Slower-Z effect in *Schistosoma japonicum*

Andrea Mrnjavac and Beatriz Vicoso

Biorxiv, doi: https://doi.org/10.1101/2024.07.02.601697

Author contributions: Beatriz Vicoso and Andrea Mrnjavac designed the study. AM performed the analysis. AM wrote the first draft of the manuscript, BV and AM the current version.

Abstract

Sex-linked and autosomal loci experience different selective pressures and evolutionary dynamics. X (or Z) chromosomes are often hemizygous, as Y (or W) chromosomes often degenerate. Such hemizygous regions can be under greater efficacy of selection, as recessive mutations are immediately exposed to selection in the heterogametic sex (the so-called Faster-X or Faster-Z effect). However, in young non-recombining regions, Y/W chromosomes often have many functional genes, and many X/Z-linked loci are therefore diploid. The sheltering of recessive mutations on the X/Z by the Y/W homolog is expected to drive a Slower-X (Slower-Z) effect for diploid X/Z loci, i.e. a reduction in the efficacy of selection. While the Faster-X effect has been studied extensively, much less is known empirically about the evolutionary dynamics of diploid X or Z chromosomes. Here, we took advantage of published population genomic data in the female-heterogametic human parasite *Schistosoma japonicum* to characterize the gene content and diversity levels of the diploid and

hemizygous regions of the Z chromosome. We used different metrics of selective pressures acting on genes to test for differences in the efficacy of selection in hemizygous and diploid Z regions, relative to autosomes. We found consistent patterns suggesting reduced N_e, and reduced efficacy of purifying selection, on both hemizygous and diploid Z regions. Moreover, relaxed selection was particularly pronounced for female-biased genes on the diploid Z, as predicted by Slower-Z theory.

Introduction

Sex chromosomes, such as the X and Y of mammals, or the Z and W of birds, originate from standard pairs of autosomes. After they are coopted for sex determination, the two chromosomes typically stop recombining and start diverging from each other (Jay et al., 2024). This leads them to evolve differently from autosomes. The most striking aspect of this is the progressive degeneration of the non-recombining Y/W that is observed in many clades (Charlesworth, 2021). However, it has become increasingly appreciated that evolutionary rates on the X chromosome (or Z, but explained in terms of the X for simplicity) are also shaped by unusual evolutionary pressures. All else being equal, the effective population size of the X chromosome is ¾ the autosomal effective population size, while Y chromosomes have a population size of only ¼ the autosomal one (Vicoso and Charlesworth, 2009). Both X and Y chromosomes exhibit sex-biased transmission: the X resides in females ⅔ of the time, while the Y is in males 100% of the time (Furman et al, 2020). Furthermore, the degeneration of the Y chromosome (Bachtrog, 2013) leaves X-linked loci hemizygous in males. Selection is more efficient for hemizygous X-linked loci despite its reduced effective population size, as recessive mutations are always exposed to selection in males. On the other hand, on the autosomal loci recessive mutations are mostly found in heterozygous form and their effect is masked by the dominant allele.. This should lead to higher rates of adaptive evolution on the X chromosome than autosomes if new beneficial mutations are on average recessive, a hypothesis known as the Faster-X effect (Charlesworth et al., 1987, Vicoso and Charlesworth, 2006). Support for the Faster-X effect comes

from the observation of elevated *dN/dS* on the X chromosome, or elevated values of α, the inferred proportion of nonsynonymous divergent sites that were fixed by positive selection (Meisel and Connallon, 2013), in various clades.

There is empirical evidence for high rates of non-synonymous evolution on the X in mammals and Drosophila, and on the Z in birds and arthropods (Meisel and Connallon, 2013, Charlesworth et al., 2018, Mank et al., 2010, Mongue et al., 2022). However, evidence suggests that this is not always driven by increased rates of adaptation. While there is evidence of increased rates of adaptive divergence on various X chromosomes (Garrigan et al. 2014, Avila et al. 2014, Veeramah et al. 2014, Kousathanas et al. 2014, Campos et al. 2014, Charlesworth et al. 2018), the Faster-Z effect has been interpreted as being the result of stronger drift on the Z chromosome of several species (Mank et al., 2010, Hayes et al., 2020, Chase et al., 2023, Mongue and Baird, 2024, but see Wanders et al., 2024). This is possibly due to the fact that the Z spends more time in males: males usually have a higher variance in reproductive success, resulting in the more extreme reduction in the effective population size for the Z chromosome than for the X (Vicoso and Charlesworth, 2009, Mank, Vicoso et al., 2010). In smaller populations, a higher proportion of mutations entering the population is effectively neutral, contributing to faster non-adaptive evolution (Mank et al., 2010, Mank, Vicoso et al., 2010). On the other hand, evidence of faster and more adaptive Z was found in some Lepidoptera, which typically have a larger population size than birds (which may make the Z chromosome less sensitive to the reduction in the effective population size) (Mongue et al., 2022, Villavicencio, 2024).

X-linked loci in young non-recombining regions, which still have a non-degenerated homologous region on the Y chromosome, are not hemizygous,

but diploid in males, as they have a functional, albeit non-recombining, gametolog on the Y. Unlike loci on older, hemizygous X chromosomes, such “diploid X” loci are not expected to adapt faster than autosomal loci. New mutations that arise on a diploid X region are always heterozygous in males, and, if (partly) recessive, are (partially) sheltered from selection by the functional copy on the Y. This is expected to cause reduced efficiency of selection in males on diploid X region, slower adaptation of male-important genes and accumulation of deleterious mutations on male-important genes, i.e. a "Slower-X" effect (Mrnjavac et al., 2023).

The evolutionary patterns of young non-recombining regions on the X or Z have been studied less often, as population data is needed to detect very young non-recombining regions with non-degenerated Y counterparts (Darolti et al., 2022), but a few have found some support for Slower-X/Z evolution. Neo-X regions (with the corresponding Y chromosomes showing intermediate levels of degeneration) in several Drosophila species experience accelerated pseudogenization, driven by the loss of male-important genes (Nozawa et al., 2016, Nozawa et al., 2021). In the plant *Silene latifolia* there is evidence of relaxed purifying selection on young X-linked genes with a non-degenerated Y homolog (Krasovec et al., 2018). Recently, a study in the butterfly genus Leptidea provided direct empirical evidence of reduced efficiency of selection for female-biased and unbiased genes on the young non-recombining region of the Z chromosome with a non-degenerated W, i.e. of a Slower-X (Slower-Z) effect (Hook et al., 2024). On the other hand, some studies have found similar rates of divergence for diploid X/Z genes as for (pseudo)autosomal genes. The young X-linked region of the plant *Salix dunni* is enriched for transposable elements and pseudogenes, but divergence of X-linked genes is similar to the autosomal divergence, possibly because the X-linked region is

very young and there was no time for non-adaptive substitutions to accumulate (He et al., 2021). Similarly, in Sylvioidea songbirds, there is no difference in evolutionary rates between the neo-Z and autosomes (Leroy et al., 2021). Darolti et al. (2023) further showed that while Faster-X correlates with hemizygosity in various species of poeciliid fishes, no evidence of increased drift or differences in divergence rates could be detected between diploid X chromosomes and their respective autosomes. Therefore, the broad relevance of the Slower-X effect in taxa with young sex-linked regions is still to be fully explored.

Blood flukes (genus Schistosoma) are a promising model for studying the evolutionary dynamics of sex-linked regions of different ages. While they all share an ancestral pair of ZW chromosomes, the non-recombining part of the sex chromosomes has been expanded independently in different lineages (Picard et al., 2018). A very young non-recombining region of the Z chromosome has been recently identified in the Asian species *Schistosoma japonicum* (Elkrewi et al. 2021, Xu et al., 2023). This region has over 700 functional W genes (ZW dS < 0.085), which makes the corresponding Z region diploid but non-recombining in females. We expect such a region to be under reduced efficiency of selection in females, compared to autosomes and hemizygous Z (Mrnjavac et al., 2023). Here, we use publicly available comparative (Protasio et al., 2012, Luo et al., 2022), population (Luo et al., 2022) and expression data (Wang et al., 2017) to test those predictions.

Methods

Strata determination

To identify hemizygous and diploid Z regions, we performed female-to-male coverage analysis and male-to-female F_{st} analysis as in Elkrewi et al. (2021), using the recently published male *Schistosoma japonicum* genome assembly (GCA_021461655.1) (Luo et al., 2022). Briefly, female (SRR6841388) and male (SRR6841389) *Schistosoma japonicum* reads were separately mapped to the *Schistosoma japonicum* male genome using *bowtie2* (Langmead and Salzberg, 2012). Only uniquely mapped reads were kept. Coverage for male and female reads was calculated with *soap.coverage* (Luo et al. 2012) per 10000 bp windows. Log2(F/M coverage) was calculated and visualised in R (R Core Team, 2023). Coordinates of the hemizygous Z region were determined as the limits of the Z chromosome region where log2(F/M coverage) values are centred at -1, meaning there are twice as many reads in males compared to females (Z chromosome coordinates: 24470001-49640001).

The F_{st} analysis also followed the approach of Elkrewi et al. (2021), but using the new chromosome-level genome assembly. F_{st} between male and female reads (PRJNA650045, sex of the individual library was determined from Elkrewi et al., 2021) was calculated with *vcftools* (Danecek et al., 2011) and visualised in R. The diploid Z region was determined as the region for which the male:female F_{st} values were consistently above the 95 percentile of the distribution across the genome (Z chromosome coordinates: 49640001-76240000). In this region 62.67% of windows had male:female F_{st} values above the 95 percentile of the genome-wide distribution

and 90.42% of reads had male:female F_{st} values above the 90 percentile of the genome-wide distribution.

Sex-biased expression analysis

Publicly available whole-body expression data was downloaded for 24 male and 24 female *Schistosoma japonicum* individuals, (PRJNA343582, Wang et al., 2017). Normalised gene expression was obtained per gene, per sample, using *kallisto* (Bray et al., 2016) and *sleuth* (Pimentel et al., 2017). We filtered out the genes with no expression in both of the sexes. We estimated sex-biased gene expression as specificity measure, or SPM (Kryuchkova-Mostacci and Robinson-Rechavi, 2017), the square of mean expression in females divided by the sum of the square of mean expression in females and the square of mean expression in males, using R. SPM=0 corresponds to male-limited expression, while SPM=1 corresponds to female-limited expression. For further analyses, genes with SPM values lower than 0.3 were assigned as male-biased, and genes with SPM values above 0.7 were assigned as female-biased. Distributions of SPM values in autosomes, hemizygous Z and diploid Z regions were visualised and compared in R. Sex-bias distribution was compared between the Z chromosome and autosomes with Kolmogorov–Smirnov test and Mann-Whitney-Wilcoxon test.

Divergence inference

We identified orthologs between *Schistosoma japonicum* and closely related species *Schistosoma mansoni* (65% median synonymous divergence, Picard et al., 2018) as the best reciprocal blat (BLAST-Like Alignment Tool), (Kent, 2002) hits between *Schistosoma japonicum* and *Schistosoma mansoni* coding sequences (assembly version GCF_000237925.1, Protasio et al., 2012) (we chose the longest coding sequence per gene for the analysis). Orthologs were aligned using TranslatorX with the “gblocks” option (Abascal et al., 2010). Divergence between orthologs was calculated with KaKs_Calculator 2.0 (Wang et al., 2010). Yang-Nielsen estimates of K_a/K_s were obtained per gene, as well as the number of nonsynonymous and synonymous substitutions per gene. These parameters were visualised and compared in R.

Polymorphism inference

We downloaded a publicly available population genomic dataset (PRJNA789681) from NCBI database (https://www.ncbi.nlm.nih.gov/), including whole genome sequences of 48 *Schistosma japonicum* adult male individuals sampled from several locations in South-East Asia (we did not include the Taiwan and the Philippines subpopulations in our analysis as those subpopulations have extremely reduced levels of diversity and could have biased our analysis), corresponding to their worldwide range (Luo et al., 2022). We trimmed the reads with *Trimmomatic* (Bolger et al., 2014). Trimmed paired reads were mapped to the *Schistosoma japonicum* male genome assembly (GCA_021461655.1) using *bowtie2* with *--end-to-end* and *--sensitive* parameters, separately for every individual.

Non-uniquely mapped reads were removed. SAM files were reformatted into sorted BAM files using samtools (Li et al., 2009). Variant calling was performed with *bcftools mpileup* option (Li et al., 2009), using (48) 72 individual bam files as input. Variants were filtered by quality, *bcftools view -i '%QUAL>=20'*, only biallelic sites were kept, *--max-alleles 2*, and indels were removed, *--exclude-types indels*. bcf file was reformatted into vcf file. Rare variants (maf<15%) were removed with *vcftools* using *--maf 0.15 --max-missing 0.9* options. Polymorphic sites were annotated as synonymous or nonsynonymous using *snpEff* and *SnpSift* (Cingolani et al., 2012).

Population genomic analyses

α denotes the proportion of nonsynonymous substitutions that are fixed by positive selection, and is based on the classic MK test. α per gene was calculated as 1-((number of nonsynonymous polymorphisms per gene (P_n) / number of synonymous polymorphisms per gene (P_s))/(number of nonsynonymous substitutions per gene (D_n) / number of synonymous substitutions per gene (D_s))) (Smith and Eyre-Walker, 2002, Charlesworth and Charlesworth, 2010, Chapter 6.4), after removing variants below 15% frequency (Fay et al., 2001, Al-Saffar and Hahn, 2022), using R. Distributions of α values for different categories of sex-bias, and different genomic regions: hemizygous Z, diploid Z and autosomal one, were visualised and statistically compared in R. Statistically significant differences between distributions were tested with the Wilcoxon-Mann-Whitney test.

In addition to α, we calculated Direction of Selection (DoS) as DoS=Dn/(Dn+Ds)-Pn/(Pn+Ps) (Stoletzki and Eyre-Walker, 2011). Direction of selection is a measure of direction and degree of departure from neutrality, based on the MK test, that corrects for biases that arise from a small number of observations

(Stoletzki and Eyre-Walker, 2011). Statistically significant differences between distributions were tested with the Wilcoxon-Mann-Whitney test.

Nucleotide diversity along the genome, in 10000 bp windows, was calculated using pixy (Korunes and Samuk, 2021). Statistically significant differences between distributions in different genomic regions were tested with the Wilcoxon-Mann-Whitney test.

Results

Hemizygous and diploid regions of the Z chromosome

Elkrewi et al. (2021) and Xu et al. (2023) recently described evolutionary strata of different ages along the Z chromosome of *S. japonicum*, and in particular the presence of a large section of the ZW pair that no longer recombines, but still exists on the W. However, a highly fragmented genome was used in Elkrewi et al. (2021), and no population genomics data was used to infer young non-recombining regions in Xu et al. (2023). We therefore set out to define precise boundaries of the diploid and hemizygous Z regions on the published chromosome-level assembly of *S. japonicum* (Luo et al., 2022). Using both coverage patterns and genetic differentiation between a population of males and females, we recovered large contiguous hemizygous and non-recombining but diploid Z regions (Xu et al., 2023, Elkrewi et al. 2021) (**Figure 1**). In particular, a large region where female coverage is consistently half of male coverage is consistent with the degeneration of the homologous region of the W chromosome, i.e. this Z chromosome region is haploid

or hemizygous in females. A second region shows no difference in coverage between males and females, but shows a high level of genetic differentiation between the Z and W, measured as male-to-female F_{st}, consistent with a recent loss of recombination between the Z and W, and a non-degenerated homologous region on the W chromosome. The hemizygous and diploid Z regions contain, respectively, 703 and 624 genes.

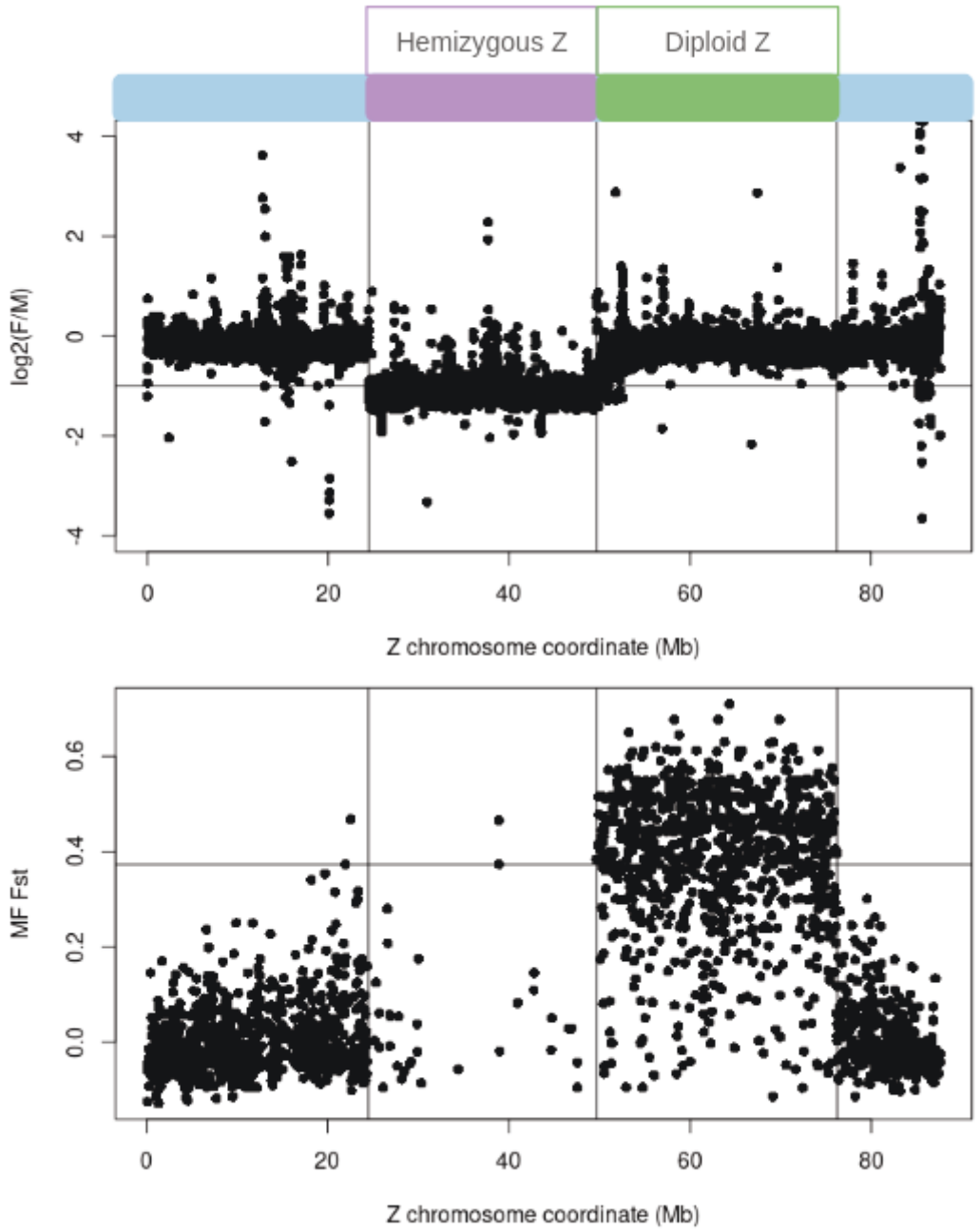


Figure 1. *Schistosoma japonicum* Z chromosome strata: hemizygous and diploid Z regions. Upper panel: Log2(Female coverage / Male coverage) along the Z chromosome. Hemizygous Z region has male coverage that is two times the female coverage. Lowe panel: Male-to-Female F_{st} along the Z chromosome. Diploid Z region has equal coverage in males and females,but has high levels of Male-to-Female F_{st}.

Lower effective population size on the Z chromosome

Since males have two Z chromosomes but females only have one (compared to two sets of autosomes in each sex), the expected effective population size of the Z is ¾ of that of the autosomes. We estimated total genetic pairwise diversity (π) from a population of 48 males from 6 sampling locations, and used it to infer the effective population size of the hemizygous and diploid Z regions relative to that of the autosomes in *Schistosoma japonicum*. Both hemizygous and diploid Z regions show lower than expected nucleotide diversity compared to autosomes, with the Z/A ratio of median nucleotide diversity 0.37 for diploid Z region and 0.09 for hemizygous Z region (p<2e-16, p<2e-16 respectively) (**Figure S1**), which suggests that the effective population size of the Z chromosome could be even lower than ¾ of the autosomal effective population size. Similar estimates were obtained when only synonymous sites were used to calculate diversity (Z:A ratios of 0.395 and 0.235 for the diploid and hemizygous regions). Given the apparent young age of the diploid Z region, we took advantage of the reduced π on the diploid Z to check that the loss of ZW recombination was found in every population. Two populations had extremely reduced levels of diversity and were excluded from further analysis. In each of the other populations, the diploid Z region had reduced levels of diversity compared to the autosomes (p<2e-16, Mann-Whitney-Wilcoxon test), confirming that it is non-recombining throughout the geographical range of the species (**Figure S1B**). Such a reduction was not observed in the pseudoautosomal region (p>0.1).

Reduced efficacy of purifying selection on both the hemizygous and diploid Z regions

We first measured the divergence between *Schistosoma japonicum* and the closely related species *S.mansoni* to estimate synonymous (Ks) and nonsynonymous (Ka) substitution rates per gene. **Figure 2** shows the distribution of K_a/K_s per gene, for hemizygous Z, diploid Z and autosomes. Distributions of K_a/K_s per gene in different genomic regions were statistically compared using the Wilcoxon-Mann-Whitney test. K_a/K_s is significantly higher on the hemizygous Z (hZ) compared to both autosomes (A) and to the diploid Z (dZ) [median(hZ)=0.1152, median(dZ)=0.0966, median(A)=0.0901, p= 1.327e-13, p= 0.0003441, respectively], while diploid Z genes show a slight increase compared to the autosomes (p = 0.02237). Median values are provided in the Supplementary Table 1. Synonymous divergence is significantly lower on the Z chromosome, with hemizygous Z exhibiting the lowest synonymous divergence [median(hZ)=0.9071, median(dZ)=0.9729, median(A)=1.121, **Table S1**., hZ vs A: p < 2.2e-16, dZ vs A: p = 1.656e-11, hZ vs dZ: p = 2.493e-5]. Overall these results support the faster protein divergence of Z-linked genes compared to the autosomes.

In order to investigate whether this fast evolution of Z-linked genes was driven by an increase in positive selection or by a decrease in the efficacy of purifying selection, we obtained estimates of synonymous and nonsynonymous polymorphism across the sampled populations (excluding the two that did not harbour any diversity). Diversity levels in different genomic regions were compared with the Wilcoxon-Mann-Whitney test. The Z chromosome has higher levels of nonsynonymous to synonymous diversity compared to the autosomes

[median(hZ)=0.3242, median(dZ)=0.2295, median(A)=0.1942, hZ vs A: *p*=5.9e-15, dZ vs A: *p*= 0.0057, **Figure 2, Table S1**]. This is in line with the reduced effective population size and the resulting reduced efficiency of selection in removing slightly deleterious mutations from the population. We also calculated α per gene, a regularly used measure of adaptive evolution based on the McDonald-Kreitman test (McDonald and Kreitman, 1991, Smith and Eyre-Walker, 2002, Charlesworth and Charlesworth, 2010, Chapter 6.4). Positive α values suggest positive selection, while negative α values mean there is an excess of nonsynonymous polymorphisms segregating in the population. This excess is usually caused by segregating slightly deleterious mutations, that is, lower efficiency of selection, or, balancing selection (Charlesworth and Charlesworth, 2010, Chapter 6.4). We removed rare polymorphic sites (with minor allele frequency below 15%) from the analysis to minimize the contribution of deleterious mutations segregating at low frequencies (Fay et al., 2001). We also removed the genes with no polymorphism for the downstream analysis, which greatly reduced the number of genes: in the hemizygous Z region up to 80% of the genes did not exhibit any polymorphism after filtering out rare variants, while in the diploid Z region and autosomes, from 20% to 60% of genes exhibited no polymorphisms (**Table S7**). This reflects extremely low levels of nonsynonymous and synonymous polymorphisms segregating on the hemizygous Z region (**Figure 2**), congruent with the extreme reduction in the population size for hemizygous Z compared to the rest of the genome (**Figure S1**). The small number of genes in some categories (**Table S7**), especially on hemizygous Z, greatly reduced our statistical power. A recent study (Al-Saffar and Hahn, 2022) shows that the Fay et al. (2001) approach underestimates the true value of α, but accurately reflects differences between the X-chromosome and autosomes. It should be noted that our

values of α likely underestimate the true proportion of nonsynonymous substitutions fixed by positive selection, however, here we are interested in relative differences in the strength of selection in different genomic regions. Furthermore, α values for the diploid Z region should be interpreted taking into account that observed higher diversity for female-biased genes is not reflected in higher divergence, as this region only recently became diploid and its divergence reflects autosomal patterns. In agreement with purifying selection being relaxed on both the hemizygous and diploid Z, genes in both regions showed reduced α values compared with autosomal genes [median(hZ)=-0.4922, median(dZ)=-0.0974, median(A)=0.0721, *p*=4.5e-10 and *p*=0.005, respectively, **Figure S3A**]. Once again, the effect was stronger for the hemizygous Z region than for the diploid Z.

In addition to α, we calculated a second metric of selection strength, the Direction of Selection (DoS, Stoletzki and Eyre-Walker, 2011), and the results were qualitatively similar (**Figure S4A**). The lower values of DoS observed for both hemizygous and diploid Z genes compared to autosomal genes [median(hZ)=-0.0913, median(dZ)=-0.0232, median(A)=0.0166, *p*=4.7e-10 and *p*=0.0043] suggest that there is an excess of nonsynonymous polymorphisms that reach high frequencies in the population (as we removed rare variants) on both regions of the Z.

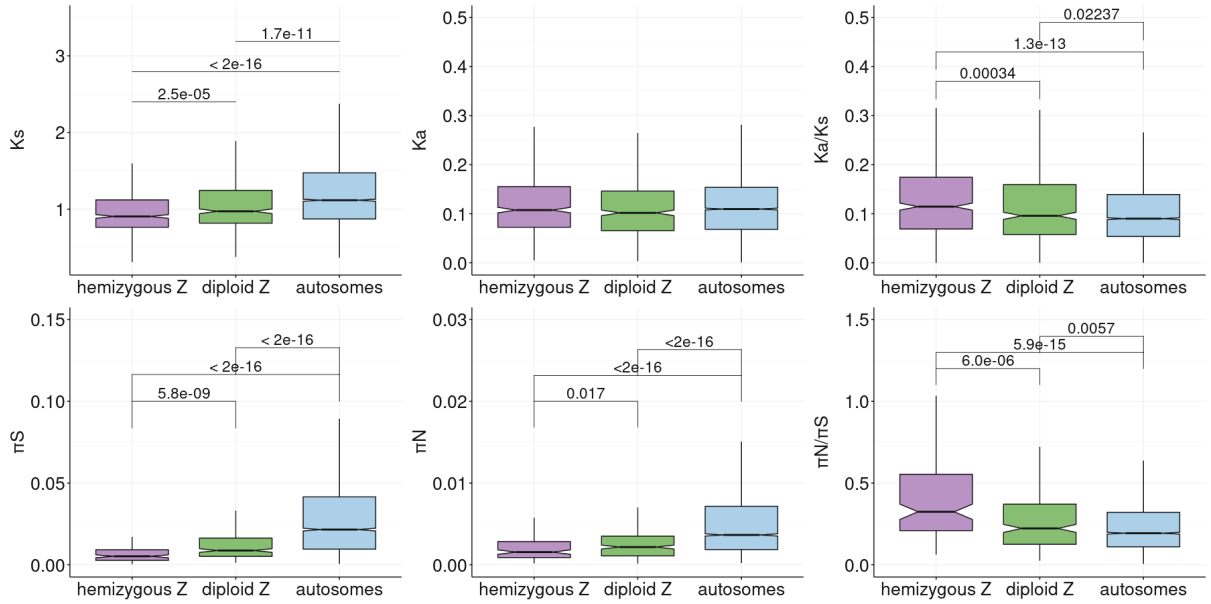


Figure 2. Synonymous (K_s) and nonsynonymous divergence (K_a) between *Schistosoma japonicum* and *Schistosoma mansoni*, and their ratio (K_a/K_s), synonymous (π_S) and nonsynonymous diversity (π_N) and their ratio (π_N/π_S) for hemizygous and diploid Z and autosomes.

Different evolutionary dynamics of hemizygous and diploid Z-linked sex-biased genes

Genes with sex-specific functions are expected to evolve differently on hemizygous and diploid Z-linked regions. To test this hypothesis, we used sex-specific patterns of expression as a proxy for function. We measured sex bias in the expression of the *Schistosoma japonicum* whole body throughout the development, as SPM (specificity metric, Kryuchkova-Mostacci and Robinson-Rechavi, 2017), where 0 corresponds to male-specific expression and 1 corresponds to female-specific expression. **Figure S2**. shows distributions of sex-biased expression on autosomes, in the diploid Z region and in the hemizygous Z region. Distributions of sex-biased expression were statistically compared using the Wilcoxon-Mann-Whitney test and Kolmogorov-Smirnov test. The hemizygous Z

region is significantly masculinized (p < 2.2e-16 for both tests), in agreement with its incomplete mechanism of dosage compensation (Picard et al., 2018), while the diploid Z region exhibits a small shift towards male-biased expression (p = 2.487e-6, p= 0.0003627, for Kolmogorov-Smirnov and Wilcoxon-Mann-Whitney tests respectively).

Figure 3. shows nonsynonymous to synonymous substitution rates and genetic diversity as a function of sex-bias and genomic region: hemizygous Z, diploid Z and autosomes. Differences in divergence and diversity patterns between different genomic regions were statistically compared within each category of sex-bias using the Wilcoxon-Mann-Whitney test. Unbiased genes generally follow the trends described above for all genes: both hemizygous Z and diploid Z genes have increased K_a/K_s [median(hZ)=0.1023, median(dZ)=0.0990, median(A)=0.0910] (though only significantly so in the case of diploid genes, p=0.025) and increased π_N/π_S compared with autosomal genes [median(hZ)=0.3371 , median(dZ)=0.2295 , median(A)=0.2008, hZ vs A: p=9.7e-7 and dZ vs A: p=0.027 ,**Table S4, Table S7**], consistent with reduced efficacy of selection on both parts of the Z. This is also supported by their reduced α values compared with autosomal values [median(hZ)=-0.5273, median(dZ)=-0.1390, median(A)=0.0280, **Figure S3B**, p=0.00038 and p=0.00781 for hemizygous and diploid Z genes].

In the hemizygous Z region, a key prediction is that genes that function primarily in females are expected to be under stronger efficacy of selection than equivalent autosomal genes, potentially leading to higher rates of adaptive divergence. The ratio of nonsynonymous to synonymous divergence (K_a/K_s) for female-biased genes on the hemizygous Z is significantly higher than for autosomal female-biased genes [median(hZ)=0.1498, median(dZ)=0.0839, median(A)=0.0827,

p = 0.03918, **Table S4**]. Female-biased genes on the hemizygous Z also have a higher median K_a/K_s than female-biased genes on the diploid Z (**Table S4.**), however, the difference is not significant (*p* = 0.1146), possibly due to low statistical power, as there are only 15 and 76 genes, respectively, in these categories. π_N/π_S values did not differ between female-biased genes on the hemizygous Z and on the autosomes [median(hZ)=0.2758, median(A)=0.2033, **Table S7**.], and α trended towards higher values for the former [median(hZ)=0.1807, median(A)=-0.0404] (though not significantly so), suggesting that positive selection acting on hemizygous mutations may contribute to the observed increase in protein coding divergence. Male-biased genes on the hemizygous Z also showed higher K_a/K_s than male-biased genes on autosomes [median(hZ)=0.1196 , median(A)=0.0868, **Table S4**, *p*= 8.587e-12] but this was in this case associated with elevated levels of π_N/π_S [median(hZ)=0.3145, median(A)=0.1597, **Table S7**, hZ vs A: *p* = 3.512e-14] and reduced values of α [median(hZ)=-0.5897, median(A)=0.2526, *p*=2.1e-15], consistent with a primary role of relaxed purifying selection.

On the diploid Z, the expectation is that female-biased genes should be under strongly reduced efficacy of selection. Neither female-biased nor male-biased genes on the diploid Z showed a significant difference in K_a/K_s when compared to their respective autosomal controls [median(dZ)=0.0839, median(dZ)=0.0883, for female- and male-biased genes respectively, **Table S4**, *p*= 0.8083, *p*= 0.239, respectively]. While π_N/π_S did not differ between diploid Z [median(dZ)=0.1597] and autosomal male-biased genes (**Table S7**), suggesting the two are under similar selective pressures, female-biased genes on the diploid Z had higher π_N/π_S [median(dZ)=0.3352, **Table S7**] and lower α [median(dZ)=-1.1602] than their autosomal counterparts (dZ vs A : *p* = 0.001672 and *p*=0.0078, **Figure 3, S. Figure**

3B). This is generally in line with our predictions that mutations can freely accumulate on female-biased genes in diploid Z, as they are sheltered from selection by the functional gemetolog on the W (Mrnjavac et al., 2023).

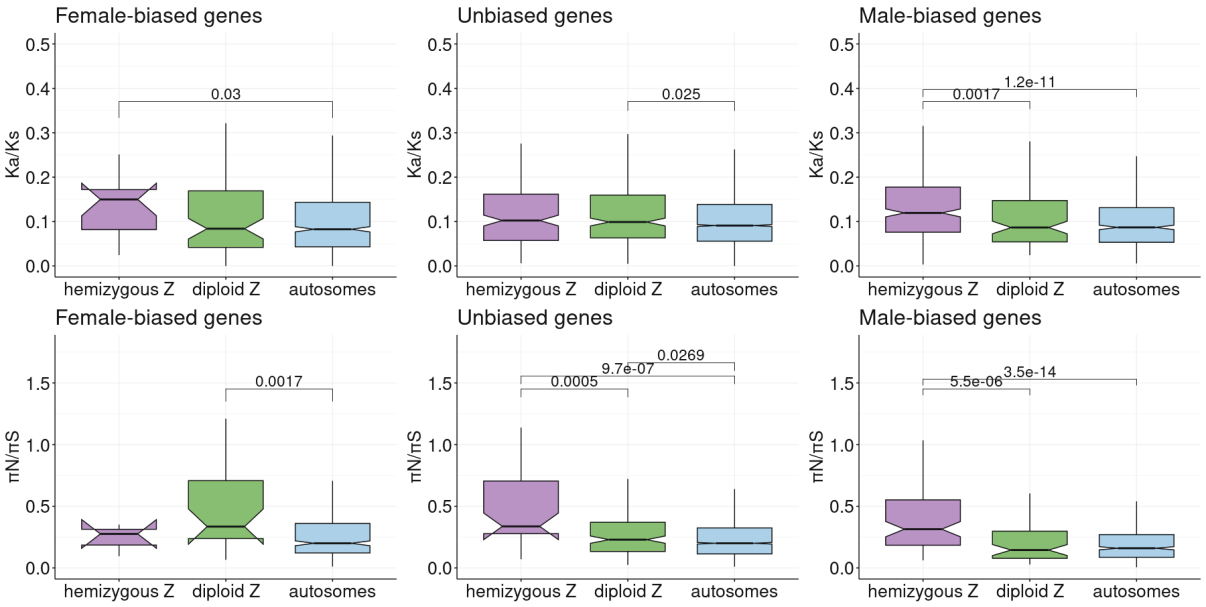


Figure 3. K_a/K_s and π_N/π_S for hemizygous and diploid Z and autosomes as a function of sex-bias.

Discussion

The sex chromosomes of *Schistosoma japonicum*, with their clearly distinguishable hemizygous Z and diploid Z regions, that is, strata with degenerated W and with non-degenerated W (Elkrewi et al., 2021, Xu et al., 2023), provide an ideal opportunity to study the evolutionary dynamics of young and old sex-linked regions in the same species. Our results suggest that the effective population size (N_e) of the Z chromosome in *Schistosoma japonicum* is much lower than the expected ¾ of the autosomal effective population size. Since neutral diversity levels are a function of N_e times the mutation rate μ, it is possible that the difference is

driven by differences in μ. Indeed, the lower synonymous divergence (Ks) and nucleotide diversity observed for the Z compared with the autosomes suggests that mutation rates may be lower on the Z. This implies female-biased mutation rates, as the Z chromosome spends less time in females. This is an unlikely explanation, as mutation rates are usually male-biased (Ellegren, 2006, de Manuel et al., 2022). To account for the potential difference in mutation rates, we also compared the distribution of π/Ks, which should control for the mutation rate. Values for the Z were still lower than ¾ of those of the autosomes (hZ/A=0.2584, dZ/A=0.4290, Figure S5), suggesting a true reduction in N_e. The effective population size of Z chromosomes is expected to typically be smaller than the effective population size of X chromosomes, because Z chromosomes spend most of the time in males, and males often have a larger variance of reproductive success, which decreases their effective population size (Caballero, 1995, Charlesworth, 2001, Laporte and Charlesworth, 2002, Vicoso and Charlesworth, 2009, Mank, Vicoso et al., 2010). While the variance in reproductive success of males and females of *S. japonicum* is not known, adult populations of adults are typically male-biased (Beltran and Boissier, 2010). Given the largely monogamous reproductive mode of schistosome parasites (Beltran and Boissier, 2008), this may lead to a substantial proportion of males remaining unpaired, thereby increasing the variance in their reproductive success.

Interestingly, the N_e of the hemizygous Z is lower than the N_e of the diploid Z. One possibility to explain this is that the hemizygous Z region has been non-recombining for a longer amount of time: if loss of recombination with the W occurred very recently, the diploid Z may still not have lost all the standing variation that it harbored when it was a pseudoautosomal region. This is however unlikely to

fully explain the pattern, as the reduction in N_e following a decrease in population size should occur fairly rapidly (as the long-term N_e is simply the harmonic mean of the population sizes over generations, Nei and Tajima, 1981, Kalinowski and Waples, 2002). The lower N_e observed on the Z chromosome could also be due to the stronger effect of linked selection. The effect of linked selection should be particularly strong on the hemizygous Z, due to recessive mutations being exposed to selection. While we did not detect evidence of stronger positive selection on the hemizygous Z than on the autosomes, a recent study did detect a few loci under strong selection (Zhou et al., 2024), which may have contributed to reducing its genetic diversity. Additionally, it is possible that the hemizygous Z region does not recombine even in males (or has very low recombination rates), which would further reduce diversity. While no linkage map is available for *S. japonicum*, the Z-specific region of its close relative *S. mansoni*, which is partly shared with *S. japonicum*, has normal levels of recombination in males (Criscione et al., 2009). It therefore seems likely that a combination of factors drives the strong reduction in N_e that we observe.

Consistent with this reduced effective population size, our results suggest that the evolution of the Z chromosome in *Schistosoma japonicum* is dominated by the effect of relaxed purifying selection. This is in line with the general pattern of faster rates of evolution on the Z chromosome, which are often caused by drift due to its smaller effective population size (Mank, Vicoso et al., 2010, Mank et al., 2010, Hayes et al, 2020, Chase et al., 2023, Mongue and Baird, 2024). Although both the hemizygous and diploid Z regions are under reduced efficacy of purifying selection, we could to some extent test the differential expectations of the “faster-Z” and “slower-Z” effects by focusing on sex-biased genes. The effect of drift is expected to be counteracted to some extent by strong haploid selection for female-biased genes

on hemizygous Z (Vicoso and Charlesworth, 2006). Consistent with this, female-biased genes located on the hemizygous Z region had a slightly increased K_a/K_s, but not π_N/π_S, when compared to autosomal female-biased genes. However, since only a handful of genes had sufficient polymorphism in our dataset, we could not obtain a significant signal of Faster-Z effect when using α inferences. On the other hand, in addition to reduced N_e, in the young diploid Z region, selective constraints should be relaxed due to the sheltering effect of functional gametologs on the W, and this effect should be stronger for genes expressed primarily in females. The effect of sheltering is supported by the fact that female-biased genes have the highest π_N/π_S, and the lowest inferred α, of the genes in the diploid Z region. Taken together, these results confirm that diploid and hemizygous sex-linked regions have different evolutionary dynamics, and that genes that function predominantly in one sex are primarily affected (assuming that sex-biased gene expression is a good proxy for sex-biased function).

Several theoretical models predict that Z-chromosomes may become “masculinized” over time, i.e. they may lose genes with female-specific functions and gain genes that work primarily in males (Gurbich and Bachtrog, 2008, Mrnjavac et al., 2023). An excess of Z-linked genes of *S. japonicum* are indeed male-biased in their expression. In the hemizygous Z region, masculinized expression can to a large extent be explained by the incomplete dosage compensation system found in this group: the Z chromosome is upregulated in both sexes, and has higher expression in males, since males have two copies of the Z (Picard et al., 2018). Whether an ancestral enrichment in genes with male-specific functions favored the evolution of such an unusual regulatory mechanism has yet to be tested. In the diploid Z region, expression patterns are more similar to autosomal ones, as we are capturing

expression from the W gametologs in females, but a significant bias towards higher male expression was observed. This could have two (non-mutually exclusive) explanations: 1. Genes on the W may have undergone some regulatory degeneration, leading to their lower expression; 2. Genes on the Z may have become masculinized, ie male-beneficial mutations may have favored their increased expression in males and/or decreased expression in females, as predicted by the Slower-Z hypothesis. A recent study found similar expression levels from the W and the Z in the diploid Z region (Elkrewi et al., 2021); however, there was very limited power as only a small subset of genes were sampled. Future work comparing the expression of the Z and the W over the whole region, as well as patterns of expression of these genes in species where they are not sex-linked, may shed light on which of these hypotheses is driving this shift.

Our study illustrates different evolutionary dynamics of old and young sex-linked regions. Together with other studies on young sex-linked regions in butterflies of genus Leptidea (Hook et al.,2023), plant *Silene latifolia* (Krasovec et al., 2018), and several Drosophila species (Nozawa et al., 2016, Nozawa et al., 2021), our study suggests that Slower-X (or Slower-Z) effect might be widespread in young sex-linked regions. This body of work also illustrates the importance of studying non-model species, where diploid Z and X regions might be common, but underreported, as well as using population data for studying ongoing evolutionary processes.

References

Abascal, F., R. Zardoya, and M. J. Telford. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research 38:W7–W13.

Al-Saffar, S. I., and M. W. Hahn. 2022. Evaluating methods for estimating the proportion of adaptive amino acid substitutions. bioRxiv.

Ávila, V., S. Marion de Procé, J. L. Campos, H. Borthwick, B. Charlesworth, and A. J. Betancourt. 2014. Faster-X Effects in two Drosophila lineages. Genome Biology and Evolution 6:2968–2982.

Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14:113–124.

Beltran, S., and J. Boissier. 2010. Male-biased sex ratio: why and what consequences for the genus *Schistosoma*? Trends in Parasitology 26:63–69.

Beltran, S., and J. Boissier. 2008. Schistosome monogamy: who, how, and why? Trends in Parasitology 24:386–391.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

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

Caballero, A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139:1007–1011.

Campos, J. L., D. L. Halligan, P. R. Haddrill, and B. Charlesworth. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. Molecular Biology and Evolution 31:1010–1028.

Charlesworth, B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. Genetics Research 77:153–166.

Charlesworth, B., J. L. Campos, and B. C. Jackson. 2018. Faster-X evolution: Theory and evidence from Drosophila. Molecular Ecology 27:3753–3771.

Charlesworth, B., and D. Charlesworth. 2010. Elements of Evolutionary Genetics. Roberts and Company.

Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. The American Naturalist 130:113–146.

Charlesworth, D. 2021. The timing of genetic degeneration of sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 376:20200093.

Chase, M. A., M. Vilcot, and C. F. Mugal. 2023. Evidence that genetic drift not adaptation drives fast-Z and large-Z effects in *Ficedula* flycatchers. Molecular Ecology n/a:e17262.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92.

Criscione, C. D., C. L. Valentim, H. Hirai, P. T. LoVerde, and T. J. Anderson. 2009. Genomic linkage map of the human blood fluke *Schistosoma mansoni*. Genome Biology 10:R71.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Darolti, I., P. Almeida, A. E. Wright, and J. E. Mank. 2022. Comparison of methodological approaches to the study of young sex chromosomes: A case study in *Poecilia*. Journal of Evolutionary Biology 35:1646–1658.

Darolti, I., L. J. M. Fong, B. A. Sandkam, D. C. H. Metzger, and J. E. Mank. 2023. Sex chromosome heteromorphism and the Fast-X effect in poeciliids. Molecular Ecology 32:4599–4609.

de Manuel, M., F. L. Wu, and M. Przeworski. 2022. A paternal bias in germline mutation is widespread in amniotes and can arise independently of cell division numbers. eLife 11:e80008.

Elkrewi, M., M. A. Moldovan, M. A. L. Picard, and B. Vicoso. 2021. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. Molecular Biology and Evolution 38:5345–5358.

Ellegren, H. 2006. Characteristics, causes and evolutionary consequences of male-biased mutation. Proceedings of the Royal Society B: Biological Sciences 274:1–10.

Fay, J. C., G. J. Wyckoff, and C.-I. Wu. 2001. Positive and negative selection on the human genome. Genetics 158:1227–1234.

Furman, B. L. S., D. C. H. Metzger, I. Darolti, A. E. Wright, B. A. Sandkam, P. Almeida, J. J. Shu, and J. E. Mank. 2020. Sex chromosome evolution: So many exceptions to the rules. Genome Biology and Evolution 12:750–763.

Garrigan, D., S. B. Kingan, A. J. Geneva, J. P. Vedanayagam, and D. C. Presgraves. 2014. Genome diversity and divergence in *Drosophila mauritiana*: Multiple signatures of Faster X evolution. Genome Biology and Evolution 6:2444–2458.

Gurbich, T. A., and D. Bachtrog. 2008. Gene content evolution on the X chromosome. Current Opinion in Genetics & Development 18:493–498.

Hayes, K., H. J. Barton, and K. Zeng. 2020. A study of Faster-Z evolution in the great tit (*Parus major*). Genome Biology and Evolution 12:210–222.

He, L., K.-H. Jia, R.-G. Zhang, Y. Wang, T.-L. Shi, Z.-C. Li, S.-W. Zeng, X.-J. Cai, N. D. Wagner, E. Hörandl, A. Muyle, K. Yang, D. Charlesworth, and J.-F. Mao. 2021. Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. Molecular Ecology Resources 21:1966–1982.

Höök, L., R. Vila, C. Wiklund, and N. Backström. 2024. Temporal dynamics of faster neo-Z evolution in butterflies. Evolution qpae082.

Jay, P., D. Jeffries, F. E. Hartmann, A. Véber, and T. Giraud. 2024. Why do sex chromosomes progressively lose recombination? Trends in Genetics 40:564–579.

Kalinowski, S. T., and R. S. Waples. 2002. Relationship of effective to census size in fluctuating populations. Conservation Biology 16:129–136.

Kent, W. J. 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Korunes, K. L., and K. Samuk. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Molecular Ecology Resources 21:1359–1368.

Kousathanas, A., D. L. Halligan, and P. D. Keightley. 2014. Faster-X adaptive protein evolution in house mice. Genetics 196:1131–1143.

Krasovec, M., B. Nevado, and D. A. Filatov. 2018. A comparison of selective pressures in plant X-linked and autosomal genes. Genes 9:234.

Kryuchkova-Mostacci, N., and M. Robinson-Rechavi. 2017. A benchmark of gene expression tissue-specificity metrics. Briefings in Bioinformatics 18:205–214.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

Laporte, V., and B. Charlesworth. 2002. Effective population size and population subdivision in demographically structured populations. Genetics 162:501–519.

Leroy, T., Y. Anselmetti, M.-K. Tilak, S. Bérard, L. Csukonyi, M. Gabrielli, C. Scornavacca, B. Milá, C. Thébaud, and B. Nabholz. 2021. A bird's white-eye view on avian sex chromosome evolution. Peer Community Journal 1.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Luo, F., W. Yang, M. Yin, X. Mo, Y. Pang, C. Sun, B. Zhu, W. Zhang, C. Yi, Z. Li, J. Wang, B. Xu, Z. Feng, Y. Huang, Y. Lu, and W. Hu. 2022. A chromosome-level genome of the human blood fluke *Schistosoma japonicum* identifies the genomic basis of host-switching. Cell Reports 39:110638.

Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.-M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, and J. Wang. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18.

Mank, J. E., K. Nam, and H. Ellegren. 2010a. Faster-Z evolution is predominantly due to genetic drift. Molecular Biology and Evolution 27:661–670.

Mank, J. E., B. Vicoso, S. Berlin, and B. Charlesworth. 2010b. Effective population size and the Faster-X effect: empirical results and their interpretation. Evolution 64:663–674.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

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

Mongue, A. J., and R. B. Baird. 2024. Genetic drift drives faster-Z evolution in the salmon louse *Lepeophtheirus salmonis*. Evolution qpae090.

Mongue, A. J., M. E. Hansen, and J. R. Walters. 2022. Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages*. Evolution 76:332–345.

Mrnjavac, A., K. A. Khudiakova, N. H. Barton, and B. Vicoso. 2023. Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. Evolution Letters 7:4–12.

Nei M, Tajima F (1981). Genetic drift and estimation of effective population size. Genetics 98: 625–640

Nozawa, M., Y. Minakuchi, K. Satomura, S. Kondo, A. Toyoda, and K. Tamura. 2021. Shared evolutionary trajectories of three independent neo-sex chromosomes in Drosophila. Genome Res. 31:2069–2079.

Nozawa, M., K. Onizuka, M. Fujimi, K. Ikeo, and T. Gojobori. 2016. Accelerated pseudogenization on the neo-X chromosome in *Drosophila miranda*. Nat Commun 7:13659.

Picard, M. A. L., C. Cosseau, S. Ferré, T. Quack, C. G. Grevelding, Y. Couté, and B. Vicoso. 2018. Evolution of gene dosage on the Z-chromosome of schistosome parasites. eLife 7:e35684.

Pimentel, H., N. L. Bray, S. Puente, P. Melsted, and L. Pachter. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14:687–690.

Protasio, A. V., I. J. Tsai, A. Babbage, S. Nichol, M. Hunt, M. A. Aslett, N. D. Silva, G. S. Velarde, T. J. C. Anderson, R. C. Clark, C. Davidson, G. P. Dillon, N. E. Holroyd, P. T. LoVerde, C. Lloyd, J. McQuillan, G. Oliveira, T. D. Otto, S. J. Parker-Manuel, M. A. Quail, R. A. Wilson, A. Zerlotini, D. W. Dunne, and M. Berriman. 2012. A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. PLOS Neglected Tropical Diseases 6:e1455.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org/*

Smith, N. G. C., and A. Eyre-Walker. 2002. Adaptive protein evolution in Drosophila. Nature 415:1022–1024.

Stoletzki, N., and A. Eyre-Walker. 2011. Estimation of the Neutrality Index. Molecular Biology and Evolution 28:63–70.

Veeramah, K. R., R. N. Gutenkunst, A. E. Woerner, J. C. Watkins, and M. F. Hammer. 2014. Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. Molecular Biology and Evolution 31:2267–2282.

Vicoso, B., and B. Charlesworth. 2009. Effective population size and the Faster-X effect: an extended model. Evolution 63:2413–2426.

Vicoso, B., and B. Charlesworth. 2006. Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet 7:645–653.

Villavicencio, M. L., J. Ledamoisel, C. Lopez-Roques, V. Debat, and V. Llaurens. 2024. Increased evolutionary rate in the Z-chromosome of *Morpho* butterflies and implications for speciation. bioRxiv.

Wanders, K., G. Chen, S. Feng, T. Székely, and A. O. Urrutia. 2024. Role-reversed polyandry is associated with faster fast-Z in shorebirds. Proceedings of the Royal Society B: Biological Sciences 291:20240397.

Wang, D., Y. Zhang, Z. Zhang, J. Zhu, and J. Yu. 2010. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. Genomics, Proteomics & Bioinformatics 8:77–80.

Wang, J., Y. Yu, H. Shen, T. Qing, Y. Zheng, Q. Li, X. Mo, S. Wang, N. Li, R. Chai, B. Xu, M. Liu, P. J. Brindley, D. P. McManus, Z. Feng, L. Shi, and W. Hu. 2017. Dynamic transcriptomes identify biogenic amines and insect-like hormonal regulation for mediating reproduction in *Schistosoma japonicum*. Nat Commun 8:14693.

Xu, X., Y. Wang, C. Wang, G. Guo, X. Yu, Y. Dai, Y. Liu, G. Wei, X. He, G. Jin, Z. Zhang, Q. Guan, A. Pain, S. Wang, W. Zhang, N. D. Young, R. B. Gasser, D. P. McManus, J. Cao, Q. Zhou, and Q. Zhang. 2023. Chromosome-level genome assembly defines female-biased genes associated with sex determination and differentiation in the human blood fluke *Schistosoma japonicum*. Molecular Ecology Resources 23:205–221.

Zhou, A., W. Zhang, X. Ge, Q. Liu, F. Luo, S. Xu, W. Hu, and Y. Lu. 2024. Characterizing genetic variation on the Z chromosome in *Schistosoma japonicum* reveals host-parasite co-evolution. Parasites & Vectors 17:207.

Supplementary tables

Supplementary Table 1. Median values of synonymous (K_s) and nonsynonymous divergence (K_a), their ratio (K_a/K_s), synonymous (π_S) and nonsynonymous diversity (π_N) and their ratio (π_N/π_S) for hemizygous and diploid Z and autosomes.

	Hemizygous Z	Diploid Z	Autosomes
K_s	0.907081	0.972942	1.12092
K_a	0.107458	0.101787	0.109628
K_a/K_s	0.11521	0.0966062	0.0900891
π_S	0.005193584	0.008731108	0.02209353
π_N	0.001556166	0.00216915	0.003773615
π_N/π_S	0.3242248	0.2295004	0.1941696

Supplementary Table 2. Median values of synonymous divergence (K_s) for different categories of sex-bias and different genomic regions.

	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	1.02167	0.963872	1.090025
Unbiased genes	0.931826	0.943244	1.07825
Male-biased genes	0.878913	1.14931	1.40383

Supplementary Table 3. Median values of nonsynonymous divergence (K_a) for different categories of sex-bias and different genomic regions.

	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	0.154339	0.102965	0.0968863
Unbiased genes	0.0964118	0.101152	0.1064175
Male-biased genes	0.111167	0.103812	0.128502

Supplementary Table 4. Median values of nonsynonymous divergence normalised with synonymous divergence (K_a/K_s) for different categories of sex-bias and different genomic regions. Number of genes is given in the parenthesis.

	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	0.149801 (15)	0.0839165 (76)	0.0827174 (722)
Unbiased genes	0.102285 (185)	0.0989523 (303)	0.0909841 (3735)
Male-biased genes	0.119556 (327)	0.0882664 (111)	0.0868252 (519)

Supplementary Table 5. Median values of synonymous diversity (π_S).

	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	0.0191522	0.008940625	0.02171105
Unbiased genes	0.006934236	0.007551501	0.01882615
Male-biased genes	0.00497688	0.01254916	0.03820391

Supplementary Table 6. Median values of nonsynonymous diversity (π_N).

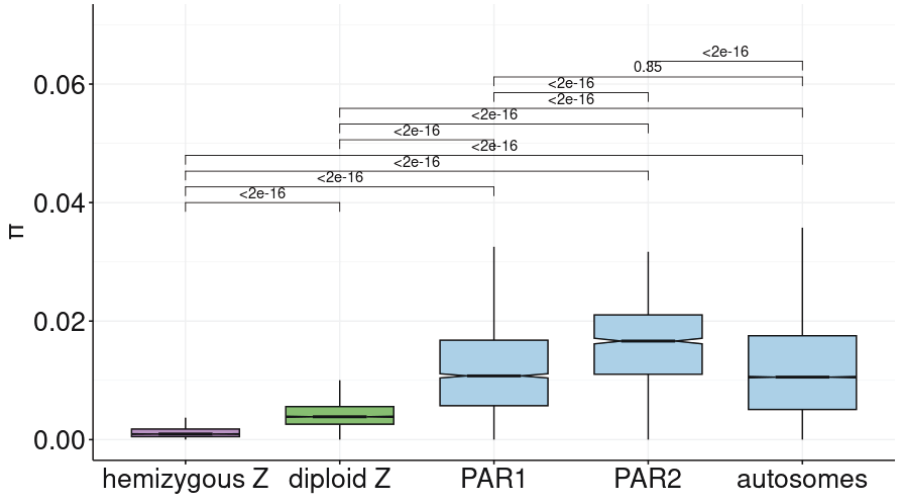
	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	0.003566143	0.003865257	0.004206034
Unbiased genes	0.002192836	0.00197378	0.003393581
Male-biased genes	0.001442663	0.002122322	0.005126339

Supplementary Table 7. Nonsynonymous diversity normalised with synonymous diversity (π_N/π_S). Number of genes is in the parentheses.

	Hemizygous Z	Diploid Z	Autosomes
Female-biased genes	0.2757828 (3)	0.3352456 (27)	0.2032735 (376)
Unbiased genes	0.3371406 (39)	0.2295004 (161)	0.2007695 (2784)
Male-biased genes	0.3144838 (88)	0.1450884 (61)	0.1596923 (531)

Supplementary figures

A)



B)

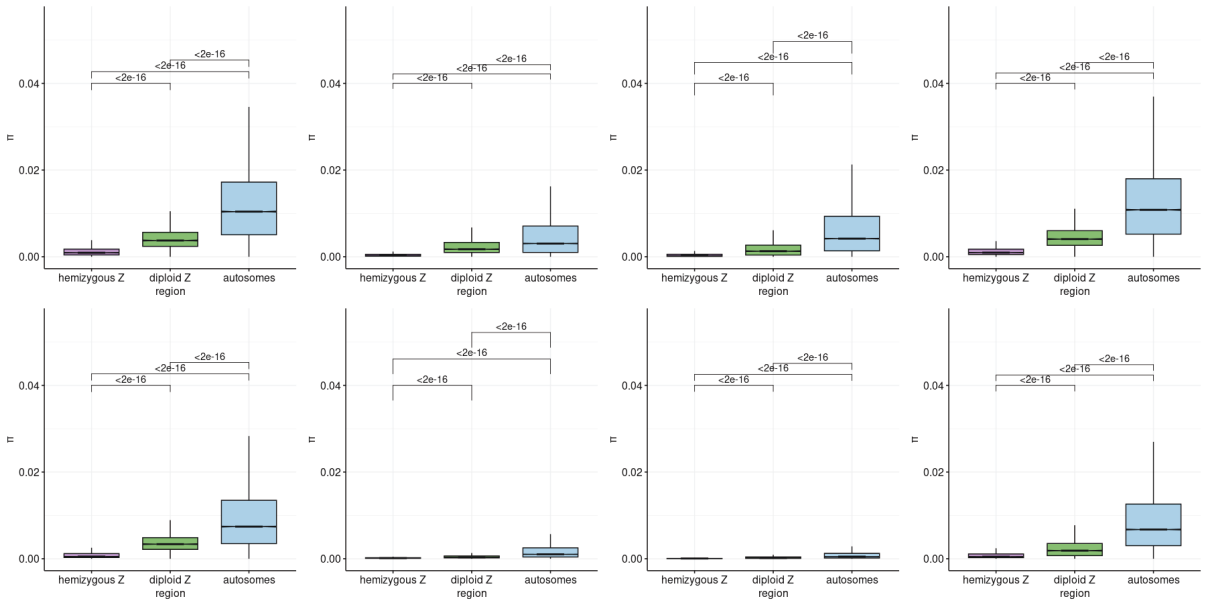
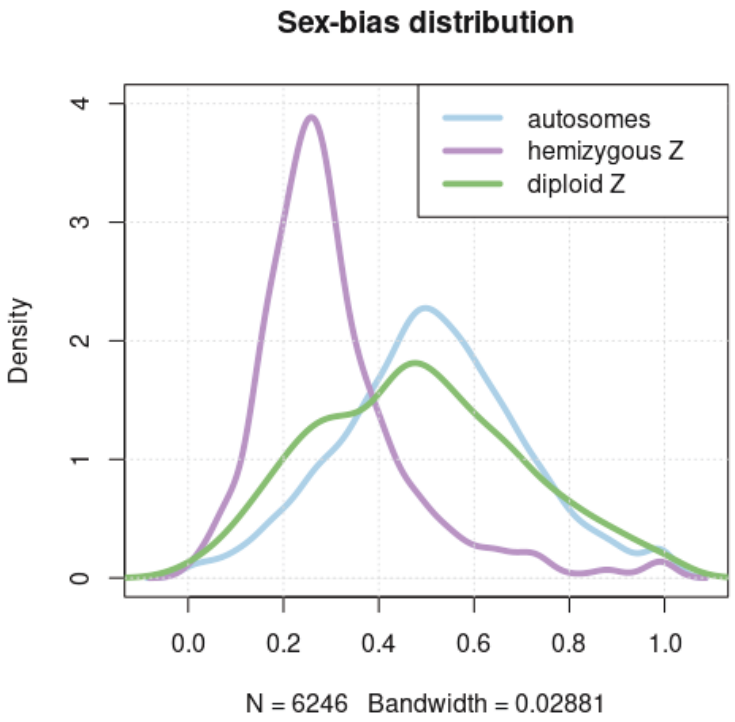
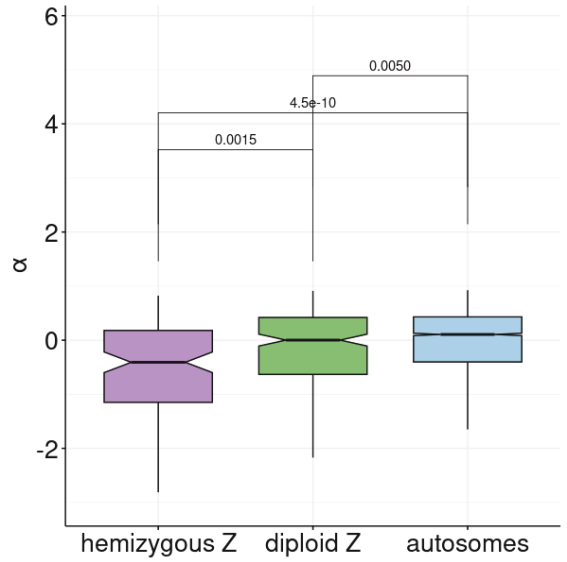


Figure S1. A) Nucleotide diversity (π) of hemizygous Z, diploid Z and pseudoautosomal regions and autosomes. B) nucleotide diversity in subpopulations.



S. Figure 2. Sex-bias distribution in autosomes, diploid Z region and hemizygous Z region.

A)



B)

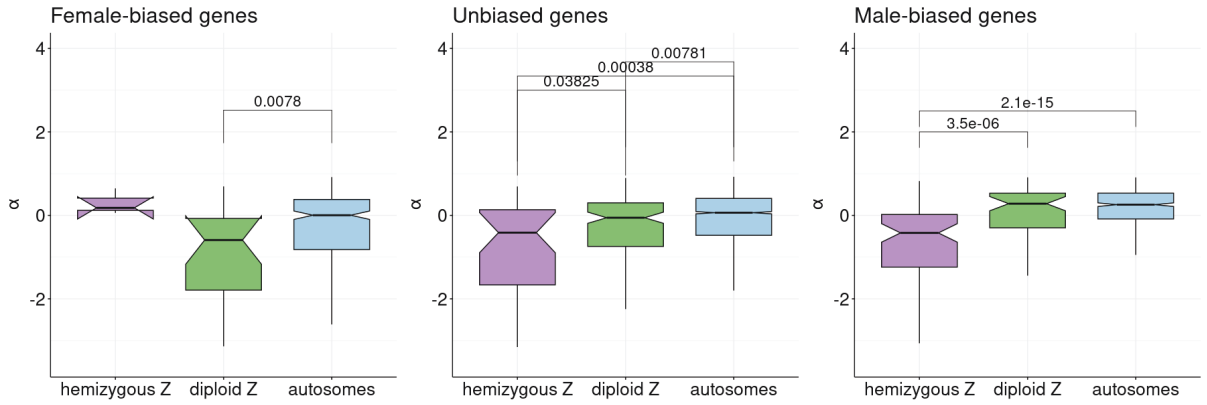
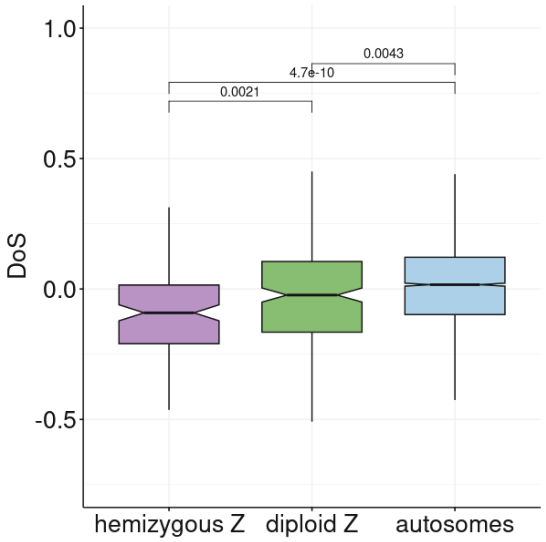


Figure S3. A α as a function of genomic location **B** α as a function of sex-bias and a genomic location.

A)



B)

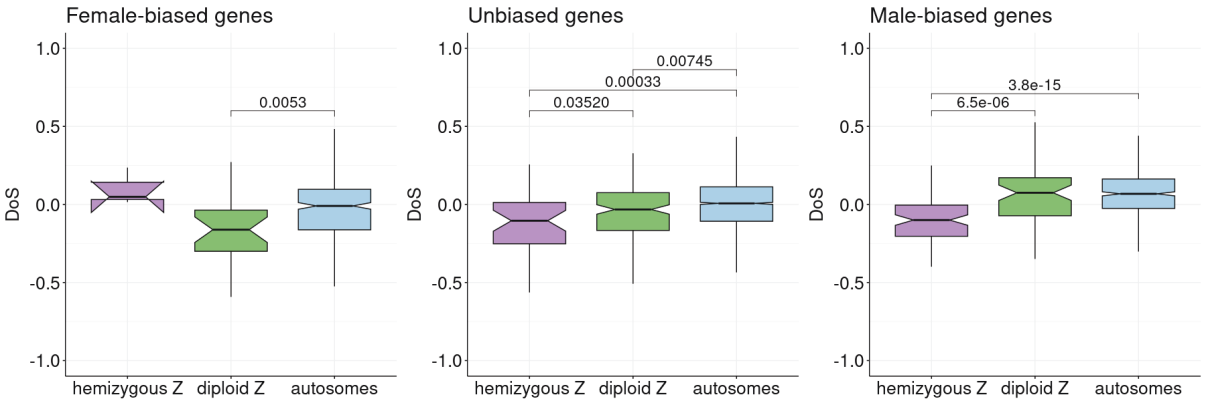


Figure S4. **A**) Direction of Selection (DoS) as a function of genomic location **B**) Direction of Selection (DoS) as a function of sex-bias and genomic location.

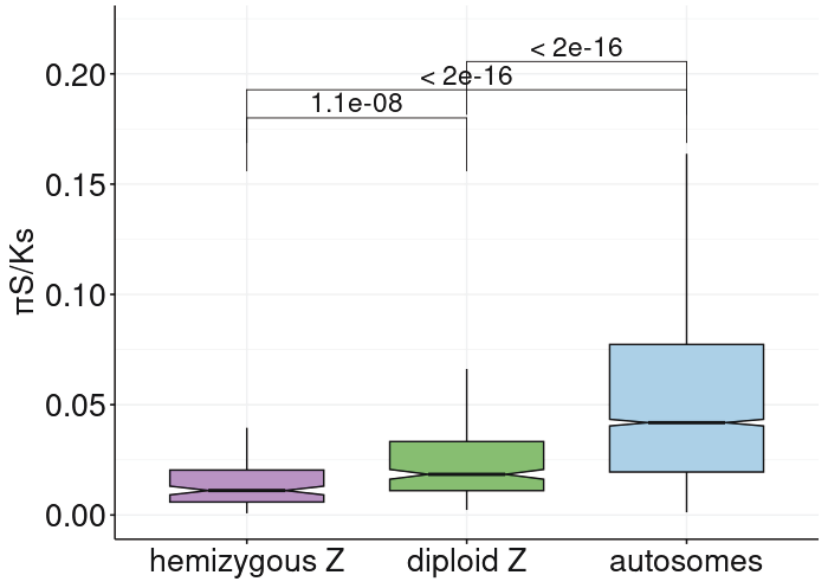


Figure S5. πs/Ks for hemizygous Z, diploid Z and autosomes

Chapter 5: Discussion

Sex chromosomes: Beyond the canonical model

The genomic revolution has brought into question many assumptions that were held about sex chromosome evolution. Before many non-model species were sequenced, most of our knowledge on sex chromosomes came from a few model species, like mammals, birds and Drosophila. These model species all have well differentiated sex chromosomes, with highly degenerated Y/Ws, consistent with the expected outcome of the canonical model of sex chromosome evolution described in Chapter 1. The Tree of Sex database (Tree of Sex Consortium, 2014), a joint effort between scientists across the world to compile the existing literature on sex determination systems and sex chromosomes in animals and plants, provided an insight into the extensive diversity of sex chromosomes across the tree of life. The sequencing of large numbers of animal and plant genomes, combined with the development of bioinformatics approaches to detect sex-linked sequences, also highlighted many examples of sex chromosomes that did not seem to fit the canonical model. Many recent reviews on sex chromosomes evolution emphasise “deconstructing myths” (Bachtrog et al., 2014), “so many exceptions to the rules” (Furman et al., 2020) and going “beyond the canonical model” (Zhu et al., 2024). This empirical diversity motivated the development of new theories and models that can help explain this variety of outcomes.

The prevailing hypothesis for the evolution of recombination suppression between sex chromosomes used to be that selection favours linkage between the sex-determining locus and sexually antagonistic (SA) loci in order to resolve sexual conflict (Rice, 1987). However, the lack of empirical support for SA hypothesis (e.g.

Dagilis et al., 2022, reviewed in Jay et al., 2024) sparked the development of many alternative models for the evolution of recombination suppression, and this field remains a hot topic until today. Ubeda et al. (2015) proposed a model in which recombination suppression is driven by selection for the linkage between an SD locus and a meiotic driver. Since meiotic drivers often carry a homozygous disadvantage, tight linkage with the sex-determining (SD) locus ensures that the driver is always heterozygous. Jeffries et al. (2021) proposed a neutral model for the evolution of recombination suppression, where permanently heterozygous SD locus initiates a positive feedback loop: the initial sequence divergence lowers the probability of recombination in the surrounding region, and low recombination rate leads to sequence divergence, which in turn leads to the expansion of the non-recombining region. Jay et al. (2022) proposed a sheltering model of recombination suppression, based on the idea that inversions that happen to carry fewer deleterious mutations than the average haplotype in the population initially have a selective advantage, but that this advantage disappears when they become frequent enough that homozygous individuals express its recessive deleterious effects. If such an inversion captures a permanently heterozygous locus, such as the SD locus, it increases in frequency and ultimately fixes in the population, because recessive deleterious mutations in the inversion remain sheltered from selection, since such mutations are always in the heterozygous form. However, this model was disputed by Olito and Charlesworth (2023), who suggested that part of the apparent selective advantage may have resulted from the exclusion of Y-linked inversions that were lost in the first 20 generations from their simulations. They also pointed out that the effect of new deleterious mutations accumulating on the spreading inversions may have been underestimated. Lenormand and Roze (2022) proposed a

“regulatory” model which explains simultaneously the evolution of recombination suppression, Y degeneration and the evolution of dosage compensation. After a “lucky” (with fewer deleterious mutations than average) inversion capturing the SD locus is fixed in the population, cis and trans regulatory elements on the X and Y chromosome start evolving independently. Deleterious mutations can accumulate on cis regulatory elements on the Y chromosome, because the levels of total gene expression in each sex are maintained by sex-specific trans regulatory elements. As cis regulators become weaker, expression is reduced, which reduces the fitness effect (we can think about it as reduction in dominance) of accumulating deleterious mutations in the coding region, in turn, further silencing is favoured (Lenormand et al., 2020). This divergence between cis and trans regulators on the X and Y prevents the reestablishment of recombination, as mismatch between regulatory elements would result in fitness decrease. Unlike traditional models of Y chromosome degeneration by selective interference, which struggle to explain observed degeneration rates and patterns in nature (Charlesworth, 2021), this model works faster and in larger populations, and can explain degeneration in very small non-recombining regions. Recent reviews like Ponnikas et al. (2018), Jay et al. (2024), and Zhu et al. (2024) summarise recent alternative models, and suggest some predictions that can be tested to disentangle the contribution of these different models. For instance, Lenormand and Roze (2022) model predicts that downregulation and dosage compensation precede the Y chromosome degeneration.

The fact that once recombination is suppressed Y or W linked genes start to degenerate is well documented (although the specific mechanisms are yet to be fully understood, see section “**Both sex chromosomes can degenerate by**

sheltering"). This loss of Y-linked gene content should cause a deficit of gene products in males (or females in the case of ZW), and select for mechanisms of dosage compensation. The fact that complex mechanisms to achieve dosage balance exist in mammals, Drosophila and nematodes suggested that dosage compensation was a crucial step in sex chromosome evolution. Today we also know that global dosage compensation does not always evolve (reviewed in Gu and Walters, Furman et al., 2020) or can be specific to some cell-types, developmental stages or environmental conditions (reviewed in Zhu et al., 2024). Dosage sensitivity of the majority of genes, which would be a driver for the evolution of dosage compensation, has also been brought into question (Gu and Walters, 2017, Furman, 2020). It remains to be understood why such a variety of solutions seem to have been found to the single problem of dosage imbalance after Y/W degeneration, and what parameters have shaped these outcomes. Finally, the evolution of sex chromosomes is not linear, with evolutionarily stable differentiated sex chromosomes such as in mammals. Instead, sex chromosomes are more or less ephemeral, with common turnovers between sex chromosomes or sex chromosome systems (reviewed in Vicoso 2019, Furman et al., 2020 and Zhu et al., 2024). Again, it remains to be seen why some sex chromosomes become differentiated enough that turnover is impossible (or impossibly rare), while other clades undergo such frequent turnover that sex chromosome differentiation is never achieved.

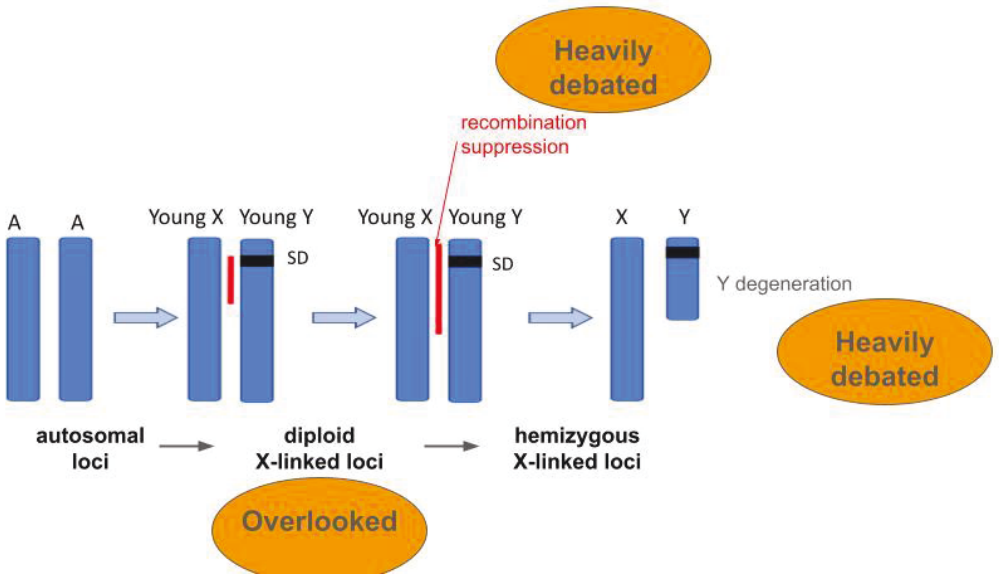


Figure 1. Hot and overlooked topics in the field of sex chromosome evolution.

Studying young sex chromosomes and young sex-linked regions can help us understand formative processes of sex chromosome evolution, and give us the opportunity to test some of the hypotheses described above. However, studying young sex chromosomes and evolutionary patterns in the early stages of sex chromosome evolution was difficult before the population genomic revolution, when many population genomic datasets on non-model species became available. Despite the growing amount of data on young sex chromosomes and young sex-linked regions, the evolutionary dynamics of the early stages of sex chromosome evolution had been largely theoretically and empirically unexplored. The theoretical work presented in this thesis provides a foundation for studying data from young sex chromosome systems, and makes several key predictions that can then be tested using populations and comparative genomics, and gene expression surveys.

Reduced efficiency of selection in the early stages of X chromosome evolution

In discussion of the evolution of sex-linked regions, non-hemizygous X chromosomes were typically thought to behave like autosomes, with unusual evolutionary features thought to arise after the degeneration of the Y. Nozawa et al. (2016) observed relaxed selective constraints on the young neo-X chromosome of *Drosophila miranda*, on which many genes still have a functional Y homolog. They put forward verbal arguments on distinct evolutionary pressures affecting a young X chromosome, but direct evidence of this and rigorous theoretical predictions were missing. Chapter 2 of my thesis provides a theoretical framework for the evolutionary dynamics of young X-linked loci. It shows that, unlike hemizygous X-linked loci which are under increased efficiency of selection (if mutations are on average partly or fully recessive), diploid X-linked loci experience reduced efficiency of selection because they are sheltered from selection by the functional gametolog on the Y chromosome. Our theory represents a shift in how we think about young X chromosomes, which are expected to evolve differently from both differentiated X chromosomes and autosomes. It explains the observed patterns of relaxed selective constraints in some young X chromosomes and provides a theoretical framework for future evolutionary studies on young sex chromosomes. Our results also have broader implications for the role of sex chromosomes in evolutionary processes. For instance, sex chromosomes might have an even greater role in speciation than previously thought, as even very young X chromosomes could contribute disproportionately to speciation between species, due to faster non-adaptive evolution. Our results emphasise the role of sex-specific selection in the evolution of

sex chromosomes. Finally, our results show that the X chromosome also has the potential to degenerate, in addition to the Y chromosome.

Since this theoretical work was motivated by molecular evolution data in Drosophila, it was *a priori* somewhat supported empirically. However, it remains key to establish whether such mal-adaptive evolution is a general feature of young X/Z chromosomes, or shaped instead by specific genomic or life history parameters of Drosophila. Chapter 4 of my thesis provides direct empirical evidence of relaxed selective constraints in young Z-linked region in *Schistosoma japonicum*, and contrasting evolutionary patterns in old and young sex-linked regions. Taken together with other studies on young sex chromosomes, our results suggest that reduced efficiency of selection is indeed widespread in young X (Z) chromosomes, and support the biological relevance of the proposed theory in Chapter 2. Our study also emphasised the importance of analysing diploid and hemizygous sex-linked regions separately, as well as using population data to explore ongoing evolutionary pressures. Finally, we analysed a non-model species using publicly available data, which illustrates how publicly available genomic datasets of non-model species allow meaningful insights.

Both sex chromosomes can degenerate by sheltering

Chapter 3 of my thesis provides a theoretical framework for X and Y chromosome degeneration by sheltering. As we saw in the previous section, sheltering of the X is the only model that explains X chromosome degeneration. But how does this play out when both the X and the Y are considered, since they are initially able to shelter each other? Sheltering was actually the first mechanism put forward to account for Y

chromosome degeneration, a hypothesis put forward by Muller over 100 years ago. However, when only mutations which affect both males and females equally are considered, this model has only limited power to explain gene loss on the Y. Our model, which extends previous work by varying the sex-specificity of deleterious effects, not only suggests that sheltering can contribute to the degeneration of sex chromosomes, but also their sexualized gene content, which results from sex-specific gene loss. This mechanism could also contribute to speciation, since X and Y chromosomes coevolve in a way that genes which are lost from the X are retained on the Y and the other way around. As in small populations it is random to some extent which genes are lost from the X and which from the Y, different genes could be lost from X (or Y) in diverged populations, and a mismatch between X and Y could contribute to hybrid incompatibilities. While other models of degeneration have been struggling to explain the very fast degeneration of the Y chromosome, in our model, gene loss can occur in less than 200000 generations. Therefore, our model could play a crucial role in explaining especially the early stages of Y chromosome degeneration. Finally, since patterns of degeneration in our model depend on the sex-bias of the initial gene content, mutation rate and the effective population size, our model greatly contributes to explaining the observed variability of sex chromosome degeneration rates and patterns observed in nature (Charlesworth, 2021). For example, it may explain why we observe X chromosome degeneration in Drosophila, and not in mammals. The key prediction of this model is that the unusual sex-biased gene content of the X and Y chromosomes, which are often seen in differentiation XY and ZW pairs, should in fact emerge very early in their history, and that it should occur through loss of function of genes with sex-specific effects. As mentioned before, different patterns are also expected for

species with various effective population sizes and mutation rates. Given the abundance of available genomic and sex-specific RNA-seq datasets, such predictions will in the future be possible to test.

Future directions

My PhD provides theoretical and empirical evaluation of the early stages of sex chromosome evolution, and it opens the door towards many future directions. Due to the emergence of data on young sex chromosomes, evolutionary dynamics of sex chromosomes in different evolutionary stages might be empirically explored in many other taxa to show if reduced efficiency of selection in the early stages of X chromosome evolution is a universal pattern. Although we provide a theoretical model for sexually dimorphic gene loss from X and Y chromosomes, future studies are needed to empirically test the relevance of our theory, i.e., if dimorphic gene loss is driven by sheltering. This could be done by exploring gene loss patterns in the early stages of sex chromosome differentiation, and by comparisons to the closely related species where recombination suppression did not occur in the corresponding region.

Also, future empirical studies are needed to test the contribution of young sex chromosomes to speciation and hybrid incompatibilities. Experimental approaches using subpopulation crosses or comparing differentiation between subpopulations could be used for these studies.

References

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, and T. T. of S. Consortium. 2014. Sex determination: why so many ways of doing it? PLOS Biology 12:e1001899.

Charlesworth, D. 2021. The timing of genetic degeneration of sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 376:20200093.

Dagilis, A. J., J. M. Sardell, M. P. Josephson, Y. Su, M. Kirkpatrick, and C. L. Peichel. 2022. Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 377:20210205.

Furman, B. L. S., D. C. H. Metzger, I. Darolti, A. E. Wright, B. A. Sandkam, P. Almeida, J. J. Shu, and J. E. Mank. 2020. Sex chromosome evolution: So many exceptions to the rules. Genome Biology and Evolution 12:750–763.

Gu, L., and J. R. Walters. 2017. Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details. Genome Biology and Evolution 9:2461–2476.

Jay, P., D. Jeffries, F. E. Hartmann, A. Véber, and T. Giraud. 2024. Why do sex chromosomes progressively lose recombination? Trends in Genetics 40:564–579.

Jay, P., E. Tezenas, A. Véber, and T. Giraud. 2022. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLOS Biology 20:e3001698.

Jeffries, D. L., J. F. Gerchen, M. Scharmann, and J. R. Pannell. 2021. A neutral model for the loss of recombination on sex chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences 376:20200096.

Lenormand, T., F. Fyon, E. Sun, and D. Roze. 2020. Sex chromosome degeneration by regulatory evolution. Current Biology 30:3001-3006.e5.

Lenormand, T., and D. Roze. 2022. Y recombination arrest and degeneration in the absence of sexual dimorphism. Science 375:663–666.

Nozawa, M., K. Onizuka, M. Fujimi, K. Ikeo, and T. Gojobori. 2016. Accelerated pseudogenization on the neo-X chromosome in *Drosophila miranda*. Nat Commun 7:13659.

Olito, C., and B. Charlesworth. 2023. Do deleterious mutations promote the evolution of recombination suppression between X and Y chromosomes? bioRxiv.

Ponnikas, S., H. Sigeman, J. K. Abbott, and B. Hansson. 2018. Why do sex chromosomes stop recombining? Trends in Genetics 34:492–503.

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 41:911–914.

Tree of Sex Consortium. 2014. Tree of Sex: a database of sexual systems. Sci Data 1:140015.

Vicoso, B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol 3:1632–1641.

Zhu, Z., L. Younas, and Q. Zhou. 2024. Evolution and regulation of animal sex chromosomes. Nat Rev Genet 1–16.