Conflict and competition in the evolution of sexual traits

Peter Dunstan Price

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
School of Biosciences

Submitted October 2024

Abstract

The living world is dominated by competition and conflict. From genes to species, units of life prioritise their reproduction over those around them. This is notably evident in the study of sexual characteristics, where conflict may occur between the sexes over traits with shared genetics, within the sexes over access to mates, and across the genome when genes preferentially force their inheritance during meiosis. This thesis investigates and discusses these varying dimensions of conflict and their roles in shaping genomic and transcriptomic diversity. How we detect selection on these traits is complex for several reasons; sex-related traits are often located on the sex chromosomes that exhibit different evolutionary environments to the autosomes, and complex traits are often housed in inversions that physically link large portions of a chromosome. As such, transcriptomics is a powerful tool in understanding the evolution and consequences of conflict. Accordingly, Chapters 2 and 3 discuss advances in our understanding of intralocus sexual conflict and the evolution of the transcriptome respectively, whilst considering the nuances of sequencing methods and models for detecting regulatory evolution. Considering these findings I then investigate the patterns of genomic and transcriptomic evolution in the zebra finch and stalk-eyed fly, models for the study of sexual selection, genomic inversions, and sperm competition. In Chapter 4, I demonstrate that within-male competition via sperm competition can maintain large structural variation in the genome, a product of associative overdominance of an inversion polymorphism. Chapter 5 then characterises the single-cell transcriptomic landscape of the stalk-eyed fly and elucidates the consequences of meiotic drive on the transcriptome, whilst providing candidate mechanisms. Together, my work highlights the diverse outcomes of conflict in the genome and transcriptome, and provides important considerations for understanding the evolution of gene expression.

Acknowledgements

I first want to thank Alison for mentoring me throughout the PhD experience. Thank you also for your enthusiasm for me to take opportunities outside of my thesis including teaching, sharing my work, and the opportunity to travel to and work in Malaysia, which was literally the experience of a lifetime (I'd recommend you go next time, I'm sure we'll always need more flies). Your academic and personal guidance and patience have also been invaluable, especially in this last year, which I managed to introduce some extra chaos into. I will be forever grateful and am looking forward to working and collaborating in the future! Secondly, I want to thank current and previous members of the lab group, including Thea Rogers, Daniela Palmer, Sylvie Parkus, Ben Alston, and Vicky Lloyd for the encouragement and feedback on my work throughout all its various forms. I would also like to thank my co-supervisor Jon Slate and his group for their motivation and support. Thank you also to Pom and Sasha for the most brilliant month in the forest. And thank you Joel and Adam for your unrelenting enthusiasm for my work and the flies I spent many days and nights with. Lastly, thank you Georgie who without I don't think I would have managed to get this far. It's been a wild final year but thanks to your mental and physical care, I think we've just about made it through. Maybe once this is out of the way we can finally write a paper together?

A Rhinoceros hornbill in the Bukit Tinggi, Malaysia, taken through my binoculars in 2023.

Declaration

I, Peter Price, confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university. Chapters 2 and 3, and Appendix A and D are published works, Chapter 4 is in review and Chapter 5 is in preparation. The author's contributions are detailed below.

Chapter 2 has been published as:
Price PD, Parkus SM & Wright AE (2023) Recent progress in understanding the genomic architecture of sexual conflict. *Current Opinion in Genetics & Development* 80: 102047

This published manuscript is included in its published form with no alterations as permitted by Andrea Oguing, Researcher Support Agent. Author contributions are as follows: PDP and AEW designed the review, all authors wrote the manuscript, and PDP created the figure.

Chapter 3 has been published as:
Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR & Wright AE (2022) Detecting signatures of selection on gene expression. *Nature Ecology and Evolution* 6:1035-1045

This published manuscript is included in its published form with no alterations as permitted by permissions assistant André Buller. Author contributions are as follows: PDP, CRC, JEM, DHPD, PDP and AEW designed the study. DWK and PDP analysed the data with input from AEW, ESP and CRC. Specifically, PDP designed and analysed the simulations. All authors wrote and edited the manuscript.

Chapter 4 is currently in review as:
Price PD*, Pepper J*, Rogers TF, Wright AE & Slate J (in review) Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches. *joint contribution

Author contributions are as follows: PDP, JP, JS & AEW designed the study. TFR collected the RNA-seq data. PDP and JP analysed the data. Specifically, PDP performed the gene

expression analyses, JP generated VCF files for the DNA sequencing data, identified karyotyping SNPs, and designed the initial population genetics analysis. PDP reanalysed the genotyping data, cleaning variants and addressing biases resulting from unaccounted ploidy differences on the Z. Additionally, PDP redesigned the approach to independently analyse different site types including 4-fold and 0-fold degenerate sites, and non-coding neutral sites. All authors wrote and edited the manuscript.

Chapter 5 is being prepared for submission for review as:
Price PD, Parkus SM, LLoyd VJ, Bradshaw SL, Bates S, Paterson S, Burke T, Darolti I, Pomiankowski A & Wright AE The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies.

Author contributions are as follows: PDP and AEW designed the study. SLB, SB and AP collected the samples. PDP conducted the laboratory work, including the dissections and single-cell dissociations. SP conducted the sequencing to generate DNA-seq and scRNA-seq data. SMP, VJL and PDP analysed the data with input from AEW and ID. Specifically, SMP generated the mitochondrial genome and VJL conducted SNP calling on the DNA-seq data to generate inbreeding values. PDP analysed the single-cell RNA-seq data. PDP and AEW wrote the manuscript with input from all authors.

Appendix A has been published as:
Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR & Wright AE (2022) Reply to Fraser: Existing methods are effective at measuring natural selection on gene expression. *Nature Ecology and Evolution* 6:1838-1839

This is a Matters Arising that PDP published during his PhD in response to another Matters Arising: Fraser HB (2022) Existing methods are effective at measuring natural selection on gene expression. *Nature Ecology and Evolution* 6, 1836–1837. This article was written in response to Price et al (2022) *Nature Ecology and Evolution* (Chapter 3).

This published manuscript is included in its published form with no alterations as permitted by permissions assistant André Buller. Author contributions are as follows: PDP and AEW designed and wrote the response with input from all authors.

Appendix D has been published as:

Smith J, Alfieri JM, Anthony N, Arensburger P, Athrey GN, Balacco J, Balic A, Bardou P, Barela P, Bigot Y, Blackmon H, Borodin PM, Carroll R, Casono MC, Charles M, Cheng H, Chiodi M, Cigan L, Coghill LM, Crooijmans R, Das N, Davey S, Davidian A, Degalez F, Dekkers JM, Derks M, Diack AB, Djikeng A, Drechsler Y, Dyomin A, Fedrigo O, Fiddaman SR, Formenti G, Frantz LAF, Fulton JE, Gaginskaya E, Galkina S, Gallardo RA, Geibel J, Gheyas A, Godinez CJP, Goodell A, Graves JAM, Griffin DK, Haase B, Han J-L, Hanotte O, Henderson LJ, Hou Z-C, Howe K, Huynh L, Ilatsia E, Jarvis E, Johnson SM, Kaufman J, Kelly T, Kemp S, Kern C, Keroack JH, Klopp C, Lagarrigue S, Lamont SJ, Lange M, Lanke A, Larkin DM, Larson G, Layos JKN, Lebrasseur O, Malinovskaya LP, Martin RJ, Martin Cerezo ML, Mason AS, McCarthy FM, McGrew MJ, Mountcastle J, Muhonja CK, Muir W, Muret K, Murphy T, Ng'ang'a I, Nishibori M, O'Connor RE, Ogugo M, Okimoto R, Ouko O, Patel HR, Perini F, Pigozzi MI, Potter KC, Price PD, Reimer C, Rice ES, Rocos N, Rogers TF, Saelao P, Schauer J, Schnabel R, Schneider V, Simianer H, Smith A, Stevens MP, Stiers K, Tiambo CK, Tixier-Boichard M, Torgasheva AA, Tracey A, Tregaskes CA, Vervelde L, Wang Y, Warren WC, Waters PD, Webb D, Weigend S, Wolc A, Wright AE, Wright D, Wu Z, Yamagata M, Yang C, Yin Z-T, Young M, Zhang G, Zhao B & Zhou H (2023) Fourth Report on Chicken Genes and Chromosomes 2022. Cytogenetic and Genome Research 162:405–527

This is an additional paper that PDP contributed to during the course of his PhD. It is relevant to the broad field of research as it outlines our current understanding of the evolution of the avian W chromosome. Author contributions are as follows: All authors wrote the manuscript. PDP co-wrote the section The avian W chromosome: Simply a Y with a different name? This section is included with no alterations as permitted by Rights & Permissions S. Karger AG.

Contents

Abstract *i*

Acknowledgements *ii*

Declaration *iii*

Contents *vi*

Figure List *xii*

Table list *xiv*

Chapter 1. Introduction ***1***

1.1 General overview *1*

1.1.1 Intralocus sexual conflict *1*

1.1.2 Intrasexual competition *3*

1.1.3 Intragenomic conflict: meiotic drive *5*

1.1.4 Sex chromosome evolution *7*

1.2 Study systems *8*

1.2.1 The zebra finch *9*

1.2.2 The stalk-eyed fly *9*

1.3 Thesis outline *10*

1.3.1 Chapter 2. Recent progress in understanding the genomic architecture of sexual conflict *10*

1.3.2 Chapter 3. Detecting signatures of selection on gene expression *10*

1.3.3 Chapter 4. Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches *11*

1.3.4 Chapter 5. The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies *11*

1.3.5 Chapter 6. Discussion *12*

1.4 References *12*

Chapter 2. Recent progress in understanding the genomic architecture of sexual conflict ***22***

2.1 Abstract *23*

2.2 Introduction *23*

2.3 Considering the developmental context of sexual conflict *25*

2.4 Reassessing how we measure differential expression *25*

2.5 Studying sex differences in gene interactions *26*

2.6 Identifying the evolutionary drivers of sexual conflict *27*

2.7 Conclusion *27*

2.8 References and recommended reading *27*

Chapter 3. Detecting signatures of selection on gene expression ***31***

3.1 Abstract *32*

3.2 Inferring the mode of gene expression evolution *33*

3.2.1 Contrasting divergence and variation in expression *33*

3.2.2 Phylogenetic comparative methods *34*

3.3.3 Fitness-based approaches. *34*

3.3 Decomposing transcriptional variation *35*

3.4 Challenges of inferring selection *36*

3.5 Future directions *38*

3.5.1 Transcriptional diversity and layers of gene regulation *38*

3.5.2 Regulatory and co-expression networks *39*

3.5.3 Developmental context *39*

3.5.4 Genotype to phenotype to adaptation *39*

3.6 Methods *39*

3.7 Data availability *40*

3.8 Code availability *40*

3.9 References *40*

Chapter 4. Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches ***43***

4.1 Significance statement *43*

4.2 Abstract *44*

4.3 Introduction *45*

4.4 Results and discussion *48*

4.4.1 Characterising the structure of the Z-linked inversion *48*

4.4.2 Coding sequences in the inversion haplotype are evolving under relaxed purifying selection *52*

4.4.3 Gene expression is conserved between Z-linked haplotypes *56*

4.5 Conclusion *59*

4.6 Materials and methods *59*

4.6.1 Identifying diagnostic SNPs for the Z-linked inversion genotype *59*

4.6.2 Variant calling *59*

4.6.3 Calculation of population genetic statistics *60*

4.6.4 Identifying synonymous and nonsynonymous SNPs *61*

4.6.5 Expression data *61*

4.6.6 Genotyping individuals with RNA-seq data *62*

4.6.7 Quantifying differential gene expression *62*

4.7 Acknowledgements *63*

4.8 Data availability *63*

4.9 References *63*

Chapter 5. The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies ***72***

5.1 Abstract *73*

5.2 Introduction *74*

5.3 Results and discussion *76*

5.3.1 Single-cell atlas of the Teleopsis dalmanni testes *76*

5.3.2 Lack of meiotic sex chromosome inactivation in Teleopsis dalmanni *79*

5.3.3 Status of dosage compensation varies across testes cell types *80*

5.3.4 Impacts of meiotic drive on the cellular landscape of the testes *82*
5.3.5 Impacts of meiotic drive on the transcriptional landscape of the testes *83*
5.4 Conclusion *86*
5.5 Methods *86*
5.5.1 Reference genome and mitochondrial genome assembly *86*
5.5.2 Sample collection *87*
5.5.3 Tissue collection, dissociation and single-cell sequencing *87*
5.5.4 Single-cell RNA-seq data processing *87*
5.5.5 Cell-type identification *88*
5.5.6 Dosage compensation analysis *89*
5.5.7 Differential abundance analysis *89*
5.5.8 Differential gene expression analysis *89*
5.5.9 Trajectory Analysis *89*
5.5.10 Gene Ontology enrichment analysis *90*
5.6 Acknowledgements *90*
5.7 Author contributions *90*
5.8 Data and code availability *91*
5.6 References *91*

Chapter 6. Discussion ***102***
6.1 Research impacts and considerations *103*
6.1.1 The impacts of inversions on gene expression in the zebra finch and stalk-eyed fly *103*
6.1.2 The effect of genomic location on inversions and intragenomic conflict *105*
6.1.3 Maintenance of genetic diversity under conflict *106*
6.1.4 Perspectives on the evolution of meiotic sex chromosome inactivation and dosage compensation *108*
6.2 Methodological advances *110*
6.2.1 The use of non-model systems for cutting-edge technologies *110*

6.2.2 The importance of Y/W reference assemblies in understanding meiotic drive and spermatogenesis *112*

6.2.3 Defining the modes of selection acting on gene expression patterns *113*

6.3 Conclusion *115*

6.4 References *115*

Appendices ***126***

Appendix A. Reply to: Existing methods are effective at measuring natural selection on gene expression ***127***

Appendix B. Chapter 4 supplementary materials ***130***

Chapter 4 supplementary figures *130*

Chapter 4 supplementary tables *130*

Appendix C. Chapter 5 supplementary materials ***138***

Chapter 5 supplementary results *138*

Expression of marker genes across cell types in Teleopsis dalmanni testes *138*

Using ploidy to distinguish cell types across spermatogenesis *139*

Chapter 5 supplementary methods *141*

Alignment, SNP calling and heterozygosity calculations *141*

Chapter 5 supplementary figures *142*

Chapter 5 supplementary tables *150*

Chapter 5 supplementary references *164*

Appendix D. The Avian W Chromosome: Simply a Y with a Different Name? ***167***

What Are the Evolutionary Dynamics of W Chromosomes across Birds? *168*

Is the Avian W Chromosome Selected for Female- Specific Functions? *169*

How Do Multi-Copy Gene Families Evolve on the W Chromosome? *170*

Is There a "Toxic W" Effect? 170

Final Remarks 171

Figure list

Figure 2.1. Detecting the genomic architecture of intralocus sexual conflict. *24*

Figure 3.1. Approaches to detect selection on gene expression. *33*

Figure 3.2. Variation in tissue composition can lead to the perception of differential expression. *36*

Figure 3.3. Inferring selection when expression level is measured from a heterogeneous tissue. *38*

Figure 3.4. The magnitude of allometric shift and covariance of expression level biases the inference of selection. *39*

Figure 4.1. Sequence divergence (d_{XY}), Fixation index (F_{ST}) and Nucleotide diversity (π) across (a) the genome, and (b) on the Z chromosome at non-coding silent sites. *49*

Figure 4.2. Neutrality (a) and Diversity (b) statistics on the Z chromosome for A and B haplotypes. *53*

Figure 4.3. Spearman's rank tests investigating the relationship between the number of fixed synonymous (Ds) and nonsynonymous (Dn) differences between A and B haplotypes and the magnitude of allele-specific expression. *58*

Figure 5.1. Single-cell atlas of the *Teleopsis dalmanni* testes. *77*

Figure 5.2. Genome wide expression patterns across *T. dalmanni* spermatogenesis. *78*

Figure 5.3. Expression of the X chromosome across *T. dalmanni* spermatogenesis. *81*

Figure 5.4. Conservation of X-linked regulation in the germline of drive males. *84*

Figure S4.1. PCA including 24 zebra finches from Singhal et al., 2015. *130*

Figure S4.2. Neutrality (a) and Diversity (b) statistics on the Z chromosome for A and B haplotypes controlled for differences in Z chromosome number. *131*

Figure S4.3. Site frequency spectrum (SFS) statistics for A and B haplotypes. *132*

Figure S4.4. Moving average of gene expression across the Z chromosome for males and females of different genotypes. *132*

Figure S5.1. Additional cell type assignment approaches. *142*

Figure S5.2. Using ploidy to distinguish cell types in ST samples. *143*

Figure S5.3. X chromosome activity across *T. dalmanni* spermatogenesis. *144*

Figure S5.4. Expression of the X chromosome and autosomes across cell types. *145*

Figure S5.5. Expression of orthologs of the *Drosophila* dosage compensation complex. *146*

Figure S5.6. Comparing expression of the X chromosome across cell types in standard (ST) and drive (SR) males. *146*

Figure S5.7. Differential gene expression across cell types. *147*

Figure S5.8. Spatial patterns of differential gene expression across cell types. *148*

Figure S5.9. Differential trajectories of selected genes across spermatogenesis. *149*

Table list

Table 4.1. Population genomic statistics for the autosomes and Z chromosome. *51*

Table 4.2. Comparison of polymorphism and fixed differences between A and B haplotypes. *54*

Table S4.1. Online material: List of diagnostic SNPs to karyotype the Z-linked inversion. *133*

Table S4.2. Inversion haplotypes of samples from Singhal et al., 2015. *133*

Table S4.3. Inversion haplotypes of samples with RNA-seq data. *134*

Table S4.4. Online material: Fixed differences between A and B haplotypes within the coding region. *135*

Table S4.5. Differential expression between genotypes. *135*

Table S4.6. Number of shared genes with and without allele-specific expression. *135*

Table S4.7. Number of genes exhibiting allele-specific expression on the Z and autosomes in each male individual. *136*

Table S4.8. Allele-specific expression in the Z-linked inversion between A and B haplotypes. *137*

Table S5.1. Cell Numbers. *150*

Table S5.2. Marker Genes. *153*

Table S5.3. Cell cycle markers and orthologs. *155*

Table S5.4. Online material: Novel cell type markers. *157*

Table S5.5. Inbreeding values. *157*

Table S5.6. Differential expression between ST & SR. *157*

Table S5.7. DGE enrichment model results. *158*

Table S5.8. Online material: Differentially expressed genes. *158*

Table S5.9. Online material: Genes with differential trajectories. *158*

Table S5.10. GO terms. *159*

Chapter 1. Introduction

1.1 General overview

Competition and conflict are ubiquitous elements of life (Rautiala & Gardner, 2023). They act at many evolutionary levels, from the genes within a genome (Levin & West, 2017), and cells within a multicellular organism (Michod & Roze, 2001), to individuals within a species (Chapman et al., 2003; Hunt et al., 2009) and species within an ecosystem (Brockhurst et al., 2014). At many of these evolutionary levels, conflict and competition are tightly intertwined with the evolution and maintenance of sex and sexual dimorphisms (Chapman et al., 2003), resulting in some of the most dramatic examples of phenotypic diversity (Andersson & Iwasa, 1996; Pitnick et al., 2009). Phenotypic diversity is often attributed to regulatory shifts, including differential expression and alternative splicing, as changes in coding sequence alone do not completely explain observed variation (King & Wilson, 1975; Brawand et al., 2011; Hill et al., 2021). Furthermore, the largely shared genome between males and females means differential regulation of the genome is key in the production of dimorphism and in the resolution of conflicts (Connallon & Knowles, 2005; Wright et al., 2018). This thesis aims to discuss and analyse these intersections across three interacting evolutionary levels: intralocus sexual conflict, intrasexual competition, and intragenomic conflicts. This is done through a variety of methodological approaches including simulations, population genetics and analyses of differential gene expression.

Each chapter in this thesis includes a detailed introduction to the topic and content. Chapter 2 is a published review on our understanding of the genomic architecture and manifestation of sexual conflict to date (Price et al. 2023). Therefore, I will avoid repeating that content in this introductory Chapter. Instead, I provide a brief overview of the key principles and current knowledge in the field to bring together the main themes of subsequent chapters, and to introduce the study systems in the thesis.

1.1.1 Intralocus sexual conflict

Males and females are bound evolutionarily by their shared genomes. We might therefore predict that sex-specific traits will be enriched in regions of the genome with sex-biased inheritance patterns, such as the sex chromosomes or mitochondrial genome. Indeed, the sex

chromosomes tend to be enriched for sex-specific functions and show feminization or masculinisation (Rice, 1984), however, many exceptions to this classic theory are being uncovered (Fry, 2010; Jaquiéry et al., 2013; Baker et al., 2016; Hitchcock & Gardner, 2020). In fact, in many systems sex chromosomes are either homomorphic or entirely absent (Ogawa et al., 1998; Stöck et al., 2011; Kamiya et al., 2012), yet these species still exhibit sex-specific phenotypes (Janzen & Phillips, 2006). Therefore, the shared genome must play an important role in sexual dimorphisms, with many sex-specific traits linked to the autosomes (Ober et al., 2008).

For autosomally linked traits, conflict will arise if a specific locus harbours alleles with opposing effects on male and female fitness, otherwise known as intralocus conflict. These traits will be subject to different evolutionary pressures depending on the sex in which the trait is expressed and selected. This might mean a ‘fit’ genotype for a male is less so for a female, and vice versa, resulting in negative correlations between parental fitness and the fitness of the opposite sex in the offspring (Pischedda & Chippindale, 2006).

Ultimately, there are two outcomes for a trait experiencing intralocus conflict. Firstly, an equilibrium may be reached between the direct and indirect fitness costs and benefits to each sex. For example, if a trait is under sexual selection in males through female mate choice, the benefit to females of having sexier sons may outweigh the cost to her daughters which express the conflicted trait. Although this can happen in both directions, male-centred sexual selection is often more prevalent due to larger gamete size (anisogamy) and higher energetic input into reproduction for females. This outcome becomes less costly when the trait is also linked to genetic condition and subsequently sexually selected as an honest sexual signal (Pomiankowski & Møller, 1997; Rowe & Houle, 1997; Tomkins et al., 2004; Higginson & Reader, 2009). Theory predicts only high-condition males will be able to afford to invest in honest sexual signals (Zahavi, 1977; Iwasa et al., 1991; Iwasa & Pomiankowski, 1994). In turn, females can use these signals to choose high-quality mates and gain fitness benefits. Here, the female’s cost is outweighed not only by the reward for her ‘sexy sons’, but also by the high genetic quality and ‘good genes’ of both her male and female offspring. Secondly, a trait such as a condition-dependent signal may become genetically decoupled between males and females, lowering the intersexual correlation (r_{MF}) and removing costs burdening the other sex. This can occur through either the migration of genes to sex-limited chromosomes (Stewart et al., 2010), or the evolution of sex-specific regulatory mechanisms including sex-biased gene expression (Wright et al., 2018) and alternative splicing (Rogers et al., 2021).

This will ultimately allow selection to independently act on male and female trait values allowing them to migrate to their sex-specific optima.

There is currently mixed evidence for the degree to which ongoing intralocus conflict is occurring. A theoretical outcome of sex-differences in viability from intralocus conflict is differences in allele frequencies between the sexes in adults. However, analyses measuring F_{ST} between the sexes have revealed both ongoing (Cheng & Kirkpatrick, 2016) and resolved conflict (Wright et al., 2018). Some of this discrepancy may be a result of autosomal genes that have duplicated to the Y, and subsequently diverging, being incorrectly mapped to the autosomes. This may in turn inflate F_{ST} between the sexes, explaining some findings of ongoing conflict (Bissegger et al., 2020; Mank et al., 2020). However, the degree to which this technical artefact applies across species is disputed, likely dependent on the quality of assembled Y and W chromosomes (Cheng & Kirkpatrick, 2020).

Despite this mixed evidence, there are signatures of balancing selection on weakly sex-biased genes (Sayadi et al., 2019). This is indicative of increased genetic variation from sexually antagonistic selection for genes whose expression is still coupled between the sexes. Together with examples of negative fitness relationships between the sexes (Pischedda & Chippindale, 2006; Brommer et al., 2007), this suggests that conflict may still play an important role in variation in reproductive success and in the evolutionary trajectory of a species.

In the study of these processes, transcriptomics provides a useful tool for detecting genes with sex-specific expression and potentially resolved or ongoing conflict. Importantly, there are two key factors affecting our detection of resolved conflict using transcriptomics. Firstly, whether the developmental stage in which we are studying expression is the stage at which conflict has occurred (e.g., adult phenotypes may be the product of conflict occurring during development). Secondly, single genes that are key to a sexual dimorphism may not be differentially expressed themselves, but instead placed differentially within the topology of a regulatory network. I review recent advances in this field in Chapter 1 and quantifying the role of gene expression in sexual traits motivates Chapters 4 and 5.

1.1.2 Intrasexual competition

Intrasexual competition generally occurs between males over reproductive access to females. The traits involved are therefore under strong selection to remain competitive between males (Dijkstra & Border, 2018). They additionally experience many other interacting dimensions of selection, such as cryptic female choice and sexually antagonistic coevolution resulting from interlocus sexual conflict (Edward et al., 2014).

One of the most common battlegrounds for intrasexual competition is sperm competition, where the differential success of fertilisation by males is realised through a variety of mechanisms. These include differences in sperm phenotypes (e.g., motility, morphology or quantity), physiological blocking of subsequent males via copulatory plugs, or behaviours such as mate guarding (Edward et al., 2014). Many of these mechanisms may have direct or indirect detrimental fitness effects on the mated females through harm (Hotzy & Arnqvist, 2009) or decreases in re-mating rates (Uhl et al., 2010). Furthermore, a female may then store and choose sperm of males from successive matings, applying yet another stage for sexual selection to act (Orr & Zuk, 2012). Consequently, sperm morphologies (Pitnick et al., 2009) and male reproductive tissues rapidly evolve and co-evolve with the female reproductive system (Presgraves et al., 1999; Murat et al., 2023).

If these processes were driven by directional or stabilising selection, there would be an erosion of genetic variation within the population (Barton & Keightley, 2002). This would be a likely outcome for reproductive traits subject to an evolutionary arms-race between the sexes, and the escalatory co-evolution of secondary sexual traits (Perry & Rowe, 2015). However, for sperm morphology, high heritability, between 40 and 95%, has been found across taxa (Edme et al., 2019). This large additive genetic component opposes the idea that either directional or stabilising selection are eroding genetic diversity. This instead suggests that other processes such as balancing selection, through overdominance or negative frequency dependent selection, may be playing an important role in maintaining the genetic variability of sperm traits. Furthermore, testis-specific genes are often also enriched on the sex chromosomes (Arunkumar et al., 2009; Fortes et al., 2020; Kim et al., 2017; Knief et al., 2017; Vockel et al., 2021), dependent on the extent to which the sex chromosomes are sexualised, and as such are also subject to their unique evolutionary pressures.

Characterising the mode of selection acting on sperm traits, especially those linked to sex chromosomes, is crucial for understanding how high levels of variation and heritability are maintained. This in turn may contribute to understanding how other traits under strong sexual selection maintain variation, and may provide alternate routes for resolving the lek paradox, the observation that variation is maintained in the light of persistent strong selection during mating (Rowe & Houle, 1997; Moore & Moore, 1999). We address this question using the zebra finch, a well-studied model of inversions, sperm competition and sexual selection, in Chapter 4.

1.1.3 Intragenomic conflict: meiotic drive

At the molecular level, genes may propagate themselves at the expense of the rest of the genome and the individual organism, leading to intragenomic conflict (Gardner & Úbeda, 2017). Intragenomic conflicts are diverse in mechanism and span from the biased inheritance of single genes, such as transposable elements (Chuong et al., 2017), to entire chromosomes such as the supernumerary B-chromosomes (Camacho et al., 2003; Ross et al., 2010). Importantly, intragenomic conflicts often occur in a sex-specific manner for both the shared genome, if there is parent-of-origin expression (Patten et al., 2014) or paternal genome elimination (Hodson et al., 2023), and for regions with sex-specific inheritance, such as cytoplasmic genes, the sex chromosomes (Cosmides & Tooby, 1981).

A particularly well-studied example of intragenomic conflict is meiotic drive. Here, selfish genes perturb Mendelian segregation during gametogenesis, causing large evolutionary effects. They have been predicted to influence the evolution of sex chromosomes (Úbeda et al., 2015) as well as the evolution and maintenance of sexual dimorphisms (Wilkinson et al., 1998; Lande & Wilkinson, 1999; Price & Wedell, 2008).

Meiotic drivers can be broadly divided into two classes, true meiotic acting drivers, and post-meiotic germ disablers or killers. Meiotic acting drivers will bias the segregation of chromosomes during meiosis II when the germline ceases being genetically homomorphic. For example, in *Drosophila simulans* the Paris driver causes non-disjunction of the Y in anaphase II (Cazemajor et al., 2000), resulting in a strongly female-biased sex ratio with a small percentage of mature sperm with no sex chromosome that result in sterile males. By contrast, germ disablers or killers act post-meiotically (Bravo Núñez et al., 2018), by killing or stunting the development of target sperm. For example in *Drosophila melanogaster,* the Segregation Distorter (*SD*) system causes incomplete maturation of spermatids carrying its target allele, Responder (*Rsp*), during the histone-to-protamine transition. Despite otherwise normal meiosis, *SD* inhibits proper individualisation of *Rsp* spermatids in the bundle (Larracuente & Presgraves, 2012).

Importantly, the target of a driver must be neither genetically linked to the driver, or be present within a drive-carrying germ cell in order to persist. This will otherwise result in self-targeting, and suicide. Meiotic drivers often act whilst sperm are still joined by ring canals or cytoplasmic bridges before they have individualised and elongated into mature spermatids (Presgraves et al., 1997). These structures allow germ cells to share cytoplasmic contents, such as X-linked transcripts, making haploid cells close to phenotypically diploid (Kaufman et al., 2020).

Ultimately, this may mitigate the differing interests of genetically-discrete gametes. A meiotic driver must therefore not target gene products likely to migrate between sister germ cells. Suicide may also be avoided through genomic inversions of the drive region, inhibiting the linkage of the driver and target allele via recombination.

Sex-linked drivers, also known as sex ratio distorters, target the opposing sex chromosome in the heterogametic sex, leading to an over-representation of male or female offspring. Normally, when sex ratios deviate, balancing selection will favour the rarer sex and an equal sex ratio will be restored (Fisher, 1930). However, in the presence of sex-chromosome drive the population sex ratio will deviate from 1:1, which may ultimately lead to population extinction (Hamilton, 1967; Lyttle, 1981). Accordingly, sex ratio distorters are easy to detect and were the first recorded example of meiotic drive (Gershenson, 1928). The skewed sex ratio results in strong selection for both genomic silencing and behavioural changes to moderate the frequency of a driver. For example, condition-dependent sexual traits may become signals that females use to discern drive-carrying males. Here, the costs of meiotic drive, either from pleiotropic effects such as a reduced germline, or from linked deleterious mutations, makes drive-carrying males lower quality and unable to afford to signal. Such is the case for eye-stalks in Diopsidae, the family of stalk-eyed flies, where drive-carrying males can not afford to invest into this costly sexual signal (Lande & Wilkinson, 1999; Cotton et al., 2014). Sex ratio distorters are also predicted to contribute to the differentiation of homomorphic sex chromosomes. When a driver is linked to a sex-determining allele, the ensuing sex-ratio skew and antagonistic coevolution between itself and its target (on the other proto sex chromosome) can catalyse their divergence into differentiated sex chromosomes (Úbeda et al., 2015; Palmer et al., 2019).

Meiotic drive commonly experiences an arms race of driver-target co-evolution, leading to large multigenic units, similar to supergenes, that are tightly linked via inversions. This makes traditional association mapping approaches less effective at identifying functional regions. In spite of this, there is a growing understanding of the mechanisms underlying drive systems in a range of species (Silver, 1993; Shin et al., 2011; Courret et al., 2019). Identified themes include heterochromatin dysregulation, due to the sensitivity of the genome during the histone-to-protamine transition (Kettaneh & Hartl, 1976; Hauschteck-Jungen & Hartl, 1982; Gingell & McLean, 2020; Herbette et al., 2021; Vedanayagam et al., 2021), and the evolution of multicopy gene families, as a result of X-Y homologous amplification (Aravin et al., 2004; Cocquet et al., 2009). These molecular descriptions are, however, largely limited to the study

of model organisms such as *Anopheles* and *Drosophila* (Courret et al., 2019), due to their research history, wealth of genomic resources and global health impacts (Simoni et al., 2020).

In Chapter 5, I will consider *Teleopsis dalmanni*, a textbook example of sexual selection that is also well known for carrying an X-linked meiotic driver. As of yet, the genomic basis of meiotic drive is sparingly described in this species, including how it acts during spermatogenesis. A comprehensive understanding of both the molecular and ecological impacts of meiotic drive in *Teleopsis dalmanni* will help advance our knowledge of the mechanisms co-opted by meiotic drivers, whether unique or shared across species.

1.1.4 Sex chromosome evolution

Sex chromosomes have asymmetric inheritance between the sexes, and a differing evolutionary environment to the autosomes. This must be considered when attempting to understand how conflict acts and is resolved into sex-linked traits, I will therefore now outline the formation of sex chromosomes and their unique evolutionary landscapes. Although I will focus mainly on XY systems, these arguments hold generally true in ZW systems.

In brief, but already reviewed in great detail (Bachtrog et al., 2014; Tree of Sex Consortium, 2014; Wright et al., 2016), differentiated sex chromosomes emerge from a sex-determining gene, such as SRY in mammals (Kashimada & Koopman, 2010), becoming linked to sexually antagonistic variation. Subsequent recombination suppression via inversions ensures these remain in linkage, which can resolve sexual conflict and lead to the formation of an early sex-specific supergene (Winge, 1927). It may also be likely that inversions create sex-specific genomic environments for sex-beneficial alleles to migrate into, rather than locking in already physically-linked sexually-antagonistic variation (Zhou & Bachtrog, 2012). Determining the ordering of these processes is hard to discern without well-understood systems with nascent sex chromosomes, and reliable methods to detect sexually antagonistic variation (Ruzicka et al., 2020). Ultimately, the continuous cycling of inversions and differentiation leads to further divergence of early sex chromosomes.

Due to the lack of recombination between the X and Y, a Muller's ratchet-like scenario may lead to the Y’s degeneration through the accumulation of deleterious mutations (Muller, 1964; Charlesworth, 1996). In turn, widespread transcriptional silencing of the Y occurs through large areas of heterochromatin (Chang & Larracuente, 2019), eventually leading to gene loss (Wilson Sayres & Makova, 2013; Beaudry et al., 2017). However, this extreme divergence is not universal. Many species including ratite birds and some species of pufferfish (Kamiya et al., 2012; Vicoso et al., 2013) exhibit homomorphic sex chromosomes, with the latter differing

by just a single nucleotide and otherwise normal recombination. Others share their genome completely and instead use environmental cues to determine sex (Janzen & Phillips, 2006).

The stepwise degeneration and divergence of the X and Y ultimately leads to shifts in evolutionary parameters between these chromosomes and the autosomes. This includes mutation rate, recombination rate, and effective population size, and thus selection, drift and diversity (Hedrick, 2007; Ellegren, 2009; Wilson Sayres, 2018). In the hemizygous sex, recessive mutations on the X are more able to respond to selection compared to those occurring on the autosomes. This means deleterious mutations can be more easily purged, and beneficial ones selected for and fixed, elevating rates of adaptive evolution. However, due to an effective population size that is 3/4 of that of the autosomes, this is countered by an increased sensitivity to drift in a finite population (Charlesworth, 2009). How the sex chromosomes respond to selection is also dependent on the functional homology between the X and Y (Mrnjavac et al., 2023), i.e., high homology in certain regions of the sex chromosomes results in them effectively experiencing selection as a diploid. Demographic factors including sexual selection (Mank et al., 2010) will likewise influence the effective population size of the X and Y as it will determine mating rates of each sex.

Ultimately, these differences in evolutionary parameters between the X, Y and autosomes lead to increased rates of evolution on the X, a phenomenon known as faster-X (Meisel & Connallon, 2013). This can subsequently lead to the large-X effect (Coyne, 1992), where the X plays a greater role in reproductive isolation than the autosomes. Sex chromosomes therefore play an important role in not just producing interspecific diversity, but also in consolidating speciation events.

All of these evolutionary features of the sex chromosomes must be considered when studying how sex-linked traits evolve. To be able to describe the relative contribution of selection and drift across the genome, it is crucial to understand how these processes act in regions where effective population size and recombination rate are reduced (Betancourt et al., 2009; Charlesworth, 2009).

1.2 Study Systems

To explore the genomic outcomes and mechanisms of conflict I use two well-established study systems in the fields of sexual selection, *Taeniopygia guttata* (the zebra finch) and *Teleopsis dalmanni* (the Malaysian stalk-eyed fly). These species are especially well suited for this due

to their pronounced sexual dimorphisms, well-documented mating behaviours, and existing genomics resources.

1.2.1 The zebra finch

The zebra finch is a historical model for the study of song (Hauber et al., 2021) and visual dimorphisms (Swaddle & Cuthill, 1997). They are highly social with sexual dimorphism in both plumage and beak colouration, with the latter an honest indicator of condition (Simons et al., 2012). And while they are generally perceived to be monogamous breeders, there is evidence for extra-pair copulation and as such, will be subject to sperm competition (Birkhead et al., 1995, 2005). Because of this they have also emerged as an important system for the field of sperm morphology and competition. Their sperm are highly variable in structure, with morphs associated with genomic inversions, influencing motility, success rates in sperm competition scenarios and fertility (Knief et al., 2017). Whilst we know the immediate impacts of sperm types on fitness, we are yet to understand the genomic processes that have driven the evolution of sperm morph diversity and the consequences of these processes. To test these questions there are good genomic resources available, due to the zebra finch's rich research history, including linkage maps (Stapley et al., 2010), and a well-annotated reference genome generated using PacBio long reads and Hi-C as part of the Vertebrate Genome Project (Rhie et al., 2021).

1.2.2 The stalk-eyed fly

Teleopsis dalmanni is a model system for the study of sexual selection, including its emergence, evolution and consequences. Belonging to the family Diopsidae within the order Diptera, they are defined by their elongated eye-stalks, known as hypercephaly, which show varying degrees of dimorphism across their clade (Shillito, 1971; Baker & Wilkinson, 2001). Diopsidae are broadly distributed across the globe but are most commonly found in Southeast Asia. Our focal species, *Teleopsis dalmanni,* is commonly found in Malaysia, and this has historically been the focal point for its study in the wild. Here, they tend to aggregate around riverbeds, forming leks at dusk and dawn, where they compete for mating rights, with the outcome highly dependent on the relative level of a male's hypercephaly (Cotton et al., 2010). From these observations, they have become a textbook example of the 'good genes' model of sexual selection, with the degree of eye-stalk exaggeration in males an honest indicator of genetic condition and environmental quality (David et al., 1998). An important factor in the study of the sexually-selected eye span is the presence of an X-linked meiotic driver that

dramatically skews wild and captive sex ratios through the destruction or inhibition of Y-linked sperm (Presgraves et al., 1997). As males pay a cost for this in their germline, their condition is impacted, and in turn, their eye span and mating success reduces (Wilkinson et al., 1998). Accordingly, sex-linked meiotic drive has been used as a predictor of levels of sexual selection in *Teleopsis dalmanni* and other Diopsidae. In the presence of drive, condition signalling and female choosiness become increasingly more important traits when sex ratios diverge from 1:1. This is reflected across the clade, with monomorphic species tending to have more even sex ratios (Lande & Wilkinson, 1999). *Teleopsis dalmanni* has a wealth of resources including descriptions of eye development (Hurley et al., 2002), genome assemblies (Reinhardt et al., 2023), and cytological descriptions of spermatogenesis (Presgraves et al., 1997), as well as several recently sequenced reference genomes (Reinhardt et al., 2023; van Rensburg et al., in prep).

1.3 Thesis Outline

This thesis aims to discuss and improve our understanding of the causes and consequences of inter-individual and intragenomic competition. I will also question the approaches we use to detect evolutionary processes acting on the transcriptomic basis of sexual traits. The thesis will be structured as follows:

1.3.1 Chapter 2. Recent progress in understanding the genomic architecture of sexual conflict

The thesis begins by outlining our current understanding of the genomic architecture of sexual conflict. Intralocus sexual conflict has often been a point of discussion in the study of conflict or competition in sexual selection, due to its clear manifestation in sexual dimorphisms. This chapter reviews recent work on the role of gene expression and gene interactions in resolving conflict, and the advent of new technologies to measure and perturb expression. In doing so, it highlights the potential for novel approaches to address outstanding knowledge gaps. This sets the stage for many of the questions addressed in subsequent chapters in this thesis.

1.3.2 Chapter 3. Detecting signatures of selection on gene expression

As many sexually selected traits evolve rapidly or emerge from a shared genome between the sexes, a key tool in understanding their genomic basis is transcriptomics. In Chapter 3, I

critically appraise current methods available for detecting selection acting on gene expression variation. Importantly, I use a series of simulations to show how the use of traditional bulk RNA-seq approaches is heavily caveated by differences in cellular composition of compared tissues. This means that changes in the composition of a tissue between species scales with the overestimation of positive and directional selection over drift. This point is vitally important in our study of sexually selected traits, which are known to evolve rapidly at the cellular level, thus confounding inter-species comparisons. The chapter finishes by highlighting the multi-dimensional nature of transcriptional variation and identifying major unanswered questions necessary for understanding how selection acts on the transcriptome.

1.3.3 Chapter 4. Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches

In Chapter 4, I examine the genomic and transcriptomic landscapes of sperm traits in the zebra finch (*Taeniopygia guttata*). This species displays multiple Z-linked inversion karyotypes associated with differing sperm morphs, whose phenotypes and fitness have been well-described. Whilst there is evidence for heterozygote advantage, the exact modes of selection acting to maintain these inversion polymorphisms are as of yet unknown (Kim et al., 2017; Knief et al., 2017). Through a series of population genetics analyses and analyses of allele-specific expression, we show that heterozygous advantage is most likely an outcome of associative overdominance rather than the previously predicted overdominance.

1.3.4 Chapter 5. The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies

In Chapter 5, I focus on intra-genomic conflicts. Meiotic drivers are often housed by inversions, limiting the use of traditional association studies, and restricting our understanding of their genomic basis. I therefore apply single-cell RNAseq to the testis of drive and standard males in stalk-eyed flies (*Teleopsis dalmanni*) to explore the causes and consequences of a meiotic driver. This chapter begins by describing the expression dynamics of the X and autosomes across spermatogenesis, highlighting their unique patterns of dosage compensation compared to other described insects. I then explore the landscape of the driver, providing several candidate genes and pathways for the inactivation of Y-bearing sperm.

1.3.5 Chapter 6. Discussion

Finally, I synthesise these findings in a general discussion, highlight the advances we have made and discuss where research should progress next. In particular, I will propose outstanding questions about how we detect selection on expression variation, the use of non-model organisms, the evolution of dosage compensation and finally the utility of well-assembled Y and W chromosomes.

1.4 References

Andersson, M., & Iwasa, Y. (1996). Sexual selection. *Trends in Ecology & Evolution*, *11*(2), 53–58.

Aravin, A. A., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., & Gvozdev, V. A. (2004). Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Molecular and Cellular Biology*, *24*(15), 6742–6750.

Arunkumar, K. P., Mita, K., & Nagaraju, J. (2009). The silkworm Z chromosome is enriched in testis-specific genes. *Genetics*, *182*(2), 493–501.

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

Baker, R. H., Narechania, A., DeSalle, R., Johns, P. M., Reinhardt, J. A., & Wilkinson, G. S. (2016). Spermatogenesis drives rapid gene creation and masculinization of the X chromosome in stalk-eyed flies (Diopsidae). *Genome Biology and Evolution*, *8*(3), 896–914.

Baker, R. H., & Wilkinson, G. S. (2001). Phylogenetic analysis of sexual dimorphism and eye-span allometry in stalk-eyed flies (Diopsidae). *Evolution*, *55*(7), 1373–1385.

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, *3*(1), 11–21.

Beaudry, F. E. G., Barrett, S. C. H., & Wright, S. I. (2017). Genomic loss and silencing on the Y chromosomes of Rumex. *Genome Biology and Evolution*, *9*(12), 3345–3355.

Betancourt, A. J., Welch, J. J., & Charlesworth, B. (2009). Reduced effectiveness of selection caused by a lack of recombination. *Current Biology*, *19*(8), 655–660.

Birkhead, T. R., Fletcher, F., Pellatt, E. J., & Staples, A. (1995). Ejaculate quality and the success of extra-pair copulations in the zebra finch. *Nature*, *377*(6548), 422–423.

Birkhead, T. R., Pellatt, E. J., Brekke, P., Yeates, R., & Castillo-Juarez, H. (2005). Genetic effects on sperm design in the zebra finch. *Nature*, *434*(7031), 383–387.

Bissegger, M., Laurentino, T. G., Roesti, M., & Berner, D. (2020). Widespread intersex differentiation across the stickleback genome - The signature of sexually antagonistic selection? *Molecular Ecology*, *29*(2), 262–271.

Bravo Núñez, M. A., Nuckolls, N. L., & Zanders, S. E. (2018). Genetic Villains: Killer Meiotic Drivers. *Trends in Genetics*, *34*(6), 424–433.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348.

Brockhurst, M. A., Chapman, T., King, K. C., Mank, J. E., Paterson, S., & Hurst, G. D. D. (2014). Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society B*, *281*(1797), 20141382.

Brommer, J. E., Kirkpatrick, M., Qvarnström, A., & Gustafsson, L. (2007). The intersexual genetic correlation for lifetime fitness in the wild and its implications for sexual selection. *PLoS One*, *2*(8), e744.

Camacho, J. P. M., Cabrero, J., López-León, M. D., Bakkali, M., & Perfectti, F. (2003). The B chromosomes of the grasshopper *Eyprepocnemis plorans* and the intragenomic conflict. *Genetica*, *117*(1), 77–84.

Cazemajor, M., Joly, D., & Montchamp-Moreau, C. (2000). Sex-ratio meiotic drive in *Drosophila simulans* is related to equational nondisjunction of the Y chromosome. *Genetics*, *154*(1), 229–236.

Chang, C.-H., & Larracuente, A. M. (2019). Heterochromatin-enriched assemblies reveal the sequence and organization of the *Drosophila melanogaster* Y chromosome. *Genetics*, *211*(1), 333–348.

Chapman, T., Arnqvist, G., Bangham, J., & Rowe, L. (2003). Sexual conflict. *Trends in Ecology & Evolution*, *18*(1), 41–47.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, *6*(2), 149–162.

Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205.

Cheng, C., & Kirkpatrick, M. (2016). Sex-specific selection and sex-biased gene expression in humans and flies. *PLoS Genetics*, *12*(9), e1006170.

Cheng, C., & Kirkpatrick, M. (2020). The signal of sex-specific selection in humans is not an

artefact: Reply to Mank et al. *Molecular Ecology*, *29*(8), 1406–1407.

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, *18*(2), 71–86.

Cocquet, J., Ellis, P. J. I., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The multicopy gene *Sly* represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biology*, *7*(11), e1000244.

Connallon, T., & Knowles, L. L. (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics*, *21*(9), 495–499.

Cosmides, L. M., & Tooby, J. (1981). Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology*, *89*(1), 83–129.

Cotton, A. J., Földvári, M., Cotton, S., & Pomiankowski, A. (2014). Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity*, *112*(4), 363–369.

Cotton, S., Small, J., Hashim, R., & Pomiankowski, A. (2010). Eyespan reflects reproductive quality in wild stalk-eyed flies. *Evolutionary Ecology*, *24*(1), 83–95.

Courret, C., Chang, C.-H., Wei, K. H.-C., Montchamp-Moreau, C., & Larracuente, A. M. (2019). Meiotic drive mechanisms: lessons from *Drosophila*. *Proceedings of the Royal Society B*, *286*(1913), 20191430.

Coyne, J. A. (1992). Genetics and speciation. *Nature*, *355*(6360), 511–515.

David, P., Hingle, A., Greig, D., Rutherford, A., Pomiankowski, A., & Fowler, K. (1998). Male sexual ornament size but not asymmetry reflects condition in stalk–eyed flies. *Proceedings of the Royal Society B*, *265*(1411), 2211–2216.

Dijkstra, P. D., & Border, S. E. (2018). How does male-male competition generate negative frequency-dependent selection and disruptive selection during speciation? *Current Zoology*, *64*(1), 89–99.

Edme, A., Zobač, P., Korsten, P., Albrecht, T., Schmoll, T., & Krist, M. (2019). Moderate heritability and low evolvability of sperm morphology in a species with high risk of sperm competition, the collared flycatcher *Ficedula albicollis*. *Journal of Evolutionary Biology*, *32*(3), 205–217.

Edward, D. A., Stockley, P., & Hosken, D. J. (2014). Sexual conflict and sperm competition. *Cold Spring Harbor Perspectives in Biology*, *7*(4), a017707.

Ellegren, H. (2009). The different levels of genetic diversity in sex chromosomes and autosomes. *Trends in Genetics*, *25*(6), 278–284.

Fisher, R. A. (1930). The genetical theory of natural selection. *Oxford, Clarendon Press.*

Fortes, M. R. S., Porto-Neto, L. R., Satake, N., Nguyen, L. T., Freitas, A. C., Melo, T. P.,

Scalez, D. C. B., Hayes, B., Raidan, F. S. S., Reverter, A., & Boe-Hansen, G. B. (2020). X chromosome variants are associated with male fertility traits in two bovine populations. *Genetics, Selection, Evolution*, *52*(1), 46.

Fry, J. D. (2010). The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution*, *64*(5), 1510–1516.

Gardner, A., & Úbeda, F. (2017). The meaning of intragenomic conflict. *Nature Ecology & Evolution*, *1*(12), 1807–1815.

Gershenson, S. (1928). A new sex-ratio abnormality in *Drosophila obscura*. *Genetics*, *13*(6), 488–507.

Gingell, L. F., & McLean, J. R. (2020). A protamine knockdown mimics the function of *Sd* in *Drosophila melanogaster*. *G3, 10*(6), 2111–2115.

Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, *156*(3774), 477–488.

Hauber, M. E., Louder, M. I., & Griffith, S. C. (2021). Neurogenomic insights into the behavioral and vocal development of the zebra finch. *eLife*, *10*, e61849

Hauschteck-Jungen, E., & Hartl, D. L. (1982). Defective histone transition during spermiogenesis in heterozygous segregation distorter males of *Drosophila melanogaster*. *Genetics*, *101*(1), 57–69.

Hedrick, P. W. (2007). Sex: differences in mutation, recombination, selection, gene flow, and genetic drift. *Evolution*, *61*(12), 2750–2771.

Herbette, M., Wei, X., Chang, C.-H., Larracuente, A. M., Loppin, B., & Dubruille, R. (2021). Distinct spermiogenic phenotypes underlie sperm elimination in the Segregation Distorter meiotic drive system. *PLoS Genetics*, *17*(7), e1009662.

Higginson, A. D., & Reader, T. (2009). Environmental heterogeneity, genotype-by-environment interactions and the reliability of sexual traits as indicators of mate quality. *Proceedings of the Royal Society B*, *276*(1659), 1153–1159.

Hill, M. S., Vande Zande, P., & Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*, *22*(4), 203–215.

Hitchcock, T. J., & Gardner, A. (2020). A gene's-eye view of sexual antagonism. *Proceedings of the Royal Society B*, *287*(1932), 20201633.

Hodson, C. N., Toon, A., Cook, L. G., & Ross, L. (2023). Are asymmetric inheritance systems an evolutionary trap? Transitions in the mechanism of paternal genome loss in the scale insect family Eriococcidae. *Genetics, 224*(3), iyad090.

Hotzy, C., & Arnqvist, G. (2009). Sperm competition favors harmful males in seed beetles.

Current Biology, *19*(5), 404–407.

Hunt, J., Breuker, C. J., Sadowski, J. A., & Moore, A. J. (2009). Male-male competition, female mate choice and their interaction: determining total sexual selection. *Journal of Evolutionary Biology*, *22*(1), 13–26.

Hurley, I., Pomiankowski, A., Fowler, K., & Smith, H. (2002). Fate map of the eye-antennal imaginal disc in the stalk-eyed fly *Cyrtodiopsis dalmanni*. *Development Genes and Evolution*, *212*(1), 38–42.

Iwasa, Y., & Pomiankowski, A. (1994). The evolution of mate preferences for multiple sexual ornaments. *Evolution*, *48*(3), 853–867.

Iwasa, Y., Pomiankowski, A., & Nee, S. (1991). The evolution of costly mate preferences II. The "Handicap" Principle. *Evolution*, *45*(6), 1431–1442.

Janzen, F. J., & Phillips, P. C. (2006). Exploring the evolution of environmental sex determination, especially in reptiles. *Journal of Evolutionary Biology*, *19*(6), 1775–1784.

Jaquiéry, J., Rispe, C., Roze, D., Legeai, F., Le Trionnaire, G., Stoeckel, S., Mieuzet, L., Da Silva, C., Poulain, J., Prunier-Leterme, N., Ségurens, B., Tagu, D., & Simon, J.-C. (2013). Masculinization of the X chromosome in the pea aphid. *PLoS Genetics*, *9*(8), e1003690.

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 Genetics*, *8*(7), e1002798.

Kashimada, K., & Koopman, P. (2010). Sry: the master switch in mammalian sex determination. *Development, 137*(23), 3921–3930.

Kaufman, R. S., Price, K. L., Mannix, K. M., Ayers, K. M., Hudson, A. M., & Cooley, L. (2020). *Drosophila* sperm development and intercellular cytoplasm sharing through ring canals do not require an intact fusome. *Development* , *147*(22), dev190140.

Kettaneh, N. P., & Hartl, D. L. (1976). Histone transition during spermiogenesis is absent in segregation distorter males of *Drosophila melanogaster*. *Science*, *193*(4257), 1020–1021.

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L. L., Griffith, S. C., Burke, T., Birkhead, T. R., & Slate, J. (2017). A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature Ecology & Evolution*, *1*(8), 1168–1176.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107–116.

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., Opatová, P., Albrechtová, J., Wittig, M., Franke, A., Albrecht, T., & Kempenaers, B. (2017). A sex-chromosome

inversion causes strong overdominance for sperm traits that affect siring success. *Nature Ecology & Evolution*, *1*(8), 1177–1184.

Lande, R., & Wilkinson, G. S. (1999). Models of sex-ratio meiotic drive and sexual selection in stalk-eyed flies. *Genetics Research*, *74*(3), 245–253.

Larracuente, A. M., & Presgraves, D. C. (2012). The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics*, *192*(1), 33–53.

Levin, S. R., & West, S. A. (2017). The evolution of cooperation in simple molecular replicators. *Proceedings of the Royal Society B*, *284*(1864), 20171967

Lyttle, T. W. (1981). Experimental population genetics of meiotic drive systems. III. Neutralization of sex-ratio distortion in *Drosophila* through sex-chromosome aneuploidy. *Genetics*, *98*(2), 317–334.

Mank, J. E., Shu, J. J., & Wright, A. E. (2020). Signature of sexual conflict is actually conflict resolved. *Molecular Ecology*, *29*(2), 215–217.

Mank, J. E., Vicoso, B., Berlin, S., & Charlesworth, B. (2010). Effective population size and the Faster-X effect: empirical results and their interpretation. *Evolution*, *64*(3), 663–674.

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

Michod, R. E., & Roze, D. (2001). Cooperation and conflict in the evolution of multicellularity. *Heredity*, *86*(Pt 1), 1–7.

Moore, A. J., & Moore, P. J. (1999). Balancing sexual selection through opposing mate choice and male competition. *Proceedings of the Royal Society B*, *266*(1420), 711–716.

Mrnjavac, A., Khudiakova, K. A., Barton, N. H., & Vicoso, B. (2023). Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. *Evolution Letters*, *7*(1), 4–12.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research*, *106*, 2–9.

Murat, F., Mbengue, N., Winge, S. B., Trefzer, T., Leushkin, E., Sepp, M., Cardoso-Moreira, M., Schmidt, J., Schneider, C., Mößinger, K., Brüning, T., Lamanna, F., Belles, M. R., Conrad, C., Kondova, I., Bontrop, R., Behr, R., Khaitovich, P., Pääbo, & S., Kaessmann, H. (2023). The molecular evolution of spermatogenesis across mammals. *Nature*, *613*(7943), 308–316.

Ober, C., Loisel, D. A., & Gilad, Y. (2008). Sex-specific genetic architecture of human disease. *Nature Reviews Genetics*, *9*(12), 911–922.

Ogawa, A., Murata, K., & Mizuno, S. (1998). The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu.

Proceedings of the National Academy of Sciences of the United States of America, *95*(8), 4415–4418.

Orr, T. J., & Zuk, M. (2012). Sperm storage. *Current Biology*, *22*(1), R8–R10.

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, *28*(21), 4709–4724.

Patten, M. M., Ross, L., Curley, J. P., Queller, D. C., Bonduriansky, R., & Wolf, J. B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity*, *113*(2), 119–128.

Perry, J. C., & Rowe, L. (2015). The evolution of sexually antagonistic phenotypes. *Cold Spring Harbor Perspectives in Biology*, *7*(6), a017558.

Pischedda, A., & Chippindale, A. K. (2006). Intralocus sexual conflict diminishes the benefits of sexual selection. *PLoS Biology*, *4*(11), e356.

Pitnick, S., Hosken, D. J., & Birkhead, T. R. (2009). 3 - Sperm morphological diversity. In *Sperm Biology* (pp. 69–149). Academic Press.

Pomiankowski, A., & Møller, A. P. (1997). A resolution of the lek paradox. *Proceedings of the Royal Society B*, *260*(1357), 21–29.

Presgraves, D. C., Baker, R. H., & Wilkinson, G. S. (1999). Coevolution of sperm and female reproductive tract morphology in stalk–eyed flies. *Proceedings of the Royal Society B*, *266*(1423), 1041–1047.

Presgraves, D. C., Severance, E., & Wilkinson, G. S. (1997). Sex chromosome meiotic drive in stalk-eyed flies. *Genetics*, *147*(3), 1169–1180.

Price, P. D., Parkus, S. M., & Wright, A. E. (2023). Recent progress in understanding the genomic architecture of sexual conflict. *Current Opinion in Genetics & Development*, *80*(102047), 102047.

Price, T. A. R., & Wedell, N. (2008). Selfish genetic elements and sexual selection: their impact on male fertility. *Genetica*, *134*(1), 99–111.

Rautiala, P., & Gardner, A. (2023). The geometry of evolutionary conflict. *Proceedings of the Royal Society B*, *290*(1992), 20222423.

Reinhardt, J. A., Baker, R. H., Zimin, A. V., Ladias, C., Paczolt, K. A., Werren, J. H., Hayashi, C. Y., & Wilkinson, G. S. (2023). Impacts of sex ratio meiotic drive on genome structure and function in a stalk-eyed fly. *Genome Biology and Evolution*, *15*(7), evad118.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*,

592(7856), 737–746.

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, *38*(4), 735–742.

Rogers, T. F., Palmer, D. H., & Wright, A. E. (2021). Sex-specific selection drives the evolution of alternative splicing in birds. *Molecular Biology and Evolution*, *38*(2), 519–530.

Ross, L., Pen, I., & Shuker, D. M. (2010). Genomic conflict in scale insects: the causes and consequences of bizarre genetic systems. *Biological Reviews of the Cambridge Philosophical Society, 85*(4), 807–828.

Rowe, L., & Houle, D. (1997). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society B*, *263*(1375), 1415–1421.

Ruzicka, F., Dutoit, L., Czuppon, P., Jordan, C. Y., Li, X.-Y., Olito, C., Runemark, A., Svensson, E. I., Yazdi, H. P., & Connallon, T. (2020). The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, *4*(5), 398–415.

Sayadi, A., Martinez Barrio, A., Immonen, E., Dainat, J., Berger, D., Tellgren-Roth, C., Nystedt, B., & Arnqvist, G. (2019). The genomic footprint of sexual conflict. *Nature Ecology & Evolution*, *3*(12), 1725–1730.

Shillito, J. F. (1971). The genera of diopsidae (Insecta: Diptera). *Zoological Journal of the Linnean Society*, *50*(3), 287–295.

Shin, D., Jin, L., Lobo, N. F., & Severson, D. W. (2011). Transcript profiling of the meiotic drive phenotype in testis of *Aedes aegypti* using suppressive subtractive hybridization. *Journal of Insect Physiology*, *57*(9), 1220–1226.

Silver, L. M. (1993). The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends in Genetics*, *9*(7), 250–254.

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G., Burt, A., Nolan, T., & Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nature Biotechnology*, *38*(9), 1054–1060.

Simons, M. J. P., Briga, M., Koetsier, E., Folkertsma, R., Wubs, M. D., Dijkstra, C., & Verhulst, S. (2012). Bill redness is positively associated with reproduction and survival in male and female zebra finches. *PLoS One*, *7*(7), e40721.

Stapley, J., Birkhead, T. R., Burke, T., & Slate, J. (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Research*, *20*(4), 496–502.

Stewart, A. D., Pischedda, A., & Rice, W. R. (2010). Resolving intralocus sexual conflict:

genetic mechanisms and time frame. *The Journal of Heredity*, *101*(Suppl 1), S94–S99.

Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., Maddalena, T., Sousa, H. C., Martinez-Solano, I., & Perrin, N. (2011). Ever-young sex chromosomes in European tree frogs. *PLoS Biology*, *9*(5), e1001062.

Swaddle, J. P., & Cuthill, I. C. (1997). Female zebra finches prefer males with symmetric chest plumage. *Proceedings of the Royal Society B*, *258*(1353), 267–271.

Tomkins, J. L., Radwan, J., Kotiaho, J. S., & Tregenza, T. (2004). Genic capture and resolving the lek paradox. *Trends in Ecology & Evolution*, *19*(6), 323–328.

Tree of Sex Consortium. (2014). Tree of Sex: a database of sexual systems. Scientific Data, 1(1), 140015.

Úbeda, F., Patten, M. M., & Wild, G. (2015). On the origin of sex chromosomes from meiotic drive. *Proceedings of the Royal Society B*, *282*(1798), 20141932.

Uhl, G., Nessler, S. H., & Schneider, J. M. (2010). Securing paternity in spiders? A review on occurrence and effects of mating plugs and male genital mutilation. *Genetica*, *138*(1), 75–104.

van Rensburg, A. J., Hipperson, H., Bradshaw, S. L., Bates, S., Burke, T., & Pomiankowski, A. (in prep). Genomics of an X-linked meiotic driver in stalk-eyed flies.

Vedanayagam, J., Lin, C.-J., & Lai, E. C. (2021). Rapid evolutionary dynamics of an expanding family of meiotic drive factors and their hpRNA suppressors. *Nature Ecology & Evolution*, *5*(12), 1613–1623.

Vicoso, B., Kaiser, V. B., & Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6453–6458.

Vockel, M., Riera-Escamilla, A., Tüttelmann, F., & Krausz, C. (2021). The X chromosome and male infertility. *Human Genetics*, *140*(1), 203–215.

Wilkinson, G. S., Presgraves, D. C., & Crymes, L. (1998). Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature*, *391*(6664), 276–279.

Wilson Sayres, M. A. (2018). Genetic diversity on the sex chromosomes. *Genome Biology and Evolution*, *10*(4), 1064–1078.

Wilson Sayres, M. A., & Makova, K. D. (2013). Gene survival and death on the human Y chromosome. *Molecular Biology and Evolution*, *30*(4), 781–787.

Winge, Ø. (1927). The Location of Eighteen Genes in Lebistes Reticulatus. *Journal of Genetics*, 18, 1-43

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

Wright, A. E., Fumagalli, M., Cooney, C. R., Bloch, N. I., Vieira, F. G., Buechel, S. D., Kolm, N., & Mank, J. E. (2018). Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters*, *2*(2), 52–61.

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology*, *67*(3), 603–605.

Zhou, Q., & Bachtrog, D. (2012). Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science*, *337*(6092), 341–345.

Chapter 2. Recent progress in understanding the genomic architecture of sexual conflict

Peter D Price[1], Sylvie M Parkus[1], Alison E Wright[1]

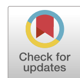
[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom

Keywords: Intralocus sexual conflict, sexual selection, scRNA-seq

Available online at www.sciencedirect.com

ScienceDirect

Recent progress in understanding the genomic architecture of sexual conflict



Peter D. Price*, Sylvie M. Parkus and Alison E. Wright*

Genomic conflict between the sexes over shared traits is widely assumed to be resolved through the evolution of sex-biased expression and the subsequent emergence of sexually dimorphic phenotypes. However, while there is support for a broad relationship between genome-wide patterns of expression level and sexual conflict, recent studies suggest that sex differences in the nature and strength of interactions between loci are instead key to conflict resolution. Furthermore, the advent of new technologies for measuring and perturbing expression means we now have much more power to detect genomic signatures of sexual conflict. Here, we review our current understanding of the genomic architecture of sexual conflict in the light of these new studies and highlight the potential for novel approaches to address outstanding knowledge gaps.

Address
Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom

Corresponding author: Wright, Alison E (a.e.wright@sheffield.ac.uk)
* Twitter account: @PeterDPrice, @alielw

Current Opinion in Genetics & Development 2023, **80**:102047

This review comes from a themed issue on **Evolutionary Genetics**

Edited by **Harmit Singh Malik** and **Judith Mank**

For complete overview of the section, please refer to the article collection, "Evolutionary Genetics (2023)"

Available online 8 May 2023

https://doi.org/10.1016/j.gde.2023.102047

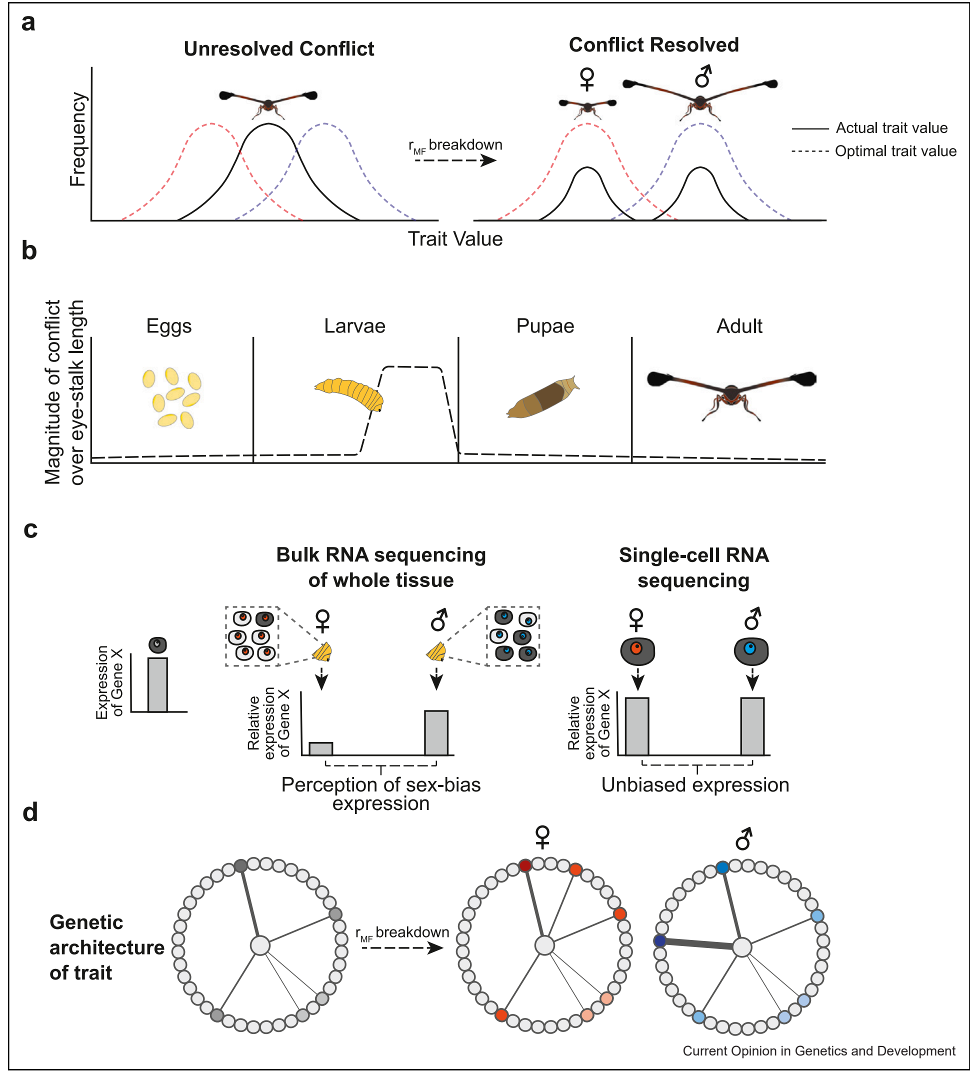
0959–437X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Males and females often experience sex-specific selection pressures towards divergent fitness optima across a range of traits [1]. However, when the genomic basis of these traits is shared between the sexes, intralocus sexual conflict can arise, where the trait in males, females, or both will be inhibited from evolving towards its fitness optima. This conflict has profound implications for multiple aspects of genome and trait evolution, including adaptability and levels of genetic variation [2–4]. Sexual conflict can ultimately be resolved via a decoupling of genetic architectures and subsequent reduction in the intersexual genetic correlation (Figure 1a), and the sex-specific loss or gain of a trait. Within a species, a single genome can encode multiple distinct phenotypes by varying expression levels of the underlying loci. Given that males and females share the majority of their genomes, transcriptomic analyses are increasingly used to study the genes that underlie sexual dimorphisms and the selective regimes acting on them [5]. In principle, sex differences in expression allow the genome to evolve in a sex-specific manner as selection can act independently on the sexes, circumventing constraints posed by sharing a genome and resolving sexual conflict. While some aspects of sexual dimorphism do result from genes located on the sex chromosomes [6], sex-linked genes are often few or absent in many species with pronounced sexually selected traits. It is therefore clear that the majority of sexual dimorphisms must arise from differential expression of genes present in both sexes.

To date, there is a large body of evidence in support of the relationship between broad, genome-wide patterns of expression and proxies for resolved sexual conflict. Differential expression both within [7–9] and across species [10–12] mirrors phenotypic sexual dimorphism with male-biased genes exhibiting greater levels of standing genetic variance in expression [13], consistent with sexual conflict theory. Male-biased gene expression also shows rapid turnover across lineages [14,15], however, this pattern might also represent relaxed pleiotropic constraints [16,17]. Furthermore, targets of ongoing sexually antagonistic selection are not typically enriched for sex differences in expression [18,19], suggesting that sex-biased genes are the footprint of resolved conflict. We also now have several examples directly linking sex-biased genes to sex-specific phenotypes [20]. In many insects, the expression of *doublesex* (*dsx*), the master regulator of sexual differentiation, is central to the evolution of sexually dimorphic traits [21]. For instance, its knockdown in male-horned beetles (*Onthophagus taurus*) reduces head horns, but induces the development of horns in females [22]. Similarly, in a closely related species (*Digitonthophagus gazella), dsx* expression is responsible for sex differences in the length of the fore tibia, used by males to grasp females during mating [23]. Whilst these examples directly link sexually dimorphic phenotypes and sex differences in expression, there are only a limited number of studies to do so and these are restricted to certain taxa, namely insects. This

Figure 1



Detecting the genomic architecture of intralocus sexual conflict. **(a)** A hypothetical pair of stalk-eyed fly species with a sexually selected trait, eye span, that exhibits a shared genomic architecture between males and females. Solid lines indicate the realised trait value, whereas the dotted lines indicate the optimal fitness value. The trait may exhibit a unimodal distribution (left) if it is under ongoing, unresolved sexual conflict or a bimodal distribution (right) where there is a breakdown of the intersexual correlation (r_{MF}) allowing males and females to evolve towards their fitness optima. **(b)** To accurately describe the molecular basis of ongoing and resolved conflict, it is essential to analyse the genetic basis of the trait at the appropriate developmental stage. Although dimorphism may be most striking in the adult phase, its manifestation can occur during development, where each sex will be subject to sex-specific selection pressures and exhibit divergent regulatory patterning. For instance, eye-stalks in stalk-eyed flies develop from eye-antennal imaginal discs and sexual conflict over cell proliferation likely manifests most strongly during the third-instar larval stage [33,81,82]. **(c)** Traditionally, to determine the molecular basis of such traits, whole tissues will be used to measure gene expression in bulk. This can, however, lead to perceptions of differential gene expression that are solely products of differences in the cellular composition of the tissue. This is especially important in the case of sexual ornamentation, where sexually selected structures may differ dramatically in size and cellular composition between the sexes. Single-cell RNA-seq accounts for this by removing the compounding effect of tissue heterogeneity and allowing the comparison of equivalent cell types. **(d)** Network-based approaches are also critical for studying sex-specific architectures. Grey circles represent loci that can contribute to phenotypic variation in eye span. Lines represent loci that do contribute to variation in the trait, and the width of the line corresponds to the size of the effect. This could be mediated by distinct male and female genetic architectures that differ in the number and identity of loci (right-hand side), where the male architecture includes a greater number or more strongly connected condition-dependent loci (e.g. hormonal and growth pathways) than in females.

is in part because the functional identification of the genetic basis of sex-specific adaptation is challenging, especially for more complex phenotypes.

On the other hand, recent research has shown that the genes responsible for some sexually selected traits are either not differentially expressed between males and females or display subtle patterns of expression change [24,25]. For instance, male water striders (*Microvelia longipes*) have exaggerated third legs used to fight and dominate egg-laying sites. Despite similar expression of *Ultrabithorax* (*Ubx*) in both male and female third legs, knockdown of *Ubx* during development results in significantly reduced leg length in males, but has only mild phenotypic effects in females [25]. This research indicates that the relationship between sex-biased expression and sexual conflict is complex, and it remains unclear what magnitude of sex-biased gene expression is necessary to fully resolve sexual conflict. For instance, it is possible that subtle expression differences between the sexes have large phenotypic effects and this is likely to differ on a gene-by-gene basis. Alternatively, sex differences in expression might be limited to specific cell types and so masked from detection using traditional RNA-seq approaches [26,27]. Here, we identify recent advances in efforts to study the role of differential expression in the resolution of sexual conflict, review our current understanding of the genomic architecture of sexual conflict and identify key outstanding questions for the field to address.

Considering the developmental context of sexual conflict

Many adult sexual dimorphisms are the product of differences in growth rate and cell-type proliferation between males and females through development, particularly for exaggerated sexual ornaments [28]. This includes rhinoceros beetle horns (*Trypoxylus dichotomus*) [29], stag beetle mandibles (*Cyclommatus metallifer*) [30], weapons in water striders (*Microvelia longipes*) [25], swordtail caudal fins (*Xiphophorus*) [31,32], and eye-stalks in stalk-eyed flies (*Teleopsis dalmanni*) [33]. Therefore, in many cases, we might expect sexual conflict to manifest most strongly over growth rates during development. For instance, eye-stalks in stalk-eyed flies)develop from eye-antennal imaginal discs. Experimental manipulation indicates that these discs are sensitive to changes in hormone signalling during the third-instar larval stage [33,34], suggesting that sexual conflict over cell proliferation is likely greatest at this point of development (Figure 1b). Such sex differences in growth and cell proliferation are likely due to differential gene expression in males and females, however, these changes in expression will not be detected if transcriptomes are measured after development is completed. This in part, may explain the inconsistencies between studies in the relationship between differential expression and signatures of sexual conflict.

An increasing number of studies are incorporating an ontogenetic perspective to the study of sexual dimorphism, either through measuring expression across multiple developmental stages [8,35,36] or perturbing expression directly during development [23,25]. Together, this research suggests that the magnitude of conflict likely varies across different developmental stages and strategies. For instance, hemimetabolous stick insects (*Timema californicum*) demonstrate a gradual increase in sex-biased expression during development, while holometabolous fruit flies (*Drosophila melanogaster*) have a burst of differential expression in the adult stage [35]. Patterns of expression in these two species closely reflect the development of sexual dimorphism, where *D. melanogaster* has monomorphic larval and pupal stages, and sexual dimorphism manifests abruptly after eclosion, whereas *T. californicum* exhibits a gradual increase in sexually dimorphic traits after its hatchling stage throughout development. Understanding the relative contribution of distinct developmental stages to adult sexual dimorphisms across species is a key priority for pinpointing the genomic architecture of conflict.

Reassessing how we measure differential expression

Key to studying the genomics of sexual conflict is the ability to distinguish whether sex-biased expression is due to regulatory differences or developmental changes in cellular composition between males and females. This is because sexual conflict can be resolved by a decoupling of male and female expression via a reduced intersexual genetic correlation, producing sex-biased genes. In turn, the resolution of conflict permits the evolution of sexual dimorphisms and sex differences in cellular composition. Therefore, only sex-biased expression arising from regulatory differences and not variation in cellular composition between males and females is informative for understanding how selection to resolve sexual conflict directly operates.

Traditional approaches of measuring expression meant it was difficult to distinguish between these two scenarios. This is because bulk RNA-seq approaches measure expression in aggregate across tissues or entire organisms, which, in practice, represents average expression across entire populations of distinct cell types. Therefore, samples that vary in tissue composition can produce patterns of differential expression that are mistaken as evidence of regulatory change or even mask genuine regulatory differences [26,27,37,38] (Figure 1c). This is especially relevant for sexual dimorphisms, which are often complex phenotypes composed of many cell types with variable expression profiles and, by definition, vary

Box 1 Outstanding questions

What magnitude of sex-biased gene expression is necessary to fully resolve sexual conflict, and how does this differ across genes?

How do genomic architectures evolve from shared to sex-specific, and vice versa, and what are the underlying regulatory networks and loci?

Are convergent patterns of sex-specific network rewiring responsible for conflict resolution across distantly diverged species?

Is loss or gain of sexually selected traits more common and how does this manifest in properties of sex-specific regulatory networks?

How do different types of sexual selection, such as Fisherian runaway selection and 'good genes' models, alter the genomic outcome of sexual conflict?

When is condition-dependent ornamentation in females a product of signalling and when is it a sign of incomplete conflict resolution?

in size, structure and composition between males and females [39]. Significant sex differences in cell type abundance seem to be the norm, even for somatic tissues [27]. Unfortunately, this makes it challenging to establish whether sex-biased genes, identified using bulk approaches, are products of regulatory change or simply sex differences in cellular composition. The problem is further confounded if the developmental perspective discussed above is not taken, as we might not expect the targets of sexual conflict to be expressed in the adult phenotype.

New advances in single-cell transcriptomics (scRNA-seq) circumvent issues of tissue composition variation by permitting direct comparisons of male and female expression across equivalent cell types (Figure 1c). To date, a handful of studies have employed single-cell approaches to test the role of differential expression in the evolution of within- [27,40,41] and across-species [42–45] phenotypic variation, however, only one explicitly addresses sex differences [27]. This study found that single-cell versus bulk approaches identify independent sets of sex-biased genes in the guppy (*Poecilia reticulata*) in both somatic and reproductive tissue and these distinct groups of genes exhibit different patterns of coding sequence evolution. Importantly, the exact proportion of genes incorrectly identified as differentially expressed is highly tissue-specific. This could explain some of the inconsistencies across studies in whether sex-biased genes exhibit genomic signatures of resolved or ongoing sexual conflict [18,19,46–48]. As single-cell approaches are increasingly applied to the study of sexual conflict, it will become possible to ascertain how many and what type of genes are truly differentially expressed, and how this relates to proxies of sexual conflict (see Box 1). This is particularly relevant for somatic tissues, where males and females are expected to contain equivalent cell types in different proportions.

Studying sex differences in gene interactions

Genes do not operate in isolation, but in multi-dimensional networks, and there is increasing evidence that sex differences in the nature and strength of interactions between loci are common [49–51]. This likely explains the growing evidence that loci expressed at similar levels in both sexes can have distinct sex-specific effects [25,52–55], consistent with separate male and female genetic architectures (Figure 1d). The evolution of sex-specific genetic architectures potentially alleviates conflict by circumventing constraints imposed by a shared genome and facilitates the evolution of sexual dimorphism [18]. Therefore, shifting focus to studying sex differences in co-expression networks is more informative for understanding how conflict can be resolved than current approaches where genes are typically studied independently. Sexual dimorphisms evolve rapidly, with frequent losses and gains [56,57], but it remains unclear if the underlying loci mirror this pattern. Important next steps include identifying the underlying regulatory networks and loci responsible for male and female genetic architectures, establishing how genetic architectures evolve from shared to sex-specific and vice versa and how frequently (see Box 1). For instance, for traits encoded by sex-specific factors expressed during the early stages of sex-determination pathways, such as the *dsx* gene, the construction of separate male and female architectures is relatively straightforward and sexual conflict could be easily mitigated.

Notably, it remains unclear whether convergent patterns of sex-specific network rewiring are responsible for conflict resolution across distantly diverged species, although the repeated involvement of *dsx* in sexual traits across insects [21,22,30,58,59] suggests similar processes might be operating. *dsx*, in particular, has distinct sex- and tissue-specific target loci due to alternative splicing into male and female isoforms. This allows regulation of the same genes in opposite directions in males and females in the dung beetle (*Onthophagus taurus*) and likely many other insects [60]. In the future, single-cell approaches are particularly important to address these questions as differences in cellular composition between males and females can affect the measurement of gene co-expression due to key differences in gene networks across cell types [61,62]. Unfortunately, many of the available methods of regulatory network inference are currently not effective for single-cell transcriptome data due to its intrinsic sparsity and high technical variation [61,63].

Identifying the evolutionary drivers of sexual conflict

Studies into the genomic basis of sexual conflict typically do not consider the mode of sexual selection. However, the type of sexual selection has important consequences for the strength of sexual conflict and how we expect conflict to manifest and be resolved across the genome. For instance, under the 'good genes' model of sexual selection, sexually dimorphic traits are predicted to evolve as honest signals of male genetic quality, where only high-condition males can afford to invest in elaborate sexual ornaments [64,65]. In contrast, under Fisherian runaway, sexually selected traits are not linked to individual condition [66]. This distinction is important as, in principle, sex-specific condition dependence restricts the expression of the costly trait to only those individuals who have sufficient resources, serving as a potential mechanism that aids in resolving sexual conflict.

There is now considerable evidence across many organisms that honest male sexual traits have evolved repeatedly in a range of phenotypes [28,67–70] and that the genetic architecture of these traits is sex-specific and condition-dependent [21], but see Ref. [71]. Recent studies have started to elucidate the precise genomic and physiological processes that link honest traits to condition and nutritional status in a sex-specific manner. For instance, conserved growth and hormonal pathways have been identified as common mechanisms regulating condition dependence of several male sexual traits, such as juvenile hormone signalling in stalk-eyed flies (*Teleopsis dalmanni*) [33] and stag beetles (*Cyclommatus metallifer*) [30], and insulin signalling in several beetle species (*O. taurus* and *T. dichotomus*) [29,72]. Often, these pathways are intrinsically linked to sex-determination factors meaning that the male but not the female trait can be linked to individual condition [23,30].

Recent evidence suggests that exaggerated traits in females can also be highly associated with individual condition [73–75]. If so, this suggests that plastic resource allocation [65] in both sexes could be key to alleviating sexual conflict for certain traits. However, whilst a 'good genes' model provides a framework for conflict resolution, it is unclear whether occurrences of female ornamentation are the active signalling of condition [74], or instead a product of correlated evolution arising from incomplete conflict resolution where male and female traits exhibit a similar genomic architecture. In this instance, high-condition females will pay a greater fitness cost [76].

Interestingly, in turn, the evolution of condition-dependent genetic architectures that exhibit variable expression across individuals may actually act to exacerbate the strength of sexual conflict [77]. This can be because high-fitness males produce low-fitness daughters and high-fitness mothers produce low-fitness sons [78,79] or because certain environments are more favourable for males than for females [80]. This sets the stage for a feedback loop, where 'good genes' processes might only be possible once sexual conflict has been resolved via the evolution of condition-dependent traits.

Together, these results suggest that the genomic architecture of sexual conflict may vary quite profoundly under different modes of sexual selection, however, this is rarely considered when testing for signatures of conflict across the genome. Establishing specific predictions for the types of loci and their interactions responsible for conflict resolution under 'good genes' versus Fisherian models of sexual selection is a major priority for the future.

Conclusion

The development of novel technologies for measuring and perturbing expression has shed new light on our understanding of how sexual conflict manifests across the genome and whether the differential gene expression we perceive is a signature of ongoing conflict or conflict resolved. It is also now apparent that sex differences in the nature and strength of gene interactions are key to conflict resolution. However, a number of outstanding questions regarding the genomic architecture of conflict remain unanswered (see Box 1). Solving these will require the effective integration of single-cell approaches across development with phenotypic studies that quantify the underlying drivers of conflict.

Data Availability

No data were used for the research described in the article.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

This work was funded by a Natural Environment Research Council (NERC), UK, Independent Research Fellowship to A.E.W. (NE/N013948/1) and an NERC ACCE DTP to P.D.P.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Parker GA: **Sexual selection and sexual conflict**. Sexual Selection and Reproductive Competition in Insects. Elsevier; 1979:166.

2. • Kaufmann P, Howie JM, Immonen E: **Sexually antagonistic selection maintains genetic variance when sexual dimorphism evolves**. *Proc R Soc B* 2023, **290**:20222484.
This paper on the seed beetle (*Callosobruchus maculatus*) demonstrates that sexually antagonistic selection has the propensity to maintain

autosomal additive genetic variance longer than sex-limited directional selection.

3. Arnqvist G, Rowe L: **Sexual Conflict**. Princeton University Press; 2005.

4. Ruzicka F, Holman L, Connallon T: **Polygenic signals of sex differences in selection in humans from the UK Biobank**. *PLoS Biol* 2022, **20**:e3001768.

5. Mank JE: **The transcriptional architecture of phenotypic dimorphism**. *Nat Ecol Evol* 2017, **1**:0006.

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

7. Pointer MA, Harrison PW, Wright AE, Mank JE: **Masculinization of gene expression is associated with exaggeration of male sexual dimorphism**. *PLoS Genet* 2013, **9**:e1003697.

8. Ingleby FC, Flis I, Morrow EH: **Sex-biased gene expression and sexual conflict throughout development**. *Cold Spring Harb Perspect Biol* 2014, **7**:a017632.

9. Zemp N, Tavares R, Widmer A: **Fungal infection induces sex-specific transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene latifolia***. *PLoS Genet* 2015, **11**:e1005536.

10. Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE: **Sexual selection drives evolution and rapid turnover of male gene expression**. *PNAS* 2015, **112**:4393-4398.

11. Parker DJ, Bast J, Jalvingh K, Dumas Z, Robinson-Rechavi M, Schwander T: **Sex-biased gene expression is repeatedly masculinized in asexual females**. *Nat Commun* 2019, **10**:1-11.

12. Wright AE, Rogers TF, Fumagalli M, Cooney CR, Mank JE: **Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict**. *Mol Ecol* 2019, **28**:2860-2871.

13. Allen SL, Bonduriansky R, Chenoweth SF: **Genetic constraints on microevolutionary divergence of sex-biased gene expression**. *Philos Trans R Soc* 2018, **373**:20170427.

14. Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B: **Constraint and turnover in sex-biased gene expression in the genus Drosophila**. *Nature* 2007, **450**:233-237.

15. Rogers TF, Palmer DH, Wright AE: **Sex-specific selection drives the evolution of alternative splicing in birds**. *Mol Biol Evol* 2021, **38**:519-530.

16. Scharmann M, Rebelo AG, Pannell JR: **High rates of evolution preceded shifts to sex-biased gene expression in Leucadendron, the most sexually dimorphic angiosperms**. *eLife* 2021, **10**:e67485.

17. Mank JE, Ellegren H: **Are sex-biased genes more dispensable?** *Biol Lett* 2009, **5**:409-412.

18. Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, Kolm N, Mank JE: **Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture**. *Evol Lett* 2018, **2**:52-61.

19. Ruzicka F, Hill MS, Pennell TM, Flis I, Ingleby FC, Mott R, Fowler K, Morrow EH, Reuter M: **Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies**. *PLoS Biol* 2019, **17**:e3000244.

20. Galouzis CC, Prud'homme B: **Transvection regulates the sex-biased expression of a fly X-linked gene**. *Science* 2021, **371**:396-400.

21. Mank JE: **Sex-specific morphs: the genetics and evolution of intra-sexual variation**. *Nat Rev Genet* 2023, **24**:44-52.

22. Kijimoto T, Moczek AP, Andrews J: **Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns**. *PNAS* 2012, **109**:20526-20531.

23. •• Rohner PT, Linz DM, Moczek AP: **Doublesex mediates species-, sex-, environment- and trait-specific exaggeration of size and shape**. *Proc R Soc B: Biol Sci* 2021, **288**:20210241.

This study applied RNAseq and RNAi-knockdown to the sexually dimorphic dung beetle and found that the gene networks and developmental pathways that underlie tibia exaggeration, as well as the *dsx* gene that differentially alters phenotypic exaggeration, function in a context-dependent manner.

24. Khila A, Abouheif E, Rowe L: **Function, developmental genetics, and fitness consequences of a sexually antagonistic trait**. *Science* 2012, **336**:585-589.

25. • Toubiana W, Armisén D, Viala S, Decaras A, Khila A: **The growth factor BMP11 is required for the development and evolution of a male exaggerated weapon and its associated fighting behavior in a water strider**. *PLoS Biol* 2021, **19**:e3001157.

This study used comparative transcriptomics and RNA interference to identify genes responsible for the exaggerated third leg found in males across many water strider species. Notably, despite similar expression of *Ubx* in both male and female third legs, knockdown of *Ubx* during development results in significantly reduced leg length in males but has only mild phenotypic effects in females, indicating that the relationship between sex-biased expression and sexual conflict is complex.

26. Price P, Palmer D, Taylor J, Kim D, Place E, Rogers T, Mank J, Cooney C, Wright A: **Detecting signatures of selection in regulatory variation**. *Nat Ecol Evol* 2022, **6**:1035-1045.

27. •• Darolti I, Mank JE: **Sex-biased gene expression at single-cell resolution: cause and consequence of sexual dimorphism**. *Evolution Letters* 2023,qrad013.

This is the first study to explicitly apply single-cell RNA sequencing to test the extent to which sex-biased expression is a consequence of regulatory change between the sexes or simply differences in cell type abundance. They found that single-cell versus bulk approaches identify independent sets of sex-biased genes, and these distinct groups of genes exhibit different patterns of coding sequence evolution.

28. Lavine L, Gotoh H, Brent CS, Dworkin I, Emlen DJ: **Exaggerated trait growth in insects**. *Annu Rev Entomol* 2015, **60**:453-472.

29. Emlen DJ, Warren IA, Johns A, Dworkin I, Lavine LC: **A mechanism of extreme growth and reliable signaling in sexually selected ornaments and weapons**. *Science* 2012, **337**:860-864.

30. Gotoh H, Miyakawa H, Ishikawa A, Ishikawa Y, Sugime Y, Emlen DJ, Lavine LC, Miura T: **Developmental link between sex and nutrition; doublesex regulates sex-specific mandible growth via juvenile hormone signaling in stag beetles**. *PLoS Genet* 2014, **10**:e1004098.

31. Powell DL, Payne C, Banerjee SM, Keegan M, Bashkirova E, Cui R, Andolfatto P, Rosenthal GG, Schumer M: **The genetic architecture of variation in the sexually selected sword ornament and its evolution in hybrid populations**. *Curr Biol* 2021, **31**:923-935 .e911.

32. Schartl M, Kneitz S, Ormanns J, Schmidt C, Anderson JL, Amores A, Catchen J, Wilson C, Geiger D, Du K, *et al.*: **The developmental and genetic architecture of the sexually selected male ornament of swordtails**. *Curr Biol* 2021, **31**:911-922 e914.

33. Fry CL: **Juvenile hormone mediates a trade-off between primary and secondary sexual traits in stalk-eyed flies**. *Evol Dev* 2006, **8**:191-201.

34. Cotton S, Fowler K, Pomiankowski A: **Condition dependence of sexual ornament size and variation in the stalk-eyed fly *Cyrtodiopsis dalmanni***. *Evolution* 2004, **58**:1038-1046.

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

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

37. Montgomery SH, Mank JE: **Inferring regulatory change from gene expression: the confounding effects of tissue scaling**. *Mol Ecol* 2016, **25**:5114-5128.

38. Hunnicutt KE, Good JM, Larson EL: **Unraveling patterns of disrupted gene expression across a complex tissue**. *Evolution* 2022, **76**:275-291.

39. Mank JE, Rideout EJ: **Developmental mechanisms of sex differences: from cells to organisms**. *Development* 2021, **148**:dev199750.
••

This review highlights the importance of incorporating both sexes for robust developmental studies in order to understand how sex differences contribute to the developmental mechanisms that underlie sexual dimorphism.

40. Fuess LE, Bolnick DI: **Single-cell RNA sequencing reveals microevolution of the stickleback immune system**. *Genome Biol Evol* 2023, **15**:evad053.

41. Wei KH-C, Chatla K, Bachtrog D: **Single cell RNA-seq in Drosophila testis reveals evolutionary trajectory of sex chromosome regulation**. *bioRxiv* 2022,519494.

42. Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, Sukhwani M, Orwig KE, Li JZ, Hammoud SS: **Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis**. *Dev Cell* 2020, **54**:529-547 e512.

43. Segami JC, Mugal CF, Cunha C, Bergin C, Schmitz M, Semon M, Qvarnström A: **The genomic basis of hybrid male sterility in Ficedula flycatchers**. *bioRxiv* 2022,508503.

44. Segami JC, Semon M, Cunha C, Bergin C, Mugal CF, Qvarnström A: **Single-cell transcriptomics reveals relaxed evolutionary constraint of spermatogenesis in two passerine birds as compared to mammals**. *bioRxiv* 2022,477241.

45. Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mößinger K, *et al*.: **The molecular evolution of spermatogenesis across mammals**. *Nature* 2023, **613**:308-316.

46. Cheng C, Kirkpatrick M: **Sex-specific selection and sex-biased gene expression in humans and flies**. *PLoS Genet* 2016, **12**:e1006170.

47. Dutoit L, Mugal CF, Bolívar P, Wang M, Nadachowska-Brzyska K, Smeds L, Yazdi HP, Gustafsson L, Ellegren H: **Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome**. *Mol Ecol* 2018, **27**:3572-3581.

48. Sayadi A, Martinez Barrio A, Immonen E, Dainat J, Berger D, Tellgren-Roth C, Nystedt B, Arnqvist G: **The genomic footprint of sexual conflict**. *Nat Ecol Evol* 2019, **3**:1725-1730.

49. Sutherland BJG, Prokkola JM, Audet C, Bernatchez L: **Sex-specific co-expression networks and sex-biased gene expression in the Salmonid Brook Charr *Salvelinus fontinalis***. *G3* 2019, **9**:955-968.

50. Lopes-Ramos CM, Chen CY, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, Platig J, Glass K, Quackenbush J, DeMeo DL: **Sex differences in gene expression and regulatory networks across 29 human tissues**. *Cell Rep* 2020, **31**:107795.
••

This paper describes the abundance of sex-specific regulatory networks in humans, highlighting that transcription factors are often not differentially expressed yet target distinct and different sets of genes in males and females.

51. Rago A, Werren JH, Colbourne JK: **Sex biased expression and co-expression networks in development, using the hymenopteran *Nasonia vitripennis***. *PLoS Genet* 2020, **16**:e1008518.

52. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, Kilpeläinen TO, Esko T, Mägi R, Li S: **Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits**. *PLoS Genet* 2013, **9**:e1003500.

53. Rawlik K, Canela-Xandri O, Tenesa A: **Evidence for sex-specific genetic architectures across a spectrum of human complex traits**. *Genome Biol* 2016, **17**:1-8.

54. Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, Brown SDM, Chesler EJ, Dickinson ME, Flenniken AM, *et al*.: **Prevalence of sexual dimorphism in mammalian phenotypic traits**. *Nat Commun* 2017, **8**:15475.

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

56. Wiens JJ: **Widespread loss of sexually selected traits: how the peacock lost its spots**. *Trends Ecol Evol* 2001, **16**:517-523.

57. Oliver JC, Monteiro A: **On the origins of sexual dimorphism in butterflies**. *Proc R Soc B: Biol Sci* 2011, **278**:1981-1988.

58. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB: **The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila**. *Cell* 2008, **134**:610-623.

59. Hendrickx F, De Corte Z, Sonet G, Van Belleghem SM, Köstlbacher S, Vangestel C: **A masculinizing supergene underlies an exaggerated male reproductive morph in a spider**. *Nat Ecol Evol* 2022, **6**:195-206.
•

There are two male morphs in *Oedothorax gibbosus* spider, the 'hunched' morph and the 'flat' morph, with the flat morph more closely resembling the female spider form. This study found that the hunched-determining allele has roughly an extra three megabases, containing several masculinizing genes, that are notably absent from the flat-determining allele and are responsible for the male reproductive hunched morph.

60. Ledón-Rettig CC, Zattara EE, Moczek AP: **Asymmetric interactions between doublesex and tissue- and sex-specific target genes mediate sexual dimorphism in beetles**. *Nat Commun* 2017, **8**:14593.

61. Cha J, Lee I: **Single-cell network biology for resolving cellular heterogeneity in human diseases**. *Exp Mol Med* 2020, **52**:1798-1808.

62. Ribeiro DM, Ziyani C, Delaneau O: **Shared regulation and functional relevance of local gene co-expression revealed by single cell analysis**. *Commun Biol* 2022, **5**:876.

63. Chen S, Mar JC: **Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data**. *BMC Bioinform* 2018, **19**:1-21.

64. Iwasa Y, Pomiankowski A, Nee S: **The evolution of costly mate preferences II. The "handicap" principle**. *Evolution* 1991, **45**:1431-1442.

65. Fromhage L, Henshaw JM: **The balance model of honest sexual signaling**. *Evolution* 2022, **76**:445-454.
••

This paper presents a new model of sexual signalling, which considers the diminishing returns of investment into individual fitness components, thereby arguing that honest sexual signals are in balance with the allocation of resources to other components of the individual's fitness.

66. Fisher RA: **The Genetical Theory of Natural Selection: A Complete Variorum Edition**. Oxford University Press; 1999.

67. Emlen DJ: **Environmental control of horn length dimorphism in the beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae)**. *Proc R Soc B: Biol Sci* 1994, **256**:131-136.

68. Warren IA, Gotoh H, Dworkin IM, Emlen DJ, Lavine LC: **A general mechanism for conditional expression of exaggerated sexually-selected traits**. *Bioessays* 2013, **35**:889-899.

69. Macartney EL, Crean AJ, Nakagawa S, Bonduriansky R: **Effects of nutrient limitation on sperm and seminal fluid: a systematic review and meta-analysis**. *Biol Rev* 2019, **94**:1722-1739.

70. Dougherty LR: **Meta-analysis reveals that animal sexual signalling behaviour is honest and resource based**. *Nat Ecol Evol* 2021, **5**:688-699.
•

This meta-analysis demonstrates that sexual signalling is overwhelmingly an honest indicator of individual state, with high-condition individuals, predominantly males, tending to signal more.

71. Bath E, Rostant W, Ostridge HJ, Smith S, Mason JS, Rafaluk-Mohr T, Mank JE, Chapman T, Perry JC: **Sexual selection and the evolution of condition-dependence: an experimental test at two resource levels**. *Evolution* 2023, **77**:776-788 Evolution, 58, 2004, 1038–1046.

72. Casasa S, Moczek AP: **Insulin signalling's role in mediating tissue-specific nutritional plasticity and robustness in the horn-polyphenic beetle *Onthophagus taurus***. *Proc R Soc B: Biol Sci* 2018, **285**:20181631.

73. Takeda K, Furuta T, Hamada M, Sato Y, Taniguchi K, Tanizawa A, Yagi T, Adachi-Yamada T: **Handicap theory is applied to females**

but not males in relation to mate choice in the stalk-eyed fly *Sphyracephala detrahens*. *Sci Rep* 2020, **10**:19684.

74. Hernández A, Martínez-Gómez M, Beamonte-Barrientos R, Montoya B: **Colourful traits in female birds relate to individual condition, reproductive performance and male-mate preferences: a meta-analytic approach**. *Biol Lett* 2021, **17**:20210283.

75. Nolazco S, Delhey K, Nakagawa S, Peters A: **Ornaments are equally informative in male and female birds**. *Nat Commun* 2022, **13**:5917.

76. Parrett JM, Chmielewski S, Aydogdu E, Łukasiewicz A, Rombauts S, Szubert-Kruszyńska A, Babik W, Konczal M, Radwan J: **Genomic evidence that a sexually selected trait captures genome-wide variation and facilitates the purging of genetic load**. *Nat Ecol Evol* 2022, **6**:1330-1342.

77. •• Flintham EO, Savolainen V, Mullon C: **Male harm offsets the demographic benefits of good genes**. *Proc Natl Acad Sci* 2023, **120**:e2211668120.

The paper models the consequences of sexual selection on a population with condition dependent sexual traits. They find that sexual conflict is stronger when individuals are in better condition and that good genes selection exacerbates sexual conflict and reduces population size.

78. Pischedda A, Chippindale AK: **Intralocus sexual conflict diminishes the benefits of sexual selection**. *PLoS Biol* 2006, **4**:e356.

79. Pischedda A, Chippindale AK: **Direct benefits of choosing a high-fitness mate can offset the indirect costs associated with intralocus sexual conflict**. *Evolution* 2017, **71**:1710-1718.

80. Harts AM, Schwanz LE, Kokko H: **Demography can favour female-advantageous alleles**. *Proc R Soc B: Biol Sci* 2014, **281**:20140005.

81. Hurley I, Pomiankowski A, Fowler K, Smith H: **Fate map of the eye-antennal imaginal disc in the stalk-eyed fly *Cyrtodiopsis dalmanni***. *Dev Genes Evol* 2002, **212**:38-42.

82. Singh A, Gogia N, Chang CY, Sun YH: **Proximal fate marker homothorax marks the lateral extension of stalk-eyed fly *Cyrtodopsis whitei***. *Genesis* 2019, **57**:e23309.

Chapter 3. Detecting signatures of selection on gene expression

Peter D. Price[1], Daniela H. Palmer Droguett[1,2], Jessica A. Taylor[1,3], Dong Won Kim[4], Elsie S. Place[5], Thea F. Rogers[1], Judith E. Mank[6,7,8], Christopher R. Cooney[1], Alison E. Wright[1]

[1] Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK.

[2] Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA.

[3] Department of Biosciences, Durham University, Durham, UK.

[4] Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[5] Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, Sheffield, UK.

[6] Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada.

[7] Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada.

[8] Centre for Ecology and Conservation, University of Exeter, Penryn, UK.

Response from the field

Following the publication of this manuscript, we received much engaging feedback from the research community, including a matters arising article from Hunter Fraser (2022). Our full response (Price et al., 2022) can be seen in Appendix A. (Reply to: Existing methods are effective at measuring natural selection on gene expression).

*Fraser, H.B. Existing methods are effective at measuring natural selection on gene expression. Nat Ecol Evol **6**, 1836–1837 (2022).*

*Price, P.D., Palmer Droguett, D.H., Taylor, J.A. et al. Reply to: Existing methods are effective at measuring natural selection on gene expression. Nat Ecol Evol **6**, 1838–1839 (2022).*

nature ecology & evolution

ANALYSIS
https://doi.org/10.1038/s41559-022-01761-8

Detecting signatures of selection on gene expression

Peter D. Price[1]✉, Daniela H. Palmer Droguett[1,2], Jessica A. Taylor[1,3], Dong Won Kim[4], Elsie S. Place[5], Thea F. Rogers[1], Judith E. Mank[6,7,8], Christopher R. Cooney[1,9] and Alison E. Wright[1,9]✉

A substantial amount of phenotypic diversity results from changes in gene expression levels and patterns. Understanding how the transcriptome evolves is therefore a key priority in identifying mechanisms of adaptive change. However, in contrast to powerful models of sequence evolution, we lack a consensus model of gene expression evolution. Furthermore, recent work has shown that many of the comparative approaches used to study gene expression are subject to biases that can lead to false signatures of selection. Here we first outline the main approaches for describing expression evolution and their inherent biases. Next, we bridge the gap between the fields of phylogenetic comparative methods and transcriptomics to reinforce the main pitfalls of inferring selection on expression patterns and use simulation studies to show that shifts in tissue composition can heavily bias inferences of selection. We close by highlighting the multi-dimensional nature of transcriptional variation and identifying major unanswered questions in disentangling how selection acts on the transcriptome.

A growing body of evidence indicates that changes in patterns of gene expression play a key role in phenotypic divergence. Within species, a single genome can encode multiple distinct traits by varying expression levels of the underlying loci[1,2]. Similarly, across species, divergence in gene expression is implicated in major phenotypic differences that underlie adaptive change[3–7]. Given the importance of gene regulation in shaping phenotypic diversity, transcriptome analyses are widely used as a genomic tool to identify the genes that underlie phenotypic variation and the selective regimes acting on them[1,7]. However, the dominant mode of selection acting on gene expression remains controversial. Current evidence supports the notion that global patterns of gene expression evolve predominantly under stabilizing selection, but the extent of neutral evolution is heavily debated[8–11].

Much of this debate is driven by the lack of a consensus neutral model of transcriptome evolution. In contrast to established models of sequence evolution that allow us to scan coding sequence data for regions of adaptive evolution, gene expression can be complex and non-additive in its phenotypic effects. This complexity has resulted in a wide range of approaches to study the evolution of gene expression[7,12,13]. Importantly, these approaches make direct assumptions about how expression evolves across species, many of which have yet to be robustly validated, and these assumptions vary extensively across models. Over the last decade, statistical frameworks developed in the field of phylogenetic comparative methods have been applied to transcriptome data to infer selection[12,14], and these have provided important insights into patterns of expression divergence. However, in recent years it has become clear that several of these phylogenetic comparative approaches suffer from biases that often lead to false inferences of stabilizing selection when applied to real phenotypic data[15,16]. Many of the root causes of these biases are even more pronounced in transcriptomic data, but the issues uncovered in the phylogenetic comparative literature[15–17] are only rarely discussed in the genomics field[18,19].

Finally, most studies make the explicit assumption that when differential gene expression is observed, it is the direct result of regulatory change. In reality, this fundamental assumption may often be flawed as most studies of expression evolution measure transcript abundance in bulk across heterogeneous tissue samples and hence cannot distinguish changes in gene expression from differences in tissue composition[20–22]. This problem undermines our current understanding of the nature and abundance of variation in gene expression across species, and how it contributes to phenotypic divergence. Although the implications of varying tissue composition across species for measuring differential expression have been discussed[20–22], the consequences of how it affects the inference of expression evolution have received less attention.

Here we examine our current understanding of the evolutionary processes generating variation in gene expression. First, we outline the main approaches for describing gene expression evolution, examine their inherent biases and synthesize findings to provide new perspectives to the debate over how selection acts on the transcriptome. Second, we attempt to bridge the gap between the fields of phylogenetic comparative methods and transcriptomics to reinforce the main pitfalls of inferring selection on expression levels. Importantly, we discuss the consequences of changes in tissue composition across taxa for the study of expression evolution, and use simulation studies to show that this issue can heavily bias inferences of selection. We close by highlighting the multi-dimensional nature of transcriptional variation and identifying major unanswered questions in disentangling how selection acts on the transcriptome.

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK. [2]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA. [3]Department of Biosciences, Durham University, Durham, UK. [4]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [5]Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, Sheffield, UK. [6]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [7]Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada. [8]Centre for Ecology and Conservation, University of Exeter, Penryn, UK. [9]These authors contributed equally: Christopher R. Cooney, Alison E. Wright. ✉e-mail: pprice3@sheffield.ac.uk; a.e.wright@sheffield.ac.uk

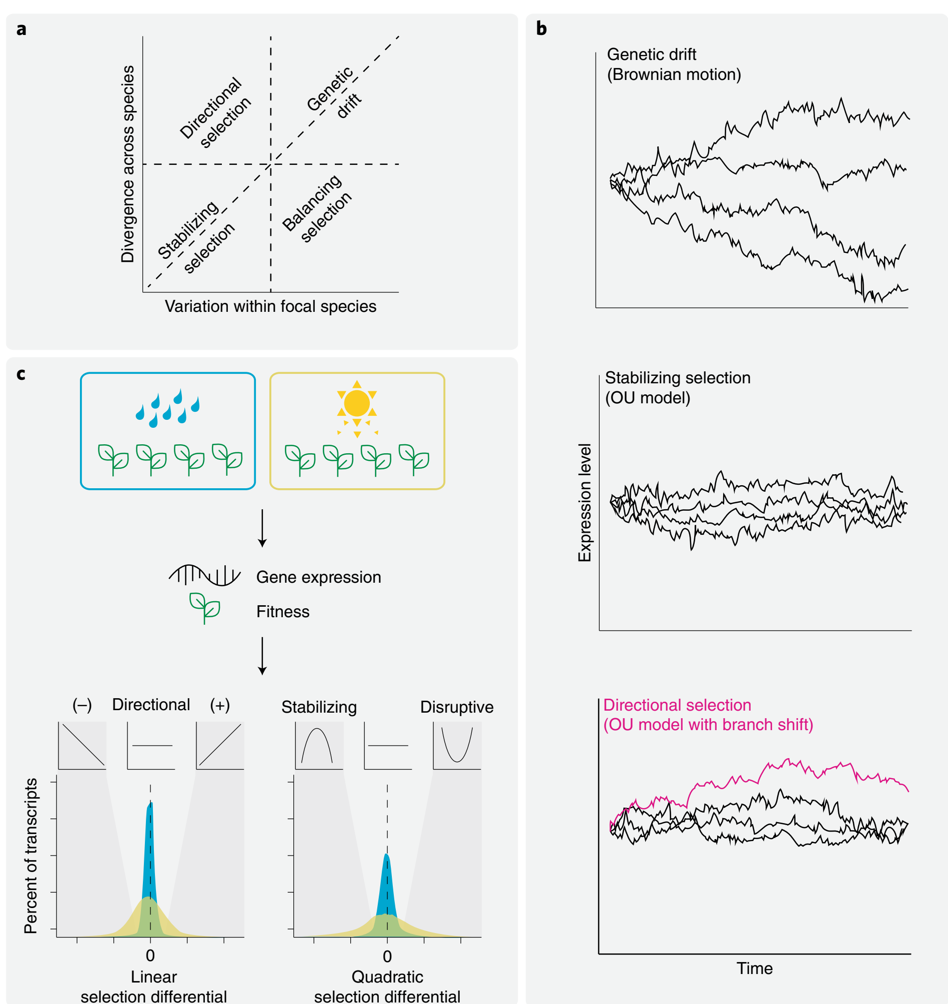


Fig. 1 | Approaches to detect selection on gene expression. a, Gene expression evolution has been inferred by contrasting levels of variation within a focal species to divergence across species in a pairwise framework. This principle is analogous to the Hudson Kreitman Aguadé (HKA) test used to detect selection at the DNA level. The neutral expectation is that divergence covaries linearly with intraspecific variance, at least over shorter evolutionary distances. Loci with the highest or lowest levels of intraspecific expression variation relative to neutrality are the best candidates for balancing or directional selection, respectively. Loci under stabilizing selection should exhibit limited biological variance and divergence. **b**, Phylogenetic comparative analyses enable comparisons across species to distinguish between evolutionary processes. Brownian motion (top) models neutral trait evolution via an unconstrained random walk. It assumes that divergence time between species (T) will describe the diversity across the phylogeny with only one parameter σ^2, the drift rate, and that variance at the tips of the phylogeny will equal $T\sigma^2$. The Ornstein–Uhlenbeck (OU) model (middle) assumes that gene regulation follows a stochastic process that is attracted towards a single optimum value, consistent with stabilizing selection. The additional parameters are therefore α (the strength of pull) and θ (the evolutionary optimum). This framework has been extended to test for branch-specific processes (bottom) by incorporating multiple optima to test for expression divergence in specific lineages (pink line). **c**, Phenotypic selection analyses have been applied to gene expression data to infer the mode and strength of selection. These employ multiple regression of relative fitness on multiple traits to calculate selection differentials that estimate total selection (direct and indirect) on gene expression. The covariance between fitness and expression is calculated to infer linear (S) and quadratic (C) selection differentials at each locus, which signify directional, stabilizing or disruptive selection. The linear selection differential estimates positive versus negative directional selection, while the quadratic selection differential estimates disruptive versus stabilizing selection. Groen et al.[67] used this approach to measure selection on gene expression in rice. Rice was grown under wet (blue) and dry (yellow) environmental conditions, and organism traits and fitness were measured. Panel **c** adapted with permission from ref. [67], Springer Nature Ltd.

Inferring the mode of gene expression evolution

Currently, several different approaches for analysing expression evolution have been proposed in the absence of a single consensus model. These can be divided into three broad categories: (1) contrasts between divergence and variation in expression (Fig. 1a), (2) phylogenetic comparative methods (Fig. 1b) and (3) fitness-based approaches (Fig. 1c). Importantly, each makes different assumptions regarding the mode of expression divergence and are subject to distinct biases. With a few exceptions[18,19,23,24], studies rarely interrogate multiple approaches, hence it remains unclear whether discrepancies between studies are biologically meaningful or are caused by inherent methodological differences. Below we synthesize results from different analytical frameworks to provide an overview on the debate concerning the importance of selection versus genetic drift in shaping divergence in gene expression levels.

Contrasting divergence and variation in expression. Many early analyses of expression evolution tested for selection by contrasting

expression divergence between species against diversity within species[5,25–28]. This method relies on the assumption that neutral changes are based solely on the underlying mutation rate[29,30], hence divergence between species relative to polymorphism within species will be equal at neutral loci[30]. When applied to expression data, mutation leads to polymorphism, which can be inferred through variation in expression level among individuals. Therefore, a neutral model of evolution can be rejected when there are deviations from a balanced ratio of within-to-between-species expression variation (Fig. 1a). Studies employing this approach are dominated by two competing viewpoints. One posits that gene expression is predominantly neutrally evolving[13,25,26,31] and the other suggests widespread conservation and purifying selection of expression levels[27,28,32,33], with evidence of positive selection acting on certain loci[34–39].

Analogous approaches using alternative neutral models of expression divergence have also found broad support for stabilizing selection[7,10]. One such approach uses mutation accumulation studies to estimate neutral expectations of expression divergence and infer selection through contrasts with natural populations[40–42]. Most recently, the distribution of expression levels of F2 offspring from a genetic cross has been used to estimate expected levels of neutral change[43]. Here, under neutrality, expression variance of the two parental populations should be equal to that of the F2 progeny as F2 expression levels result from random combinations of segregating alleles. Following this logic, directional selection can be inferred when parental divergence is markedly greater than the neutral expectation, and stabilizing selection can be inferred when expression of parental populations is significantly less diverged than expected. This study found widespread stabilizing selection on expression level across a range of species, the magnitude of which was dependent on the species' effective population size, consistent with population genetics theory that selection is more effective in species with larger effective population sizes. Selection has also been inferred through comparisons of additive genetic variance of expression (Q_{ST}) with sequence divergence in neutral molecular markers (F_{ST}) across populations[44]. However, while Q_{ST}:F_{ST} approaches have been successfully applied to gene expression variation in a few instances[45–49], accurately estimating the additive genetic basis of gene expression level can be challenging[50], and there is a tendency for dominance variance to bias Q_{ST} estimates, potentially leading to incorrect inferences of neutrality[44].

Nonetheless, the broad approach of contrasting inter- and intraspecific expression variation offers a tractable method to investigate selective forces shaping expression levels. However, one drawback is that these tests assume species or populations are phylogenetically independent and do not account for shared and often complex evolutionary histories. Therefore, in cases where more than one pair of species are compared, these methods can produce evolutionary patterns that are generated by the structure of the underlying phylogeny[51,52]. Furthermore, the neutral expectation that expression divergence equals diversity tends to break down over longer evolutionary time periods. This is because gene expression divergence cannot accumulate indefinitely due to upper limits on the rate of transcription. With increasing genetic distance, changes in expression among taxa may become nonlinear, leading to instances of genetic drift being mistaken as directional selection[13,19]. To test for selection across multiple species and evolutionary distances, approaches that take a phylogenetic perspective are required.

Phylogenetic comparative methods. Phylogenetic comparative methods have been widely adopted for several decades to infer selection acting on phenotypic traits[14,52–54]. By incorporating phylogenetic information, these methods account for shared ancestry and therefore can overcome issues of statistical non-independence. Recently, these approaches have been widely applied to transcriptome data to infer selection acting on gene expression by fitting several evolutionary models to expression data for a given gene[12,55,56] (Fig. 1b). A commonly used model, Brownian Motion (BM), assumes that expression divergence between species will be a function of divergence time and evolutionary rate (σ^2), and as such, is often seen as analogous to genetic drift. A second model, the Ornstein–Uhlenbeck (OU) model, adds an 'elastic band' element (α) drawing expression values towards an optimum (θ) across the phylogeny, akin to stabilizing selection[12,57]. The OU model can be extended to allow for branch-specific events, such as shifts in optimum trait values[12,58], analogous to directional selection in particular lineages.

Until now, comparative transcriptomic analyses have found overwhelming support for stabilizing selection on expression levels across a wide range of species, including *Drosophila*[12,59], African cichlids[60] and mammals[61]. While this appears consistent with past work[27,28,32,33], using OU models to infer selection has received repeated criticism within the phylogenetic comparative literature (Box 1). In essence, any factor that leads to a reduction in phylogenetic signal of species' trait values will favour the inference of an OU process over BM, regardless of the underlying evolutionary process. Importantly, failing to account for biological intraspecific variance or methodological measurement error by running these models on a mean species expression value has been shown to erode phylogenetic signal and lead to false inferences of stabilizing selection[15,16,18] or branch-specific selection[19]. These issues are particularly relevant to expression data, which can be noisy (that is, subject to a high degree of measurement error), particularly when environmental and developmental variance is not strictly controlled for. The OU framework has been adapted to specifically include within-species expression variability as an error term[18,58,62], and while it has been shown to reduce false inferences of stabilizing selection, this approach has only been employed by a handful of studies[24,63].

Recently, Rohlfs et al.[19] built on this approach with the Expression Variance and Evolution (EVE) model for testing expression evolution. This approach is grounded in the OU framework but incorporates contrasts of expression variance within versus between species, analogous to divergence–diversity ratio comparisons (Fig. 1a). This is a major advance as it accounts for evolutionary relationships between species and incorporates a neutral expectation for expression divergence that is dataset-specific. Interestingly, the few studies that have employed this approach have typically revealed a higher proportion of genes evolving under directional than under stabilizing selection[19,24], and evidence for elevated rates of expression evolution consistent with adaptive evolution[63–66]. This contrasts with past evidence for stabilizing selection, outlined above, and may reveal the inherent biases of simpler OU models. However, it should be noted that the studies that employed EVE were primarily focused on contrasts between stabilizing versus directional selection, not stabilizing versus neutral evolution, hence do not explicitly rule out neutral processes. Finally, EVE also relies on accurately estimating parameters of the OU process, so it is still probably subject to similar pitfalls identified by the phylogenetic comparative literature (Box 1).

Fitness-based approaches. Most recently, fitness-based approaches have been applied to study contemporary patterns of selection acting on gene expression[67,68]. One classical approach, which has been used to study a wide range of morphological traits, uses regression-based methods to estimate the strength of selection[69]. In this approach, the covariance between fitness and gene expression is calculated to infer selection differentials at each locus, which signify the mode of selection[67,68] (Fig. 1c). To reduce noise and computation time, as well as increase robustness of model prediction, expression data can be transformed to reduce dimensionality (that is, by principal component analysis) and selection gradients can then be obtained to estimate direct selection on suites of correlated

Box 1 | Common pitfalls of inferring selection using OU models

Recent work from the phylogenetic comparative methods field has revealed inherent biases in estimating OU processes, often leading to false inferences of stabilizing selection. As these have already been discussed elsewhere[15–17], we summarize the main pitfalls in relation to transcriptome studies.

Small phylogenetic samples. Recent work has shown that the ability to accurately estimate parameters of the OU model is strongly influenced by the number of species. Cooper et al.[15] simulated a range of phylogenies of varying size under Brownian motion and compared the fit of BM and OU models to test how often stabilizing selection was falsely inferred. They found a high type 1 error rate, especially when the number of sampled taxa was limited. For example, with a phylogeny of 25 species, stabilizing selection was falsely inferred ~10% of the time. This is especially concerning for transcriptomic studies, which are frequently composed of far fewer species due to sampling and computational costs and employ thousands of model comparisons to infer selection at each orthologous locus separately. We anticipate this concern will diminish as expression data become available for more species. However, even with phylogenies of 100 species, Cooper et al.[15] still estimate a type 1 error rate >0.05.

Measurement error. Here we use the term measurement error to broadly refer to any factor that adds noise to heritable expression values. This includes (1) data quality problems, such as RNA degradation, sequencing and assembly issues, (2) low sample sizes and (3) unwanted biological variance arising from the failure to control for environmental variation across samples. Measurement error across lineages can erode phylogenetic signal in the data, falsely biasing model selection away from BM models and towards OU processes and the inference of stabilizing selection[15,16]. Recent work has shown that even small amounts of measurement error can be problematic, particularly when the number of taxa sampled is small. For instance, Cooper et al.[15] estimate that with a phylogeny of 25 species and a 10% trait measurement error, stabilizing selection will be falsely concluded ~50% of the time. This is a particular concern for gene expression studies, as the environment can strongly influence gene regulation. Studies should endeavour to control environmental conditions so that differences in expression across samples reflect the heritable genetic component of expression, as has been discussed previously[8,116]. Second, it is clear that using a single mean expression value for each species can lead to spurious inferences of selection[18], making multiple replicates essential. Importantly, the OU framework has been extended to parameterize within-species variance as an error term[18,58,62] and this appears to be a promising approach. Finally, there are methods to control for technical problems that can introduce noise into measurements of expression, such as controlling for batch effects[117–119].

Complex patterns of trait evolution. Many phenotypic traits exhibit complex patterns of evolution and evolve at different rates across lineages[120]. While few studies have directly tested the tempo of expression change across species[60], it seems probable that gene expression does not evolve at a constant rate but instead shifts as mutation rate, selective pressures and pleiotropic constraints vary[47,121,122]. However, many evolutionary models, including BM and OU, assume a homogeneous process of trait change across lineages and/or through time. This is analogous to fitting a fixed d_N/d_S (the ratio of substitutions at non-synonymous and synonymous sites) across all branches when estimating selection on coding sequences. Recent work has shown that fitting single-process models masks complexity and leads to inaccurate inferences about the underlying evolutionary process[120]. Comparative methods that account for rate heterogeneity are available (discussed in ref. [120]), analogous to allowing d_N/d_S to vary across branches, but to our knowledge have not been widely applied in the context of gene expression evolution.

transcripts. Recent studies have used these principles to measure selection on gene expression in experimental contexts (for example, by quantifying flowering success and fecundity of rice grown in wet versus drought conditions[67]) and in natural settings (for example, by measuring parasite load and survivorship of wild trout using mark–recapture[68]). In contrast to comparative approaches, neither of these studies found strong support for stabilizing selection, and in one case, the dominant mode of selection was disruptive[68]. Causes of this discrepancy require further investigation, particularly whether this reflects methodological biases or difficulties in accurately estimating fitness. However, it is possible that selection pressures vary over short- versus long-term evolutionary time frames, and these approaches are capturing different snapshots of the evolutionary process. Furthermore, unlike these fitness-based approaches, phylogenetic comparative studies primarily rely on contrasting expression across highly conserved orthologous genes, often between very distantly related species, which probably biases our understanding of how gene expression evolves. Gene duplicates are probably key to the evolution of tissue-specific expression patterns[70], hence further work in this area might shed new light on how selection on gene expression varies across genes.

Decomposing transcriptional variation

Approaches designed to test for selection on gene expression all make the explicit assumption that differential expression is the direct result of regulatory change. However, in most cases, it is unclear whether this assumption is valid as processes other than regulatory evolution can generate apparent gene expression differences among taxa. For example, until now, studies have primarily used bulk sequencing approaches to measure expression across aggregate tissues or even entire body regions, which are often composed of many different cell types with variable expression profiles. In doing so, these ‘bulk’ expression values represent an average of expression across entire populations of distinct cell types. Here we use existing single-cell expression data (scRNA-seq) for the developing chicken hypothalamus[71] to illustrate this (Fig. 2a). The developing hypothalamus at Hamburger-Hamilton stage 10 is composed of three major cell types, where the FOXA1 cell type represents the greatest proportion of cells. Each cell type exhibits a distinct gene expression profile, but average expression estimated across all cells, analogous to a bulk RNA-seq approach for the whole hypothalamus, is not reflective of genuine variation in gene expression. The magnitude of this effect varies across genes, consistent with recent work in the mouse gonad[21] and primate heart tissue[22].

Within species, dramatic changes in tissue composition are well documented throughout development[71–73] and between the sexes[72]. This is exemplified by gonadal tissue, which exhibits sex-specific cell types[72] as well as a mix of haploid and diploid cells at various stages of differentiation[73–76]. Similarly, changes in cell type abundance between homologous tissues are common across species, particularly in the brain[77–79] and testes[21,80,81], the latter probably a result of varying levels of sperm competition and sexual selection. For instance, New World blackbird species under more intense sperm competition exhibit a greater proportion of sperm-producing tissue in the testes[80]. Importantly, this means that samples that vary in tissue composition can produce patterns of differential expression

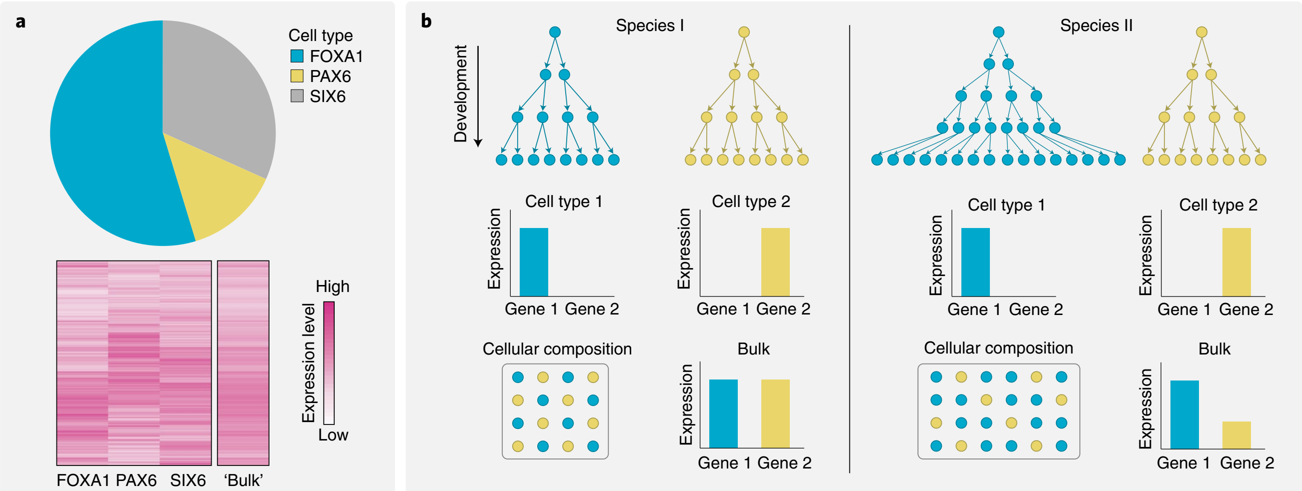


Fig. 2 | Variation in tissue composition can lead to the perception of differential expression. a,**b**, Schematic illustrating how variation in tissue composition can bias perception of expression measured from bulk RNA-seq within (**a**) and across species (**b**). **a**, The chicken hypothalamus is composed of 3 major cell types at developmental stage HH10. Pie chart (top) shows the proportion of cells in each cell type. Heatmap (bottom) shows gene expression measured across cells in each cell type and average 'bulk' expression estimated across all cells, equivalent to generating RNA-seq data from the whole tissue. Each cell type exhibits a distinct gene expression profile and bulk expression does not accurately reflect this. Data from ref. [71]. **b**, Illustration of how differences in tissue composition between species can lead to the false perception of differential expression. Here we illustrate a single tissue composed of two cell types, type 1 (blue) and type 2 (yellow), in two species. During development in species I (left), cell types 1 and 2 have the same rate of cell proliferation. The resulting tissue is evenly composed of each cell type. Cell type 1 only expresses gene 1 and cell type 2 only expresses gene 2. Bulk RNA-seq expression reflects single-cell expression. In species II (right), an increase in the rate of cellular proliferation for cell type 1 results in a greater proportion of cells of type 1 in the resulting adult tissue. Although there has been no change in per-cell expression of either gene 1 or 2, the relative expression from bulk RNA-seq of the entire tissue results in the perception of higher expression of gene 1 and lower expression of gene 2 compared with species I.

that are often mistaken as evidence of changes in gene regulation. Conversely, this approach can also dampen or mask genuine differences in expression within or between populations and species[20–22]. Of course, changes in tissue composition, which encompass both changes in cell type abundance within tissues and allometric scaling across them, are probably due to changes in gene expression across development. However, these changes in expression will not be detected if transcriptomes are measured after development is completed. Instead, the resulting differences in gene abundance will be mistaken as causative adaptive changes (Fig. 2b).

To our knowledge, only a handful of studies have directly accounted for the consequences of varying tissue allometry when studying modes of expression evolution[20–22,82,83]. Addressing this is a major priority for the field. Recent advances in single-cell transcriptomics permit direct comparisons of expression across homologous cell types in a comparative framework, overcoming issues of tissue composition variation. However, scRNA-seq currently presents its own set of challenges both in terms of expense and robustly identifying homologous cell types across species[84] (Box 2). Importantly, several fields, including cancer and developmental biology, have developed methods to deconvolve expression data from complex tissues, and these will probably be extremely valuable to evolutionary genomics studies. We urge future studies to carefully consider these points in project design (Box 2).

Challenges of inferring selection

While the implications of varying tissue allometry for measuring gene expression change across species have been discussed[20–22] (Fig. 2), the consequences of tissue composition on inferences of expression evolution have received less attention. Most studies that test for selection on the transcriptome use expression data generated from heterogeneous tissue, except for recent work that used cell sorting to isolate distinct cell types in mouse testes[66]. As discussed, there is a tendency for phylogenetic comparative methods to falsely infer stabilizing selection or more complex adaptive processes if non-evolutionary processes (such as measurement error) reduce phylogenetic signal. Perceived changes in expression that are driven by variation in tissue composition across species represent a prominent source of non-evolutionary expression variance and could therefore bias inferences of selection. This possibility has yet to be formally examined, hence using a series of simulated scenarios, we directly explore how shifts in tissue composition can bias the inference of evolutionary processes in a phylogenetic framework.

We simulated three distinct scenarios to explore how asymmetry in tissue composition across a phylogeny can drive false model inferences of expression evolution when applying comparative methods (Fig. 3). We imagine a simple situation where a tissue is composed of two distinct cell types. We estimate bulk expression values as a function of expression level in each cell type and their relative abundances in the tissue, and fitted a set of evolutionary models to this bulk expression.

First, we describe a scenario of extreme stabilizing selection on gene expression of a single locus. This locus is highly expressed in one cell type and lowly expressed in the other, but importantly, expression values are identical (that is, not evolving) across species. However, the relative abundance of each cell type is evolving under genetic drift and thus varies across species (Fig. 3a(i)). As predicted, the bulk expression value is neither reflective of single-cell expression levels nor consistent with extreme stabilizing selection (Fig. 3b(i)). A phylogenetic comparative approach consistently rejects a 'static' model of expression evolution and finds the greatest support for genetic drift as the dominant mode of evolution (Fig. 3c(i)). In this instance, the false positive rate is around ~86% relative to when these models are run on single-cell expression levels. This suggests that shifts in tissue composition can lead to false inferences of evolutionary processes acting on gene expression

Box 2 | Best practices for inferring selection in a comparative framework

Best practices for inferring selection on traits using comparative approaches have been discussed at length in the phylogenetic literature[15–17]. Briefly, to avoid false inferences of stabilizing selection (Box 1), studies should (1) strive to minimize measurement error, (2) maximize the number of species sampled and (3) use comparative approaches that parameterize within-species variance as an error term. Below, we discuss additional recommendations.

Validation of model fit. As discussed, many factors can bias model inference to conclude stabilizing selection over genetic drift. The best-fitting model is often chosen by comparing the relative fit of different models. However, studies rarely examine the absolute model fit[120]. This simple step, performed using existing methods such as ARBUTUS[123] or RevBayes[124], can be used to assess confidence in model selection. This approach relies on the process of posterior predictive simulations, in which datasets are simulated on the estimated parameters, and then a series of test statistics are run on the simulated data. Similarly, parametric bootstrapping approaches can be applied, resampling the data to generate a bootstrapped sampling distribution from which test statistics are calculated. These results can then be compared to the empirical data to assess the adequacy of the model. Using such approaches for model estimation has been shown to outperform maximum likelihood approaches in specific cases[125].

Multiple testing and false discovery rate (FDR). Comparative transcriptomics studies perform multiple statistical tests across thousands of genes, making them susceptible to the effects of multiple testing. Procedures including FDR and Bonferroni correction can easily manage this phenomenon[126,127], yet are frequently not included as standard in phylogenetic comparative transcriptomic approaches. Neutral simulations under predicted parameters permit the estimation of the FDR to account for the inflation of false positives and can be implemented in EVE[65].

Single-cell transcriptomics. By directly comparing gene expression levels across equivalent cell types, comparative single-cell transcriptomics (scRNA-seq) can circumvent problems arising when expression is measured from heterogeneous tissue (Fig. 2). However, scRNA-seq is more challenging for many non-model organisms, especially those sampled from the wild, as scRNA-seq performs optimally when single cells are isolated and processed immediately after harvesting tissue. Although tissue dissociation and storage techniques are being developed, such as methanol fixation and cryopreservation, there are concerns that these can either bias expression profiles[128] or lead to cell death. However, we anticipate that these challenges will be overcome as the field progresses and the costs of scRNA-seq decrease.

Consider tissue composition. We suggest that, where possible, studies should quantify cellular composition of the tissue in question and how this varies across species. For instance, if a single cell type dominates or expression level is dominated by one cell type, then our simulations suggest that the potential for bias is reduced. Importantly, if scRNA-seq data are available for the tissue, it is possible to use these to directly test for biases in cellular composition in bulk RNA-seq data[83,129]. Deconvolution methods, such as Decon2[130], BayesPrism[131] or ABIS[132], can be used to estimate cell type abundances and subsequently resolve expression profiles closer to those observed from purified cell subpopulations or scRNA-seq. Such methods have been widely implemented[22,133,134], and may prove valuable if they can be co-opted into evolutionary genomic studies. Finally, we urge the use of sampling techniques to directly isolate specific regions or cells of interest using microdissection or cell sorting to greatly reduce cell composition complications, as discussed by Hunnicutt et al.[21].

in the complete absence of any change in expression level within each cell type.

Second, we assume that gene expression is evolving under genetic drift. The two cell types are of equal abundance in all species with one exception in which a lineage-specific change in cellular composition occurs so that one cell type dominates (Fig. 3a(ii)). After model fitting, we find that this type of composition shift in one lineage leads to false inferences of a shift in gene expression, consistent with adaptive evolution (Fig. 3b(ii),c(ii)). The scale of this bias is highly dependent on the size of the allometric shift (Fig. 4a). Where the shift leads to a single cell type dominating, the actual mode of evolution (that is, genetic drift), will be rejected in ~35% of instances. While this extreme situation is arguably biologically unrealistic, our simulations show that even marginal shifts in relative cell type proportion result in elevated type 1 error rates. For example, across New World blackbirds, the proportion of seminiferous tissue in the testes ranges from 87% to 96%[80]. This equates to a shift in the proportion of ~9%. Although our simulations use different starting conditions, it is clear that shifts of a similar magnitude can result in increased type 1 errors.

Finally, we simulated a scenario where gene expression and cell type abundance both evolve under genetic drift (Fig. 3(iii)). Here we recover the true signal of genetic drift more reliably (Fig. 3c(iii)). However, in all instances so far, we have assumed that gene expression at a single locus is evolving independently in each cell type. While this is probably a reasonable assumption for some loci that have evolved tissue- or cell-specific regulatory machinery[85,86], expression changes are probably correlated in many instances. Interestingly, we find that this has implications for which evolutionary processes are inferred (Fig. 4b). When tissue composition evolves across the phylogeny, the type 1 error rate is highly dependent on the level of expression covariance between the cellular components of that tissue. In particular, if expression across cell types negatively covaries, where an increase in expression in one cell type is associated with a decrease in expression in another cell type at a single locus, the type 1 error rate can exceed 40% (Fig. 4b). The extent to which gene regulation is decoupled across cell types is, in and of itself, an interesting question. But here we have shown that gene expression covariation across cell types can also have profound implications for how we infer which selective processes are operating.

These scenarios demonstrate the potential challenges of inferring selection on expression level using data from heterogeneous tissues. It is also worth noting that our simulations are conservative as we do not model other non-evolutionary sources of variation (such as measurement error and tree topology error) that are probably common in transcriptome studies. We believe this highlights an urgent need to reappraise our current understanding of expression evolution in light of these underlying methodological issues. In particular, establishing (1) how often and by what magnitude changes in tissue composition occur and (2) the extent to which transcriptional variation is correlated across cell types are important factors to consider when studying expression evolution using phylogenetic comparative approaches with bulk RNA-seq. Unfortunately, we are not aware of a simple solution for correcting the biases we have uncovered, beyond recommending the use

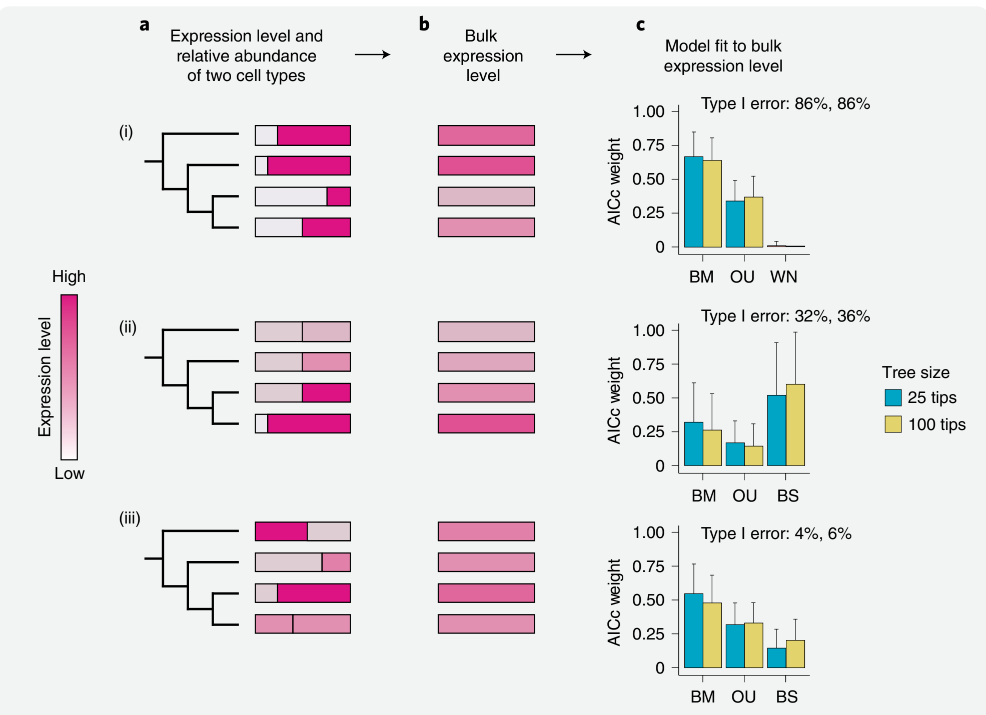


Fig. 3 | Inferring selection when expression level is measured from a heterogeneous tissue. a–c, Three scenarios illustrating potential pitfalls of inferring selection on gene expression level at a single locus using phylogenetic approaches when expression is measured from bulk sequencing. **a**, Expression level of a single gene in two different cell types across a phylogeny. High levels of expression are in dark pink and low expression in light pink. The relative proportion of each cell type is indicated by the size of the rectangle where cell type A is on the left and cell type B is on the right. **b**, Composite expression level of the gene as a function of cell type proportion and gene expression in each species. This would be analogous to measuring expression in bulk from a heterogeneous tissue. **c**, Results of simulated phylogenetic comparative analyses for each scenario with a phylogeny of 25 (blue) or 100 (yellow) tips on 1,000 unique trees. BM, Brownian motion; WN, white noise; OU, Ornstein–Uhlenbeck model; BS, OU model with a branch shift. These models were fitted on the simulated bulk expression values and the relative support for each model was calculated using Akaike information criterion (AIC) weights while accounting for varying sample size (AICc). Error bars show standard deviation around the mean across simulations. Type 1 error rates (for trees of 25 and 100 tips, respectively) for each scenario relative to when these models were fitted to expression at the single-cell level are shown. In scenario (i), expression values are static across the phylogeny for each cell type but cell type abundance is evolving under BM. However, phylogenetic approaches falsely infer that expression is evolving under BM. For (ii), expression in both cell types is evolving under BM, whereas tissue composition is stable across the phylogeny except for one tip that has undergone an allometric shift. Here, phylogenetic approaches falsely infer an adaptive shift in expression on a single branch. For (iii), gene expression in both cell types, as well as cell-type abundance, is evolving under BM. However, phylogenetic approaches increasingly falsely infer stabilizing selection on expression level.

of single-cell data to study expression evolution where possible. However, while single-cell approaches are increasingly available, the technical demands of this approach mean that they remain challenging for many species. In the meantime, we urge caution when using phylogenetic comparative approaches with bulk RNA-seq and recommend some steps to minimize other sources of error (Box 2).

Future directions

Given the importance of changes in gene expression to phenotypic divergence, studying transcriptome evolution is key to understanding adaptive change. As we discussed, we currently lack a consensus neutral model of transcriptome evolution and it is debatable whether we expect this to be universal across all loci due to the complex transcriptional architecture of many phenotypes. Here we argue that our understanding of the evolution of gene expression will permit critical advances as we continue to link insights across layers of the genotype-to-phenotype map, developmental contexts and evolutionary timescales, with organismal biology as our foundation. Below we identify major, unanswered questions in disentangling how selection acts on the entire transcriptome. We note that a complete understanding of how the transcriptome evolves also requires detailed knowledge of how regulatory elements combine to facilitate expression change and how selection acts on these non-coding regions[87,88], recently discussed elsewhere[7].

Transcriptional diversity and layers of gene regulation. Variation in splicing, whereby the same gene can express different RNA variants that produce distinct proteins or isoforms, is a common source of transcriptional variation across species[33,89–91], with important phenotypic effects (recently reviewed[92,93]). For genes with constraints on expression levels (for example, because of pleiotropic effects), alternative splicing may act as another adaptive mechanism of gene regulation[94]. Long-read sequencing methods have the advantage of producing full-length transcript sequences[95], which can be a more reliable way to identify alternatively spliced variants in transcriptomic datasets. Understanding the evolution of gene regulation will ultimately require an integrated understanding of how and when differences in expression level and splicing contribute to phenotypes under selection.

For transcriptional variation – whether in terms of expression level or alternative splicing – to be selected upon, it must contribute to variation at the protein layer of the genotype-to-phenotype

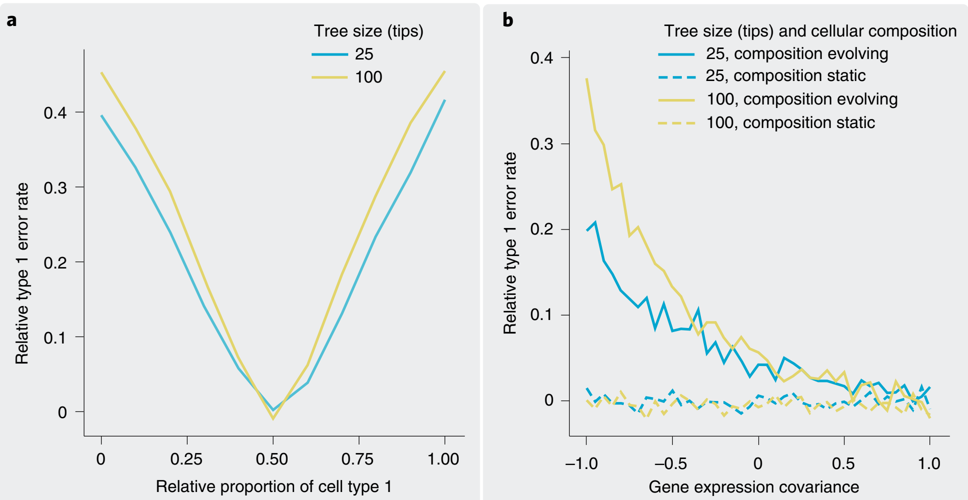


Fig. 4 | The magnitude of allometric shift and covariance of expression level biases the inference of selection. a, The probability that selection on gene expression is incorrectly inferred increases substantially with the magnitude of an allometric shift. This plot is a more detailed representation of Fig. 3(ii), where one species undergoes a shift in tissue composition, ranging from when a tissue is composed of two cell types at equal proportion to when only a single cell type is present. All other species have a tissue composition of 50:50 and expression is evolving under BM in each cell type. **b**, Covariance of expression between cell types biases inferences of selection. This plot is an extension of Fig. 3(iii). Expression is evolving under BM but cell type composition is either static (dashed lines) or also evolving under BM (solid lines). We varied the extent to which gene expression is correlated between cell types, ranging from negative covariance, where expression levels increase in one cell type while at the same time decreasing in the other cell type, to positive covariance, where expression levels decrease or increase in both cell types in a correlated manner. The relative type 1 error rate was calculated as the rate at which a BM model was not best fitted to the composite expression value relative to the equivalent error rate when models were fitted to single-cell simulations.

map. Due to difficulties in assaying proteins in comparison to RNA, the links between transcription and translation are underexplored, particularly in non-model organisms. Recent methodological advances that measure rates of protein synthesis to assay the translatome show that protein expression levels evolve under stronger evolutionary constraint than transcript levels[96], and report a higher correlation between the translatome and proteome than between the transcriptome and proteome[97]. However, this effect tends to decrease for functionally relevant loci, such as differentially expressed genes[98]. This indicates that in many cases, mRNA abundance does not fully capture transcriptional variation, and more work is needed to understand the complex relationship between transcription and translation (for example, mechanisms of buffering, feedback, degradation)[9,99].

Regulatory and co-expression networks. The intrinsically correlated nature of gene expression means that identifying selection at a single locus is hard to disentangle from the expression patterns at loci with shared architectures. To account for this, we must either take on network-based approaches and try to account for connectivity or covariance between loci, or we must reduce the dimensionality of our data. Furthermore, recent work identifying key nodes in gene regulatory networks of health and disease phenotypes between sexes also established that genes that appear architecturally central to a phenotype may also not appear differentially expressed[100]. Similarly, genome-wide association studies have revealed that complex phenotypes are often the product of many different loci where regulatory networks are probably highly interconnected and heritability is distributed across the entire genome[101–104]. Together, this means that studying expression on a locus-by-locus basis and not through inter-locus interactions may limit our ability to understand the transcriptional architectures underlying adaptive phenotypes, and how this impacts the mode and strength of selection on gene expression[101].

Developmental context. Phenotypic variation is produced by dynamic developmental changes through space and time. While gene regulation is highly context-dependent in terms of tissue identity and developmental stage[105–107], studies primarily test for expression evolution in a single snapshot, most often in adult tissues. Single-cell transcriptomic methods offer a promising path to better understand how these sources of variation interface with gene expression through development and inform models of gene expression evolution.

Genotype to phenotype to adaptation. If our goal is to uncover how gene regulation underlies adaptation, we must link transcriptional variation with organismal ecology and natural history. This effort is twofold, as it requires understanding when and how selection acts on organisms, and how transcriptional variation contributes to phenotypic responses to selection. Methods of surveying variation in gene expression offer increasing precision and resolution. However, our ability to identify the evolutionary processes causing this variation ultimately depends on our understanding of the organisms in question. Model systems such as yeast continue to enable high-throughput analyses that have yielded pivotal insights into the evolution of the transcriptome[7,108–111], but non-model systems also hold promise for studying how gene expression evolves under natural settings, which may yield novel and more ecologically relevant findings[68,112]. Furthermore, it remains to be seen how results from microevolutionary studies within or across a single generation integrate with those from macroevolutionary studies comparing diverged lineages, and the relative roles of stabilizing versus directional selection across these scales.

Methods

Single-cell transcriptomics. We analysed existing scRNA-seq data for the developing chicken hypothalamus[71]. Cell types expressing ‘PAX6’, ‘FOXA1’ or ‘SIX6’ at Hamburger-Hamilton (HH)10 were used in this study. Methods to

identify cell types and estimate expression levels have been previously published[71]. Pseudo-bulk datasets were generated at HH10 by calculating the average expression across cells in the three cell types.

Simulations. For the first scenario (Fig. 3a(i)), expression values were set at 1 and 2 in two cell types (A and B), respectively. The relative proportion of each cell type (pr) was simulated under BM for 1,000 unique trees of either 25 or 100 tips, using fastBM from phytools[113] in R v4.1.1. The resultant values were normalized between 0 and 1. Composite expression values for each tip (i) were calculated as follows:

$$\mathrm{exp}_i = (\mathrm{pr}_i \times 1) + ((1 - \mathrm{pr}_i) \times 2). \quad (1)$$

For the second scenario (Fig. 3a(ii)), expression values were evolved under BM over 1,000 unique phylogenies of 25 and 100 tips. The relative proportion of each cell type (pr) was set to 0.5 across the phylogeny, except for one randomly chosen tip (t). For this tip, the relative abundance of one cell type was shifted to an alternate value within the range 0 to 1 in 0.05 increments (Fig. 4a). The composite expression value of the shifted tip (t) was calculated as follows:

$$\mathrm{exp}_t = (\mathrm{pr}_t \times \mathrm{exp}_{\mathrm{A}t}) + ((1 - \mathrm{pr}_t) \times \mathrm{exp}_{\mathrm{B}t}). \quad (2)$$

Expression for the other tips (i) was calculated as above using equation (1).

For the third scenario (Fig. 3a(iii)), expression values were evolved under BM with varying covariances between cell types A and B, with covariance values varying from −1 to 1 in increments of 0.05. In all cases, σ^2 was set to 1.0001, and trees of 25 and 100 tips were examined. Simulations for each covariance value were run 1,000 times on unique trees. This scenario was run with both a fixed cell type proportion (pr), where pr = 0.5 in A and B at all tips, and with proportion values evolving under BM (Fig. 4b). The composite expression value at each tip (i) was calculated as follows:

$$\mathrm{exp}_i = (\mathrm{pr}_i \times \mathrm{exp}_{\mathrm{A}i}) + ((1 - \mathrm{pr}_i) \times \mathrm{exp}_{\mathrm{B}i}). \quad (3)$$

Fitting evolutionary models to composite expression levels. We fitted evolutionary models in R using phylolm[114] for scenario 1 (Fig. 3a(i)) and OUwie[115] for scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)). For the first scenario (Fig. 3a(i)), a static evolutionary model was rejected if the 95% bootstrapped confidence interval for σ^2 crossed 0. If rejected, a BM, an OU (Ornstein–Uhlenbeck) and a WN (White Noise) model were fitted and their Akaike weights calculated. The WN model was fitted by suppressing the phylogenetic signal by fixing Pagel's λ to 0. We calculated the type 1 error rate for scenario 1 (Fig. 3a(i)) as the rate at which a non-static model was accepted in favour of the static model, relative to when the same set of models was applied to a single-cell-type simulation. For scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)), we fitted a BM, an OU and an OU-shift model, where in the latter the optimum value of the trait was allowed to vary on a single tip. For scenario 2 (Fig. 3a(ii)), the OU-shift model was fitted so that the tip with the proportion shift was allowed the alternate optima, whereas for scenario 3 (Fig. 3a(iii)), a random tip was allocated. For scenarios 2 (Fig. 3a(ii)) and 3 (Fig. 3a(iii)), the type 1 error rate was calculated as the rate at which a non-BM model was favoured (that is, where ΔAICc > 2) relative to BM, relative to when the same set of models was applied to a single-cell-type simulation.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data have been previously published[71].

Code availability

All code is publicly available at https://github.com/Wright-lab-2021-Transcriptome-Evo/Inferring_expression_evolution_review.

Received: 2 November 2021; Accepted: 1 April 2022;
Published online: 12 May 2022

References

1. Mank, J. E. The transcriptional architecture of phenotypic dimorphism. *Nat. Ecol. Evol.* **1**, 6 (2017).
2. Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* **14**, 83–87 (2013).
3. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
4. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
5. Gilad, Y., Oshlack, A. & Rifkin, S. A. Natural selection on gene expression. *Trends Genet.* **22**, 456–461 (2006).
6. Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* **15**, 734–748 (2014).
7. Hill, M. S., Vande Zande, P. & Wittkopp, P. J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215 (2021).
8. Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505–516 (2012).
9. Signor, S. A. & Nuzhdin, S. V. The evolution of gene expression in *cis* and *trans*. *Trends Genet.* **34**, 532–544 (2018).
10. Fay, J. C. & Wittkopp, P. J. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199 (2008).
11. Khaitovich, P., Enard, W., Lachmann, M. & Pääbo, S. Evolution of primate gene expression. *Nat. Rev. Genet.* **7**, 693–702 (2006).
12. Bedford, T. & Hartl, D. L. Optimization of gene expression by natural selection. *Proc. Natl Acad. Sci. USA* **106**, 1133–1138 (2009).
13. Whitehead, A. & Crawford, D. L. Neutral and adaptive variation in gene expression. *Proc. Natl Acad. Sci. USA* **103**, 5425–5430 (2006).
14. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).
15. Cooper, N., Thomas, G. H., Venditti, C., Meade, A. & Freckleton, R. P. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc. Lond.* **118**, 64–77 (2016).
16. Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B. & Salamin, N. Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods Ecol. Evol.* **6**, 340–346 (2015).
17. Ho, L. S. T. & Ané, C. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* **5**, 1133–1146 (2014).
18. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. *Mol. Biol.* **31**, 201–211 (2014).
19. Rohlfs, R. V. & Nielsen, R. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst. Biol.* **64**, 695–708 (2015).
20. Montgomery, S. H. & Mank, J. E. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* **25**, 5114–5128 (2016).
21. Hunnicutt, K. E., Good, J. M. & Larson, E. L. Unraveling patterns of disrupted gene expression across a complex tissue. *Evolution* **76**, 275–291 (2021).
22. Fair, B. J. et al. Gene expression variability in human and chimpanzee populations share common determinants. *eLife* **9**, e59929 (2020).
23. Nourmohammad, A. et al. Adaptive evolution of gene expression in *Drosophila*. *Cell Rep.* **20**, 1385–1395 (2017).
24. Catalán, A., Briscoe, A. D. & Höhna, S. Drift and directional selection are the evolutionary forces driving gene expression divergence in eye and brain tissue of *Heliconius* butterflies. *Genetics* **213**, 581–594 (2019).
25. Oleksiak, M. F., Churchill, G. A. & Crawford, D. L. Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266 (2002).
26. Khaitovich, P. et al. A neutral model of transcriptome evolution. *PLoS Biol.* **2**, e132 (2004).
27. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144 (2003).
28. Lemos, B., Meiklejohn, C. D., Cáceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137 (2005).
29. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
30. Kimura, M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res* **11**, 247–270 (1968).
31. Staubach, F., Teschke, M., Voolstra, C. R., Wolf, J. B. W. & Tautz, D. A test of the neutral model of expression change in natural populations of house mouse subspecies. *Evolution* **64**, 549–560 (2010).
32. Somel, M. et al. Transcriptional neoteny in the human brain. *Proc. Natl Acad. Sci. USA* **106**, 5743–5748 (2009).
33. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
34. Moghadam, H. K., Pointer, M. A., Wright, A. E., Berlin, S. & Mank, J. E. W chromosome expression responds to female-specific selection. *Proc. Natl Acad. Sci. USA* **109**, 8207–8211 (2012).
35. Gilad, Y., Oshlack, A., Smyth, G. K., Speed, T. P. & White, K. P. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242–245 (2006).
36. Enard, W. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).

37. Blekhman, R., Oshlack, A., Chabot, A. E., Smyth, G. K. & Gilad, Y. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genet.* **4**, e1000271 (2008).
38. Warnefors, M. & Eyre-Walker, A. A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLoS ONE* **7**, e34935 (2012).
39. Ometto, L., Shoemaker, D., Ross, K. G. & Keller, L. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**, 1381–1392 (2011).
40. Rifkin, S. A., Houle, D., Kim, J. & White, K. P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
41. Denver, D. R. et al. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548 (2005).
42. Huang, W. et al. Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *eLife* **5**, e14625 (2016).
43. Fraser, H. B. Detecting selection with a genetic cross. *Proc. Natl Acad. Sci. USA* **117**, 22323–22330 (2020).
44. Leinonen, T., McCairns, R. J. S., O'Hara, R. B. & Merilä, J. Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* **14**, 179–190 (2013).
45. Mähler, N. et al. Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet.* **13**, e1006402 (2017).
46. Kohn, M. H., Shapiro, J. & Wu, C.-I. Decoupled differentiation of gene expression and coding sequence among *Drosophila* populations. *Genes Genet. Syst.* **83**, 265–273 (2008).
47. Papakostas, S. et al. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat. Commun.* **5**, 4071 (2014).
48. Leder, E. H. et al. The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. *Mol. Biol. Evol.* **32**, 674–689 (2015).
49. Blanc, J., Kremling, K. A. G., Buckler, E. & Josephs, E. B. Local adaptation contributes to gene expression divergence in maize. *G3* **11**, jkab004 (2021).
50. Pujol, B., Wilson, A. J., Ross, R. I. C. & Pannell, J. R. Are QST-FST comparisons for natural populations meaningful? *Mol. Ecol.* **17**, 4782–4785 (2008).
51. Dunn, C. W., Zapata, F., Munro, C., Siebert, S. & Hejnol, A. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proc. Natl Acad. Sci. USA* **115**, e409–e417 (2018).
52. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
53. Pennell, M. W. & Harmon, L. J. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* **1289**, 90–105 (2013).
54. Felsenstein, J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471–492 (1973).
55. Oakley, T. H., Gu, Z., Abouheif, E., Patel, N. H. & Li, W.-H. Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data. *Mol. Biol. Evol.* **22**, 40–50 (2005).
56. Gu, X. Statistical framework for phylogenomic analysis of gene family expression profiles. *Genetics* **167**, 531–542 (2004).
57. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
58. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
59. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
60. El Taher, A. et al. Gene expression dynamics during rapid organismal diversification in African cichlid fishes. *Nat. Ecol. Evol.* **5**, 243–250 (2021).
61. Chen, J. et al. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res.* **29**, 53–63 (2019).
62. Pal, S., Oliver, B. & Przytycka, T. M. Modeling gene expression evolution with EvoGeneX uncovers differences in evolution of species, organs and sexes. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.06.895615 (2021).
63. Greenway, R. et al. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proc. Natl Acad. Sci. USA* **117**, 16424–16430 (2020).
64. Vegesna, R. et al. Ampliconic genes on the great ape Y chromosomes: rapid evolution of copy number but conservation of expression levels. *Genome Biol. Evol.* **12**, 842–859 (2020).
65. Gillard, G. B. et al. Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
66. Kopania, E. E. K., Larson, E. L., Callahan, C. & Keeble, S. Molecular evolution across mouse spermatogenesis. *Mol. Biol. Evol.* **39**, msac023 (2022).
67. Groen, S. C. et al. The strength and pattern of natural selection on gene expression in rice. *Nature* **578**, 572–576 (2020).
68. Ahmad, F. et al. The strength and form of natural selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2724–2737 (2021).
69. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
70. Guschanski, K., Warnefors, M. & Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).
71. Kim, D. W. et al. Single-cell analysis of early chick hypothalamic development reveals that hypothalamic cells are induced from prethalamic-like progenitors. *Cell Rep.* **38**, 110251 (2022).
72. Estermann, M. A. et al. Insights into gonadal sex differentiation provided by single-cell transcriptomics in the chicken embryo. *Cell Rep.* **31**, 107491 (2020).
73. Niu, W. & Spradling, A. C. Two distinct pathways of pregranulosa cell differentiation support follicle formation in the mouse ovary. *Proc. Natl Acad. Sci. USA* **117**, 20015–20026 (2020).
74. Witt, E., Benjamin, S., Svetec, N. & Zhao, L. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. *eLife* **8**, e47138 (2019).
75. Hermann, B. P. et al. The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep.* **25**, 1650–1667.e8 (2018).
76. Green, C. D. et al. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev. Cell* **46**, 651–667.e10 (2018).
77. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016).
78. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* **360**, 881–888 (2018).
79. Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
80. Lüpold, S., Linz, G. M., Rivers, J. W., Westneat, D. F. & Birkhead, T. R. Sperm competition selects beyond relative testes size in birds. *Evolution* **63**, 391–402 (2009).
81. Shami, A. N. et al. Single-cell RNA sequencing of human, macaque, and mouse testes uncovers conserved and divergent features of mammalian spermatogenesis. *Dev. Cell* **54**, 529–547.e12 (2020).
82. Harrison, P. W. et al. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl Acad. Sci. USA* **112**, 4393–4398 (2015).
83. Bauernfeind, A. L. et al. Tempo and mode of gene expression evolution in the brain across primates. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.21.440670 (2021).
84. Shafer, M. E. R. Cross-species analysis of single-cell transcriptomic data. *Front. Cell Dev. Biol.* **7**, 175 (2019).
85. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481–487 (2005).
86. Prud'homme, B. et al. Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* **440**, 1050–1053 (2006).
87. Liu, J., Mosti, F. & Silver, D. L. Human brain evolution: emerging roles for regulatory DNA and RNA. *Curr. Opin. Neurobiol.* **71**, 170–177 (2021).
88. Sarropoulos, I. et al. Developmental and evolutionary dynamics of *cis*-regulatory elements in mouse cerebellar cells. *Science* **373**, eabg4696 (2021).
89. Brown, J. B. et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).
90. Gibilisco, L., Zhou, Q., Mahajan, S. & Bachtrog, D. Alternative splicing within and between *Drosophila* species, sexes, tissues, and developmental stages. *PLoS Genet.* **12**, e1006464 (2016).
91. Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat. Genet.* **53**, 925–934 (2021).
92. Gómez-Redondo, I., Planells, B., Navarrete, P. & Gutiérrez-Adán, A. Role of alternative splicing in sex determination in vertebrates. *Sex. Dev.* **15**, 381–391 (2021).
93. Singh, P. & Ahi, E. P. The importance of alternative splicing in adaptive evolution. *Mol. Ecol.* **31**, 1928–1938 (2022).
94. Rogers, T. F., Palmer, D. H. & Wright, A. E. Sex-specific selection drives the evolution of alternative splicing in birds. *Mol. Biol. Evol.* **38**, 519–530 (2021).
95. Naftaly, A. S., Pau, S. & White, M. A. Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. *Genome Res.* **31**, 1486–1497 (2021).
96. Khan, Z. et al. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).
97. Wang, Z.-Y. et al. Transcriptome and translatome co-evolution in mammals. *Nature* **588**, 642–647 (2020).
98. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Smith, V. A. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci. Rep.* **5**, 10775 (2015).
99. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
100. Lopes-Ramos, C. M. et al. Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep.* **31**, 107795 (2020).

101. Liu, X., Li, Y. I. & Pritchard, J. K. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034.e6 (2019).
102. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
103. Mathieson, I. The omnigenic model and polygenic prediction of complex traits. *Am. J. Hum. Genet.* **108**, 1558–1563 (2021).
104. O'Connor, L. J. et al. Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* **105**, 456–476 (2019).
105. Pantalacci, S. & Sémon, M. Transcriptomics of developing embryos and organs: a raising tool for evo-devo. *J. Exp. Zool. B* **324**, 363–371 (2015).
106. Liu, J. & Robinson-Rechavi, M. Developmental constraints on genome evolution in four bilaterian model species. *Genome Biol. Evol.* **10**, 2266–2277 (2018).
107. Cardoso-Moreira, M. et al. Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
108. Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F. & Wittkopp, P. J. Selection on noise constrains variation in a eukaryotic promoter. *Nature* **521**, 344–347 (2015).
109. Metzger, B. P. H. et al. Contrasting frequencies and effects of *cis*- and *trans*-regulatory mutations affecting gene expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).
110. Hodgins-Davis, A., Duveau, F., Walker, E. & Wittkopp, P. J. Empirical measures of mutational effects define neutral models of regulatory evolution in *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci. USA* **116**, 21085–21093 (2019).
111. Vaishnav, E. D. et al. A comprehensive fitness landscape model reveals the evolutionary history and future evolvability of eukaryotic *cis*-regulatory DNA sequences. Preprint at *bioRxiv* https://doi.org/10.1101/2021.02.17.430503 (2021).
112. Josephson, M. P. & Bull, J. K. Innovative mark–recapture experiment shows patterns of selection on transcript abundance in the wild. *Mol. Ecol.* **30**, 2707–2709 (2021).
113. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2011).
114. Ho, L., Si, T. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
115. Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B. C. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
116. Harrison, P. W., Wright, A. E. & Mank, J. E. The evolution of gene expression and the transcriptome–phenotype relationship. *Semin. Cell Dev. Biol.* **23**, 222–229 (2012).
117. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
118. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
119. Jacob, L., Gagnon-Bartsch, J. A. & Speed, T. P. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics* **17**, 16–28 (2016).
120. Chira, A. M. & Thomas, G. H. The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *J. Evol. Biol.* **29**, 2502–2518 (2016).
121. Allen, S. L., Bonduriansky, R. & Chenoweth, S. F. Genetic constraints on microevolutionary divergence of sex-biased gene expression. *Phil. Trans. R. Soc. B* **373**, 20170427 (2018).
122. Dean, R. & Mank, J. E. Tissue specificity and sex-specific regulatory variation permit the evolution of sex-biased gene expression. *Am. Nat.* **188**, e74–e84 (2016).
123. Pennell, M. W., FitzJohn, R. G., Cornwell, W. K. & Harmon, L. J. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* **186**, e33–e50 (2015).
124. Höhna, S. et al. Probabilistic graphical model representation in phylogenetics. *Syst. Biol.* **63**, 753–771 (2014).
125. Slater, G. J. & Pennell, M. W. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Syst. Biol.* **63**, 293–308 (2014).
126. Barr, W. A. & Scott, R. S. Phylogenetic comparative methods complement discriminant function analysis in ecomorphology. *Am. J. Phys. Anthropol.* **153**, 663–674 (2014).
127. Brzyski, D. et al. Controlling the rate of GWAS false discoveries. *Genetics* **205**, 61–75 (2017).
128. Wang, X., Yu, L. & Wu, A. R. The effect of methanol fixation on single-cell RNA sequencing data. *BMC Genomics* **22**, 420 (2021).
129. Jew, B. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
130. Aguirre-Gamboa, R. et al. Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* **21**, 243 (2020).
131. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Bayesian cell-type deconvolution and gene expression inference reveals tumor-microenvironment interactions. Preprint at *bioRxiv* https://doi.org/10.1101/2020.01.07.897900 (2021)
132. Monaco, G. et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640.e7 (2019).
133. Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).
134. Shen-Orr, S. S. et al. Cell type-specific gene expression differences in complex tissues. *Nat. Methods* **7**, 287–289 (2010).

Acknowledgements

This work was funded by an NERC Independent Research Fellowship to C.R.C. (NE/T01105X/1); an NERC Independent Research Fellowship to A.E.W. (NE/N013948/1); a grant from the European Research Council (grant agreement 680951) and a Canada 150 Research Chair to J.E.M.; an NERC ACCE DTP to P.D.P.; and an NSF Postdoctoral Research Fellowship and an MSU Presidential Postdoctoral Fellowship to D.H.P.D. We thank M. Placzek, P. E. Pifarré, E. Josephs, A. Platts, M. Roberts, R. Panko, M. Wilson Brown and S. Buysse for helpful comments and suggestions on the manuscript.

Author contributions

A.E.W., C.R.C., D.H.P.D., P.D.P. and J.E.M. designed the study. D.W.K., E.S.P., A.E.W., C.R.C. and P.D.P. analysed the data. All authors wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01761-8.

Correspondence and requests for materials should be addressed to Peter D. Price or Alison E. Wright.

Peer review information *Nature Ecology & Evolution* thanks Camille Berthelot and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Chapter 4. Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches

Peter D Price[1]*, Jake Pepper[1]*, Thea F Rogers[2], Alison E Wright[1] & Jon Slate[1]

* joint contribution

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom

[2]Department of Neuroscience and Developmental Biology, Division of Molecular Evolution and Development, University of Vienna, Vienna, Austria

Key words: inversion, supergene, balancing selection, zebra finch, sperm competition

4.1 SIGNIFICANCE STATEMENT

Inversion polymorphisms are often found associated with complex traits, acting as supergenes, however, the exact mode of selection that acts to maintain them is highly varied. We demonstrate that a sex-linked supergene in the zebra finch, a model for sperm competition and inversion polymorphisms, is likely maintained via associative overdominance, a process where inversion heterozygosity masks the fitness costs seen in homozygotes. This result illustrates how inversion polymorphisms can be maintained, specifically as an outcome of a reduced capacity to respond to selection, and the molecular costs of reduced recombination

4.2 ABSTRACT

Recent years have seen an explosion in examples of supergenes, where recombination is suppressed between haplotypes, often via inversion polymorphisms, to control complex traits. However, an enduring problem in evolutionary biology is understanding how these inversion polymorphisms are maintained, often for long periods of time. Here, we test alternate theories for the role of balancing selection by exploring the molecular evolution and expression of a large, sex-linked inversion polymorphism in a passerine bird, the zebra finch (*Taeniopygia guttata*) responsible for most of the genetic variation in sperm morphology. We find evidence for degeneration within the inversion characterised by relaxed purifying selection and an excess of nonsynonymous differences which is indicative of associative overdominance protecting heterozygotes from deleterious recessive alleles. Furthermore, we find that the transcriptome acts to mitigate the negative consequences of relaxed purifying selection through the more equal expression of those most highly diverged, and potentially degenerated, alleles.

4.3 INTRODUCTION

Complex phenotypes are often associated with regions of high linkage disequilibrium (LD). This segregation of co-adapted alleles allows them to be treated as a single Mendelian unit, otherwise known as a supergene (Charlesworth and Charlesworth 1975; Schwander et al. 2014; Thompson and Jiggins 2014; Charlesworth 2016). Supergenes are present across eukaryotic life, with prominent examples in plants (Lowry and Willis 2010; Li et al. 2016; Shore et al. 2019; Todesco et al. 2020; Gutiérrez-Valencia et al. 2022), invertebrates (Joron et al. 2006; Wang et al. 2013; Kunte et al. 2014; Nosil et al. 2018; Li et al. 2020; Koch et al. 2021), fish (Pearse et al. 2019; Meyer et al. 2024), mammals (Harringmeyer and Hoekstra 2022), and birds (Thomas et al. 2008; Huynh et al. 2011; Küpper et al. 2016; Lamichhaney et al. 2016; Tuttle et al. 2016). An enduring problem in evolutionary biology is understanding how supergenes are maintained (Booker et al. 2015; Charlesworth 2016; Berdan et al. 2022), in some cases for millions of years (Hill et al. 2023). However, until recently, most work on loci responsible for complex traits has focused on describing their function or on the phenotypes they affect, rather than their genetic architecture or the molecular evolutionary processes acting on them. Work has now begun to explore supergene sequence evolution in a handful of species, such as the white throated sparrow (Tuttle et al. 2016; Maney et al. 2020; Jeong et al. 2022), ruff (McGrath 2023), redpoll (Funk et al. 2021), fire ant (Pracana et al. 2017; Martinez-Ruiz et al. 2020), primrose (Potente et al. 2022), wild flaxseed (Gutiérrez-Valencia et al. 2022), water snowflake (Yang et al. 2023), Atlantic cod (Matschiner et al. 2022) and smut fungus (Carpentier et al. 2022), yielding insight into their evolutionary history and patterns of selection (Berdan et al. 2023). In many, but not all of these examples, supergenes are housed by inversion polymorphisms causing linked loci to be inherited as single non-recombining loci (Wellenreuther and Bernatchez 2018).

The picture emerging from these recent studies is that multiple, often non-exclusive processes are responsible for the maintenance of inversions, and the supergenes that are housed by them (Berdan et al. 2023). It is generally assumed that if beneficial dominant mutations arise on a given haplotype, they should rapidly go to fixation. However, balancing selection via overdominance, negative frequency dependent selection, or spatial and/or temporal variance in selection can facilitate the maintenance of alleles at intermediate frequencies (Charlesworth 1974; Berdan et al. 2023). Similarly, admixture of populations with an independent rearrangement may also enable a fitter heterozygous genotype of dominant beneficial mutations. These processes can effectively function to maintain inversion polymorphisms and

have been proposed to explain the maintenance of supergenes in a number of species, including seaweed flies (Mérot et al. 2020) and *Drosophila melanogaster* (Kapun et al. 2023).

However, an old but alternative possible explanation for the maintenance of inversion polymorphisms has recently received growing attention (Sturtevant and Mather 1938; Dobzhansky 1950; Zhang et al. 2017; Faria et al. 2019; Berdan et al. 2022). Although supergenes can facilitate adaptive evolution, recombination suppression does not come without costs (Smith 1978). When recombination is absent or suppressed, as in heterozygotes for inversion polymorphisms, the efficacy of purifying selection will be strongly reduced, especially on a haplotype that is rare. As a result, a process similar to a Muller's Ratchet (Muller 1964) begins where deleterious recessive alleles can accumulate on each haplotype, resulting in the degeneration of the genes within the haplotype and a reduction in fitness of homozygous individuals (Berdan et al. 2021; Jay et al. 2021; Berdan et al. 2022). This effect is particularly pronounced in populations with small effective population sizes and limited migration. Under this scenario, the expression of deleterious recessive alleles is avoided in heterozygotes, and so the polymorphism may be maintained through balancing selection via associative overdominance. Associative overdominance is thought to act on supergenes in fire ants (Pracana et al. 2017), *Heliconius* butterflies (Jay et al. 2021) and possibly white throated sparrows (Jeong et al. 2022) and seahorses (Meyer et al. 2024). The process can be initiated from the emergence of a novel inversion polymorphism if the inversion has a low mutational load compared to the ancestral form (Nei et al. 1967), followed by heterozygous advantage from beneficial dominant mutations or the hiding of recessive deleterious alleles. At its most extreme, this can ultimately lead to a balanced lethal system (Berdan et al. 2022) where only heterozygotes are viable.

Both overdominance and associative overdominance can maintain inversion polymorphisms as they result in heterozygotes being fitter than homozygotes, but overdominance is driven by heterozygote advantage at a single locus, while associative overdominance is caused by a relaxation of purifying selection such that deleterious mutations that arise on both haplotypes are masked in heterozygous individuals. However, given that the molecular evolution of inversion polymorphisms, and the supergenes they house, has only been tested in a handful of species, it remains unclear if one mechanism dominates or both are equally as likely (they can both be operating in the same inversion polymorphism). Furthermore, even fewer studies have tested whether similar evolutionary forces govern the expression of genes within supergenes and how differential expression of each haplotype might mitigate the

consequences of relaxed purifying selection arising from reduced recombination (Martinez-Ruiz et al. 2020; Gutiérrez-Valencia et al. 2022; Jeong et al. 2022).

Notably, inversion polymorphisms and their associated supergenes are often sex-linked (Kim et al. 2017; Baird et al. 2023) and this can alter the evolutionary dynamics acting on them. Firstly, they are predicted to accumulate alleles with sex-specific benefits (Rice 1984). In these cases, sex-beneficial alleles that arise on a specific haplotype will increase in frequency on that haplotype, but cannot recombine onto the alternative haplotype. For instance, in ZZ/ZW (male/female) systems, dominant male-benefit alleles are predicted to accumulate on the Z chromosome due to its male-biased inheritance pattern (and the opposite is expected for the X chromosome in XX/XY female/male systems) (Rice 1984). Thus, over time, each Z-linked haplotype will accumulate more beneficial alleles, but only heterozygous males reap the benefit of carrying all of those beneficial alleles, resulting in the maintenance of the inversion polymorphism. However, sex chromosomes frequently have lower effective population sizes and reduced recombination relative to the autosomes, which can lead to relaxed purifying selection (Charlesworth et al. 1987; Vicoso and Charlesworth 2009; Meisel and Connallon 2013). It remains unclear how this unique evolutionary environment intersects with the selective forces acting to maintain inversion polymorphisms across the genome.

In this paper, we test theories for how balancing selection, either via dominance or associative overdominance, maintains inversion polymorphisms through heterozygote advantage. Specifically, we explore the molecular evolution and gene expression of a large, Z-linked inversion polymorphism in a passerine bird, the zebra finch (*Taeniopygia guttata*). The Z chromosome harbours a large pericentric inversion polymorphism (Itoh and Arnold 2011; Knief et al. 2016), which encompasses the majority of the Z chromosome, and completely suppresses crossing over in heterozygotes besides the tips. There are at least three distinct inversion haplotypes (A, B & C), which are found in populations that diverged 1.2 to 2.8 million years ago (Balakrishnan and Edwards 2009; Knief et al. 2016), suggesting the polymorphism is not recent. Despite representing just 7% of the genome, these karyotypes are responsible for 67-90% of the additive genetic variance in sperm morphology within male finches in both wild and captive populations (Kim et al. 2017; Knief et al. 2017). The sperm of heterozygous males, specifically AB and AC, have the greatest motility, determined by an intermediate midpiece-to-tail ratio relative to homozygotes (Bennison et al. 2016; Kim et al. 2017; Knief et al. 2017). Although the zebra finch produces monogamous breeding pairs, they still sparingly partake in extra-pair copulations (Birkhead et al. 1988), with extra-pair paternity at around 2.4% in wild populations (Birkhead et al. 1990). Although much lower than most other

songbirds, this infrequent sperm competition appears enough to maintain a balanced polymorphism through heterozygote advantage (Knief et al. 2017). Additionally, and regardless of low extra-pair paternity, and so limited opportunity for sperm competition, heterozygote karyotypes also show greater fertility rates (Knief et al. 2017). Importantly, attempts to find associations between haplotype and other life history or morphological traits have not identified any convincing effects of the polymorphism (Knief et al. 2016; Assersohn et al. 2023), despite the large number of genes within it.

4.4 RESULTS AND DISCUSSION

Using a high-density genotyping dataset that we had previously generated for the zebra finch (Kim et al. 2017) (Table S4.1), we identified 2,123 SNPs to karyotype the Z-linked inversion and distinguish three haplotypes, A, B and C. We karyotyped 24 individuals for which whole genome sequencing data was available using these diagnostic SNPs (Singhal et al 2015) (Table S4.2). In total, we scored 8 birds as inversion haplotype A or AA (4 females, 4 males), 7 birds were scored as inversion haplotype B or BB (6 females, 1 male), 1 female bird as haplotype C, 4 male birds as AB heterozygotes, and another 4 males as AC heterozygotes (Figure S4.1). Additionally, we generated a separate gene expression dataset with another 10 birds, sequenced using Illumina paired-end RNA-seq and genotyped, to identify three female B birds, two female A birds, three AB heterozygous males, and two AA homozygous males (Table S4.3).

4.4.1 Characterising the structure of the Z-linked inversion

To first characterise patterns of Z chromosome diversity, we calculated π for silent sites, using a combination of intronic and intergenic sites, across the bTG1.4 reference genome for the 15 AA/A and BB/B individuals (Singhal et al., 2015), and compared patterns of diversity within and between haplotypes. Previous work showed that the inversion regions encompass the majority of the ~75Mb Z chromosome, apart from approximately 5Mb at either end (Knief et al 2016, Knief et al 2017, Kim et al 2017). These end regions outside of the inversion can act as a useful control to test for deviation from patterns of Z chromosome evolution under neutrality. This is because they are small and so experience elevated rates of recombination, therefore, we do not expect them to experience a pronounced Faster-Z effect. As expected, we found that nucleotide diversity was markedly higher at the ends of the Z (0-6.50Mb and 70.10-75.39Mb) (Figure 4.1a) - more than triple the mean across the intervening region and similar to the genome-wide average (Table 4.1). Since reduced recombination of the inversion

polymorphism is expected to reduce genetic diversity (Charlesworth 2023) when the region is not under neutrality, these boundaries are likely the limits of the inversion. We also found that mean d_{XY} and F_{ST} between the A and B haplotype at the ends of the Z chromosome in the regions 0Mb-6.50Mb and 70.10Mb and above were consistent with autosomal values (Table 4.1, Figure 4.1a & b).

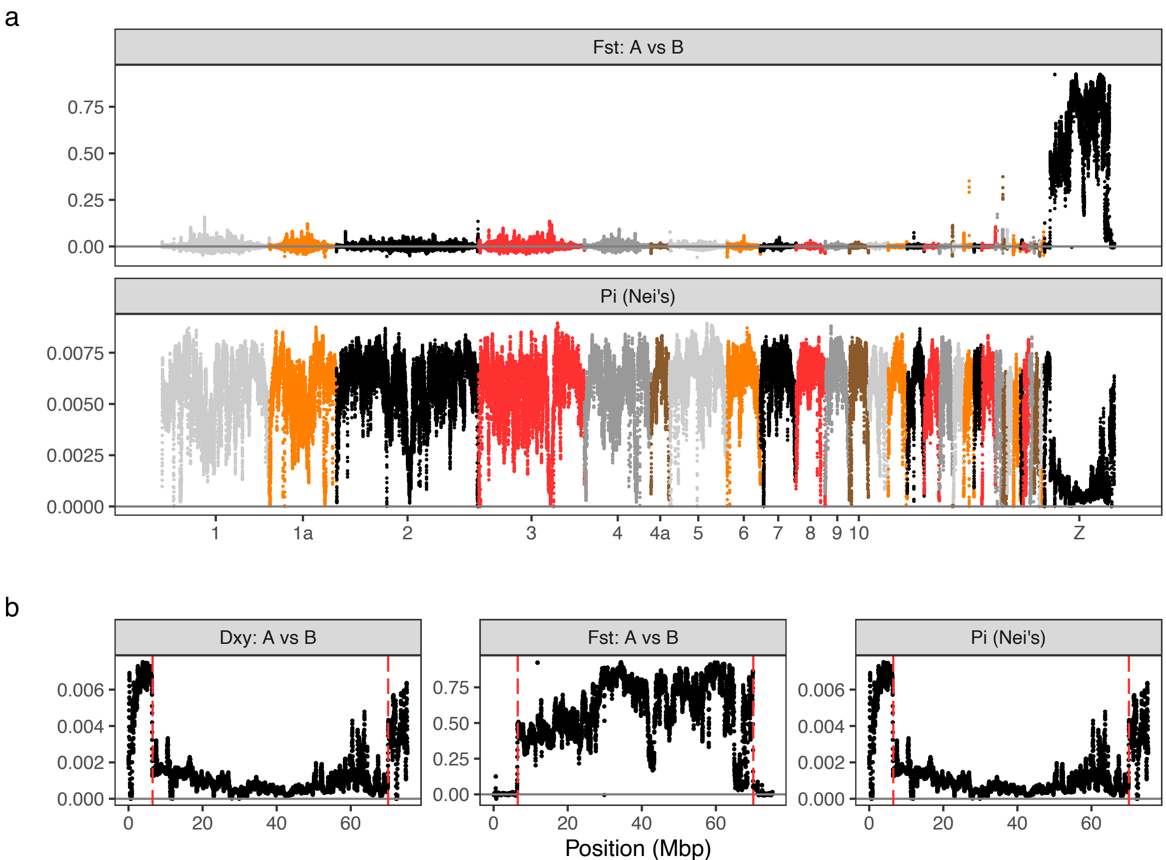


Figure 4.1. Sequence divergence (d_{XY}), Fixation index (F_{ST}) and Nucleotide diversity (π) across (a) the genome, and (b) on the Z chromosome at non-coding silent sites. (a) *d_{XY}, F_{ST} and π were calculated across autosomes 1-30 and Z in 100kbp sliding windows with 10kbp steps. d_{XY}, F_{ST} and π were calculated between A or AA (n=8) and B or BB (n=7) birds.* ***(b)*** *d_{XY}, F_{ST} and π are shown for the Z separately, with inversion breakpoints indicated by red vertical dashed lines. All statistics were calculated using SNPs from intronic and intergenic regions as non-coding sites.*

Next, within the inversion, when diversity was measured in the A and B birds separately it was lower still, especially in birds with the B haplotype (Table 4.1). Mean F_{ST} between A and B birds within the inversion region was exceptionally elevated relative to the autosomes or the Z regions flanking the inversion. Population differentiation as pronounced as this might be interpreted as positive selection acting on advantageous alleles on one or both haplotypes.

However, d_{XY}, an absolute measure of divergence, was not elevated between the two haplotypes - in fact it was ~6 times lower than on the autosomes (Table 4.1). This is indicative of a region of reduced within-population diversity, especially in the B haplotype, suggesting either its recent origin or a selective sweep. This is a similar pattern to that seen in hummingbirds (Henderson and Brelsford 2020), where regions of high F_{ST} between diverging chromosomes of low diversity show low d_{XY}. These results hold true when we analyse only a subset of A and AA haplotypes to avoid biases arising from unequal numbers of Z chromosomes when comparing groups of birds (Figure S4.2).

Next, we tested for sequence differences that have accumulated between the A and B inversion. We identified a total of 363 fixed differences (Table S4.4) within the coding regions of 203 of the 802 protein-coding genes on the Z chromosome. All of the fixed coding region differences are found in genes within the inversion polymorphism, further supporting the estimated inversion boundaries. Because there are no fixed differences outside of the inversion, it is not possible to perform a chi-square test to compare the numbers outside and inside the inversion. However, there are a total of 648 protein-coding genes inside and 154 protein-coding genes outside the inversion. If the rate of nonsynonymous and synonymous fixed differences outside of the inversion was the same as the observed rate inside, then we would expect to see 33-54 (95% confidence interval) nonsynonymous and 32-54 synonymous fixed differences outside of the inversion. Unsurprisingly, the observed number (zero of each) is significantly lower than this null expectation ($P = 8.9*10^{-23}$, nonsynonymous; $P = 1.2*10^{-22}$, synonymous). Taken together, the patterns of diversity and divergence on the Z suggest that the region outside of the inversion is similar to the autosomes and that the inversion region is very different. Thus, the patterns within the inversion cannot be regarded as typical of general features of the Z chromosome including signatures of Faster-Z (Vicoso and Charlesworth 2009; Oyler-McCance et al. 2015).

Table 4.1. Population genomic statistics for the autosomes and Z chromosome

Measure	Variable Site	Autosomes (Mean±SD)	Z outside inversion (Mean±SD)	Z within inversion (Mean±SD)	A birds within inversion (Mean±SD)	B birds within inversion (Mean±SD)
π (*10^{-3})	0-fold	1.32 (0.0013)	1.177 (0.0017)	0.37 (0.00058)	0.22 (0.00038)	0.057 (0.00018)
	4-fold	8.27 (0.0050)	6.91 (0.0047)	1.14 (0.0024)	0.65 (0.0013)	0.14 (0.00089)
	Non-Coding	5.67 (0.0017)	4.69 (0.0020)	0.91 (0.00057)	0.60 (0.00055)	0.11 (0.00013)
dxy (*10^{-3}): A vs B	0-fold	1.32 (0.0013)	1.177 (0.0017)	0.37 (0.00058)		
	4-fold	8.27 (0.0049)	6.91 (0.0047)	1.14 (0.0024)		
	Non-Coding	5.67 (0.0017)	4.69 (0.0020)	0.91 (0.00057)		
FST: A vs B	0-fold	-3 * 10^{-04} (0.031)	-0.0012 (0.045)	0.39 (0.38)		
	4-fold	-8 * 10^{-04} (0.031)	0.0035 (0.051)	0.39 (0.40)		
	Non-Coding	5 * 10^{-04} (0.011)	0.012 (0.043)	0.61 (0.20)		

4.4.2 Coding sequences in the inversion haplotype are evolving under relaxed purifying selection

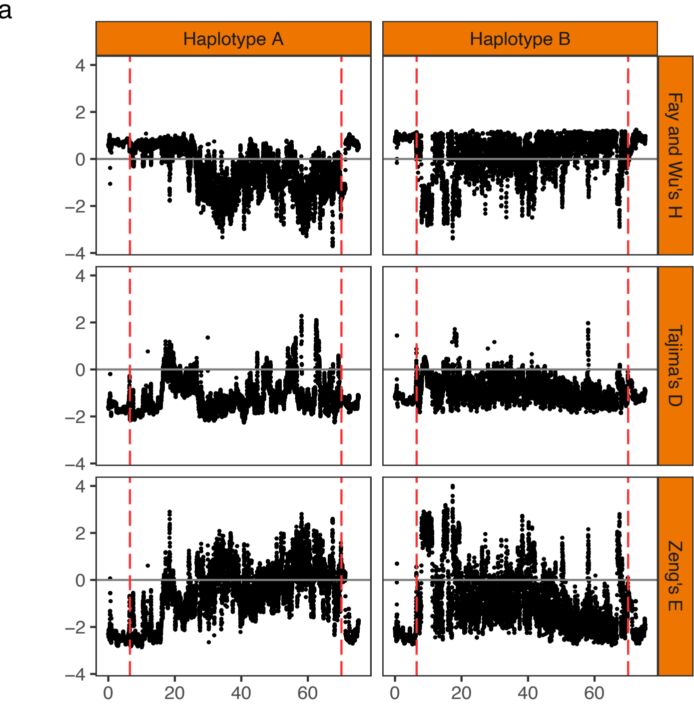
Next, we tested the evolutionary forces acting on sequences in the inversion haplotypes using a series of population genetic site frequency selection (SFS) statistics including Fay and Wu's H, Tajima's D, and Zeng's E. D and H identify regions under positive selection by respectively searching for regions with a depletion or enrichment of rare variants, and by comparing the level of high frequency to intermediate frequency variants. E overcomes D and H's lack of sensitivity to other factors affecting the site frequency spectrum by looking at intermediate frequency variants (Zeng et al. 2006). These stats were then compared between the inversion and outside of the inversion as well as to those autosomes of a similar size and sharing a similar recombination landscape (Chromosomes 1, 1a, 2, 4, 5, and 6) (Backström et al. 2010).

Given we observe low nucleotide diversity (π), elevated relative divergence (F_{ST}) but reduced absolute divergence (d_{XY}) in the inversion region, which is known to have a low recombination rate, background (purifying) selection may have been acting independently on the haplotypes (Cruickshank and Hahn 2014; Wolf and Ellegren 2016). However, given the very low recombination rate following an inversion event, we might also expect a relaxed response to purifying selection and the polymorphism being maintained through associative overdominance. A scenario that is unlikely here is that background selection and recombination between haplotypes may purge recessive alleles and mutation load, as has been suggested at a large inversion polymorphism affecting colour morph in redpoll (Funk et al. 2021).

To test these hypotheses, we first estimated Tajima's D, Fay and Wu's H and Zeng's E on the Z chromosome for A and B haplotypes independently and found less deviation from 0 (neutrality) within the inversion relative to the autosomes and outside of the inversion (Figure 4.2 & S4.3). Thus, there is, at best, limited evidence for recent selective sweeps on the A or B haplotypes. Given the similarity of these metrics across the different genomic regions (autosomes, outside inversion, within inversion), the most likely explanation for any departures from zero are demographic causes, such as population expansions or contractions.

Next, we compared patterns of polymorphic and fixed differences across A (AA males/A females) and B haplotypes (BB males/B females) (Table 4.2). First, we found that patterns of coding region polymorphism were significantly different between the inversion and non-inversion region for both A and B haplotypes (χ^2=275.01, P=<0.00001; χ^2=217.68, P=<0.00001 respectively). Specifically, in the inversion region of A and B haplotypes, nonsynonymous polymorphisms were more prevalent than synonymous ones, whereas the

opposite was observed outside of the inversion breakpoints (Table 4.2). Nonsynonymous mutations are usually selected against by purifying selection, and so this excess in the inversion region could be caused by a relaxation of purifying selection or, if these changes are adaptive, by ongoing positive selection acting on each haplotype.



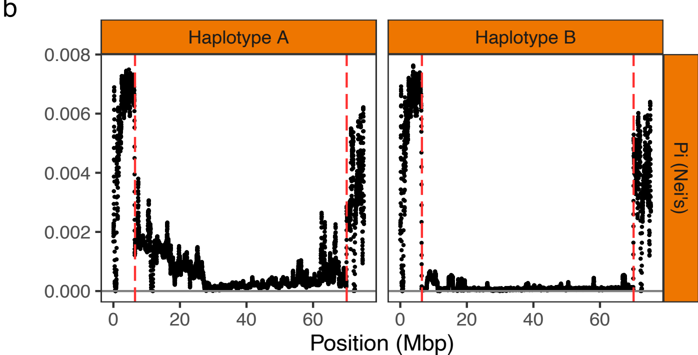


Figure 4.2. Neutrality (a) and Diversity (b) statistics on the Z chromosome for A and B haplotypes. *All statistics were calculated with SNPs called from sequence data of 8 A or AA birds and 7 B or BB birds. Reads were aligned against the reference genome (bTG1.4) to give their physical positions. Sequence from a female long-tailed finch was used as an outgroup sequence for calculating H and E. Dashed vertical lines indicate the approximate breakpoints of the Z chromosome inversion region. All statistics were calculated from 100Kbp sliding windows with a 10Kbp step between windows and using SNPs from intronic and intergenic regions as non-coding silent sites.*

To distinguish between these two alternate models, we employed the McDonald Kreitman test (McDonald and Kreitman 1991). If positive selection were acting, we would expect a greater proportion of nonsynonymous fixed differences relative to nonsynonymous polymorphisms. However, we found no significant relationship between the ratio of nonsynonymous to synonymous polymorphisms in A haplotypes relative to fixed differences between A and B haplotypes (χ^2=0.84, P=0.360). Whilst we did find a significant relationship in B haplotypes (χ^2=8.06, P=0.005), it was opposite to that expected under positive selection with more nonsynonymous polymorphisms than fixed differences.

Table 4.2. Comparison of polymorphism and fixed differences between A and B haplotypes

	Inside Inversion		Outside Inversion		Whole Z Chromosome		Inversion vs Outside Inversion
	N	S	N	S	N	S	
A polymorphisms	1150	1024	1584	3351	2734	4375	$\chi^2 = 275.01$, $p<0.00001$
B polymorphisms	463	318	1220	2663	1683	2981	$\chi^2 = 217.68$, $p<0.00001$
A vs B fixed differences	182	181	0	0	182	181	
Polymorphism in A compared to A vs B fixed differences	$\chi^2 = 0.84$, $p = 0.3600$						
Polymorphism in B compared to A vs B fixed differences	$\chi^2 = 8.06$, $p = 0.0045$						
Polymorphism in A compared to polymorphism in B	$\chi^2 = 9.19$, $p = 0.0024$		$\chi^2 = 0.43$, $p = 0.5100$		$\chi^2 = 6.67$, $p = 0.0098$		

N = nonsynonymous changes, S = synonymous changes.

Notably, this inversion is pericentric (Itoh and Arnold 2011) and evidence from various species indicates that the suppression of crossing over in pericentromeric regions results in reduced variability and reduced efficacy of natural selection, due to various types of Hill-Robertson interference processes (Charlesworth and Jensen 2021). In principle, this could contribute to the patterns of relaxed purifying selection we observe. However, if this was the case, then we would expect to see patterns of diversity and the ratios of nonsynonymous to synonymous

polymorphisms to be similar between the middle of the Z chromosome and the middle of the autosomes in our analysis (chromosomes 1, 1A, 2, 3, and 4) which were specifically chosen as they have very similar recombination patterns to the Z chromosome (Stapley et al. 2008; Backström et al. 2010; Stapley et al. 2010).

Taken together, these independent lines of evidence point to some role of associative overdominance in maintaining the Z-linked supergene in zebra finches. None of the population genetic tests (Tajima's D, Fay and Wu's H, Zeng's E and McDonald Kreitman test) we ran gave strong evidence for recent or ongoing positive selection and instead pointed towards relaxed purifying selection acting on the A and B haplotypes. We do however observe low nucleotide diversity within the inversion. This, at first glance, might appear inconsistent with our conclusion of associative overdominance which instead might be expected to increase nucleotide diversity due to neutral mutations linked with slightly deleterious recessive mutations (Zhao and Charlesworth 2016; Gilbert et al. 2020). However, as this inversion is Z-linked, we expect it to have a significantly lower effective population size and reduced recombination rate, which will result in lower genetic diversity. This is particularly the case for B which is present at a lower frequency than A (Kim et al. 2017). If recombination in heterozygotes is completely suppressed, then recombination in B haplotypes will only happen around 1/10th as often as in A haplotypes, based on the expected frequencies of AA and BB. Following this, we predict that the B haplotype is subject to even more relaxed purifying selection. Consistent with our prediction, whilst the relative number of Z-linked nonsynonymous polymorphisms outside of the inversion boundary does not differ between A and B haplotypes (χ^2=0.43, P=0.510), B haplotypes exhibit relatively more nonsynonymous polymorphisms than A within the inversion region (χ^2=9.19, P=0.002).

Interestingly, in Arctic cod (Matschiner et al. 2022), genetic exchange between haplotypes of inversions responsible for migratory lifestyle has been shown to counter the accumulation of deleterious mutations, leading to long term maintenance. However, the zebra finch supergene is located in a region of low recombination, with linkage mapping studies showing almost no recombination in this part of the Z chromosome, despite the mapping panels containing many males with two copies of the same haplotype (Stapley et al. 2008; Backström et al. 2010; Stapley et al. 2010). i.e. recombination suppression is not restricted to heterozygous birds. Instead, our data are consistent with recent work in the fire ant (Pracana et al. 2017), where the supergene appears to be maintained as a result of a relaxation of purifying selection, the accumulation of deleterious alleles (in fire ants Sb homozygotes are usually lethal), and associative overdominance protecting heterozygotes from recessive lethals. The zebra finch

haplotypes are not recessive lethal, and so theoretically they experience a lesser effect of the Muller's ratchet than fire ants do. Our results indicate that weaker mutational load can still maintain supergenes through associative overdominance. Conditions under which associative overdominance alone can maintain an inversion polymorphism for a long time are restricted (see Berdan et al. 2022), but they are helped by large population sizes, which zebra finches do have, and an initial establishment of the polymorphism through balancing selection, such as we see through AB heterozygotes having advantageous sperm traits.

4.4.3 Gene expression is conserved between Z-linked haplotypes

Given the reduced efficacy of purifying selection acting on the coding sequence of the Z-linked haplotypes, we next tested whether expression is evolving to compensate for the accumulation of deleterious mutations. Differential expression of supergene haplotypes has been observed in several species (Sun et al. 2018; Martinez-Ruiz et al. 2020; Berdan et al. 2021; Jeong et al. 2022) as well as cis- and trans-regulatory effects of the inversion across the whole genome (Arsenault et al. 2023). We might expect selection to downregulate the expression of A and B alleles that have independently accumulated deleterious nonsynonymous mutations. Indeed, in fire ants, alleles of the Sb haplotype tend to be more highly expressed than the SB allele mirroring patterns of sequence degeneration (Martinez-Ruiz et al. 2020), sequence degeneration is greater for differentially expressed than unbiased genes on mating-type chromosomes of the anther-smut fungus *Microbotryum* (Ma et al. 2020), and that reduced gene expression is associated with the early stages of sex chromosome decay (Pucholt et al. 2017).

We did find a handful of genes with significant differential gene expression between the testes of AB and AA males (2.51%) or ovaries of A and B females (1.35%) (Table S4.5 & Figure S4.4) within the inversion, consistent with previous expression studies in the zebra finch (Kim et al. 2017; Viitaniemi et al. 2023). However, this proportion relative to the autosomes was only weakly significant for the comparison between AB and AA males (χ^2 = 7.55, P_{MCMC} = 0.010) and was borderline significant for the A and B female comparison (χ^2 = 4.36, P_{MCMC}= 0.049). Together, this is consistent with a lack of large-scale expression divergence on the Z.

Comparisons of bulk expression values between samples across heterogeneous tissues such as the gonad can be problematic due to differences in cellular composition (Montgomery and Mank 2016; Price et al. 2022), and either mask differential expression or produce false signatures of expression change. This is particularly relevant for our male AB and AA comparisons as we might expect variation in the cellular composition of the testes (Lüpold et

al. 2020). Therefore, we tested for expression divergence between Z inversion haplotypes by identifying patterns of allele-specific expression. As we now compare expression of different alleles within the same sample, this goes some way to minimise biases arising from variation in tissue structure between individuals. However, an enrichment of allele-specific expression of the Z could be a product of the inactivation of a single Z chromosome due to dosage compensation, although there is limited evidence that this occurs in birds (Mank and Ellegren 2009; Segami et al. 2022). We did find a weak significant enrichment of genes exhibiting allele-specific expression on the Z relative to the autosomes 1-10 in AA males (Table S4.6). This pattern held when analysing individuals independently or when only genes exhibiting allele-specific expression in the two AA males were analysed (Table S4.6 & S4.7). However, the proportion of Z-linked genes with allele-specific expression was very small, where only 17 genes had allele-specific expression in both AA birds. This is consistent with previous results in the chicken gonad (Zimmer et al. 2016) and supports a lack of global mechanism of dosage compensation or Z chromosome inactivation in zebra finches.

Having shown that Z chromosome inactivation is unlikely, we then tested for allele-specific expression between A and B haplotypes in heterozygous males. We found only 8 Z-linked genes that exhibited significant allele-specific expression in all three AB males, none of which had known functions consistent with sperm morphology or fertility (Table S4.7 & S4.8), and only one of which (*MRPL50*), overlapped with the differentially expressed genes associated with sperm morphology identified by Kim et al. (2017). Interestingly, a recent study identified 3 candidate genes with differential expression between A and B haplotypes in the testes of young zebra finches (Viitaniemi et al. 2023); however, two of these were not expressed in our adult testes dataset and the other mapped to the W chromosome in our reference genome. Together this suggests that the expression of the supergene might vary through development. Importantly, in our study, the proportion of expressed genes with significant allele-specific expression on the Z inversion did not differ between homozygous and heterozygous males ($\chi^2 = 3.51$, $P_{MCMC} = 0.07$).

Finally, to test for the mode of selection on expression level we compared patterns of allele-specific expression with sequence divergence between A and B haplotypes. If selection is acting to downregulate the expression of A and B alleles that have independently accumulated deleterious nonsynonymous mutations, we predict the magnitude of allele-specific expression to increase with nonsynonymous divergence between A and B haplotypes. In fact, we find a significant but weakly negative relationship between the magnitude of allele-specific expression and both synonymous ($r_s = -0.070$, $P = 0.013$, Figure 4.3a) and nonsynonymous

divergence (r_s = -0.078, P = 0.006, Figure 4.3b) in the inversion. In contrast, homozygous individuals showed no association between allele-specific expression and synonymous divergence (P > 0.05) and a very weakly significant association between allele-specific expression and nonsynonymous divergence (r_s = -0.08, P = 0.047) (Figures 4.3c & 4.3d). This may suggest that the limited expression divergence we observe between A and B alleles may actually be selection acting to maintain equal expression of the two haplotypes, presumably to compensate for the independent accumulation of deleterious mutations.

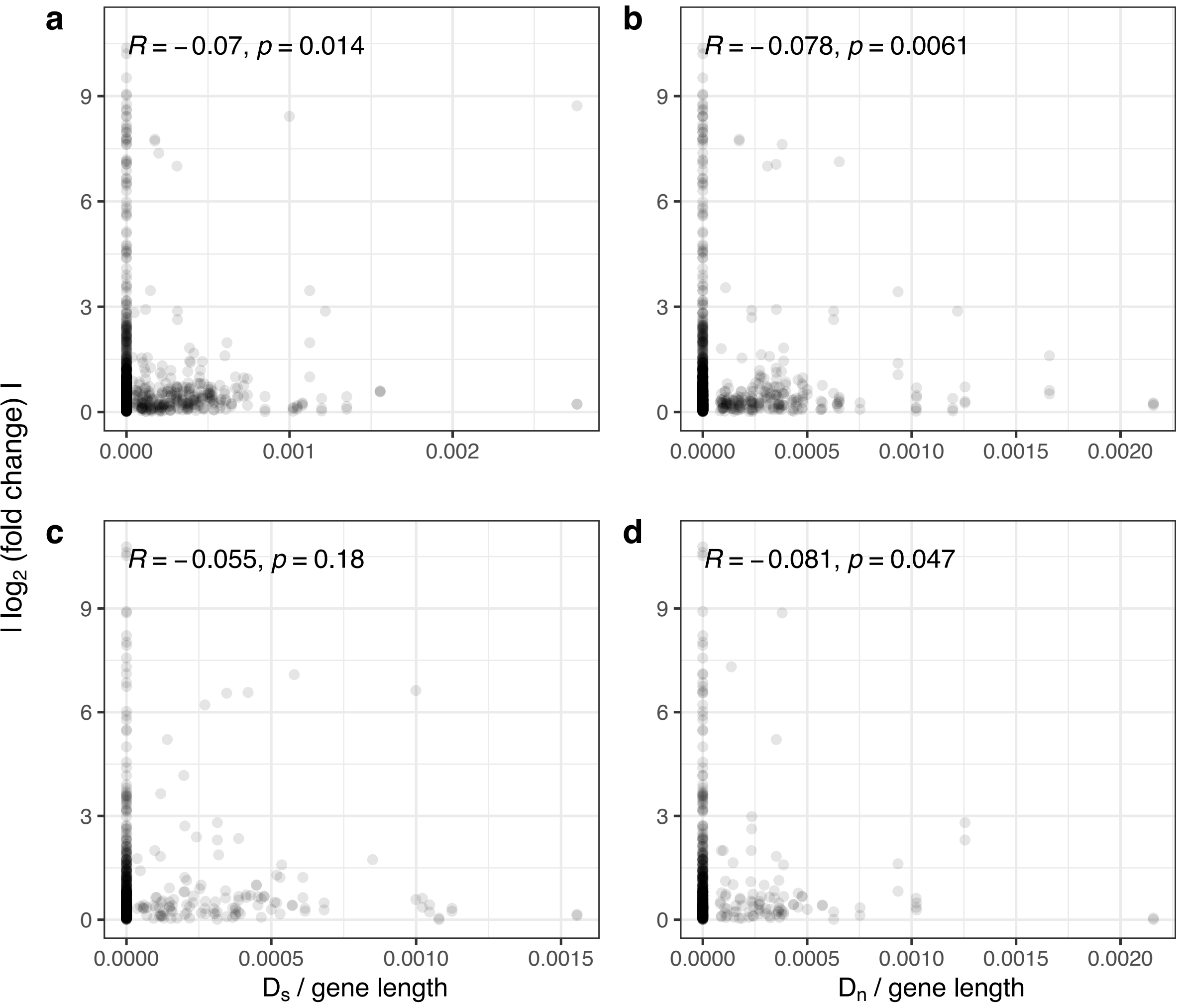


Figure 4.3. Spearman's rank tests investigating the relationship between the number of fixed synonymous (D_s) and nonsynonymous (D_n) differences between A and B haplotypes and the magnitude of allele-specific expression. *Absolute measures of allele-specific expression between A and B haplotypes in AB heterozygote birds (**panel a** and **b**) and between heterozygous sites within AA birds (**panel c** and **d**). For all tests, the number of fixed differences is normalised by dividing by total coding sequence length.*

4.5 CONCLUSION

Our analyses highlight the impact of inversion polymorphisms on the evolution of the coding and regulatory sequence of the Z chromosome, whilst controlling for the unique pressures that the Z alone experiences. Importantly, we show that while the coding sequence is subject to relaxed purifying selection, and subsequently an excess of nonsynonymous differences that may lead to a fitness cost through degeneration, the transcriptome may somewhat mitigate this through the more equal expression of those most highly diverged, and potentially degenerated genes.

4.6 MATERIALS & METHODS

4.6.1 Identifying diagnostic SNPs for the Z-linked inversion genotype

We previously genotyped male zebra finches (n=1202) from a population that was previously maintained at the University of Sheffield (Kim et al. 2017). Genotypes included 3056 SNPs spanning the Z chromosome (Kim et al. 2017). With these data, we identified markers with the most significant difference in genotype frequency between the haplotypes and were therefore highly diagnostic of the inversion karyotype. Specifically, each marker was given a weighted chi-square score ranging from 0 to 1; any SNPs with a weighted chi-square score greater than 0.9 (90% of maximum score) were included in a list of 2123 “diagnostic SNPs” (Table S4.1).

4.6.2 Variant calling

We downloaded publicly available (www.ebi.ac.uk/ena/data/view/PRJEB10586) sequence reads from 24 zebra finches (11 female, 13 male), and one female long-tailed finch (*Poephila acuticauda acuticauda*) (Table S4.2). All birds were originally sequenced using Illumina HiSeq 2000 paired end sequencing (read length=100bp) in 2012 and 2013 (Singhal et al. 2015).

These reads were aligned to a male zebra finch reference genome (bTG1.4) (Rhie et al. 2021) with Bowtie v2.3.4.3, using the default settings (Langmead and Salzberg, 2012). This reference genome was constructed using DNA from a male zebra finch known as “Black17”, which has been shown to be an AB heterokaryotype for the Z chromosome inversion polymorphism (Pepper 2022).

Variants were called using genome analysis toolkit (GATK) version 4.2.5.0 (van der Auwera and O’Connor 2020). Specifically, duplicate reads were removed for each individual, using GATK MarkDuplicates, before GATK HaplotypeCaller was run using aligned reads.

Genotypes were then called using GATK GenotypeGVCFs. Indels were discarded whilst called SNPs were retained and then filtered for quality using GATK VariantFiltration. Parameters that were used in variant quality filtering were: variant confidence (QD) < 2.0, which is intended to normalise the variant quality in order to avoid inflation caused when there is deep coverage; phred-scaled probability of strand bias (FS) > 60.0 and symmetric odds ratio test for strand bias (SOR) > 4.0, which both describe whether the alternate allele is seen more or less often on the forward or reverse strand than the reference allele; mapping quality of reads (MQ) < 40.0, and the compared mapping qualities of reads supporting the reference and alternate allele (MQRankSum) <- 12.5; and a comparison of positions of the reference and alternate alleles within different reads (ReadPosRankSum) < -8.0. Sites were then filtered on a per sample basis with vcftools v0.1.16 (Danecek et al. 2011), keeping variants with depth >= 5, genotyping quality >= 20 and missing data in <= 50% of samples.

4.6.3 Calculation of population genetic statistics

We calculated summary statistics using R v4.2.1 and the package "PopGenomeR" v2.7.5 (Pfeifer et al. 2014). Specifically, we calculated the fixation index (F_{ST}) from minor allele frequencies (Hudson et al. 1992), as well as between haplotype diversity (d_{XY}) between the 8 zebra finches of inversion haplotype A or AA and the 7 zebra finches of inversion haplotype B or BB. Nucleotide diversity (π), Tajima's D, Fay and Wu's H and Zeng's E (a composite statistic of D and H) were calculated for birds of each haplotype. In each case, the statistic was calculated for sliding windows of 100Kb of sequence, with each window overlapping by 10Kb, across the Z chromosome. For the three neutrality test statistics (D, H & E), sliding windows were also calculated for the 30 largest autosomes (Chromosomes 1-30). For calculating Fay and Wu's H and Zeng's E, a sequence from a female long-tailed finch was used as a closely related outgroup (~6MY) (Hooper and Price 2015) to distinguish the derived and ancestral alleles at sites which are polymorphic in zebra finches. Each statistic was independently calculated with four-fold degenerate sites, zero-fold degenerate sites and non-coding silent sites (from introns or outside of genic regions).

Since there were an unequal number of A or AA birds and B or BB birds, nucleotide diversity, Tajima's D, Fay and Wu's H and Zeng's E were also all calculated using just 8 chromosomes from A or AA birds so that the number of A haplotype Z chromosome sequences was equal to the number of B haplotype Z chromosome sequences (n=8).

4.6.4 Identifying synonymous and nonsynonymous SNPs

Genome annotations from the NCBI *Taeniopygia guttata* annotation release 106 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Taeniopygia_guttata/106/) were used to determine synonymous and nonsynonymous sites within coding regions on the zebra finch (bTG1.4) Z chromosome. Using PopGenomeR v2.7.5 (Pfeifer et al. 2014), SNPs found through variant calling with GATK v4.2.6.1 were sorted into those falling within and those outside of coding sequence. Then, SNPs within coding regions were sorted into those which were fixed between A and B and at either a nonsynonymous ($N_{Dnonsyn}$) or a synonymous (N_{Dsyn}) position, and those which were polymorphic within either the A or B haplotype and at either a nonsynonymous ($N_{Pnonsyn}$) or synonymous (N_{Psyn}) position. The number of SNPs within each of these categories was then counted, and these counts were used to perform a series of chi-square tests to test for nonsynonymous and synonymous variation between the A and B haplotypes inside and outside of the inversion region. Chi-square tests of independence (df=1) were performed to examine differences in proportions of $N_{Dnonsyn}$ and N_{Psyn} polymorphisms: (a) between the two haplotypes, and (b) from the proportion of N_{Dnonsy} and N_{Psyn} fixed differences between the two haplotypes. Tests were performed within the inversion, outside of the inversion and across the entire Z chromosome (Table 4.2). These tests are analogous to McDonald Kreitman tests (McDonald and Kreitman 1991) except they compare divergence between the A and B haplotypes rather than between two related species.

4.6.5 Expression data

RNA-seq data were obtained from a captive population of zebra finches at Queen Mary University of London. Individuals were in their first breeding season, having formed breeding pairs and produced fertile eggs. Samples were collected in accordance with national guidelines. The left gonad was dissected from five males and five females, homogenised and stored in RNAlater until preparation. We used the Animal Tissue RNA Kit (Qiagen) to extract RNA. Dual-indexed, strand-specific RNA-seq libraries were prepared at the NERC Environmental Omics Facility (NEOF) Liverpool using the NEBNext polyA selection and Ultra II Directional RNA library preparation kits. RNA was sequenced on the Illumina NovaSeq using S4 chemistry, resulting in an average of 74 million 150bp paired-end reads per sample.

The data were quality assessed using FastQC 0.38 and filtered using Trimmomatic v0.38. Specifically, we removed reads containing adaptor sequences and trimmed reads if the sliding window average Phred score over four bases was <15 or if the leading/trailing bases had a Phred score <3. Reads were removed post filtering if either read pair was <95 bases in length.

4.6.6 Genotyping individuals with RNA-seq data

Trimmed reads were aligned against the indexed bTG1.4 reference (Rhie et al. 2021) using HISAT2 v2.1.0. Using GATK v4.1.4 (van der Auwera and O'Connor 2020), reads in each BAM file were then assigned a new shared readgroup and duplicated reads were marked. Reads were then split if they contained Ns in their cigar strings. Variants were called for chromosomes 1-10 and the Z chromosome, using GATK HaplotypeCaller and filtered for a minimum depth (DP) of 10 reads, and genotyping quality (GQ) of 30. Then samples were genotyped using the set of diagnostic SNPs we previously identified.

4.6.7 Quantifying differential gene expression

Salmon v1.8.0 (Patro et al. 2017) was used to quantify expression for all samples. Briefly, trimmed reads were pseudo-aligned against the bTG1.4 transcriptome creating read count matrices. Outputs were analysed using EdgeR v3.34.1 (Robinson et al. 2010). Gene level counts were generated and any gene with log rpkm < 2 in less than half of the individuals in each genotype was removed, following our previously described approaches (Harrison et al. 2015; Wright et al. 2018).

Next, we quantified allele-specific expression (ASE) in all samples. WASP v0.3.4 (van de Geijn et al. 2015) was used to identify reads overlapping differentiating sites, following the WASP documentation. Upon read remapping, mapping stringency was reduced to 10 mismatches (MX=10) to reduce mapping bias in regions of high divergence between haplotypes. Variants were then recalled using HaplotypeCaller and genotyped before being filtered for a minimum depth (DP) of 10 reads, and genotyping quality (GQ) of 30. Reads with more than 4 SNPs in a 95bp window were also removed (Stevenson et al. 2013) to reduce further mapping bias to the reference allele. Allele-specific expression was then quantified at the genic level using phASER v1.1.1 (Castel et al. 2016). phASER advances on SNP-based ASE detection approaches by phasing variants across an entire gene, leveraging the linkage between variants, giving a single value for the gene as opposed to for each individual SNP. For final analyses, any heterozygous gene with log rpkm < 2, from Salmon, or with a total read count from phASER < 20 was removed. Genes were defined as having allele-specific expression if the log fold change of the expression value (plus one to handle zero count data for the A or B allele) between alleles >=1 with a fdr adjusted p-value < 0.05.

4.7 ACKNOWLEDGEMENTS

This work was funded by a NERC Independent Research Fellowship to AEW (NE/N013948/1), a NERC ACCE DTP to PDP and JP, and a BBSRC grant to JS (BB/I02185X/1). The laboratory work was supported by the UK Natural Environment Research Council (NERC) Environmental Omics Facility. We thank Julia George and David Clayton for raising zebra finches at Queen Mary University for the expression analyses, Kai Zeng for valuable input on interpretation of population genetic analyses, and Judith Mank and Roger Butlin for helpful comments and suggestions on the manuscript.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

4.8 DATA AVAILABILITY

Code to reproduce analyses can be found:
https://github.com/petedprice/SupergeneEvolutionZebraFinch

Gene expression data is deposited at: DOI: 10.5061/dryad.7pvmcvf

4.9 REFERENCES

Arsenault, S. V., Riba-Grognuz, O., Shoemaker, D., Hunt, B. G., & Keller, L. (2023). Direct and indirect genetic effects of a social supergene. *Molecular Ecology, 32*(5), 1087–1097.

Assersohn, K., Morton, O., Slate, J., & Hemmings, N. (2024). A sex-linked supergene with large effects on sperm traits has little impact on reproductive traits in female zebra finches. *Proceedings of the Royal Society B, 291*(2019), 20232796.

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., Webster, M. T., Ost, T., Schneider, M., Kempenaers, B., & Ellegren, H. (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research, 20*(4), 485–495.

Balakrishnan, C. N., & Edwards, S. V. (2009). Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). *Genetics, 181*(2), 645–660.

Bennison, C., Hemmings, N., Brookes, L., Slate, J., & Birkhead, T. (2016). Sperm morphology, adenosine triphosphate (ATP) concentration and swimming velocity: unexpected

relationships in a passerine bird. *Proceedings of the Royal Society B*, 283(1837), 20161558.

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., Mérot, C., Durmaz Mitchell, E., Pascual, M., Peichel, C. L., Rafajlović, M., Westram, A. M., Schaeffer, S. W., Johannesson, K., & Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. Journal of *Evolutionary Biology, 36*(12), 1761–1782.

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics, 17*(3), e1009411.

Berdan, E. L., Blanckaert, A., Butlin, R. K., Flatt, T., Slotte, T., & Wielstra, B. (2022). Mutation accumulation opposes polymorphism: supergenes and the curious case of balanced lethals. *Proceedings of the Royal Society B, 377*(1856), 20210199.

Booker, T., Ness, R. W., & Charlesworth, D. (2015). Molecular evolution: breakthroughs and mysteries in Batesian mimicry. *Current Biology, 25*(12), R506–R508.

Carpentier, F., Rodríguez de la Vega, R. C., Jay, P., Duhamel, M., Shykoff, J. A., Perlin, M. H., Wallen, R. M., Hood, M. E., & Giraud, T. (2022). Tempo of degeneration across independently evolved nonrecombining regions. *Molecular Biology and Evolution, 39*(4), msac060.

Castel, S. E., Mohammadi, P., Chung, W. K., Shen, Y., & Lappalainen, T. (2016). Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nature Communications, 7*, 12817.

Charlesworth, B. (2023). The effects of inversion polymorphisms on patterns of neutral genetic diversity. *Genetics, 224*(4), iyad116.

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist, 130*(1), 113–146.

Charlesworth, D. (2016). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications, 9*(1), 74–90.

Charlesworth, D., & Charlesworth, B. (1975). Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *Journal of Theoretical Biology, 55*(2), 305–324.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23*(13), 3133–3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker,

R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158.

Dobzhansky, T. (1950). Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura. Genetics*, 35(3), 288–302.

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving Inversions. *Trends in Ecology & Evolution, 34*(3), 239–248.

Funk, E. R., Mason, N. A., Pálsson, S., Albrecht, T., Johnson, J. A., & Taylor, S. A. (2021). A supergene underlies linked variation in color and morphology in a Holarctic songbird. *Nature Communications, 12*(1), 6833.

Gilbert, K. J., Pouyet, F., Excoffier, L., & Peischl, S. (2020). Transition from background selection to associative overdominance promotes diversity in regions of low recombination. *Current Biology, 30*(1), 101–107.e3.

Gutiérrez-Valencia, J., Fracassetti, M., Berdan, E. L., Bunikis, I., Soler, L., Dainat, J., Kutschera, V. E., Losvik, A., Désamoré, A., Hughes, P. W., Foroozani, A., Laenen, B., Pesquet, E., Abdelaziz, M., Pettersson, O. V., Nystedt, B., Brennan, A. C., Arroyo, J., & Slotte, T. (2022). Genomic analyses of the *Linum distyly* supergene reveal convergent evolution at the molecular level. *Current Biology, 32*(20), 4360–4371.e6.

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution, 6*(12), 1965–1979.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America, 112*(14), 4393–4398.

Hill, J., Enbody, E. D., Bi, H., Lamichhaney, S., Lei, W., Chen, J., Wei, C., Liu, Y., Schwochow, D., Younis, S., Widemo, F., & Andersson, L. (2023). Low Mutation Load in a Supergene Underpinning Alternative Male Mating Strategies in Ruff (*Calidris pugnax*). *Molecular Biology and Evolution, 40*(12), msad224.

Hooper, D. M., & Price, T. D. (2015). Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution, 69*(4), 890–903.

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics, 132*(2), 583–589.

Huynh, L. Y., Maney, D. L., & Thomas, J. W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Heredity, 106*(4), 537–546.

Itoh, Y., & Arnold, A. P. (2011). Zebra finch cell lines from naturally occurring tumors. *In Vitro Cellular & Developmental Biology. Animal, 47*(4), 280–282.

Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics, 53*(3), 288–293.

Jeong, H., Baran, N. M., Sun, D., Chatterjee, P., Layman, T. S., Balakrishnan, C. N., Maney, D. L., & Yi, S. V. (2022). Dynamic molecular evolution of a supergene with suppressed recombination in white-throated sparrows. *eLife, 11*, e79387.

Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S. J., Rogers, J., Beasley, H., Barlow, K., ffrench-Constant, R. H., Mallet, J., McMillan, W. O., & Jiggins, C. D. (2006). A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology, 4*(10), e303.

Kapun, M., Mitchell, E. D., Kawecki, T. J., Schmidt, P., & Flatt, T. (2023). An Ancestral Balanced Inversion Polymorphism Confers Global Adaptation. *Molecular Biology and Evolution, 40*(6), msad118.

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L. L., Griffith, S. C., Burke, T., Birkhead, T. R., & Slate, J. (2017). A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature Ecology & Evolution, 1*(8), 1168–1176.

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., Opatová, P., Albrechtová, J., Wittig, M., Franke, A., Albrecht, T., & Kempenaers, B. (2017). A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nature Ecology & Evolution, 1*(8), 1177–1184.

Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Kempenaers, B., & Forstmeier, W. (2016). Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biology, 17*(1), 199.

Koch, E. L., Morales, H. E., Larsson, J., Westram, A. M., Faria, R., Lemmon, A. R., Lemmon, E. M., Johannesson, K., & Butlin, R. K. (2021). Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evolution Letters, 5*(3), 196–213.

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., & Kronforst, M. R. (2014). doublesex is a mimicry supergene. *Nature, 507*(7491), 229–232.

Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene

determines highly divergent male reproductive morphs in the ruff. *Nature Genetics, 48*(1), 79–83.

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S. M.-Y., Xu, X., Höglund, J., Liu, X. & Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics, 48*(1), 84–88.

Li, B., Bickel, R. D., Parker, B. J., Saleh Ziabari, O., Liu, F., Vellichirammal, N. N., Simon, J.-C., Stern, D. L., & Brisson, J. A. (2020). A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *eLife, 9*, e50608.

Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., van Oosterhout, C., & Gilmartin, P. M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris. Nature Plants, 2*(12), 16188.

Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology, 8*(9), e1000500.

Lüpold, S., de Boer, R. A., Evans, J. P., Tomkins, J. L., & Fitzpatrick, J. L. (2020). How sperm competition shapes the evolution of testes and sperm: a meta-analysis. *Proceedings of the Royal Society B*, *375*(1813), 20200064.

Maney, D. L., Merritt, J. R., Prichard, M. R., Horton, B. M., & Yi, S. V. (2020). Inside the supergene of the bird with four sexes. *Hormones and Behavior, 126*, 104850.

Mank, J. E., & Ellegren, H. (2009). All dosage compensation is local: gene-by-gene regulation of sex-biased expression on the chicken Z chromosome. *Heredity, 102*(3), 312–320.

Martinez-Ruiz, C., Pracana, R., Stolle, E., Paris, C. I., Nichols, R. A., & Wurm, Y. (2020). Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *eLife, 9*, e55862.

Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Brieuc, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., & Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution, 6*(4), 469–481.

Ma, W.-J., Carpentier, F., Giraud, T., & Hood, M. E. (2020). Differential gene expression between fungal mating types is associated with sequence degeneration. *Genome Biology and Evolution, 12*(4), 243–258.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature, 351*(6328), 652–654.

McGrath, C. (2023). Highlight: Unmasking the supergene beneath the ruff's mating strategies.

Molecular Biology and Evolution, 40(12), msad249.

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

Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L., & Wellenreuther, M. (2020). Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications, 11*(1), 670.

Meyer, L., Barry, P., Riquet, F., Foote, A., Der Sarkissian, C., Cunha, R. L., Arbiol, C., Cerqueira, F., Desmarais, E., Bordes, A., Bierne, N., Guinand, B., & Gagnaire, P.-A. (2024). Divergence and gene flow history at two large chromosomal inversions underlying ecotype differentiation in the long-snouted seahorse. *Molecular Ecology,* e17277.

Montgomery, S. H., & Mank, J. E. (2016). Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Molecular Ecology, 25*(20), 5114–5128.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research, 106*, 2–9.

Nosil, P., Villoutreix, R., de Carvalho, C. F., Farkas, T. E., Soria-Carrasco, V., Feder, J. L., Crespi, B. J., & Gompert, Z. (2018). Natural selection and the predictability of evolution in *Timema* stick insects. *Science, 359*(6377), 765–770.

Oyler-McCance, S. J., Cornman, R. S., Jones, K. L., & Fike, J. A. (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity, 115*(5), 452–459.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods, 14*(4), 417–419.

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., … Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution, 3*(12), 1731–1742.

Pepper, J. (2022). Evolutionary genomics of the zebra finch Z chromosome inversion polymorphism. University of Sheffield.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. Molecular Biology and *Evolution, 31*(7), 1929–1936.

Potente, G., Léveillé-Bourret, É., Yousefi, N., Choudhury, R. R., Keller, B., Diop, S. I., Duijsings, D., Pirovano, W., Lenhard, M., Szövényi, P., & Conti, E. (2022). Comparative

genomics elucidates the origin of a supergene controlling floral heteromorphism. *Molecular Biology and Evolution, 39*(2), msac035.

Pracana, R., Priyam, A., Levantis, I., Nichols, R. A., & Wurm, Y. (2017). The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Molecular Ecology, 26*(11), 2864–2879.

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R., & Wright, A. E. (2022). Detecting signatures of selection on gene expression. *Nature Ecology & Evolution, 6*(7), 1035–1045.

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*(8), 1991–2001.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature, 592*(7856), 737–746.

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

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139–140.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. Current Biology, 24(7), R288–R294.

Segami, J. C., Semon, M., Cunha, C., Bergin, C., Mugal, C. F., & Qvarnström, A. (2022). Single-Cell Transcriptomics reveals relaxed evolutionary constraint of spermatogenesis in two passerine birds as compared to mammals. bioRxiv, 2022.01.22.477241.

Shore, J. S., Hamam, H. J., Chafe, P. D. J., Labonne, J. D. J., Henning, P. M., & McCubbin, A. G. (2019). The long and short of the S-locus in Turnera (Passifloraceae). *The New Phytologist, 224*(3), 1316–1329.

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., & Przeworski, M. (2015). Stable recombination hotspots in birds. *Science, 350*(6263), 928–932.

Smith, J. M. (1978). The Evolution of Sex. *CUP Archive*.

Stapley, J., Birkhead, T. R., Burke, T., & Slate, J. (2008). A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics,*

179(1), 651–667.

Stapley, J., Birkhead, T. R., Burke, T., & Slate, J. (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Research, 20*(4), 496–502.

Stevenson, K. R., Coolon, J. D., & Wittkopp, P. J. (2013). Sources of bias in measures of allele-specific expression derived from RNA-sequence data aligned to a single reference genome. *BMC Genomics, 14*, 536.

Sturtevant, A. H., & Mather, K. (1938). The interrelations of inversions, heterosis and recombination. *The American Naturalist, 72*(742), 447–452.

Sun, D., Huh, I., Zinzow-Kramer, W. M., Maney, D. L., & Yi, S. V. (2018). Rapid regulatory evolution of a nonrecombining autosome linked to divergent behavioral phenotypes. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2794–2799.

Thomas, J. W., Cáceres, M., Lowman, J. J., Morehouse, C. B., Short, M. E., Baldwin, E. L., Maney, D. L., & Martin, C. L. (2008). The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics, 179*(3), 1455–1468.

Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity, 113*(1), 1–8.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature, 584*(7822), 602–607.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology, 26*(3), 344–350.

van de Geijn, B., McVicker, G., Gilad, Y., & Pritchard, J. K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods, 12*(11), 1061–1063.

van der Auwera, G., & O'Connor, B. D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media, Incorporated.

Vicoso, B., & Charlesworth, B. (2009). Effective population size and the faster-X effect: an extended model. *Evolution, 63*(9), 2413–2426.

Viitaniemi, H. M., Leder, E. H., Kauzál, O., Stopková, R., Stopka, P., Lifjeld, J. T., & Albrecht, T. (2023). Impact of Z chromosome inversions on gene expression in testis and liver tissues in the zebra finch. *Molecular Ecology*, mec.17236.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature, 493*(7434), 664–668.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution, 33*(6), 427–440.

Wolf, J. B. W., & Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics, 18*(2), 87–100.

Wright, A. E., Fumagalli, M., Cooney, C. R., Bloch, N. I., Vieira, F. G., Buechel, S. D., Kolm, N., & Mank, J. E. (2018). Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters, 2*(2), 52–61.

Yang, J., Xue, H., Li, Z., Zhang, Y., Shi, T., He, X., Barrett, S. C. H., Wang, Q., & Chen, J. (2023). Haplotype-resolved genome assembly provides insights into the evolution of S-locus supergene in distylous *Nymphoides indica. The New Phytologist, 240*(5), 2058–2071.

Zeng, K., Fu, Y.-X., Shi, S., & Wu, C.-I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics, 174*(3), 1431–1439.

Zhang, W., Westerman, E., Nitzany, E., Palmer, S., & Kronforst, M. R. (2017). Tracing the origin and evolution of supergene mimicry in butterflies. *Nature Communications, 8*(1), 1269.

Zhao, L., & Charlesworth, B. (2016). Resolving the conflict between associative overdominance and background selection. *Genetics, 203*(3), 1315–1334.

Zimmer, F., Harrison, P. W., Dessimoz, C., & Mank, J. E. (2016). Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome. *Genome Biology and Evolution, 8*(4), 1233–1242.

Chapter 5. The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies

Peter D Price[1*], Sylvie M Parkus[1], Victoria J Lloyd[1], Sasha L Bradshaw[2], Sadé Bates[2], Steve Paterson[3], Terry Burke[1], Iulia Darolti[4], Andrew Pomiankowski[2], Alison E Wright[1*]

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom

[2]Department of Genetics, Evolution and Environment, University College London, United Kingdom

[3]School of Biological Sciences, University of Liverpool, United Kingdom

[4]Department of Ecology and Evolution, University of Lausanne, Switzerland

*Corresponding authors: pprice3@sheffield.ac.uk, a.e.wright@sheffield.ac.uk

Keywords: X chromosome, scRNA-seq, dosage compensation, meiotic sex chromosome inactivation, spermatogenesis

Short title: Genomics of meiotic drive in *Teleopsis dalmanni*

5.1 ABSTRACT

Sex-linked meiotic drivers limit the inheritance of the alternate sex chromosome in the heterogametic sex, subsequently skewing the ratio of males to females in the offspring. They consequently have large impacts on genome evolution, adaptation, and the emergence and maintenance of sexually selected traits. Detecting the molecular basis of meiotic drive is complex. It is affected by both the common occurrence of inversions housing a driver as well as the transcriptomic and cell type heterogeneity of the gonads. Herein, we utilise single-cell RNA sequencing to investigate the molecular basis of a sex-linked meiotic driver in the Malaysian stalk-eyed fly, *Teleopsis dalmanni*. We first characterise the regulatory dynamics of the X chromosome during spermatogenesis, providing evidence for incomplete dosage compensation during meiosis and a lack of complete meiotic sex chromosome inactivation. We then show that the presence of a meiotic driver does not disturb broad patterns of X-linked expression, with the majority of expression differences enriched in the later stages of spermatogenesis. Finally, we provide candidates for drive function and effect, including the disruption of sperm development and motility.

5.2 INTRODUCTION

Following Mendelian genetics, the expectation at meiosis is that maternal and paternal alleles segregate equally. However, meiosis is often a battleground for inheritance. Intragenomic conflicts emerge through selfish genetic elements forcing unequal segregation of alleles, skewing their chances of being represented in the mature germline (Sandler & Novitski, 1957; Jaenike, 2001; Lindholm et al., 2016). These selfish genes, known as meiotic drivers, are widespread across eukaryotic life (Sandler et al., 1959; Turner & Perkins, 1979; Ardlie, 1998; Taylor et al., 1999; Fishman & Willis, 2005), and have large consequences for the ecology and evolution of populations (Hamilton, 1967; Wilkinson et al., 2014; Zhang et al., 2015; Lindholm et al., 2016).

Sex-chromosome meiotic drive is the most common form of meiotic drive (Hurst & Pomiankowski, 1991), also known as sex-ratio distorter (SR). Here, X-linked drivers (in XY systems) reduce the inheritance of the Y chromosome and due to the altering of offspring sex ratios from 1:1, are easily detectable. Consequently, they have profound effects on reproductive traits (Holman et al., 2015; Herbette et al., 2021), genome evolution (Úbeda et al., 2015; Blackmon et al., 2019; Eickbush et al., 2019; Reinhardt et al., 2023), adaptation (Dyer et al., 2007; Mackintosh et al., 2021), sexual selection (Wilkinson et al., 1998; Pinzone & Dyer, 2013; Cotton et al., 2014), and population persistence (Hamilton, 1967; Jaenike, 2001; Mackintosh et al., 2021).

Characterising the molecular mechanisms of meiotic drivers and their consequences is therefore key to understanding a range of biological processes. However, despite meiotic drive having been identified almost a century ago (Gershenson, 1928), the evolutionary origins and the general mechanisms by which drivers operate remain unclear. Further, at the molecular level, they have been well studied in only a handful of species, including *Drosophila* (Courret et al., 2019), *Anopheles* (M. Li et al., 2020; Simoni et al., 2020), house mouse (Silver, 1985; Cocquet et al., 2012; Kelemen et al., 2022), monkeyflower (Fishman & Willis, 2005), and *Neurospora* (Vogan et al., 2022). Recent work has started to highlight both the diverse mechanisms that are utilised by drivers to disrupt meiosis, but also shared elements and time points that are targeted (Zimmering et al., 1970; Courret et al., 2019; Kruger & Mueller, 2021). As such, meiotic drivers have been broadly divided into two classes. Meiotic-acting drivers disrupt proper segregation at meiosis, such as the Paris driver in *Drosophila simulans* that leads to improper segregation of the Y in anaphase II (Cazemajor et al., 1997, 2000). Whereas, post-meiotic drivers disrupt motility of sperm or poison them, for example the *Winters* driver in the same species leads to a defect in nuclear condensation of Y sperm (Tao

et al., 2007). In the latter class, the histone-protamine transition, an essential checkpoint in spermatid elongation, has been repeatedly identified as a post-meiotic target (Kettaneh & Hartl, 1976; Hauschteck-Jungen & Hartl, 1982; Gingell & McLean, 2020; Herbette et al., 2021; Vedanayagam et al., 2021).

Meiotic drivers are often challenging to study as they are frequently housed by inversions that guard against the subsequent breaking up of these complex molecular phenotypes (Sandler et al., 1959; Lyttle, 1993; Silver, 1993; Reinhardt et al., 2023) and prevent the formation of suicidal haplotypes bearing both driver and target (Dyer et al., 2007). The resulting high level of linkage disequilibrium between the driver and neutral variation across the inversion limits the use of traditional genetic mapping approaches (Dyer et al., 2007). In addition, the processes they disrupt, such as spermatogenesis, are complex and operate alongside unique regulatory mechanisms in the gonads, including dosage compensation and meiotic sex chromosome inactivation, that are often poorly characterised in non-model organisms. Transcriptomics therefore provides an important avenue for understanding the molecular underpinnings and consequences of drivers.

Here, we combine single-cell RNA-sequencing (scRNA-seq) approaches with a classic sex-ratio distorter in *Teleopsis dalmanni*, the Malaysian stalk-eyed fly, to test how drivers increase their transmission during spermatogenesis and affect the transcriptomic landscape of the testes. *T. dalmanni* harbours an X-linked meiotic driver in both wild and captive populations (Presgraves et al., 1997; Wilkinson et al., 1998) where drive males produce in excess of 90% female offspring (Presgraves et al., 1997). The driver X chromosome has multiple impacts on individual fitness (Cotton et al., 2014; Finnegan et al., 2019; Meade et al., 2019; Bradshaw et al., 2022; Bates et al., 2023). Most notably, it reduces eye-stalk length in drive males, a sexual ornament used by females to choose males, and so makes them less attractive to females (Wilkinson et al., 1998; Cotton et al., 2014). There does appear to be compensatory evolution for drive individuals to match the ejaculate size of standard individuals (Meade et al., 2019; Bates et al., 2023), most likely via increased testes size (Bradshaw et al., 2022). However, exactly how this is achieved is not clear. Recent work has shown that the driver X harbours a number of inversions relative to the standard X, and bulk expression analyses have revealed significant differential expression between driver and standard male testes (Reinhardt et al., 2014, 2023). However, little is known about the molecular mechanism of the driver and its consequences for spermatogenesis and gene regulation in the testes more broadly.

We first characterise the genomic landscape of the testes throughout spermatogenesis in standard males. X-chromosome expression dynamics across spermatogenesis are somewhat

in dispute, with a lack of consensus on the processes of dosage compensation and meiotic sex chromosome inactivation across insects (Page et al., 2023; Robben et al., 2024; Wei et al., 2024). Both of these have significant implications for how evolution acts on the X throughout spermatogenesis (Witt et al., 2019; Xia et al., 2020), as they dictate which portions of the genome are exposed to selection. We then show how these patterns persist in the presence of the X-linked meiotic driver. Finally, we provide several candidate genes that show diverged expression patterns across spermatogenesis, utilising the time-series nature of the scRNA-seq dataset. These data both remove the bias of bulk RNA-seq approaches and afford us a high-dimensional perspective of the driver's impact on spermatogenesis, tissue structure and X-linked expression.

5.3 RESULTS AND DISCUSSION

We generated eight scRNA-seq datasets from the testes of four standard *T. dalmanni* males and four males carrying the X-linked meiotic driver, referred to as ST and SR respectively. Following quality control and filtering, we recovered 12,546 cells, in which 12,452 genes were expressed, with 4,548 cells from standard individuals and 7,998 cells from drive individuals (Table S5.1).

5.3.1 Single-cell atlas of the Teleopsis dalmanni testes

Following the clustering of cells via expression patterns, we used orthologs of cell-type-specific markers for *Drosophila melanogaster* testes (Witt et al., 2019; Mahadevaraju et al., 2021; Li et al., 2022) (Figures 5.1b, S5.1c, & Table S5.2, Appendix C Supplementary Results) to identify seven distinct cell types, four of these comprising different stages of spermatogenesis (Figure 5.1a). We identified somatic muscle and two groups of cyst cells, the latter of which supports germline development. Another identified cluster corresponded to the germline stem cells (GSC) and the spermatogonia they produce. Finally, we were able to distinguish the primary and secondary spermatocytes, which enter meiosis to produce haploid spermatids.

We then used additional approaches to validate these cell types. Firstly, we used the number of expressed genes to confirm the stages of the germline across spermatogenesis. Previous studies in insect testes have shown that the total number of genes expressed varies significantly across spermatogenesis. Transcriptional activity in the germline peaks before the onset of meiosis, in primary spermatocytes, following which transcription dramatically reduces in spermatids (Barreau et al., 2008; Witt et al., 2019; Page et al., 2023; Raz et al., 2023; Wei et al., 2024). Consistent with this, we find a clear decrease in the number of expressed

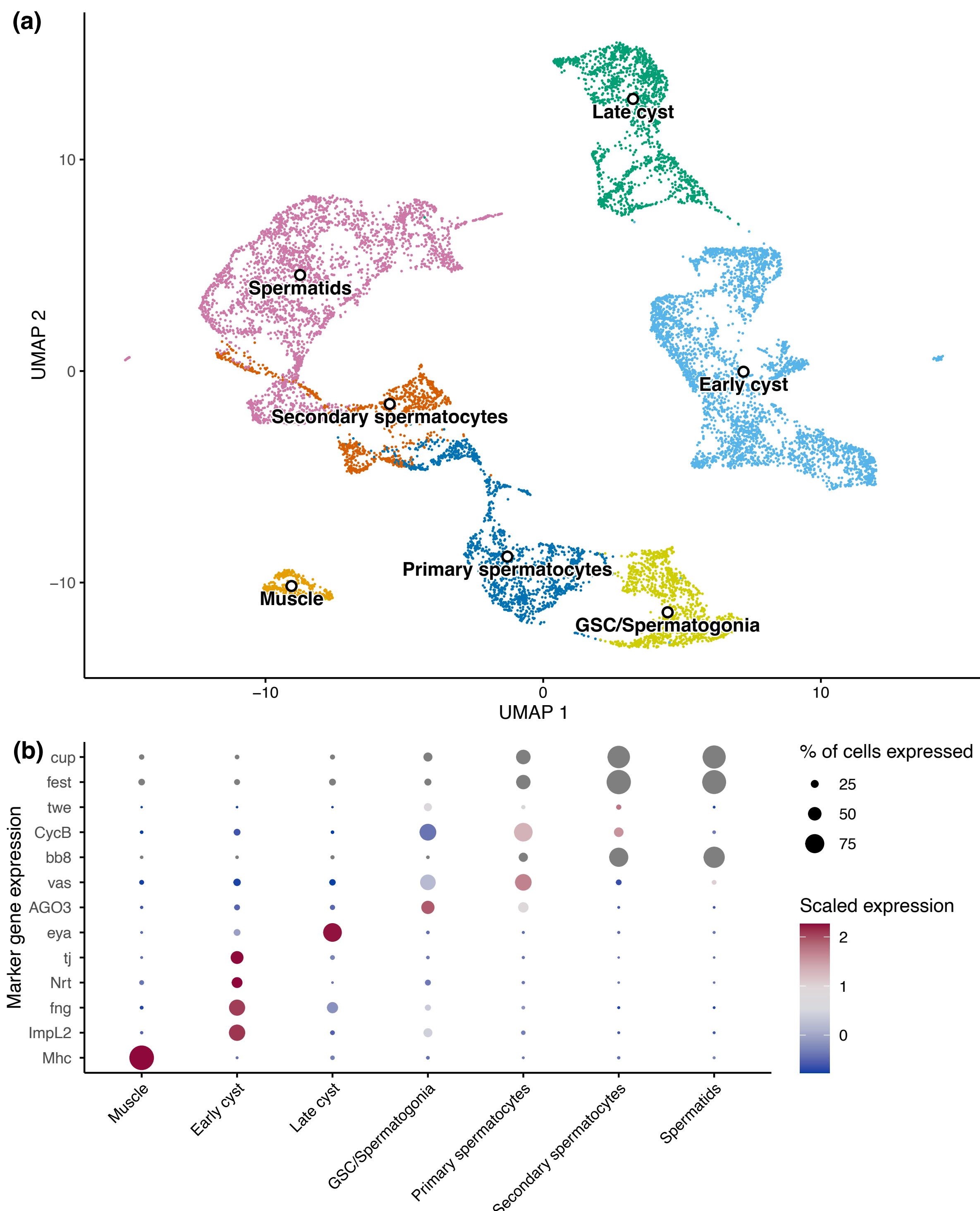


Figure 5.1: Single-cell atlas of the Teleopsis dalmanni testes

__(a)__ Uniform Manifold Approximation and Projection (UMAP) of identified cell types from the Teleopsis dalmanni testes. __(b)__ Dot plot of relative expression of orthologs of key Drosophila melanogaster cell-type-specific testes markers. Size of dots indicates the relative number of cells expressing the marker in a cluster and colour indicates the scaled level of expression (blue lowest and red highest).

autosomal genes over spermatogenesis (Figures 5.2a, 5.2b & S5.1b), supporting our separation of spermatocytes into primary and secondary spermatocytes. Our trajectory analysis, where cells are assigned pseudotimes across a developmental trajectory, further supports this pattern of expression change over developmental time (Figure 5.2c). Secondly, we used eukaryotic classifiers of the mitotic cycle stage to corroborate our classification of primary and secondary spermatocytes (Figure S5.1a, Table S5.3, Appendix C Supplementary Results). Interestingly, we were unable to use ploidy to distinguish pre- from post-meiotic cell types as proposed by a recent study (Robben et al., 2024). We hypothesise several reasons for this and present data urging caution when undertaking this approach with scRNA-seq data (Appendix C Supplementary Results). Finally, we generated a comprehensive list of markers which are robustly differentially expressed between these cell types for future studies (Table S5.4).

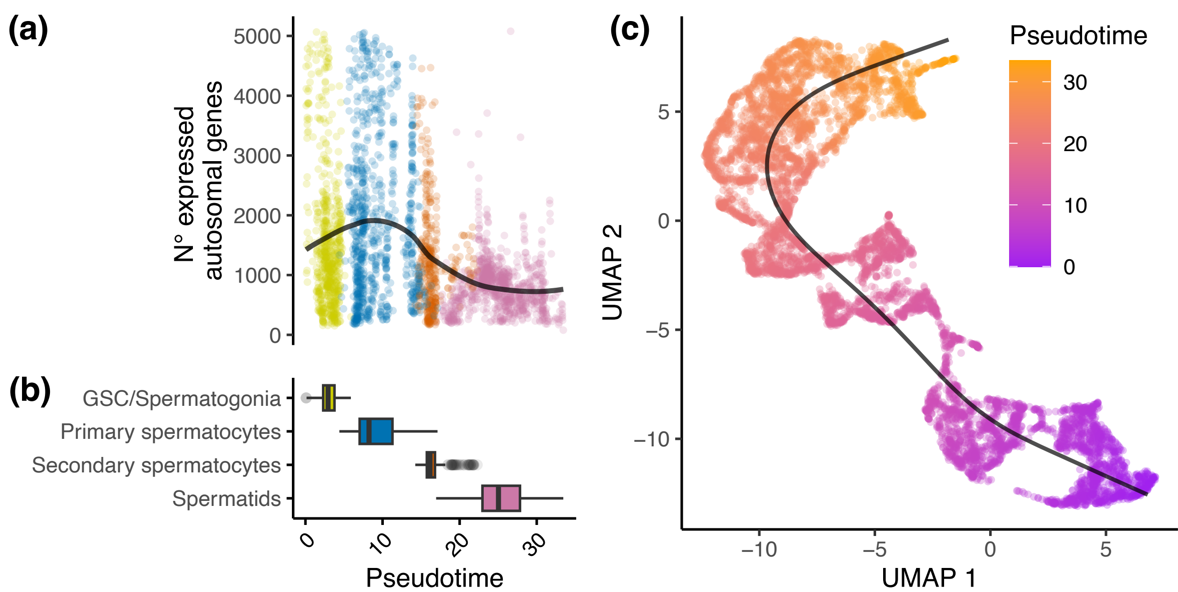


Figure 5.2. Genome wide expression patterns across T. dalmanni spermatogenesis

(a) *Number of autosomal genes expressed across spermatogenesis per cell (gene classified as expressed if counts > 1). Data shown for standard (ST) males. Colours indicate different cell types as shown in panel (b).* ***(b)*** *Boxplot of cell type abundances across pseudotime.* ***(c)*** *UMAP of germline cells, coloured by pseudotime. Plotted line is the principal curve fitted through the centre of the data by Slingshot.*

5.3.2 Lack of meiotic sex chromosome inactivation in Teleopsis dalmanni

Next, we characterised patterns of expression across *T. dalmanni* testes cell types, with a particular focus on the X chromosome. Due to their unique inheritance pattern and characteristics, the X chromosomes frequently exhibit sex- and cell-type specific gene regulation compared to the rest of the genome (Rice, 1984; Charlesworth, 1996; Parisi et al., 2003; Carrel & Willard, 2005; Turner, 2007; Mank, 2013; Bachtrog et al., 2014; Lucchesi & Kuroda, 2015).

Firstly, using our scRNA-seq data, we examined expression of the X chromosome across the four stages of *T. dalmanni* spermatogenesis that we identified to test for meiotic sex chromosome inactivation. Meiotic sex chromosome inactivation acts in many species, inhibiting expression of the X chromosome during the meiotic stages of spermatogenesis (Turner, 2007, 2015). A number of theories have been suggested to explain its evolution (Turner, 2007), including as a mechanism to ensure proper segregation of the X and Y during meiosis by preventing their recombination, as a result of the demasculinisation and loss of spermatogenesis related genes on the X (Wu & Xu, 2003), or due to the inactivation of selfish genetic elements (Meiklejohn & Tao, 2010). However, despite its suggested evolutionary importance, the status of meiotic sex chromosome inactivation in insects has remained controversial (Wu & Xu, 2003; Hense et al., 2007; Meiklejohn et al., 2011; Mikhaylova & Nurminsky, 2011; Vibranovski, 2014; Landeen et al., 2016; Mahadevaraju et al., 2021; Djordjevic et al., 2024). In part, this uncertainty arises from the methodological challenges of manually dissecting specific cell populations from testes of multiple individuals. Recent scRNA-seq data in a handful of model species has circumvented these challenges and provided new insights into the status of meiotic sex chromosome inactivation (Mahadevaraju et al., 2021; Witt et al., 2021; Page et al., 2023; Wei et al., 2024), however, it remains unclear how conserved this regulatory process is across insects more generally.

We find no evidence for meiotic sex chromosome inactivation in *T. dalmanni*, with a substantial number of expressed X-linked genes relative to autosomal genes across all stages of spermatogenesis (Figures 5.3a & 5.3b). Indeed, in meiotic and post-meiotic germline cells we find this ratio to be equivalent to or greater than in somatic cells (Figure S5.3). This lack of meiotic sex chromosome inactivation is consistent with recent scRNA-seq data in other Brachycera dipterans including *Drosophila melanogaster* (Witt et al., 2021, but see Mahadevaraju et al., 2021) and *Drosophila miranda (Wei et al., 2024)*.

Interestingly, recent observations of meiotic sex chromosome inactivation in insects have been limited to species exhibiting chiasmatic male meiosis, including *Anopheles gambiae* (Benedict

et al., 2003; Page et al., 2023), *Tribolium castaneum* (Johnson, 1966; Robben et al., 2024) and *Timema poppense* (Parker et al., 2022; Djordjevic et al., 2024). The Brachycera suborder of diptera exhibit achiasmatic meiosis, and therefore no recombination in males (Gethmann, 1988). This apparent difference in the status of meiotic sex chromosome inactivation between achiasmatic and chiasmatic species lends further support to a prominent theory that inactivation occurs to protect against harmful recombination between heteromorphic sex chromosomes (McKee & Handel, 1993)

5.3.3 Status of dosage compensation varies across testes cell types

We then examined patterns of dosage compensation across testes cell types. Dosage compensation is predicted to evolve on the X chromosome when the X and Y diverge in sequence (Ohno, 1966). This is thought to equalise the expression of sex chromosomes and autosomes in both sexes and mitigate the costs of hemizygous X expression in the heterogametic sex. However, the completeness of dosage compensation varies substantially across species and tissues, particularly between gonadal and somatic tissue (Julien et al., 2012; Mank, 2013; Djordjevic et al., 2024). For instance, previous work in the stalk-eyed fly has suggested equal expression between the X and autosomes in somatic tissue but incomplete compensation in the testes (Wilkinson et al., 2013; Baker et al., 2016). This pattern is found across many other species (Julien et al., 2012; Djordjevic et al., 2024) but the exact reasons still remain unclear (Gu & Walters, 2017). It has been suggested that differences in the magnitude of sexual conflict over optimal expression levels could be responsible (Mullon et al., 2015). However, with a handful of recent exceptions (Witt et al., 2021; Page et al., 2023; Robben et al., 2024; Wei et al., 2024), studies of dosage compensation measure aggregate expression across entire tissues or body regions, potentially masking variability in dosage compensation status across cell types or drawing inaccurate conclusions about the presence or absence of compensation. This is particularly consequential for the testes which are composed of both somatic and germline cell types.

Our scRNA-seq dataset reveals a complex pattern of dosage compensation in the *T. dalmanni* testes across spermatogenesis (Figure 5.3c). We show that somatic cell types exhibit equal expression of the autosomes and the X chromosome in males, consistent with complete dosage compensation. Interestingly, the early stages of spermatogenesis appear to lack dosage compensation, with expression of the X close to half that of the autosomes, whilst equal expression is restored during the later stages (Figure 5.3c).

When this is broken down by chromosome, we show that the expression of both autosomes is constant across testes cell types, but the X exhibits a clear reduction in expression in the GSC/spermatogonia and primary spermatocytes (Figure S5.4a). This is opposite to the early germline pattern of incomplete dosage compensation of the X in other insect species, including *Drosophila* (Wei et al., 2024), and *Anopheles* (Page et al., 2023). In these species, the X is instead expressed at comparable levels between somatic cell types and the pre-meiotic germline before becoming progressively downregulated through spermatogenesis from meiosis onwards.

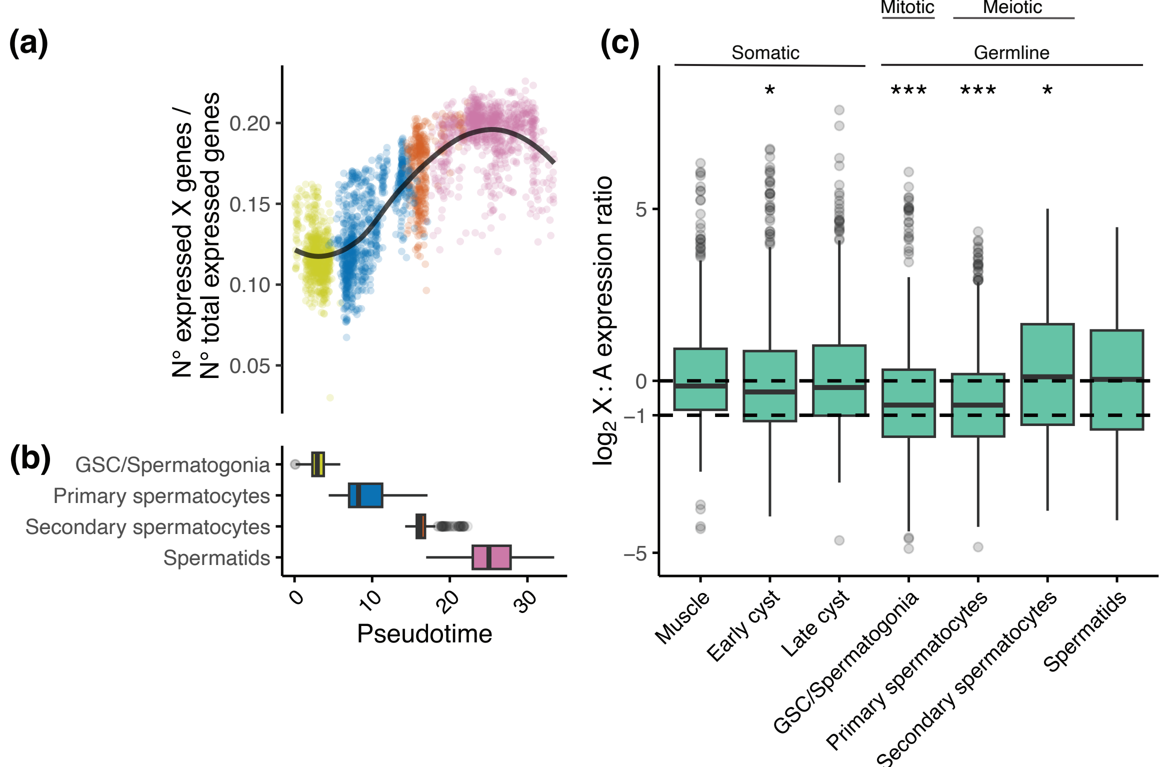


Figure 5.3. Expression of the X chromosome across T. dalmanni spermatogenesis

(a) *Relative number of X-linked genes expressed across spermatogenesis in standard (ST) males. For each cell, the number of expressed X-linked genes divided by the number of expressed autosomal genes is shown (gene classified as expressed if counts > 1). Colours indicate different cell types as shown in panel (b).* ***(b)*** *Boxplot of cell type abundances across pseudotime. The GSC and spermatogonia are mitotic germline cell types whereas spermatocytes are meiotically active.* ***(c)*** *Box plots of X-linked gene expression compared to median autosomal expression across cell types. Line at 0 represents even expression of autosomal and X-linked genes and at -1 represents 50% X-linked expression. A two-sided*

*Wilcox test was used to determine if $log_2(X:A)$ values for each cell type deviate from 0. $p < 0.00001$ = ***, $p < 0.001$ = **, $p < 0.05$ = *.*

Previous work has suggested that the *T. dalmanni* X chromosome is highly enriched for testes-specific genes, with almost twice as many as expected on the basis of its size. This appears to be the result of the migration of testes-specific genes from the autosomes, the evolution of testes-specific expression for ancestral X-linked genes, and the emergence of new genes on the X with expression limited to the testes (Baker et al., 2016). This pattern of masculinisation is not found on the *Drosophila* X (Sturgill et al., 2007; Vibranovski et al., 2009) and so could explain the relative increase in X-linked expression we observe later in *T. dalmanni* spermatogenesis, where male-benefit genes are disproportionately expressed in mature sperm. It is also possible that the upregulation of the X we observe here is actually a false signal driven by Y-linked genes that share sequence similarity to the X and are expressed later in spermatogenesis. However, this is unlikely as the *T. dalmanni* sex chromosomes are thought to be highly diverged (Baker & Wilkinson, 2010), where only one Y-linked gene has been identified to date (ORF-126) (Baker & Wilkinson, 2010) which was not expressed in our dataset. Furthermore, the increase in expression appears not to be limited to a handful of highly expressed genes on the X but looks to be evenly distributed across the entire chromosome (Figure S5.4b), hinting at a chromosome-wide mechanism of upregulation.

Interestingly, whilst we found differences in patterns of germline dosage compensation between *Drosophila* and *T. dalmanni*, orthologs of male-specific lethal (MSL) genes, key components of the dosage compensation complex (DCC) in *Drosophila* (Lucchesi & Kuroda, 2015), show similar expression (Witt et al., 2021; Wei et al., 2024). Specifically, as in *Drosophila*, we find that MSL genes exhibit a gradual reduction in expression during stalk-eyed fly spermatogenesis (Figure S5.5). There is currently mixed evidence for the role of this complex in regulating dosage compensation in the germline (Conrad & Akhtar, 2012; Witt et al., 2021; Robben et al., 2024; Wei et al., 2024), particularly as the MSL complex does not localise to the X in the male germline (Rastelli & Kuroda, 1998). Given the lack of concordance we find between MSL expression and patterns of dosage compensation, our results further support a non-canonical mechanism of dosage compensation in insect testes.

5.3.4 Impacts of meiotic drive on the cellular landscape of the testes

The mechanisms by which drivers bias their transmission to gametes have only been studied in a handful of species (Courret et al., 2019), however, they appear to operate through two main approaches, either by killing gametes directly or halting their maturation. We compared

the cellular composition of the testes across spermatogenesis to test which of these mechanisms is operating in *T. dalmanni*. In this species, sperm are formed in bundles, composed of 128 germ cells housed in two cyst cells (Presgraves et al., 1997). If the driver acts by killing Y bearing sperm directly, we expect to see a relative depletion in the number of germ cells found in later stages of spermatogenesis in drive males. This is because if Y bearing sperm were killed or not formed during meiosis, we would expect bundles to contain less than 128 germ cells, and thus a relatively smaller number of germ cells relative to cyst cells in drive (SR) compared to standard (ST) males. Alternatively, if the driver prevents Y bearing sperm from fully maturing, we expect to see no difference or even a relative increase in cell numbers towards the end of spermatogenesis. This is because immobilised or improperly elongated Y (or O if improper segregation at meiosis) sperm may be unable to migrate to the seminal vesicle and so temporarily accumulate in the testes.

Notably, we see no significant effect of drive on cell type abundance when comparing the size of the cyst to the germline (Table S5.1, $p = 0.48$). This lack of a clear difference in the number of cells progressing through spermatogenesis between standard and drive males suggests there is no sudden sperm cull, but instead that the driver causes incomplete spermatid maturation. Therefore, once Y bearing sperm are immobilised, they may simply build up in the distal end of the testes before being eliminated by standard cellular programs. Indeed, we do observe a non-significant enrichment for post-meiotic germ cells in drive (SR) individuals (Table S5.1, $p = 0.09$). Together, our findings are consistent with cytological work in *T. dalmanni* where sperm of drive males reach the later stages of spermatogenesis, but, just before individualization, sperm heads either deteriorate before leaving the bundle or appear overextended (Presgraves et al., 1997). Together, this pattern is analogous to the *Segregation Distorter* (SD) male meiotic drive system in *Drosophila melanogaster*, where the driver operates post-meiotically to prevent sperm maturation (Herbette et al., 2021).

5.3.5 Impacts of meiotic drive on the transcriptional landscape of the testes

We expect standard and drive males to exhibit differential gene regulation due to both direct and indirect consequences of meiotic drive. Meiotic drivers are frequently housed by inversions (Stalker, 1961; Dyer et al., 2007; Pieper & Dyer, 2016) and the *T. dalmanni* driver X is no exception (Johns et al., 2005). A consequence of these inversions is reduced recombination for both the standard and driver X chromosomes, leading to high sequence divergence between X types and low diversity within the drive X (Reinhardt et al., 2023). Furthermore, inversions may directly disrupt cis-regulation by physically shuffling promoters

and enhancers within a chromosome (Kleinjan & Coutinho, 2009). Therefore, we tested the consequences of the meiotic driver for the evolution of gene expression across spermatogenesis (Figure 5.4).

First, as in standard males, we find no evidence for meiotic sex chromosome inactivation (Figures 5.4a, 5.4b & S5.6) and conserved patterns of dosage compensation (Figure 5.4c). The only exception is a marginal increase in relative X-wide expression in primary spermatocytes in drive relative to standard males ($p < 0.05$). Together, these patterns suggest that there is no widespread dysregulation of the X chromosome across spermatogenesis as a consequence of drive.

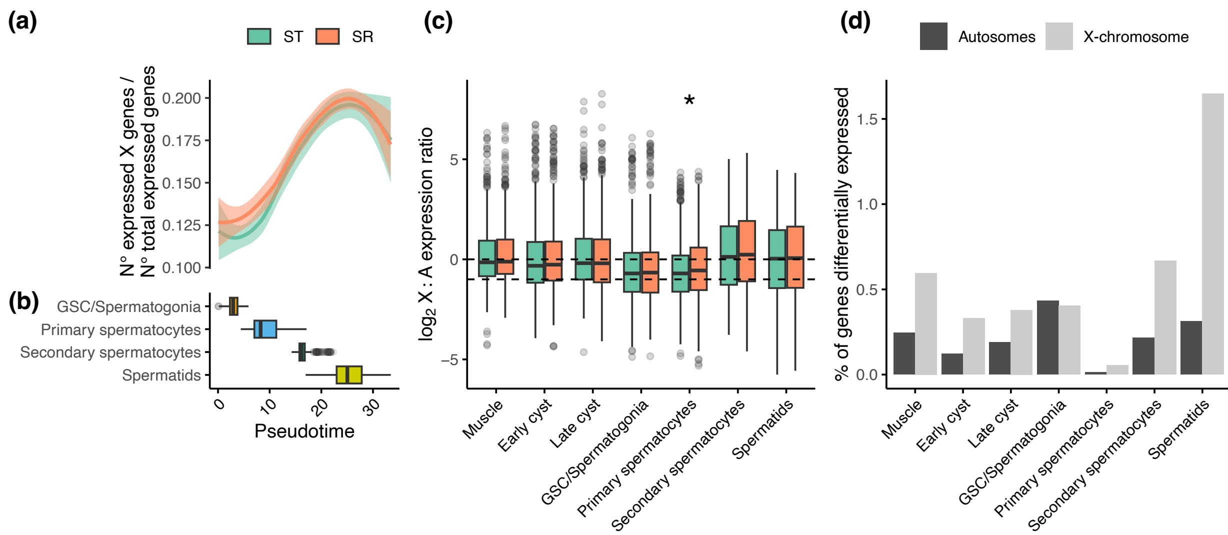


Figure 5.4. Conservation of X-linked regulation in the germline of drive males

(a) *Loess curves fit to the relative number of X-linked genes expressed across spermatogenesis for standard (ST) and drive (SR) male cell types separately. For each cell, the number of detected X-linked genes divided by the number of expressed autosomal genes (gene classified as expressed if counts > 1) is shown. Filled area is the standard deviation.* ***(b)*** *Boxplot of cell type abundances across pseudotime.* ***(c)*** *Boxplots of log_2(X:A) ratios across cell types in standard (ST) and drive (SR) males. Line at 0 represents even expression of autosomal and X-linked genes and at -1 represents 50% X-linked expression and complete lack of dosage compensation. A two-sided Wilcox test was used to determine if values for each cell type varied between ST and SR individuals. $p < 0.05$ = ** ***(d)*** *Proportions of expressed autosomal (dark grey) or X-linked (light grey) genes in each cell type that were differentially expressed between ST and SR individuals.*

Strikingly, we observe only a limited number of genes that are differentially expressed between standard and driver males across either the autosomes or X chromosome in each cell type (Figure S5.7, Tables S5.6 & S5.8). This contrasts previous bulk RNA-seq studies in *T. dalmanni* (Reinhardt et al., 2014, 2023) that suggest a significant portion of the genome has diverged in expression (~600 genes) in response to the meiotic driver. Notably, only one of the genes, *mcm10*, that Reinhardt et al. (2023) identified was also differentially expressed in our dataset. This discrepancy is most likely an outcome of measuring differential expression from bulk approaches, which represent an average of expression across entire populations of distinct cell types and can lead to false inferences of regulatory variation (Montgomery & Mank, 2016; Price et al., 2022a, 2022b).

The differentially expressed genes we do observe are disproportionately located on the X chromosome across most cell types (Figure 5.4d, Tables S5.6, S5.7 & S5.8). Additionally, these genes are evenly distributed across the entire X and not localised to a specific region (Figure S5.8). Whilst we cannot distinguish whether X-linked enrichment is a result of indirect or direct effects of the driver, interestingly, this pattern is most pronounced in the spermatids (p = 0.006), with the vast majority of differentially expressed genes being X-linked (42/50). This is consistent with the earlier finding that the driver likely acts through improper maturation of spermatids (Presgraves et al., 1997).

Similarly, we also observe few genes with significantly differential trajectories between standard (ST) and drive (SR) spermatogenesis. Trajectory analyses allow us to test whether genes are differentially regulated across developmental time, rather than at distinct, self-assigned timepoints that include cells spanning developmental states. Briefly, we assigned pseudotime points to each germ cell and fitted a Generalised Additive Model (GAM) for standard (ST) and drive (SR) cells independently for expression of each gene. To ensure high confidence in the identified trajectories, genes with no association between pseudotime and expression were removed. Of the 5047 genes maintained for this analysis, only 260 had significantly differential trajectories between standard (ST) and drive (SR) spermatogenesis (Table S5.9). These genes are enriched for gene ontology terms including cilia, axoneme, and cell projection assembly (Table S5.10), suggesting that the driver may interfere with proper sperm development by affecting sperm motility.

Using our analyses of differential expression and trajectory, we sought to identify candidates for functional aspects of the driver. Of those we found, many are associated with sperm motility and function (Figure S5.9). Notably, *Grip75* is required for tethering of microtubules, and

Drosophila mutants of this gene are sterile with defects in meiosis and sperm motility (Vogt et al., 2006). Consistent with this phenotype, *Grip75* is X-linked and expressed at significantly lower levels in *T. dalmanni* drive individuals, particularly in the later stages of spermatogenesis. Similarly, we identify differential expression of *Tsr, Drosophila* mutants of which are unable to perform proper cytokinesis at meiosis I and II (Gunsalus et al., 1995). This gene is X-linked in *T. dalmanni* and has significantly lower expression in spermatocytes of drive individuals and exhibits a different trajectory between standard and drive spermatogenesis. Finally, *Ced-12* is required for apoptotic cell clearance (Van Goethem et al., 2012) with *Drosophila* mutants showing significantly increased spermatogonia volume (Zohar-Fux et al., 2022). *Ced-12* is X-linked in *T. dalmanni* and downregulated in the later stages of spermatogenesis in drive individuals. With strong selection pressure for increased germline size acting in drive males (Bradshaw et al., 2022), this shift in regulation of germline growth is a clear mechanism to mitigate the loss of Y-bearing sperm.

5.4 CONCLUSION

In conclusion, we describe the cellular and transcriptional landscape of the testes of the stalk-eyed fly. Specifically, we show limited evidence for meiotic sex chromosome inactivation and unique patterns of dosage compensation across spermatogenesis, relative to both other dipterans and insects in general. We show that the driver likely acts by interfering with proper sperm development, rather than directly killing gametes, and provide evidence that it does so via affecting motility of sperm cells.

5.5 METHODS

5.5.1 Reference genome and mitochondrial genome assembly

A reference genome for *Teleopsis dalmanni* is publicly available (van Rensburg et al., 2024) and consists of two autosomes and an X chromosome. However, it lacks a mitochondrial sequence. We therefore assembled a mitochondrial genome using publicly available PacBio Hifi reads generated from pooled *T. dalmanni* larvae (van Rensburg et al., 2024) and MitoHifi v3.01 (Uliano-Silva et al., 2023). Specifically, we used raw Hifi reads as input, the rust fly (*Loxocera sinicia*) mitochondrial genome as a reference, and MitoFinder to annotate the genome, to produce a circularised assembly 20,708bp in length containing 37 genes. This *T. dalmanni* mitochondrial reference is available at https://doi.org/10.5061/dryad.brv15dvk3. The

reference genome also lacks a Y chromosome, but it is thought to be highly diverged from the X and contain only a handful of genes (Baker & Wilkinson, 2010).

5.5.2 Sample collection

Flies were reared at University College London from a wild-caught population originating from the Gombak Valley, Malaysia. All flies and larvae were incubated and reared at 25°C and fed on a diet of sweetcorn. To ensure known genotypes of samples, a homozygous drive (SR) population is maintained through a series of crosses as previously described (Presgraves et al., 1997). Eight adult males, four standard (ST) and four with drive (SR), were sacrificed before the dissection of both testes in iced phosphate-buffered saline (PBS). These adults were all virgins, reproductively mature and reared from egg lays collected on the same day.

5.5.3 Tissue collection, dissociation and single-cell sequencing

Testes pairs were individually dissociated by incubation in a collagenase-TrypLE lysis solution (10mg/ml collagenase in 10X TrypLE) at 37.5°C for one hour with three sets of mechanical dissociation by trituration of wide then narrow bore Pasteur pipettes. Digestion was inhibited by the addition of iced Schneider's Serum. The solution was then gently triturated with a narrow-bore Pasteur pipette before filtering through a 35µm filter pre-rinsed with Schneider's Serum. The sample was then spun in a swing bucket centrifuge for 5 minutes at 1000xg and 4°C. The supernatant was removed and the pellet resuspended in 50µl of iced PBS with gentle pipetting of a wide-bore pipette. To count cells, 10µl of the suspension was combined with 10µl of trypan blue and placed onto a humidified haemocytometer plate before counting in triplicate.

10X Genomics Chromium transcriptome libraries were generated at the NERC Environmental Omics Facility (NEOF) Liverpool before sequencing with Illumina NovaSeq using S2 chemistry, aiming for recovery of ~10,000 cells per sample and ~20,000 reads per cell. Raw scRNA-seq data for eight males is available at https://doi.org/10.5061/dryad.zkh1893kb.

5.5.4 Single-cell RNA-seq data processing

Sequencing data for each sample was processed using Cell Ranger v7.2.0 (Zheng et al., 2017). First, a custom reference genome was built with the *T. dalmanni* reference genome using mkref. Using cellrangers count function, fastq reads were then aligned against the custom index and counted, creating gene-by-cell count matrices. Data filtering and downstream analyses were performed using Seurat v5.0.3 (Hao et al., 2024) in R v4.3.2 (R

Core Team, 2021). Cells in each sample were filtered for a minimum of 200 features expressed and less than 20% mitochondrial expression. Count data for each sample was also filtered by only keeping genes with expression in at least three cells. We used DoubletFinder v2.0 (McGinnis et al., 2019) in R (R Core Team, 2021) with default parameters to identify and remove doublets. The filtered dataset consisted of 12,546 cells across the eight samples, expressing 12,452 genes, with 7,998 cells from drive individuals and 4,548 cells from standard individuals (Table S5.1). Seurat objects from all eight samples were then integrated post-filtering using the 'SCTransform' function (Hafemeister & Satija, 2019).

5.5.5 Cell-type identification

After running a PCA on the integrated Seurat object, we used the ElbowPlot function to identify how many PCs were necessary to describe a significant amount of variation. Subsequently, a nearest neighbour graph was created using FindNeighbors, and clusters at varying resolutions identified with FindClusters. From this, an appropriate resolution for the number of clusters was determined using the clustree package v0.5.1 (Zappia & Oshlack, 2018), giving a final cluster number of 16.

Using a series of cell-type-specific markers for *Drosophila melanogaster* testes (Table S5.2) (Witt et al., 2019; Mahadevaraju et al., 2021; Li et al., 2022), clusters were assigned into biological groupings. Orthology between *T. dalmanni* and *D. melanogaster* reference genome (dm6) was established using OrthoFinder v2.5.5 with default parameters (Emms & Kelly, 2019) giving a total of 9,883 reciprocal orthologs. Distinguishing cell populations in non-model organisms relies primarily on databases of marker genes from model species, which are often distantly related. Our comparison between *T. dalmanni* and *D. melanogaster* (with a divergence time of ~150 MY) is within the range of species pairs previously employed to identify cell types using orthologous marker genes in recent single-cell RNA-seq studies (Segami et al., 2022; Robben et al., 2024; Darolti & Mank, 2023).

To further validate cell types, we used information on the number of features expressed and classifiers of the mitotic cycle stage (Table S5.3). Finally, to remove unwanted or unknown clusters, we cleaned the data by removing clusters that (a) had no clear biological classification, (b) were predominantly represented by a single sample (a cluster must have at least two samples from a treatment representing >12.5% of the cells, weighted by total cell number for each sample) and (c), had doublet-like expression profiles (high numbers of features and expression of markers from divergent cell types). After identifying cell types, new

markers were identified on the basis of differential expression using FindAllMarkers from the Seurat package.

5.5.6 Dosage compensation analysis

For each sample and cell type, expression values were aggregated using a pseudobulk approach with scuttle v1.14.0 (McCarthy et al., 2017). Specifically, the expression counts of all cells belonging to a cell type were summed for each gene across the genome. Using this approach instead of the expression of each cell reflects that the sample is the biological replicate and not the cell itself (Lun and Marioni 2017). Genes for each cell type were then filtered in two ways. First, genes were kept if they were expressed in >= 5% of cells (> 1 count). Second, genes had to have a pseudo-bulk $\log_2$(CPM) (counts per million) > 2 in more than half of standard (ST) or drive (SR) males. Dosage was measured as the ratio of expression of X-linked genes to the median autosomal expression ($\log_2$(X:A). In each cell type, a non-parametric two-sided Wilcoxon test was used to test for deviations of $\log_2$(X:A) from 0 in standard (ST) males, with μ set to 0, and differences in $\log_2$(X:A) between SR and ST males.

5.5.7 Differential abundance analysis

To test for differences in cell type abundance between standard (ST) and drive (SR) males, a series of binomial models were fit comparing cell counts of germline to cyst, early cyst to late cyst, and pre-meiotic germline to post-meiotic germline. All models were run using glmer from lme4 (D. Bates et al., 2015) in R (R Core Team, 2021) with sample as a random effect (count~treatment + (1|sample)).

5.5.8 Differential gene expression analysis

A quasi-likelihood (QL) approach from EdgeR v4.0.16 (Robinson et al., 2010) was used to identify differentially expressed genes between standard (ST) and drive (SR) males in each cell type ($|\log_2$(fold-change)$| > 1$ and FDR < 0.05). Enrichment of the number of differentially expressed genes across chromosomes and cell types was modelled with a generalised linear model of family 'binomial' where genes were classified as biased or unbiased and regressed against cell type and chromosome. Nested models were then compared using chi-squared in anova.glm from the R 'stats' package (R Core Team, 2021).

5.5.9 Trajectory Analysis

To identify genes with differential trajectories between standard (ST) and drive (SR) males across spermatogenesis, data was subset to include only germline cells (GSC/spermatogonia, primary and secondary spermatocytes, and spermatids). The Seurat object was then converted into a SingleCellExperiment class for downstream analysis using tradeSeq v1.18.0 (Van den Berge et al., 2020) and slingshot v2.12.0 (Street et al., 2018). First, pseudotimes were assigned to each cell within the germline. Then a negative binomial generalized additive model (NB-GAM) with 10 knots was fit to each gene for ST and SR individuals separately. Genes were kept if they were expressed in at least 10% of either ST or SR cells with 2 or more counts (Van den Berge et al., 2020), and if their expression was significantly associated with pseudotime in either ST or SR cells ($p < 0.05$ & $\log_2$ fold-change > 1). GAM smoothers were then compared between ST and SR cells to identify genes with significantly different trajectories using the conditionTest function. Genes were classed as significant if false discovery rate $p < 0.05$ & $\log_2$ fold-change > 2. Additionally, the remaining genes were then only kept if they were expressed in at least 10% of either ST or SR cells with 2 or more counts (Van den Berge et al., 2020). The ST and SR NB-GAMs are then compared to find differential trajectories.

5.5.10 Gene Ontology enrichment analysis

A gene ontology term enrichment of the genes with significant differential trajectories was performed using the clusterProfiler package v4.12.2 (Yu et al., 2012). The background gene set used was the genes that were previously identified as having a significant association with pseudotime. The org.Dm.eg.db v3.19.1 (Carlson, 2019) package for *D. melanogaster* was used as a reference database.

5.6 ACKNOWLEDGEMENTS

This work was funded by a NERC Independent Research Fellowship to AEW (NE/N013948/1), a NERC Exploring the Frontiers Grant to AEW and AP (NE/X009734/1), and a NERC ACCE DTP to PDP. The laboratory work was supported by the UK Natural Environment Research Council (NERC) Environmental Omics Facility. We thank Georgia Milne for helpful comments and suggestions on the manuscript. We acknowledge IT Services at The University of Sheffield for the provision of services for High Performance Computing.

5.7 AUTHOR CONTRIBUTIONS

AEW and PDP designed the research. SLB, SB and AP reared the flies and provided expertise with dissections. PDP collected and processed the data that was sequenced by SP and NEOF Liverpool. PDP, SP, VJL analysed the data with assistance from AEW, ID, SB and AP. AEW and PDP wrote the manuscript with input from all authors.

5.8 DATA AND CODE AVAILABILITY STATEMENT

T. dalmanni mitochondrial reference is available at https://doi.org/10.5061/dryad.brv15dvk3.

Raw scRNA-seq data for 8 males is available at https://doi.org/10.5061/dryad.zkh1893kb.

Code is available at https://github.com/petedprice/scStalkie_Drive

5.9 REFERENCES

Ardlie, K. G. (1998). Putting the brake on drive: meiotic drive of t haplotypes in natural populations of mice. *Trends in Genetics*, *14*(5), 189–193.

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

Baker, C. C., Gim, B. S., & Fuller, M. T. (2015). Cell type-specific translational repression of Cyclin B during meiosis in males. *Development, 142*(19), 3394–3402.

Baker, R. H., Narechania, A., DeSalle, R., Johns, P. M., Reinhardt, J. A., & Wilkinson, G. S. (2016). Spermatogenesis drives rapid gene creation and masculinization of the X chromosome in stalk-eyed flies (Diopsidae). *Genome Biology and Evolution*, *8*(3), 896–914.

Barreau, C., Benson, E., & White-Cooper, H. (2008). Comet and cup genes in *Drosophila* spermatogenesis: the first demonstration of post-meiotic transcription. *Biochemical Society Transactions*, *36*(Pt 3), 540–542.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bates, S., Meade, L., & Pomiankowski, A. (2023). Meiotic drive does not impede success in sperm competition in the stalk-eyed fly, *Teleopsis dalmanni*. *Evolution*, *77*(10), 2326–2333.

Benedict, M. Q., McNitt, L. M., & Collins, F. H. (2003). Genetic traits of the mosquito *Anopheles gambiae*: red stripe, frizzled, and homochromy1. *The Journal of Heredity*, *94*(3), 227–235.

Blackmon, H., Justison, J., Mayrose, I., & Goldberg, E. E. (2019). Meiotic drive shapes rates of karyotype evolution in mammals. *Evolution*, *73*(3), 511–523.

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience*, *10*(2), giab007.

Bradshaw, S. L., Meade, L., Tarlton-Weatherall, J., & Pomiankowski, A. (2022). Meiotic drive adaptive testes enlargement during early development in the stalk-eyed fly. *Biology Letters*, *18*(11), 20220352.

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., & Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature*, *337*(6205), 373–376.

Carlson, M. (2019). *org.Dm.eg.db: Genome wide annotation for Fly. R package version 3.8.2*.

Carrel, L., & Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, *434*(7031), 400–404.

Cazemajor, M., Joly, D., & Montchamp-Moreau, C. (2000). Sex-ratio meiotic drive in *Drosophila simulans* is related to equational nondisjunction of the Y chromosome. *Genetics*, *154*(1), 229–236.

Cazemajor, M., Landré, C., & Montchamp-Moreau, C. (1997). The sex-ratio trait in *Drosophila simulans*: genetic analysis of distortion and suppression. *Genetics*, *147*(2), 635–642.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, *6*(2), 149–162.

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLoS Genetics*, *8*(9), e1002900.

Conrad, T., & Akhtar, A. (2012). Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nature Reviews Genetics*, *13*(2), 123–134.

Cotton, A. J., Földvári, M., Cotton, S., & Pomiankowski, A. (2014). Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity*, *112*(4), 363–369.

Courret, C., Chang, C.-H., Wei, K. H.-C., Montchamp-Moreau, C., & Larracuente, A. M. (2019). Meiotic drive mechanisms: lessons from *Drosophila*. *Proceedings of the Royal Society B*, *286*(1913), 20191430.

Courtot, C., Fankhauser, C., Simanis, V., & Lehner, C. F. (1992). The *Drosophila cdc25* homolog *twine* is required for meiosis. *Development*, *116*(2), 405–416.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.

Darolti, I., & Mank, J. E. (2023). Sex-biased gene expression at single-cell resolution: cause and consequence of sexual dimorphism. *Evolution Letters*, *7*(3), 148–156.

Djordjevic, J., Tran Van, P., Toubiana, W., Labédan, M., Dumas, Z., Aury, J.-M., Cruaud, C., Istace, B., Labadie, K., Noel, B., Parker, D. J., & Schwander, T. (2024). Dynamics of X chromosome hyper-expression and inactivation in male tissues during stick insect development. *bioRxiv*, 2024.07.01.601468.

Dyer, K. A., Charlesworth, B., & Jaenike, J. (2007). Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1587–1592.

Eickbush, M. T., Young, J. M., & Zanders, S. E. (2019). Killer meiotic drive and dynamic evolution of the *wtf* gene family. *Molecular Biology and Evolution*, *36*(6), 1201–1214.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Erickson, R. P. (1990). Post-meiotic gene expression. *Trends in Genetics*, *6*, 264–268.

Finnegan, S. R., White, N. J., Koh, D., Camus, M. F., Fowler, K., & Pomiankowski, A. (2019). Meiotic drive reduces egg-to-adult viability in stalk-eyed flies. *Proceedings of the Royal Society B*, *286*(1910), 20191414.

Fishman, L., & Willis, J. H. (2005). A novel meiotic drive locus almost completely distorts segregation in *mimulus* (monkeyflower) hybrids. *Genetics*, *169*(1), 347–353.

Gershenson, S. (1928). A new sex-ratio abnormality in *Drosophila obscura*, *Genetics*, *13*(6), 488–507.

Gethmann, R. C. (1988). Crossing over in males of higher Diptera (Brachycera). *The Journal of Heredity*, *79*(5), 344–350.

Gingell, L. F., & McLean, J. R. (2020). A Protamine Knockdown Mimics the Function of Sd in *Drosophila* melanogaster. *G3, 10*(6), 2111–2115.

Greenbaum, M. P., Iwamori, T., Buchold, G. M., & Matzuk, M. M. (2011). Germ Cell

Intercellular Bridges. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a005850.

Gu, L., & Walters, J. R. (2017). Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. *Genome Biology and Evolution*, *9*(9), 2461–2476.

Gunsalus, K. C., Bonaccorsi, S., Williams, E., Verni, F., Gatti, M., & Goldberg, M. L. (1995). Mutations in *twinstar*, a *Drosophila* gene encoding a cofilin/ADF homologue, result in defects in centrosome migration and cytokinesis. *The Journal of Cell Biology*, *131*(5), 1243–1259.

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, *20*(1), 296.

Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, *156*(3774), 477–488.

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, *42*(2), 293–304.

Hauschteck-Jungen, E., & Hartl, D. L. (1982). Defective histone transition during spermiogenesis in heterozygous segregation distorter males of *Drosophila melanogaster*. *Genetics*, *101*(1), 57–69.

Hense, W., Baines, J. F., & Parsch, J. (2007). X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biology*, *5*(10), e273.

Herbette, M., Wei, X., Chang, C.-H., Larracuente, A. M., Loppin, B., & Dubruille, R. (2021). Distinct spermiogenic phenotypes underlie sperm elimination in the Segregation Distorter meiotic drive system. *PLoS Genetics*, *17*(7), e1009662.

Hess, N. K., Singer, P. A., Trinh, K., Nikkhoy, M., & Bernstein, S. I. (2007). Transcriptional regulation of the *Drosophila melanogaster* muscle myosin heavy-chain gene. *Gene Expression Patterns*, *7*(4), 413–422.

Holman, L., Price, T. A. R., Wedell, N., & Kokko, H. (2015). Coevolutionary dynamics of polyandry and sex-linked meiotic drive. *Evolution*, *69*(3), 709–720.

Hurst, L. D., & Pomiankowski, A. (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics*, *128*(4), 841–858.

Jaenike, J. (2001). Sex Chromosome Meiotic Drive. *Annual Review of Ecology, Evolution, and*

Systematics, *32*(Volume 32, 2001), 25–49.

Johnson, G. R. (1966). Recombination differences with reciprocal crosses in *Tribolium castaneum*. *Genetics*, *53*(1), 111–115.

Johns, P. M., Wolfenbarger, L. L., & Wilkinson, G. S. (2005). Genetic linkage between a sexually selected trait and X chromosome meiotic drive. *Proceedings of the Royal Society B*, *272*(1576), 2097–2103.

Julien, P., Brawand, D., Soumillon, M., Necsulea, A., Liechti, A., Schütz, F., Daish, T., Grützner, F., & Kaessmann, H. (2012). Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biology*, *10*(5), e1001328.

Kaufman, R. S., Price, K. L., Mannix, K. M., Ayers, K. M., Hudson, A. M., & Cooley, L. (2020). *Drosophila* sperm development and intercellular cytoplasm sharing through ring canals do not require an intact fusome. *Development* , *147*(22), dev190140.

Kelemen, R. K., Elkrewi, M., Lindholm, A. K., & Vicoso, B. (2022). Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. *Proceedings of the Royal Society B*, *289*(1968), 20211985.

Kettaneh, N. P., & Hartl, D. L. (1976). Histone transition during spermiogenesis is absent in segregation distorter males of *Drosophila melanogaster*. *Science*, *193*(4257), 1020–1021.

Kleinjan, D.-J., & Coutinho, P. (2009). Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. *Briefings in Functional Genomics & Proteomics*, *8*(4), 317–332.

Kruger, A. N., & Mueller, J. L. (2021). Mechanisms of meiotic drive in symmetric and asymmetric meiosis. *Cellular and Molecular Life Sciences*, *78*(7), 3205–3218.

Landeen, E. L., Muirhead, C. A., Wright, L., Meiklejohn, C. D., & Presgraves, D. C. (2016). Sex chromosome-wide transcriptional suppression and compensatory Cis-regulatory evolution mediate gene expression in the *Drosophila* male germline. *PLoS Biology*, *14*(7), e1002499.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760.

Li, H., Janssens, J., De Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W., David, F. P. A., Brbić, M., Spanier, K., Leskovec, J., McLaughlin, C. N., Xie, Q., Jones, R. C., Brueckner, K., Shim, J., Tattikota, S. G., Schnorrer, F., Rust, K., … Zinzen, R. P. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science*, *375*(6584), eabk2432.

Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sánchez C, H. M., Lanzaro, G. C., Schmidt, H., Lee, Y., Marshall, J. M., & Akbari, O. S. (2020). Development

of a confinable gene drive system in the human disease vector *Aedes aegypti*. *eLife*, *9*, e51701.

Lindholm, A. K., Dyer, K. A., Firman, R. C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A. M., Manser, A., Montchamp-Moreau, C., Petrosyan, V. G., Pomiankowski, A., Presgraves, D. C., Safronova, L. D., Sutter, A., Unckless, R. L., Verspoor, R. L., … Price, T. A. R. (2016). The ecology and evolutionary dynamics of meiotic drive. *Trends in Ecology & Evolution*, *31*(4), 315–326.

Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage compensation in *Drosophila*. *Cold Spring Harbor Perspectives in Biology*, *7*(5), a019398.

Lyttle, T. W. (1993). Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends in Genetics*, *9*(6), 205–210.

Mackintosh, C., Pomiankowski, A., & Scott, M. F. (2021). X-linked meiotic drive can boost population size and persistence. *Genetics*, *217*(1), 1–11.

Mahadevaraju, S., Fear, J. M., Akeju, M., Galletta, B. J., Pinheiro, M. M. L. S., Avelino, C. C., Cabral-de-Mello, D. C., Conlon, K., Dell'Orso, S., Demere, Z., Mansuria, K., Mendonça, C. A., Palacios-Gimenez, O. M., Ross, E., Savery, M., Yu, K., Smith, H. E., Sartorelli, V., Yang, H., … Oliver, B. (2021). Dynamic sex chromosome expression in *Drosophila* male germ cells. *Nature Communications*, *12*(1), 892.

Mank, J. E. (2013). Sex chromosome dosage compensation: definitely not for everyone. *Trends in Genetics*, *29*(12), 677–683.

McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, *33*(8), 1179–1186.

McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*, *8*(4), 329–337.e4.

McKee, B. D., & Handel, M. A. (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma*, *102*(2), 71–80.

Meade, L. C., Dinneen, D., Kad, R., Lynch, D. M., Fowler, K., & Pomiankowski, A. (2019). Ejaculate sperm number compensation in stalk-eyed flies carrying a selfish meiotic drive element. *Heredity*, *122*(6), 916–926.

Meiklejohn, C. D., Landeen, E. L., Cook, J. M., Kingan, S. B., & Presgraves, D. C. (2011). Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biology*, *9*(8), e1001126.

Meiklejohn, C. D., & Tao, Y. (2010). Genetic conflict and sex chromosome evolution. *Trends in Ecology & Evolution*, *25*(4), 215–223.

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

Montgomery, S. H., & Mank, J. E. (2016). Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Molecular Ecology*, *25*(20), 5114–5128.

Mullon, C., Wright, A. E., Reuter, M., Pomiankowski, A., & Mank, J. E. (2015). Evolution of dosage compensation under sexual selection differs between X and Z chromosomes. *Nature Communications*, *6*(1), 7720.

Ohlstein, B., & McKearin, D. (1997). Ectopic expression of the *Drosophila* Bam protein eliminates oogenic germline stem cells. *Development, 124*(18), 3651–3662.

Ohno, S. (1966). *Sex Chromosomes and Sex-Linked Genes*. Springer Berlin Heidelberg.

Page, N., Taxiarchi, C., Tonge, D., Kuburic, J., Chesters, E., Kriezis, A., Kyrou, K., Game, L., Nolan, T., & Galizi, R. (2023). Single-cell profiling of *Anopheles gambiae* spermatogenesis defines the onset of meiotic silencing and premeiotic overexpression of the X chromosome. *Communications Biology*, *6*(1), 850.

Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S., & Oliver, B. (2003). Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science*, *299*(5607), 697–700.

Parker, D. J., Jaron, K. S., Dumas, Z., Robinson-Rechavi, M., & Schwander, T. (2022). X chromosomes show relaxed selection and complete somatic dosage compensation across *Timema* stick insect species. *Journal of Evolutionary Biology*, *35*(12), 1734–1750.

Pieper, K. E., & Dyer, K. A. (2016). Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology*, *29*(11), 2229–2241.

Pinzone, C. A., & Dyer, K. A. (2013). Association of polyandry and sex-ratio drive prevalence in natural populations of *Drosophila neotestacea*. *Proceedings of the Royal Society B*, *280*(1769), 20131397.

Presgraves, D. C., Severance, E., & Wilkinson, G. S. (1997). Sex chromosome meiotic drive in stalk-eyed flies. *Genetics*, *147*(3), 1169–1180.

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R., & Wright, A. E. (2022a). Detecting signatures of selection on gene expression. *Nature Ecology & Evolution*, *6*(7), 1035–1045.

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F.,

Mank, J. E., Cooney, C. R., & Wright, A. E. (2022b). Reply to: Existing methods are effective at measuring natural selection on gene expression. *Nature Ecology & Evolution*, *6*(12), 1838–1839.

Raz, A. A., Vida, G. S., Stern, S. R., Mahadevaraju, S., Fingerhut, J. M., Viveiros, J. M., Pal, S., Grey, J. R., Grace, M. R., Berry, C. W., Li, H., Janssens, J., Saelens, W., Shao, Z., Hu, C., Yamashita, Y. M., Przytycka, T., Oliver, B., Brill, J. A., … Fuller, M. T. (2023). Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of *Drosophila* testes. *eLife*, *12*, e82201.

R Core Team. (2021). *R: A language and environment for statistical computing*.

Reinhardt, J. A., Baker, R. H., Zimin, A. V., Ladias, C., Paczolt, K. A., Werren, J. H., Hayashi, C. Y., & Wilkinson, G. S. (2023). Impacts of Sex Ratio Meiotic Drive on Genome Structure and Function in a Stalk-Eyed Fly. *Genome Biology and Evolution*, *15*(7), evad118.

Reinhardt, J. A., Brand, C. L., Paczolt, K. A., Johns, P. M., Baker, R. H., & Wilkinson, G. S. (2014). Meiotic drive impacts expression and evolution of x-linked genes in stalk-eyed flies. *PLoS Genetics*, *10*(5), e1004362.

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

Robben, M., Ramesh, B., Pau, S., Meletis, D., Luber, J., & Demuth, J. (2024). scRNA-seq Reveals Novel Genetic Pathways and Sex Chromosome Regulation in *Tribolium* Spermatogenesis. *Genome Biology and Evolution*, *16*(3), evae059.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140.

Sandler, L., Hiraizumi, Y., & Sandler, I. (1959). Meiotic drive in natural populations of *Drosophila melanogaster*. I. the cytogenetic basis of segregation-distortion. *Genetics*, *44*(2), 233–250.

Sandler, L., & Novitski, E. (1957). Meiotic Drive as an Evolutionary Force. *The American Naturalist*, *91*(857), 105–110.

Segami, C. J., Semon, M., Cunha, C., Bergin, C., Mugal, C. F., & Qvarnström, A. (2022). Single-Cell Transcriptomics reveals relaxed evolutionary constraint of spermatogenesis in two passerine birds as compared to mammals. *bioRxiv,* 2022.01.22.477241.

Silver, L. M. (1985). Mouse t haplotypes. *Annual Review of Genetics*, *19*(1), 179–208.

Silver, L. M. (1993). The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends in Genetics*, *9*(7), 250–254.

Simoni, A., Hammond, A. M., Beaghton, A. K., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D.,

Gribble, M., Morselli, G., Burt, A., Nolan, T., & Crisanti, A. (2020). A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nature Biotechnology*, *38*(9), 1054–1060.

Stalker, H. D. (1961). The genetic systems modifying meiotic drive in *Drosophila paramelanica*. *Genetics*, *46*(2), 177–202.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*(1), 477.

Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H., & Hartl, D. L. (2007). A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biology*, *5*(11), e293.

Taylor, D. R., Saur, M. J., & Adams, E. (1999). Pollen performance and sex-ratio evolution in a dioecious plant. *Evolution*, *53*(4), 1028–1036.

Terry, N. A., Tulina, N., Matunis, E., & DiNardo, S. (2006). Novel regulators revealed by profiling *Drosophila* testis stem cells within their niche. *Developmental Biology*, *294*(1), 246–257.

Turner, B. C., & Perkins, D. D. (1979). Spore killer, a chromosomal factor in neurospora that kills meiotic products not containing it. *Genetics*, *93*(3), 587–606.

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, *134*(10), 1823–1831.

Turner, J. M. A. (2015). Meiotic silencing in mammals. *Annual Review of Genetics*, *49*(1), 395–412.

Úbeda, F., Patten, M. M., & Wild, G. (2015). On the origin of sex chromosomes from meiotic drive. *Proceedings of the Royal Society B*, *282*(1798), 20141932.

Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Darwin Tree of Life Consortium, Formenti, G., Abueg, L., Torrance, J., Myers, E. W., Durbin, R., Blaxter, M., & McCarthy, S. A. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*, *24*(1), 288.

Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications*, *11*(1), 1201.

van der Auwera, G., & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated.

Vandereyken, K., Sifrim, A., Thienpont, B., & Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics*, *24*(8), 494–515.

Van Goethem, E., Silva, E. A., Xiao, H., & Franc, N. C. (2012). The *Drosophila* TRPP cation

channel, PKD2 and Dmel/Ced-12 act in genetically distinct pathways during apoptotic cell clearance. PLoS One, 7(2), e31488.

van Rensburg, A. J., Hipperson, H., Bradshaw, S. L., Bates, S., Burke, T., & Pomiankowski, A. (in prep). Genomics of an X-linked meiotic driver in stalk-eyed flies.

Vedanayagam, J., Lin, C.-J., & Lai, E. C. (2021). Rapid evolutionary dynamics of an expanding family of meiotic drive factors and their hpRNA suppressors. *Nature Ecology & Evolution*, *5*(12), 1613–1623.

Vedelek, V., Laurinyecz, B., Kovács, A. L., Juhász, G., & Sinka, R. (2016). Testis-Specific Bb8 Is Essential in the Development of Spermatid Mitochondria. *PLoS One*, *11*(8), e0161289.

Ventelä, S., Toppari, J., & Parvinen, M. (2003). Intercellular Organelle Traffic through Cytoplasmic Bridges in Early Spermatids of the Rat: Mechanisms of Haploid Gene Product Sharing. *Molecular Biology of the Cell*, *14*(7), 2768–2780.

Vibranovski, M. D. (2014). Meiotic sex chromosome inactivation in *Drosophila*. *Journal of Genomics*, *2*, 104–117.

Vogan, A. A., Svedberg, J., Grudzinska-Sterno, M., & Johannesson, H. (2022). Meiotic drive is associated with sexual incompatibility in *Neurospora*. *Evolution*, *76*(11), 2687–2696.

Vogt, N., Koch, I., Schwarz, H., Schnorrer, F., & Nüsslein-Volhard, C. (2006). The gammaTuRC components *Grip75* and *Grip128* have an essential microtubule-anchoring function in the *Drosophila* germline. *Development*, *133*(20), 3963–3972.

Wei, K. H.-C., 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*(4), e3002605.

Wilkinson, G. S., Christianson, S. J., Brand, C. L., Ru, G., & Shell, W. (2014). Haldane's rule is linked to extraordinary sex ratios and sperm length in stalk-eyed flies. *Genetics*, *198*(3), 1167–1181.

Wilkinson, G. S., Johns, P. M., Metheny, J. D., & Baker, R. H. (2013). Sex-biased gene expression during head development in a sexually dimorphic stalk-eyed fly. *PLoS One*, *8*(3), e59826.

Wilkinson, G. S., Presgraves, D. C., & Crymes, L. (1998). Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature*, *391*(6664), 276–279.

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*. *eLife*, *8, e47138.*

Witt, E., Shao, Z., Hu, C., Krause, H. M., & Zhao, L. (2021). Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in *Drosophila* testis. *PLoS*

Genetics, *17*(8), e1009728.

Wu, C. I., & Xu, E. Y. (2003). Sexual antagonism and X inactivation--the SAXI hypothesis. *Trends in Genetics*, *19*(5), 243–247.

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S. Y., Keefe, D. L., Alukal, J. P., Boeke, J. D., & Yanai, I. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*, *180*(2), 248–262.e21.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–287.

Zappia, L., & Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience*, *7*(7), giy083.

Zhang, L., Sun, T., Woldesellassie, F., Xiao, H., & Tao, Y. (2015). Sex ratio meiotic drive as a plausible evolutionary mechanism for hybrid male sterility. *PLoS Genetics*, *11*(3), e1005073.

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049.

Zimmering, S., Sandler, L., & Nicoletti, B. (1970). Mechanisms of meiotic drive. *Annual Review of Genetics*, *4*(1), 409–436.

Zohar-Fux, M., Ben-Hamo-Arad, A., Arad, T., Volin, M., Shklyar, B., Hakim-Mishnaevski, K., Porat-Kuperstein, L., Kurant, E., & Toledano, H. (2022). The phagocytic cyst cells in *Drosophila* testis eliminate germ cell progenitors via phagoptosis. *Science Advances*, *8*(24), eabm4937.

Zoller, R., & Schulz, C. (2012). The *Drosophila* cyst stem cell lineage: Partners behind the scenes? *Spermatogenesis*, *2*(3), 145–157.

Chapter 6. Discussion

This thesis explores the genetic bases and consequences of competition and conflict within sexual traits. Sexual traits are often complex and polygenic (Lande, 1980; Chenoweth & McGuigan, 2010; Civetta & Ranz, 2019), and commonly linked to sex chromosomes or wrapped into large linked genetic units such as supergenes and inversions (Charlesworth & Hartl, 1978; Rice, 1984; Itoh et al., 2011; Pearse et al., 2019; Reinhardt et al., 2023). Tracing the evolution of these traits is therefore affected by high levels of linkage disequilibrium between functional and linked loci on the sex chromosomes or within inversions, blunting the use of association studies (Uffelmann et al., 2021). Additionally, the phenotypic diversity of sexual traits is large, and unlikely explained by sequence alone (Harrison et al., 2015; King & Wilson, 1975). I therefore applied a series of transcriptomic and genomics analyses to search for the modes of selection acting on the genome and transcriptome in response to intermale or intragenomic conflicts.

In Chapter 2 (Price et al., 2023), I collated recent advances in our understanding of intralocus sexual conflict, considering important factors that may hinder how we perform research in this field and how we can progress into the future. In Chapter 3 (Price et al., 2022a), I quantified biases in current transcriptomic methods used to identify the genes responsible for producing phenotypic traits and their evolution. I used simulation and empirical studies of single-cell expression to show that traditional methods can lead to the incorrect identification of candidate genes as they produce false inferences of expression change and signatures of selection. In turn, I proposed guidelines for strengthening future analyses, with a particular focus on the use of single-cell approaches. In Chapter 4, I unpick the general assumption that the genomic architecture of reproductive traits such as sperm competition is under strong directional selection. I instead find evidence for balancing selection in zebra finches for sperm traits on the Z chromosome that is likely an outcome of associative overdominance, where heterozygote advantage masks deleterious alleles. Here, the transcriptome showed minimal evidence for relaxed purifying selection, and instead strong selection for equal expression of alternate haplotypes to mitigate degeneration of the coding sequence. In Chapter 5, I explored the molecular basis of meiotic drive in *Teleopsis dalmanni*, which although well studied experimentally, lacks information on the genetics underlying its driver. I characterised the gene expression patterns over spermatogenesis and found no impact of drive on broad patterns of genome wide expression. Instead, I provide evidence for key pathways affected by drive and supply several candidate genes that have interesting potential for function and impacts of

drive. In this discussion, I will expand on some of my key findings. Each chapter has a distinct discussion section outlining the results so I will avoid repeating this material here. Notably, I will identify key themes and links between chapters and discuss how the methodologies I used are to be applied with caution, providing guidance for the use of non-model systems. Finally, I will also briefly touch on interesting avenues for the future of this field, by myself or others.

6.1 Research impacts and considerations

Conflict has broad impacts on adaptation, speciation and genome structure. In this section I will dissect how it may contribute to variation in the coding genome and transcriptome in the species I have studied. I will then discuss conflict in the context of the persistence of polymorphism, and how the genomic location of a trait impacts its evolutionary trajectory. Finally, I will discuss the impacts of conflict on patterns of X-linked expression in spermatogenesis.

6.1.1 The impacts of inversions on gene expression in the zebra finch and stalk-eyed fly

In Chapter 4 and Chapter 5, I aimed to provide insight into the molecular bases and consequences of conflicts in the zebra finch and stalk-eyed fly. Both species share similar genomic landscapes in certain regards. Their sex chromosomes house inversion polymorphisms that effectively act as supergenes (Reinhardt et al., 2023) and are maintained by balancing selection (Wilkinson et al., 2006; Kim et al., 2017; Knief et al., 2017; Finnegan et al., 2019). This reduces recombination in heterozygotes, and ultimately leads to reductions in genetic diversity, elevating relative sequence divergence between either X (Reinhardt et al., 2023) or Z haplotypes (Chapter 4). Under these conditions, there are clear hypotheses for the impacts on the evolution of coding sequences, namely the reduced capacity to purge deleterious mutations and subsequent degeneration (Muller, 1964; Berdan et al., 2021), which we see in the zebra finch and is likely ongoing in the stalk-eyed fly. However, the impacts of reduced recombination for the evolution of the transcriptome, both on the sex chromosomes and for trans-regulation of the autosomes, is less clear.

We might expect the transcriptome to show similar patterns of evolution to the coding genome, with expression changes explained by divergence in cis-regulatory variation between X or Z haplotypes (Wittkopp et al., 2004; Juneja et al., 2016; Berdan, Mérot, et al., 2021). Indeed, we show that the transcriptome has enriched differential expression on the X chromosome

between standard (ST) and drive (SR) stalk-eyed fly males. Similarly in the zebra finch, there is elevated Z-linked differential expression between heterozygote (AB) and homozygote (AA) males for the inversion polymorphism, as well as between A and B females, compared to the autosomes. Both results suggest that the predominant effect of the divergence of sex chromosome haplotypes on expression is cis-acting, limited to sex-chromosome specific genetic architectures.

However, this does not inform us to the extent of which observed expression divergence is a product of relaxed purifying selection or directional selection. Due to the covariance of expression of genes within a network, changes in expression of one gene may shift that of other genes in a network in a non-adaptive manner (Lynch, 2007). If selection is inefficient on the X or Z, we would expect a proportion of genes to be biased due to regulatory linkage. This may be particularly evident in Z systems with high levels of promiscuity as the Z's effective population size shifts towards half of that of the autosomes and the relative strength of drift over selection increases (Wright & Mank, 2013; Wright et al., 2015). To discern the mode of selection acting on the transcriptome, we therefore must describe the indirect effects of these inversion polymorphisms on expression levels.

In the stalk-eyed fly, we have recorded expression patterns across a variety of cell types, including cells where expression of the driver is not expected (muscle and cyst cells). As such we can examine the degree of indirect regulatory change caused by the driving X. Two predicted outcomes of selection being able to mitigate the indirect fitness effects of drive in these cell types are similar expression patterns between SR and ST males, or the downregulation of drive-related genes in SR males. We in fact see an enrichment for X-linked differential expression across these cell types, which is predominantly SR-biased, suggesting pleiotropic effects of drive in other cell types. These effects may likely be an outcome of changes to regulatory networks by the driver, as well as a reduced capacity to respond to purifying selection from recombination suppression. This is in keeping with the pleiotropic costs experienced by males and females carrying a drive X (Finnegan et al., 2019), and what appears to be degeneration of the drive X (Reinhardt et al., 2023).

In the zebra finch, we tested for the mode of selection by analysing the degree of allele-specific expression (ASE) within heterozygous males. Under relaxed selection, we expected an enrichment for allele-specific expression compared to homozygous males, as the two Z haplotypes diverge in cis-regulatory variation and thus expression. We did not see this, however, suggesting either a lack of divergence in cis-regulatory variation on the Z, or a high degree of robustness in the transcriptome when exposed to non-adaptive processes. If the

transcriptome is in fact mitigating the effects of relaxed selection on the Z, as in the social supergene of the fire ant where degenerate genes are down-regulated (Martinez-Ruiz et al., 2020), we predicted allele-specific expression to relate to the degree of coding changes in a gene. Accordingly, we show greater balanced expression for genes that are more diverged between Z haplotypes. These results suggest that the transcriptome of the zebra finch has been robust to relaxed purifying selection, and that allele-specific expression mitigates costs from deleterious coding changes.

6.1.2 The effect of genomic location on inversions and intragenomic conflict

The zebra finch and stalk-eyed fly, as discussed in Chapter 4, Chapter 5 and Chapter 6.1.1, provide insights into how inversions can directly and indirectly shape the transcriptome, and how this may differ to how selection acts on the coding genome. An important factor in determining the generalisable impacts of inversions across life, as well as their evolutionary trajectories, is their location in the genome. As the inversions in both species are sex-linked, they likely experience different evolutionary processes to the autosomes (Charlesworth et al., 1987; Ellegren & Parsch, 2007). Ultimately, this can lead to differing outcomes both for the trans-effects and persistence of inversions, compared to those on the autosomes.

The location and gene content of an inversion will determine its trans-regulatory effects throughout the rest of the genome (Y.-C. Huang et al., 2018; Said et al., 2018; Villoutreix et al., 2021). For a trait that is not sex-specific, i.e., with shared functions between the sexes, the genes composing the underlying regulatory network may be more spread throughout the genome (Boyle et al., 2017; Rodríguez-Montes et al., 2023). In this case, we would expect greater trans effects of autosomal inversion polymorphisms, as divergence between haplotypes will lead to regulatory change in other portions of the genome. Accordingly, there are several examples of autosomal inversions with large trans-acting effects outside of the chromosome on which they occur (W. Huang et al., 2015; Lavington & Kern, 2017; Said et al., 2018). Contrastingly, for sexualised sex-chromosomes, the cis- and trans-regulatory effects of their inversions will likely be largely limited to these chromosomes as they will be enriched for sex-specific traits compared to elsewhere in the genome. The Z and X chromosomes of the zebra finch and stalk-eyed fly respectively, are masculinized, with an enrichment for testis-specific genes (Baker et al., 2016; Kim et al., 2017). Accordingly, the vast majority of differentially expressed genes between inversion polymorphisms in these species are Z/X-linked. Similar localised patterns of expression divergence are also seen for autosomal

inversions that house supergene polymorphisms (Berdan et al., 2021). This suggests that the trans-regulatory effects of an inversion polymorphism in other portions of the genome are less accentuated if the inversion has a high density of genes with interconnected function.

The long-term persistence of an inversion polymorphism will also be affected by its location within the genome. If an inversion is autosomal instead of sex-linked, it will experience a larger effective population size than the X or Z and a higher recombination rate as homozygosity can occur in both sexes. This may somewhat mitigate the degeneration of the inversion polymorphisms and contribute to their long term persistence (Connallon et al., 2018). An autosomal meiotic driver, which are often housed in inversions, will therefore likewise benefit from an increased effective population size and recombination rate. Unlike sex-linked drivers, autosomal drivers do not skew the population sex ratio, which can ultimately cause population extinction (Hamilton, 1967). As such, when fixed in a population, these drivers will not destabilise population dynamics via a female or male-bias population. At fixation, autosomal drivers exist cryptically. Here, no target alleles remain and normal recombination can resume. Through this, the autosomal driver can purge new and historic deleterious mutations.

Differentiated Y chromosomes exhibit minimal recombination, if at all. This may explain the infrequency of Y-linked drivers in XY systems, where the majority of meiotic drivers are located on the X (Hurst & Pomiankowski, 1991; Helleu et al., 2014; Courret et al., 2019). However, in several Aedes and Culex mosquitoes species including *Aedes aegypti*, this is not the case. These species do not have heteromorphic sex chromosomes, and instead sex is determined by heterozygosity of a sex determining allele (Toups & Hahn, 2010; Turner et al., 2018). *Aedes aegypti's* sex ratio-distorter is tightly linked to the male copy of this allele (Shin et al., 2011), potentially overcoming some of the restrictions of being located on a non-recombining degenerate Y chromosome. The location in the genome of inversions and intragenomic conflicts clearly plays an important role in their evolution. Accordingly, these factors will be essential to consider as drivers are increasingly applied in the context of disease vector control.

6.1.3 Maintenance of genetic diversity under conflict

Conflict on many evolutionary levels plays a clear role in generating phenotypic diversity between species and consolidating those differences through reproductive barriers (Arnqvist et al., 2000). However, how it maintains genetic polymorphisms within populations is less clear. Intralocus selection between males and females can hypothetically promote balancing selection, however the degree to which this occurs is uncertain and debated (see Chapter

1.1.1). Competition between males and intragenomic conflicts however, have much better described impacts on genetic diversity (Charlesworth & Hartl, 1978; Pomiankowski & Moller, 1995; Mackintosh et al., 2021).

One might predict that strong sexual selection via male-male competition would erode genetic variation through limiting the number of males contributing to the gene pool. This however would paradoxically remove variation on which sexual selection may act. Theories have now resolved this, namely through proposing the genic-capture of condition-dependent sexually-selected traits (Pomiankowski & Moller, 1995; Rowe & Houle, 1996; Kotiaho et al., 2001). Here, the expression of a costly sexual signal is dependent on genome-wide variation associated with individual condition. Accordingly, traits involved in mate competition are often highly heritable, attributed to their condition dependence. This includes sperm traits (Simmons & Kotiaho, 2002), weaponry (Johns et al., 2014), colouration (Roulin, 2016), ornamentation (David et al., 2000) and behavioural phenotypes (Massot et al., 2002). If sexually selected traits such as sperm morphology are also favoured via heterozygote advantage, or other modes of balancing selection (Moore & Moore, 1999; Yassin et al., 2016; Kim et al., 2019), this may further contribute to the genetic variation underlying both the sexual phenotype and genetic condition.

Sperm competition and mate choice can also promote the maintenance of drive as a polymorphism within a population. Irrespective of sexual selection, the maintenance of a drive polymorphism is dependent on the fitness costs of linked deleterious alleles, and the pleiotropic effects such as reduced fecundity, significantly reducing male or female fitness (Curtsinger & Feldman, 1980; Taylor & Jaenike, 2002; Finnegan et al., 2019). This will help slow the invasion of a drive allele, with the costs particularly evident when drive frequency, and therefore homozygosity in females, is high (Dyer & Hall, 2019; Larner et al., 2019). In systems with pre- or post-copulatory sexual selection, such as mate choice or polyandry respectively, fixation can become less likely due to males' reduced ability to compete in mating. For example, in *Drosophila pseudoobscura*, the drive allele leads to reductions in ejaculate size and therefore reduced competitiveness in sperm competition in wild populations, and in double-mating trials (Mackintosh et al., 2021; Price et al., 2008a, 2008b; Hodgson, et al., 2008). In comparison, in the stalk-eyed fly, although drive carries pleiotropic costs associated with the loss of Y-bearing sperm, drive males are comparable to standard males in sperm competition (Cotton et al., 2010; Bradshaw et al., 2022; Bates et al., 2023). The mating costs are instead a trade-off with competitiveness in pre-copulatory sexual selection, with reduced relative eye span lowering drive-male mating frequencies (Wilkinson

et al., 1998; Cotton et al., 2014). In the absence of pre- or post-copulatory sexual selection, drivers may instead lead to unstable frequencies and population extinction.

Interestingly, sex-ratio drive has been predicted to bolster population size under certain conditions (Mackintosh et al., 2021). Here, sex-ratio skew optimises male mating rates and boost birth rate beyond non-drive populations. This growth is capped when females begin to go unmated as the sex-ratio skew surpasses peak male fecundity (Hatcher et al., 1999). These effects are most impactful in small populations with increased extinction risk (Mackintosh et al., 2021), suggesting that under drive can in fact help to maintain local pockets of variation. Importantly, this mechanism of drive persistence is dependent on the sex-determination system. While an X-linked driver leads to a female-biased population, a Z-linked driver will lead to male-bias and female mating rates will rapidly saturate.

6.1.4 Perspectives on the evolution of meiotic sex chromosome inactivation and dosage compensation

The X chromosome exhibits dynamic expression patterning across spermatogenesis and differing expression between the sexes, a product of differences in copy number, varying levels of sexually antagonistic variation, and heteromorphy with the Y (Wu & Xu, 2003; Turner, 2007; Robben et al., 2024; Wei et al., 2024). The predominant resulting expression patterns are meiotic sex chromosome inactivation (MSCI) and dosage compensation (DC). MSCI is the transcriptional silencing of the X and Y during the meiotic phases of spermatogenesis, enabling proper segregation and avoiding aneuploidy (Turner, 2007), whilst DC is the matching of the X's expression between the sexes (Disteche, 2012). Given the similarity in presentation between incomplete DC and MSCI, their relative role is debated across species. Similar pre-meiotic and meiotic down-regulation of the X in *Anopheles* (Taxiarchi et al., 2019) and *Drosophila* (Vibranovski, 2014) has been historically attributed to MSCI, likely due to the similarity of their X activity to that which was previously described in mammals (Turner, 2007). However, as *Drosophila* males do not experience recombination and therefore aneuploidy is not a risk, other mechanisms have been debated for reductions in X-activity in spermatogenesis, for example, the silencing of sex-ratio distorters (Tao et al., 2007), the inhibition of recombination with the Y (McKee & Handel, 1993), or the suppression of expression of sexually antagonistic variation (Wu & Xu, 2003). Recent advances in single-cell sequencing have enabled the description of a diverse array of X-linked expression patterns during spermatogenesis in insects. These include MSCI (Page et al., 2023; Djordjevic et al., 2024; Robben et al., 2024), pre-meiotic X overactivation (Page et al., 2023), incomplete DC

(Witt et al., 2021; Wei et al., 2024), and developmental variation in dosage (Djordjevic et al., 2024).

Adding to this landscape of X-linked dynamics, our findings of incomplete pre-meiotic and meiotic DC and post-meiotic re-upregulation of the X in *Teleopsis dalmanni*, highlight that X regulation in spermatogenesis is a diverse and evolving process in insects. These observations bring up interesting avenues for research into how and why MSCI and incomplete DC evolve in spermatogenesis. Currently, there is consensus around the drivers of both meiotic and somatic sex chromosome inactivation (Turner, 2007). Likewise, there is a good theoretical and empirical understanding of how and why DC evolves in somatic tissues of XY systems (Gu & Walters, 2017), namely to avoid the cost of hemizygous expression in the heterogametic sex (Ohno, 1966). In ZW systems, DC is often incomplete in females. This masculinisation can be explained by a greater receptiveness to sexual selection and higher male-mutation bias in the Z chromosome, compared to the X (Naurin et al., 2010). The role of DC in the germline cells of the testis is an exception to the above, with many different hypotheses for their unique patterns. Diopsidae may play an interesting role in understanding the evolution and impacts of complete and incomplete DC.

Firstly, incomplete DC is often attributed to an enrichment of female-biased variation on the X (Rice, 1984; Wu & Xu, 2003; Allen et al., 2013). As the *T. dalmanni* X is masculinized, this is unlikely the case for this species, and so alternate causes of incomplete DC should be considered. Secondly, the Diopsidae exhibit varying levels of sexual selection and sperm competition (Presgraves et al., 1999) and, therefore, most likely exhibit covarying levels of sexual antagonism (Cox & Calsbeek, 2009). This can affect the gene content of the X across the family, and the degree to which it is masculinized and enriched for testes-specific genes (Gurbich & Bachtrog, 2008). Examining patterns of dosage, as well as broader X-activity in spermatogenesis, across the Diopsidae, may therefore provide further insight into the role of sexual antagonism in the evolution of the transcriptomic landscape of the X. Thirdly, the Diopsidae harbour meiotic drivers, the silencing of which has been predicted to lead to MSCI (Meiklejohn & Tao, 2010). Whilst descriptions of the ongoing evolution of sex chromosome silencing in response to drive are lacking, current reports of incomplete DC in *T. dalmanni* may represent an intermediary step towards drive silencing. As such, this system could prove valuable in determining the response of X-activity to the presence of meiotic drive. Finally, the consistent expression of large numbers of genes throughout meiosis, on both the X and autosomes, makes spermatogenesis in *T. dalmanni* a valuable tool for studying rates of sex-chromosome evolution. Large numbers of, often truncated, proteins are expressed during the

early stages of spermatogenesis. This has been attributed to transcriptional scanning, which facilitates DNA repair. Accordingly, genes with low expression tend to show elevated divergence rates in *Drosophila* (Xia et al., 2020; Xia & Yanai, 2022; Witt et al., 2023). The continued expression of the *T. dalmanni* X throughout the early stages of spermatogenesis may mitigate the elevated rate of evolution on the X compared to the autosomes. This is known as the faster-X effect, a product of lower recombination rate, elevated mutation rate, and greater ability to respond to selection on recessive mutations on the X (Charlesworth et al., 2018).

It is not yet clear what mechanism underlies the patterns of DC I uncovered in *T. dalmanni*, or how they impact other evolutionary processes. However, Diopsidae may provide a brilliant tool for understanding the interactions between the regulation of the X during spermatogenesis, sexual selection and sex-ratio distorter evolution.

6.2 Methodological advances

Studying evolution across broad taxa is invaluable for determining its generalisable patterns. However, the use of organisms with limited genomic resources, compared to those available in models such as *Drosophila*, comes with important methodological considerations. In this section, I will discuss the application of genomics to less-studied organisms, especially in the context of the hard-to-sequence sex chromosomes. Finally, I will extend upon the important discussions of Chapter 3 (Price et al., 2022a), in detecting broad patterns of regulatory evolution.

6.2.1 The use of non-model systems for cutting-edge technologies

In Chapter 5, I highlighted some of the difficulties of working with non-model systems when applying cutting-edge technologies such as single-cell RNA-seq. Whilst *Drosophila* is a useful reference point for the study of other dipterans and insects, the overreliance on comparisons to distantly related model systems is problematic. Specifically, our ability to isolate and define cell types is dependent on several key factors. Not only do we need to re-tune the wet-lab process. For example, laboratory protocols optimised for model systems have to be finessed to ensure proper isolation of cell types in non-model species, and whilst tissues of *Drosophila* are readily available and can go straight from organism to dissociation to sequencing with minimal time gaps, researchers working with wild species must maximise cell concentrations

from limited samples whilst minimising transport time. We must also consider the accuracy of ortholog identification and differences in cell type and tissue biology differences.

Firstly, for species lacking descriptions of cell type marker genes, we rely on model organisms where these are well described. The ability to biologically classify cells in our target species therefore requires high confidence in orthology identification between these species. In recently annotated species such as *Teleopsis dalmanni*, this confidence may be lower. In this case we can instead use other approaches to classify cell types such as broad-scale expression patterns, for example, the relative number of expressed genes, and how these vary along a predicted developmental trajectory. This is especially relevant to the testis, where there is conservation across taxa in the expression of large numbers of genes, with the magnitude differing across stages of spermatogenesis (Shami et al., 2020; Page et al., 2023; Wei et al., 2024). Another approach uses mitochondrial expression which, whilst often used to filter dying cells, can also aid the differentiation between cell types with differing energetic output (Mercer et al., 2011).

Secondly, markers for a cell type of interest may not be consistent across species, and cellular populations in one species may not be present in the other. This is especially relevant in the testis, or other sexually selected tissues, due to their rapid evolution (Ramm & Schärer, 2014; Murat et al., 2023). For example, whilst sperm are some of the most diverse cell types across eukaryotes (Lüpold et al., 2020), their developmental stage is identifiable through expression of conserved markers. However, for less studied species, we may be unable to accurately classify cell populations that have evolved in response to sperm competition or in tandem with the female reproductive tract. These include heteromorphic sperm types (Snook, 1998; Presgraves et al., 1999) and varying somatic composition include cyst size (Ramm & Schärer, 2014). Furthermore, as sex-specific traits are often enriched on the sex chromosomes, the unique evolutionary dynamics they experience can lead to rapid gene evolution across species (Mank et al., 2010; Charlesworth et al., 2018). This may accelerate the level of functional divergence that sex-linked markers have in more distantly related taxa. These points highlight why exploratory genomics cannot replace traditional cytology and histochemistry approaches, but instead should be used, where possible, in tandem.

6.2.2 The importance of Y/W reference assemblies in understanding meiotic drive and spermatogenesis

The Y and W chromosomes often house large ampliconic regions, including tandem repeats and palindromes (Betrán et al., 2012). This is an outcome of limited recombination which, if present, is isolated to the pseudoautosomal regions (Mangs & Morris, 2007). The evolutionary consequences of this have been widely discussed (Bachtrog, 2013) and we have summarised recent updates for the avian W chromosome in Appendix D (Smith et al., 2022). However, a common problem in the study of sex chromosome evolution and the evolution of sexual dimorphisms is confidently assembling the repeat rich Y and W chromosomes (Tomaszkiewicz et al., 2017), with the human Y only last year seeing a full telomere to telomere sequence (Rhie et al., 2023).

This is especially relevant in the study of sex-linked meiotic drive, where missing the sequence of the Y reduces the ability to characterise both the timing and target of a driver. Firstly, a sequence of the Y may allow the detection of evolutionary strata (Palmer et al., 2019), where regions of the chromosome have discretely diverged, attributable to inversions linking suppressor mechanisms to the target. Dating these strata and comparing them to those of the driving-X, as well as to non-coadapted Y chromosomes, may inform us of their evolutionary history including the sequential emergence and linkage of new drive and suppression mechanisms. Secondly, the gene content of the Y may show the genetic footprint of the drive-suppressor co-evolution as homologous amplification of multi-copy gene families between the X and Y is often involved in evolving drive systems (Soh et al., 2014). In this case, mapped differential expression of these families, and other genes on the Y, will further describe the genetic basis. Thirdly, given we expect the Y to be under strong selection, we can utilise evolutionary statistics such as dN/dS (the ratio of nonsynonmous to synonymous differences) to detect genes with inflated coding differences and estimate how selection has responded to drive on the Y. Finally, given the common role of open chromatin structures in driver targeting (Courret et al., 2019), the use of ATAC-Seq can help elucidate physical targets in the genome and the degree to which Y chromosome chromatin content has evolved as a result of drive-suppressor co-evolution.

A robust Y assembly would also be invaluable in the identification of meiotic sex chromosome inactivation (MSCI), where in mammals the Y and the X are packaged into a nuclear subdomain of pachytene stage cells (Turner et al., 2006; Turner, 2007). Being able to detect Y-linked expression, or lack thereof, would be insightful into how the sex chromosomes are regulated during meiosis. This is especially relevant when discerning MSCI from incomplete

dosage compensation (DC). Incomplete DC is often predicted to mitigate the expression of sexually antagonistic variation (Wu & Xu, 2003), whilst MSCI allows proper sex-chromosome segregation (Turner, 2007). Therefore, if Y expression is similar between germline and somatic cell types, but the X is downregulated, we could conclude that this expression pattern is more likely a mitigation of sexually antagonistic variation than protection against non-disjunction.

6.2.3 Defining the modes of selection acting on gene expression patterns

In Chapter 3 (Price et al., 2022a), I highlight the importance of considering tissue composition when determining patterns of gene expression evolution, specifically, the overestimation of stabilising selection in more rapidly diverging tissue types. These findings suggest that the relative role of expression in adaptive evolution has been largely misrepresented and that drift or non-modelled processes may be playing important roles in shaping the transcriptome.

Given the implication of the above, it was not surprising to receive feedback from the research community, including a Matters Arising article from Hunter B. Fraser (Fraser, 2022). Fraser offered resolutions for the effects of tissue differences when detecting signatures of expression evolution. This included approaches that harness purified cell lines or measure allele-specific-expression in hybrid crosses. My full response can be seen in Appendix A (Price et al., 2022b). The main consideration of the response is that the proposed alternative approaches are only accessible for either model organisms or cell types that are more readily cultured. Most importantly, the overreliance on cultured single cell types ignores the fact that a tissue is a sum of its cellular parts, meaning whole tissue comparisons are often most appropriate for understanding phenotypic evolution. This is especially true when determining the level at which selection is acting, i.e., the cell or the tissue, and how it may be constrained.

Adaptation through differences in protein abundance can arise via two main pathways; an increase in the cellular expression of a protein or the number of protein-expressing cells (Vogel & Marcotte, 2012). Both of these are functionally significant in the evolution of phenotypic change and emergence of novelty but will be bound by different biological limitations.

Firstly, the individual cell has biophysical limitations in transcription and translation rate, and mRNA and protein decay (Hausser et al., 2019; Metzl-Raz et al., 2020). If restricted by these conditions, the expression of a gene may be trapped at a fitness peak, unable to cross to a higher peak. This can be overcome through routes such as duplication events to increase functional copies (Ohno, 1970). This, however, relies on inhibiting divergence and neo-functionalisation of novel copies through strong purifying selection (Nei et al., 2000) or gene

conversion within the gene family (Hurst & Smith, 1998). Alternatively, constraint can be overcome through changes in the composition of the cellular cytoplasm so as to increase expression or translation rates. For example, ribosomal composition can play an important role in both translation rate and post-translational modification (Genuth & Barna, 2018).

Secondly, the physiology and ontogeny of an organism may be spatially and energetically bound, inhibiting increases in relative numbers of cell types within a tissue or body part (Gould, 1989). The evolution of a trait may also be limited by robust developmental pathways that ensure proper allometric scaling between body parts (Frankino et al., 2005). To resolve this developmental conflict, pleiotropic relationships between developmental trajectories must be broken down. Likewise, to facilitate changes in cell-level protein abundance, expression covariance with other regions of the genome must reduce. Much like the resolution of sexually antagonistic selection, transcriptional evolution may often require the rewiring of regulatory networks to decouple shared regulation patterns between genes (Arendt et al., 2016).

Understanding the relative roles of the above limitations within the micro and macro-evolution of protein abundances is an important factor in determining how phenotypic novelty is generated, as well as detecting evolvable pathways within systems. Measuring protein abundance or gene expression will give differing results depending on the quantification method used, for example, if measuring RNA or protein levels, or if using bulk or single-cell RNA-seq. Protein abundance does not always perfectly correlate with gene expression level; as the proteome experiences more stringent selection than the transcriptome, it thus shows less variation (Khan et al., 2013; Bathke et al., 2019; Wang et al., 2020). This is exceptionally relevant in the testis; whilst large numbers of genes are expressed here, they are often truncated and potentially non-functional (Kleene, 2003; D. Wang et al., 2019; Xia et al., 2020).

Trying to explain how increases in cell type abundance or protein expression contribute to adaptation is therefore non-trivial. For single-cell approaches, I advise measuring differences in both cellular-abundance and transcripts to understand their relative roles in a trait. Additionally, discrepancies between the transcriptome and proteome can be resolved though multi-omics approaches, which are becoming increasingly available, allowing the joint sequencing of proteome, transcriptome and epigenome (Flynn et al., 2023). Finally, following an ontogenetic approach to researching adaptation will allow us to accurately describe when and where selection acts and novelty originates (Kalinka et al., 2010; Piasecka et al., 2013).

6.3 Conclusion

In summary, it has been a massive privilege to work in this complex and highly dimensioned field. I feel now, more than ever, that I have barely scratched the surface of understanding how these processes intersect and interact and am excitedly looking forward to many more years working towards our understanding, or at least describing a small section, of evolution and the many components of sex and conflict within it.

6.4 References

Allen, S. L., Bonduriansky, R., & Chenoweth, S. F. (2013). The genomic distribution of sex-biased genes in *Drosophila serrata*: X chromosome demasculinization, feminization, and hyperexpression in both sexes. *Genome Biology and Evolution*, *5*(10), 1986–1994.

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D., & Wagner, G. P. (2016). The origin and evolution of cell types. *Nature Reviews Genetics*, *17*(12), 744–757.

Arnqvist, G., Edvardsson, M., Friberg, U., & Nilsson, T. (2000). Sexual conflict promotes speciation in insects. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10460–10464.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113–124.

Baker, R. H., Narechania, A., DeSalle, R., Johns, P. M., Reinhardt, J. A., & Wilkinson, G. S. (2016). Spermatogenesis drives rapid gene creation and masculinization of the X chromosome in stalk-eyed flies (Diopsidae). *Genome Biology and Evolution*, *8*(3), 896–914.

Bates, S., Meade, L., & Pomiankowski, A. (2023). Meiotic drive does not impede success in sperm competition in the stalk-eyed fly, *Teleopsis dalmanni*. *Evolution*, *77*(10), 2326–2333.

Bathke, J., Konzer, A., Remes, B., McIntosh, M., & Klug, G. (2019). Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides* throughout growth. *BMC Genomics*, *20*(1), 358.

Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics*, *17*(3), e1009411.

Berdan, E. L., Mérot, C., Pavia, H., Johannesson, K., Wellenreuther, M., & Butlin, R. K. (2021).

A large chromosomal inversion shapes gene expression in seaweed flies (*Coelopa frigida*). *Evolution Letters*, *5*(6), 607–624.

Betrán, E., Demuth, J. P., & Williford, A. (2012). Why chromosome palindromes? *International Journal of Evolutionary Biology*, *2012*, 207958.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*(7), 1177–1186.

Bradshaw, S. L., Meade, L., Tarlton-Weatherall, J., & Pomiankowski, A. (2022). Meiotic drive adaptive testes enlargement during early development in the stalk-eyed fly. *Biology Letters*, *18*(11), 20220352.

Charlesworth, B., Campos, J. L., & Jackson, B. C. (2018). Faster-X evolution: Theory and evidence from *Drosophila*. *Molecular Ecology*, *27*(19), 3753–3771.

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, *130*(1), 113–146.

Charlesworth, B., & Hartl, D. L. (1978). Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics*, *89*(1), 171–192.

Chenoweth, S. F., & McGuigan, K. (2010). The genetic basis of sexually selected variation. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 81–101.

Civetta, A., & Ranz, J. M. (2019). Genetic factors influencing sperm competition. *Frontiers in Genetics*, *10*, 820.

Connallon, T., Olito, C., Dutoit, L., Papoli, H., Ruzicka, F., & Yong, L. (2018). Local adaptation and the evolution of inversions on sex chromosomes and autosomes. *Proceedings of the Royal Society B*, *373*(1757), 20170423.

Cotton, A. J., Földvári, M., Cotton, S., & Pomiankowski, A. (2014). Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity*, *112*(4), 363–369.

Cotton, S., Small, J., Hashim, R., & Pomiankowski, A. (2010). Eyespan reflects reproductive quality in wild stalk-eyed flies. *Evolutionary Ecology*, *24*(1), 83–95.

Courret, C., Chang, C.-H., Wei, K. H.-C., Montchamp-Moreau, C., & Larracuente, A. M. (2019). Meiotic drive mechanisms: lessons from *Drosophila*. *Proceedings of the Royal Society B*, *286*(1913), 20191430.

Cox, R. M., & Calsbeek, R. (2009). Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist*, *173*(2), 176–187.

Curtsinger, J. W., & Feldman, M. W. (1980). Experimental and theoretical analysis of the “sex-ratio” polymorphism in *Drosophila pseudoobscura*. *Genetics*, *94*(2), 445–466.

David, P., Bjorksten, T., Fowler, K., & Pomiankowski, A. (2000). Condition-dependent

signalling of genetic variation in stalk-eyed flies. *Nature*, *406*(6792), 186–188.

Disteche, C. M. (2012). Dosage compensation of the sex chromosomes. *Annual Review of Genetics*, *46*(1), 537–560.

Djordjevic, J., Van, P. T., Toubiana, W., Labédan, M., Dumas, Z., Aury, J.-M., Cruaud, C., Istace, B., Labadie, K., Noel, B., Parker, D. J., & Schwander, T. (2024). Dynamics of X chromosome hyper-expression and inactivation in male tissues during stick insect development. *bioRxiv*, 2024.07.01.601468.

Dyer, K. A., & Hall, D. W. (2019). Fitness consequences of a non-recombining sex-ratio drive chromosome can explain its prevalence in the wild. *Proceedings of the Royal Society B*, *286*(1917), 20192529.

Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*(9), 689–698.

Finnegan, S. R., White, N. J., Koh, D., Camus, M. F., Fowler, K., & Pomiankowski, A. (2019). Meiotic drive reduces egg-to-adult viability in stalk-eyed flies. *Proceedings of the Royal Society B*, *286*(1910), 20191414.

Flynn, E., Almonte-Loya, A., & Fragiadakis, G. K. (2023). Single-cell multiomics. *Annual Review of Biomedical Data Science*, *6*, 313–337.

Frankino, W. A., Zwaan, B. J., Stern, D. L., & Brakefield, P. M. (2005). Natural selection and developmental constraints in the evolution of allometries. *Science*, *307*(5710), 718–720.

Fraser, H. B. (2022). Existing methods are effective at measuring natural selection on gene expression. *Nature Ecology & Evolution*, *6*(12), 1836–1837.

Genuth, N. R., & Barna, M. (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Molecular Cell*, *71*(3), 364–374.

Gould, S. J. (1989). A developmental constraint in Cerion, with comments on the definition and interpretation of constraint in evolution. *Evolution*, *43*(3), 516–539.

Gu, L., & Walters, J. R. (2017). Evolution of sex chromosome dosage compensation in animals: A beautiful theory, undermined by facts and bedeviled by details. *Genome Biology and Evolution*, *9*(9), 2461–2476.

Gurbich, T. A., & Bachtrog, D. (2008). Gene content evolution on the X chromosome. *Current Opinion in Genetics & Development*, *18*(6), 493–498.

Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, *156*(3774), 477–488.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene

expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4393–4398.

Hatcher, M. J., Taneyhill, D. E., Dunn, A. M., & Tofts, C. (1999). Population dynamics under parasitic sex ratio distortion. *Theoretical Population Biology*, *56*(1), 11–28.

Hausser, J., Mayo, A., Keren, L., & Alon, U. (2019). Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*, *10*(1), 68.

Helena Mangs, A., & Morris, B. J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Current Genomics*, *8*(2), 129–136.

Helleu, Q., Gérard, P. R., & Montchamp-Moreau, C. (2014). Sex chromosome drive. *Cold Spring Harbor Perspectives in Biology*, *7*(2), a017616.

Huang, W., Carbone, M. A., Magwire, M. M., Peiffer, J. A., Lyman, R. F., Stone, E. A., Anholt, R. R. H., & Mackay, T. F. C. (2015). Genetic basis of transcriptome diversity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), E6010–E6019.

Huang, Y.-C., Dang, V. D., Chang, N.-C., & Wang, J. (2018). Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. *Proceedings of the Royal Society B*, *285*(1878), 20180221.

Hurst, L. D., & Pomiankowski, A. (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics*, *128*(4), 841–858.

Hurst, L. D., & Smith, N. G. C. (1998). The evolution of concerted evolution. *Proceedings of the Royal Society B*, *265*(1391), 121–127.

Itoh, Y., Kampf, K., Balakrishnan, C. N., & Arnold, A. P. (2011). Karyotypic polymorphism of the zebra finch Z chromosome. *Chromosoma*, *120*(3), 255–264.

Johns, A., Gotoh, H., McCullough, E. L., Emlen, D. J., & Lavine, L. C. (2014). Heightened condition-dependent growth of sexually selected weapons in the rhinoceros beetle, *Trypoxylus dichotomus* (Coleoptera: Scarabaeidae). *Integrative and Comparative Biology*, *54*(4), 614–621.

Juneja, P., Quinn, A., & Jiggins, F. M. (2016). Latitudinal clines in gene expression and cis-regulatory element variation in *Drosophila melanogaster*. *BMC Genomics*, *17*(1), 981.

Kalinka, A. T., Varga, K. M., Gerrard, D. T., Preibisch, S., Corcoran, D. L., Jarrells, J., Ohler, U., Bergman, C. M., & Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. *Nature*, *468*(7325), 811–814.

Khan, Z., Ford, M. J., Cusanovich, D. A., Mitrano, A., Pritchard, J. K., & Gilad, Y. (2013). Primate transcript and protein expression levels evolve under compensatory selection

pressures. *Science*, *342*(6162), 1100–1104.

Kim, K.-W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L. L., Griffith, S. C., Burke, T., Birkhead, T. R., & Slate, J. (2017). A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature Ecology & Evolution*, *1*(8), 1168–1176.

Kim, K.-W., Jackson, B. C., Zhang, H., Toews, D. P. L., Taylor, S. A., Greig, E. I., Lovette, I. J., Liu, M. M., Davison, A., Griffith, S. C., Zeng, K., & Burke, T. (2019). Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nature Communications*, *10*(1), 1852.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107–116.

Kleene, K. C. (2003). Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenetic and Genome Research*, *103*(3-4), 217–224.

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., Opatová, P., Albrechtová, J., Wittig, M., Franke, A., Albrecht, T., & Kempenaers, B. (2017). A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nature Ecology & Evolution*, *1*(8), 1177–1184.

Kotiaho, J. S., Simmons, L. W., & Tomkins, J. L. (2001). Towards a resolution of the lek paradox. *Nature*, *410*(6829), 684–686.

Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*, *34*(2), 292.

Larner, W., Price, T., Holman, L., & Wedell, N. (2019). An X-linked meiotic drive allele has strong, recessive fitness costs in female *Drosophila pseudoobscura*. *Proceedings of the Royal Society B*, *286*(1916), 20192038.

Lavington, E., & Kern, A. D. (2017). The effect of common inversion polymorphisms in(2L)t and in(3R)Mo on patterns of transcriptional variation in *Drosophila melanogaster*. *G3, 7*(11), 3659–3668.

Lüpold, S., de Boer, R. A., Evans, J. P., Tomkins, J. L., & Fitzpatrick, J. L. (2020). How sperm competition shapes the evolution of testes and sperm: a meta-analysis. *Proceedings of the Royal Society B*, *375*(1813), 20200064.

Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*, *8*(10), 803–813.

Mackintosh, C., Pomiankowski, A., & Scott, M. F. (2021). X-linked meiotic drive can boost population size and persistence. *Genetics*, *217*(1), 1–11.

Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic

drift. *Molecular Biology and Evolution*, *27*(3), 661–670.

Martinez-Ruiz, C., Pracana, R., Stolle, E., Paris, C. I., Nichols, R. A., & Wurm, Y. (2020). Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *eLife*, *9*, e55862.

Massot, M., Clobert, J., Lorenzon, P., & Rossi, J.-M. (2002). Condition-dependent dispersal and ontogeny of the dispersal behaviour: an experimental approach. *The Journal of Animal Ecology*, *71*(2), 253–261.

McKee, B. D., & Handel, M. A. (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma*, *102*(2), 71–80.

Meiklejohn, C. D., & Tao, Y. (2010). Genetic conflict and sex chromosome evolution. *Trends in Ecology & Evolution*, *25*(4), 215–223.

Mercer, T. R., Neph, S., Dinger, M. E., Crawford, J., Smith, M. A., Shearwood, A.-M. J., Haugen, E., Bracken, C. P., Rackham, O., Stamatoyannopoulos, J. A., Filipovska, A., & Mattick, J. S. (2011). The human mitochondrial transcriptome. *Cell*, *146*(4), 645–658.

Metzl-Raz, E., Kafri, M., Yaakov, G., & Barkai, N. (2020). Gene Transcription as a Limiting Factor in Protein Production and Cell Growth. *G3*, *10*(9), 3229–3242.

Moore, A. J., & Moore, P. J. (1999). Balancing sexual selection through opposing mate choice and male competition. *Proceedings of the Royal Society B*, *266*(1420), 711–716.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research*, *106*, 2–9.

Murat, F., Mbengue, N., Winge, S. B., Trefzer, T., Leushkin, E., Sepp, M., Cardoso-Moreira, M., Schmidt, J., Schneider, C., Mößinger, K., Brüning, T., Lamanna, F., Belles, M. R., Conrad, C., Kondova, I., Bontrop, R., Behr, R., Khaitovich, P., Pääbo, S., … Kaessmann, H. (2023). The molecular evolution of spermatogenesis across mammals. *Nature*, *613*(7943), 308–316.

Naurin, S., Hansson, B., Bensch, S., & Hasselquist, D. (2010). Why does dosage compensation differ between XY and ZW taxa? *Trends in Genetics*, *26*(1), 15–20.

Nei, M., Rogozin, I. B., & Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 10866–10871.

Ohno, S. (1966). *Sex Chromosomes and Sex-Linked Genes*. Springer Berlin Heidelberg.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer Berlin Heidelberg.

Page, N., Taxiarchi, C., Tonge, D., Kuburic, J., Chesters, E., Kriezis, A., Kyrou, K., Game, L., Nolan, T., & Galizi, R. (2023). Single-cell profiling of *Anopheles gambiae* spermatogenesis defines the onset of meiotic silencing and premeiotic overexpression of

the X chromosome. *Communications Biology*, *6*(1), 850.

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, *28*(21), 4709–4724.

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Moser, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., … Lien, S. (2019). Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, *3*(12), 1731–1742.

Piasecka, B., Lichocki, P., Moretti, S., Bergmann, S., & Robinson-Rechavi, M. (2013). The hourglass and the early conservation models — co-existing patterns of developmental constraints in vertebrates. *PLoS Genetics*, *9*(4), e1003476.

Pomiankowski, A., & Moller, A. P. (1995). A resolution of the lek paradox. *Proceedings of the Royal Society B*, *260*(1357), 21–29.

Presgraves, D. C., Baker, R. H., & Wilkinson, G. S. (1999). Coevolution of sperm and female reproductive tract morphology in stalk–eyed flies. *Proceedings of the Royal Society B*, *266*(1423), 1041–1047.

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R., & Wright, A. E. (2022a). Detecting signatures of selection on gene expression. *Nature Ecology & Evolution*, *6*(7), 1035–1045.

Price, P. D., Palmer Droguett, D. H., Taylor, J. A., Kim, D. W., Place, E. S., Rogers, T. F., Mank, J. E., Cooney, C. R., & Wright, A. E. (2022b). Reply to: Existing methods are effective at measuring natural selection on gene expression. *Nature Ecology & Evolution*, *6*(12), 1838–1839.

Price, P. D., Parkus, S. M., & Wright, A. E. (2023). Recent progress in understanding the genomic architecture of sexual conflict. *Current Opinion in Genetics & Development*, *80*(102047), 102047.

Price, T. A. R., Bretman, A. J., Avent, T. D., Snook, R. R., Hurst, G. D. D., & Wedell, N. (2008a). Sex ratio distorter reduces sperm competitive ability in an insect. *Evolution*, *62*(7), 1644–1652.

Price, T. A. R., Hodgson, D. J., Lewis, Z., Hurst, G. D. D., & Wedell, N. (2008b). Selfish genetic elements promote polyandry in a fly. *Science*, *322*(5905), 1241–1243.

Ramm, S. A., & Schärer, L. (2014). The evolutionary ecology of testicular function: size isn't everything. *Biological Reviews of the Cambridge Philosophical Society*, *89*(4), 874–888.

Reinhardt, J. A., Baker, R. H., Zimin, A. V., Ladias, C., Paczolt, K. A., Werren, J. H., Hayashi, C. Y., & Wilkinson, G. S. (2023). Impacts of Sex Ratio Meiotic Drive on Genome Structure

and Function in a Stalk-Eyed Fly. *Genome Biology and Evolution*, *15*(7), evad118.

Rhie, A., Nurk, S., Cechova, M., Hoyt, S. J., Taylor, D. J., Altemose, N., Hook, P. W., Koren, S., Rautiainen, M., Alexandrov, I. A., Allen, J., Asri, M., Bzikadze, A. V., Chen, N.-C., Chin, C.-S., Diekhans, M., Flicek, P., Formenti, G., Fungtammasan, A., … Phillippy, A. M. (2023). The complete sequence of a human Y chromosome. *Nature*, *621*(7978), 344–354.

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, *38*(4), 735–742.

Robben, M., Ramesh, B., Pau, S., Meletis, D., Luber, J., & Demuth, J. (2024). scRNA-seq Reveals Novel Genetic Pathways and Sex Chromosome Regulation in *Tribolium* Spermatogenesis. *Genome Biology and Evolution*, *16*(3), evae059.

Rodríguez-Montes, L., Ovchinnikova, S., Yuan, X., Studer, T., Sarropoulos, I., Anders, S., Kaessmann, H., & Cardoso-Moreira, M. (2023). Sex-biased gene expression across mammalian organ development and evolution. *Science*, *382*(6670), eadf1046.

Roulin, A. (2016). Condition-dependence, pleiotropy and the handicap principle of sexual selection in melanin-based colouration: Melanin and sexual selection. *Biological Reviews of the Cambridge Philosophical Society*, *91*(2), 328–348.

Rowe, L., & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society B*, *263*(1375), 1415–1421.

Said, I., Byrne, A., Serrano, V., Cardeno, C., Vollmers, C., & Corbett-Detig, R. (2018). Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(21), 5492–5497.

Shami, A. N., Zheng, X., Munyoki, S. K., Ma, Q., Manske, G. L., Green, C. D., Sukhwani, M., Orwig, K. E., Li, J. Z., & Hammoud, S. S. (2020). Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell*, *54*(4), 529–547.e12.

Shin, D., Jin, L., Lobo, N. F., & Severson, D. W. (2011). Transcript profiling of the meiotic drive phenotype in testis of *Aedes aegypti* using suppressive subtractive hybridization. *Journal of Insect Physiology*, *57*(9), 1220–1226.

Simmons, L. W., & Kotiaho, J. S. (2002). Evolution of ejaculates: patterns of phenotypic and genotypic variation and condition dependence in sperm competition traits. *Evolution*, *56*(8), 1622–1631.

Smith, J., Alfieri, J. M., Anthony, N., Arensburger, P., Athrey, G. N., Balacco, J., Balic, A., Bardou, P., Barela, P., Bigot, Y., Blackmon, H., Borodin, P. M., Carroll, R., Casono, M.

C., Charles, M., Cheng, H., Chiodi, M., Cigan, L., Coghill, L. M., … Zhou, H. (2022). Fourth report on chicken genes and chromosomes 2022. *Cytogenetic and Genome Research*, *162*(8-9), 405–528.

Snook, R. R. (1998). The risk of sperm competition and the evolution of sperm heteromorphism. *Animal Behaviour*, *56*(6), 1497–1507.

Soh, Y. Q. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., Fulton, R. S., Kremitzki, C., Koutseva, N., Mueller, J. L., Rozen, S., Hughes, J. F., Owens, E., Womack, J. E., Murphy, W. J., Cao, Q., de Jong, P., Warren, W. C., Wilson, R. K., … Page, D. C. (2014). Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, *159*(4), 800–813.

Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H., & Hartl, D. L. (2007). A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biology*, *5*(11), e293.

Taxiarchi, C., Kranjc, N., Kriezis, A., Kyrou, K., Bernardini, F., Russell, S., Nolan, T., Crisanti, A., & Galizi, R. (2019). High-resolution transcriptional profiling of *Anopheles gambiae* spermatogenesis reveals mechanisms of sex chromosome regulation. *Scientific Reports*, *9*(1), 14841.

Taylor, J. E., & Jaenike, J. (2002). Sperm competition and the dynamics of X chromosome drive: stability and extinction. *Genetics*, *160*(4), 1721–1731.

Tomaszkiewicz, M., Medvedev, P., & Makova, K. D. (2017). Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends in Genetics*, *33*(4), 266–282.

Toups, M. A., & Hahn, M. W. (2010). Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics*, *186*(2), 763–766.

Turner, J., Krishna, R., Van't Hof, A. E., Sutton, E. R., Matzen, K., & Darby, A. C. (2018). The sequence of a male-specific genome region containing the sex determination switch in *Aedes aegypti*. *Parasites & Vectors*, *11*(1), 549.

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation. *Development*, *134*(10), 1823–1831.

Turner, J. M. A., Mahadevaiah, S. K., Ellis, P. J. I., Mitchell, M. J., & Burgoyne, P. S. (2006). Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Developmental Cell*, *10*(4), 521–529.

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(59).

Vibranovski, M. D. (2014). Meiotic sex chromosome inactivation in *Drosophila*. *Journal of Genomics*, *2*, 104–117.

Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, *30*(12), 2738–2755.

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227–232.

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., Zecha, J., Asplund, A., Li, L.-H., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., & Kuster, B. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology*, *15*(2), e8503.

Wang, Z.-Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M. E., Peters, A. H. F. M., Anders, S., & Kaessmann, H. (2020). Transcriptome and translatome co-evolution in mammals. *Nature*, *588*(7839), 642–647.

Wei, K. H.-C., 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*(4), e3002605.

Wilkinson, G. S., Johns, P. M., Kelleher, E. S., Muscedere, M. L., & Lorsong, A. (2006). Fitness effects of X chromosome drive in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. *Journal of Evolutionary Biology*, *19*(6), 1851–1860.

Wilkinson, G. S., Presgraves, D. C., & Crymes, L. (1998). Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature*, *391*(6664), 276–279.

Witt, E., Langer, C. B., Svetec, N., & Zhao, L. (2023). Transcriptional and mutational signatures of the *Drosophila* ageing germline. *Nature Ecology & Evolution*, *7*(3), 440–449.

Witt, E., Shao, Z., Hu, C., Krause, H. M., & Zhao, L. (2021). Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in *Drosophila* testis. *PLoS Genetics*, *17*(8), e1009728.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88.

Wright, A. E., Harrison, P. W., Zimmer, F., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. *Molecular Ecology*, *24*(6), 1218–1235.

Wright, A. E., & Mank, J. E. (2013). The scope and strength of sex-specific selection in genome evolution. *Journal of Evolutionary Biology*, *26*(9), 1841–1853.

Wu, C. I., & Xu, E. Y. (2003). Sexual antagonism and X inactivation--the SAXI hypothesis. *Trends in Genetics*, *19*(5), 243–247.

Xia, B., & Yanai, I. (2022). Gene expression levels modulate germline mutation rates through the compound effects of transcription-coupled repair and damage. *Human Genetics*, *141*(6), 1211–1222.

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S. Y., Keefe, D. L., Alukal, J. P., Boeke, J. D., & Yanai, I. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*, *180*(2), 248–262.e21.

Yassin, A., Bastide, H., Chung, H., Veuille, M., David, J. R., & Pool, J. E. (2016). Ancient balancing selection at tan underlies female colour dimorphism in *Drosophila erecta*. *Nature Communications*, *7*(1), 10400.

Appendices

The following section contains additional publications and materials that complement the first six chapters of this thesis.

Appendix A. Reply to: Existing methods are effective at measuring natural selection on gene expression

Peter D. Price[1], Daniela H. Palmer Droguett[1,2], Jessica A. Taylor[1,3], Dong Won Kim[4], Elsie S. Place[5], Thea F. Rogers[1], Judith E. Mank[6,7,8], Christopher R. Cooney[1], Alison E. Wright[1]

[1] Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK.

[2] Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA.

[3] Department of Biosciences, Durham University, Durham, UK.

[4] Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[5] Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, Sheffield, UK.

[6] Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada.

[7] Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada.

[8] Centre for Ecology and Conservation, University of Exeter, Penryn, UK.

Price, Peter D., Daniela H. Palmer Droguett, Jessica A. Taylor, Dong W. Kim, Elsie S. Place, Thea F. Rogers, Judith E. Mank, Christopher R. Cooney, and Alison E. Wright. 2022. “Reply to: Existing Methods Are Effective at Measuring Natural Selection on Gene Expression.” Nature Ecology & Evolution 6 (12): 1838–39.

nature ecology & evolution

Matters arising

https://doi.org/10.1038/s41559-022-01916-7

Reply to: Existing methods are effective at measuring natural selection on gene expression

Received: 22 June 2022

Accepted: 23 September 2022

Published online: 7 November 2022

Check for updates

Peter D. Price[1], Daniela H. Palmer Droguett[1,2], Jessica A. Taylor[1], Dong W. Kim[3], Elsie S. Place[4], Thea F. Rogers[1], Judith E. Mank[5,6,7], Christopher R. Cooney[1] & Alison E. Wright[1]

REPLYING TO: Hunter B. Fraser *Nature Ecology & Evolution* https://doi.org/10.1038/s41559-022-01889-7 (2022).

We read with great interest the commentary by Fraser[1] and wholeheartedly agree that understanding how selection acts on patterns of gene expression is key to identifying mechanisms of adaptive change. In Price et al.[2], we identified significant challenges to testing how the transcriptome evolves, specifically that shifts in tissue composition can bias inferences of selection over long evolutionary time frames. However, Fraser[1] suggests the problems we outline have been 'largely solved by the research community' in two ways—through studies of cell lines and interspecific hybrids. We are in complete agreement that both approaches circumvent issues arising from varying cell type abundance that we highlight and so are useful tools to accurately measure expression change. However, neither represents a panacea for detecting natural selection on gene expression that Fraser[1] suggests.

Contrasts of cell lines can be used to accurately identify regulatory variation and in principle can be applied over greater evolutionary distances to quantify the mode of gene expression evolution among distantly related species. However, creating cell lines is non-trivial and likely not feasible for many species. Importantly for multicellular model systems, the diversity of cell types that can be cultured is severely limited and the costs in doing so prohibitive if all cell types in a tissue are to be included. This means that this approach is unlikely to extend across the tree of life in the near future.

Most importantly, organisms are far more than the sum of their parts. Changes in tissue composition are key to the evolution of many adaptive phenotypes (for example, refs. [3–6]) and likely the product of differences in expression across development[7]. Therefore, by their very nature, cell types analysed individually have limited potential for studying the developmental regulatory changes that produce variation in cell type abundance and complex adaptive traits. Consistent with these limitations, the vast majority of the cell line research cited by Fraser[1] does not test for selection on gene expression levels, with the exception of three studies using primate cell lines[8–10]. Examining cell types one at a time is therefore unlikely to yield a comprehensive picture of differences between species.

The second approach highlighted by Fraser[1] is the sign test of selection[11], which was extended by Fraser et al.[12] to test for selection on gene expression. This method has provided important insight into how gene expression evolves[13], including the first known example of polygenic gene expression adaptation[12], and we have no wish to diminish this important contribution to the field. However, since this approach relies on prior knowledge of the directionality of genetic changes affecting a quantitative trait, to our knowledge, it has exclusively been applied to species that can produce viable hybrids, namely very closely related species[14–16] or subspecies[12,17,18]. Therefore, while informative for understanding expression evolution over very short evolutionary time frames, its potential to study many instances of adaptive change over the full breadth of evolutionary time is limited.

Together, neither approach suggested by Fraser[1] is widely applicable outside of model systems, limited cell types that are readily cultured, or relatively narrow evolutionary windows, leaving large gaps in both the scope of questions that can be addressed and the range of organisms that can be studied. Instead, we see a number of important points emerging from Fraser[1]. First, developmental context matters for the evolution of many adaptive phenotypes, particularly in multicellular organisms. For these traits we should focus not on eliminating differences in cellular composition but instead properly

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK. [2]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA. [3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [4]Development, Regeneration and Neurophysiology, School of Biosciences, University of Sheffield, Sheffield, UK. [5]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [6]Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada. [7]Centre for Ecology and Conservation, University of Exeter, Penryn, UK. e-mail: pprice3@sheffield.ac.uk; a.e.wright@sheffield.ac.uk

accounting for such differences when testing for selection on gene expression, potentially through the use of single-cell RNA sequencing. Second, it is likely that selection pressures vary over short versus long evolutionary time frames and limiting our analyses to closely related species will bias our understanding of how the transcriptome evolves. Comparative approaches that sample a range of evolutionary scales are clearly essential to understand the full spectrum of evolutionary responses to selection. Therefore, addressing the confounding issues of cellular composition, as discussed in Price et al.[2], is a major priority for the field.

References

1. Fraser, H. B. Existing methods are effective at measuring natural selection on gene expression. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-022-01889-7 (2022).
2. Price, P. D. et al. Detecting signatures of selection on gene expression. *Nat. Ecol. Evol.* **6**, 1035–1045 (2022).
3. Lüpold, S., Linz, G. M., Rivers, J. W., Westneat, D. F. & Birkhead, T. R. Sperm competition selects beyond relative testes size in birds. *Evolution* **63**, 391–402 (2009).
4. Montgomery, S. H. & Merrill, R. M. Divergence in brain composition during the early stages of ecological specialization in *Heliconius* butterflies. *J. Evol. Biol.* **30**, 571–582 (2017).
5. Firman, R. C. et al. Evolutionary change in testes tissue composition among experimental populations of house mice. *Evolution* **69**, 848–855 (2015).
6. Kaskan, P. M. et al. Peripheral variability and central constancy in mammalian visual system evolution. *Proc. Biol. Sci.* **272**, 91–100 (2005).
7. Mank, J. E. & Rideout, E. J. Developmental mechanisms of sex differences: from cells to organisms. *Development* **148**, dev199750 (2021).
8. Khan, Z. et al. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104 (2013).
9. Gokhman, D. et al. Human–chimpanzee fused cells reveal *cis*-regulatory divergence underlying skeletal evolution. *Nat. Genet.* **53**, 467–476 (2021).
10. Agoglia, R. M. et al. Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. *Nature* **592**, 421–427 (2021).
11. Orr, H. A. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**, 2099–2104 (1998).
12. Fraser, H. B., Moses, A. M. & Schadt, E. E. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl Acad. Sci. USA* **107**, 2977–2982 (2010).
13. Fraser, H. B. Genome-wide approaches to the study of adaptive gene expression evolution: systematic studies of evolutionary adaptations involving gene expression will allow many fundamental questions in evolutionary biology to be addressed. *Bioessays* **33**, 469–477 (2011).
14. Tirosh, I., Reikhav, S., Levy, A. A. & Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**, 659–662 (2009).
15. Singh-Babak, S. D., Babak, T., Fraser, H. B. & Johnson, A. D. Lineage-specific selection and the evolution of virulence in the *Candida* clade. *Proc. Natl Acad. Sci. USA* **118**, e2016818118 (2021).
16. Naranjo, S. et al. Dissecting the genetic basis of a complex *cis*-regulatory adaptation. *PLoS Genet.* **11**, e1005751 (2015).
17. Fraser, H. B. et al. Systematic detection of polygenic *cis*-regulatory evolution. *PLoS Genet.* **7**, e1002023 (2011).
18. Fraser, H. B. et al. Polygenic *cis*-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res.* **22**, 1930–1939 (2012).

Author contributions

All authors contributed equally.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Peter D. Price or Alison E. Wright.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Appendix B: Chapter 4 supplementary materials

Supplementary materials from Chapter 4. Relaxed purifying selection maintains a sex-linked supergene polymorphism in zebra finches

Chapter 4 supplementary figures

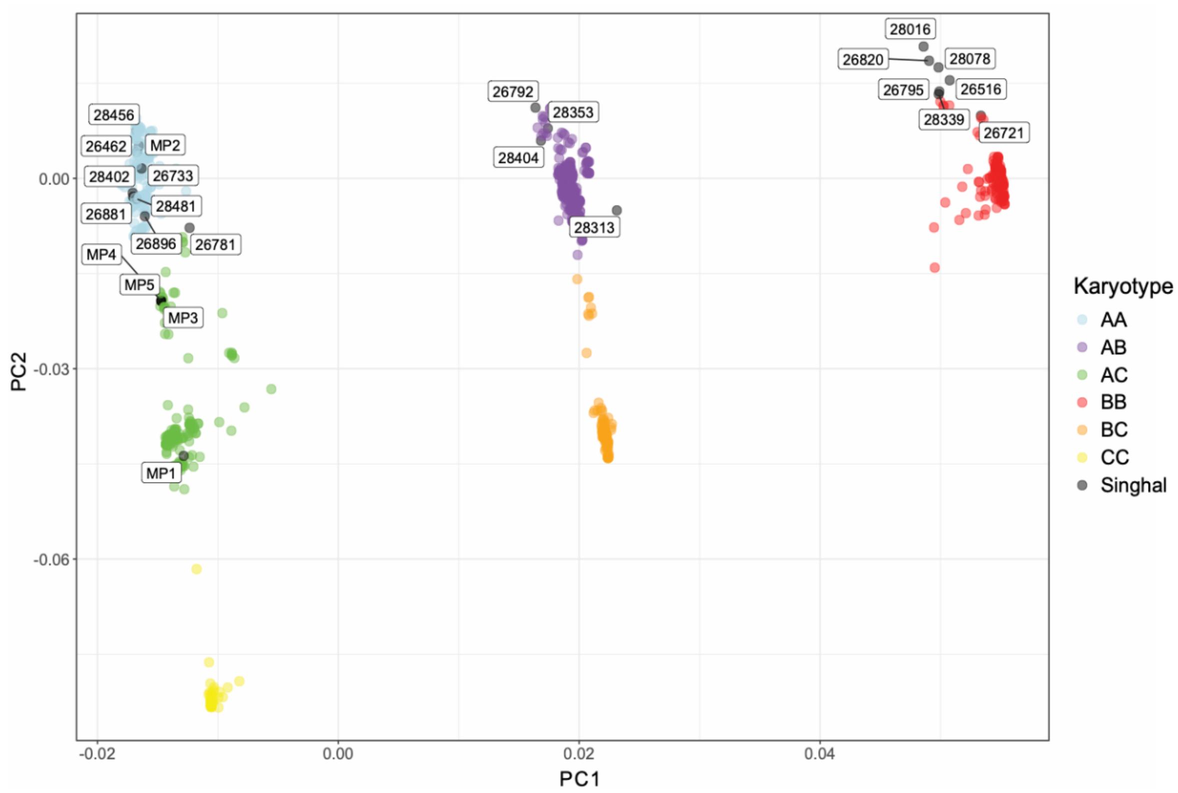


Figure S4.1. PCA including 24 zebra finches from Singhal et al., 2015. *Principal component analysis (PCA) of 1202 male zebra finches from a population that was previously maintained at the University of Sheffield along with 24 zebra finches (11 female, 13 male) sequenced by Singhal et al., and published in 2015. Birds from the University of Sheffield were of known inversion karyotypes and are labelled accordingly. 2330 SNP positions were used, each with weighted chi-square statistics >0.9 in at least one pairwise comparison between haplotypes.*

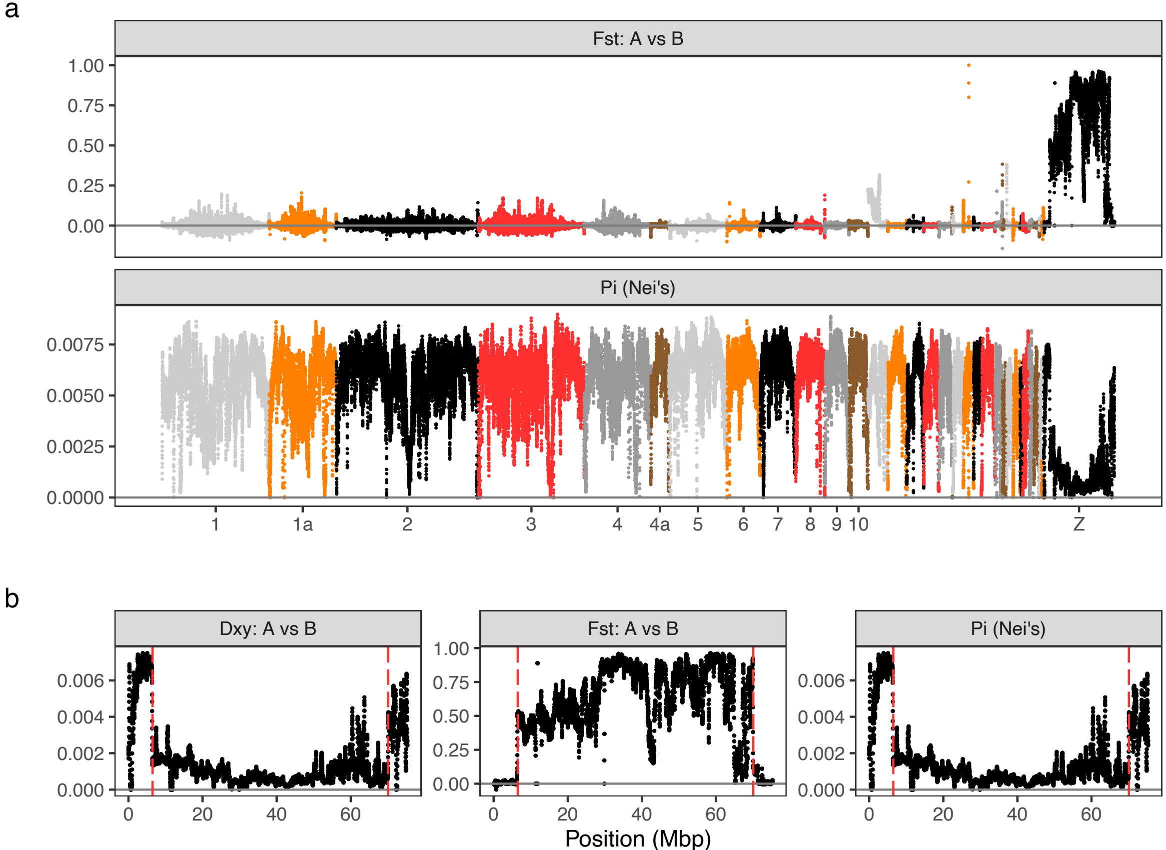


Figure S4.2. Neutrality (a) and Diversity (b) statistics on the Z chromosome for A and B haplotypes controlled for differences in Z chromosome number. *All statistics were calculated as in Figure 4.1 but controlled for chromosome number, including only 8 chromosomes to represent A or B on the autosomes and Z chromosome.*

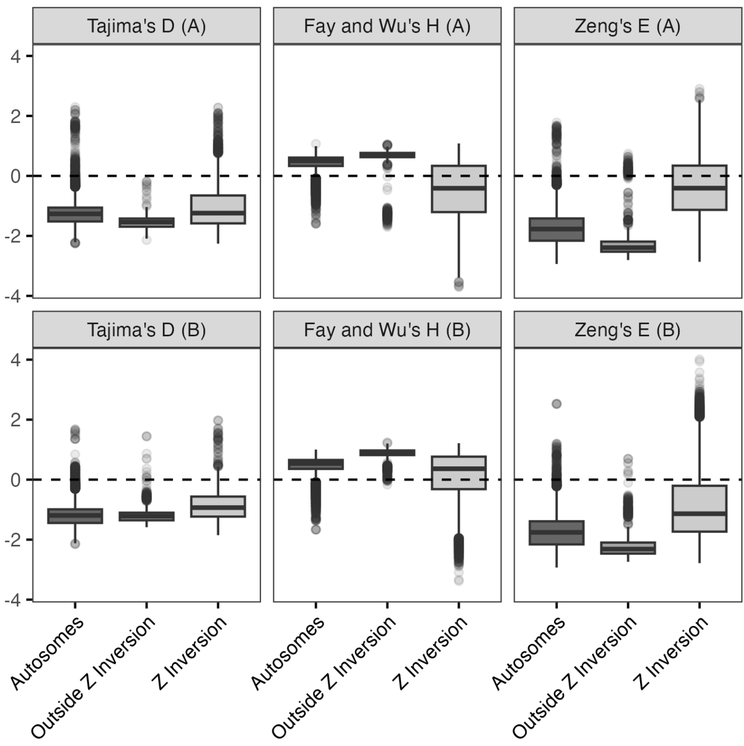


Figure S4.3. Site frequency spectrum (SFS) statistics for A and B haplotypes. *Tajima's D, Fay and Wu's H, and Zeng's E were calculated for AA/A (top row) and BB/B (bottom row) birds across the autosomes and Z chromosome. SFS statistics were calculated in 100kbp sliding windows with 10kbp steps and divided into windows from the autosomes, regions outside the inversion, and the inversion.*

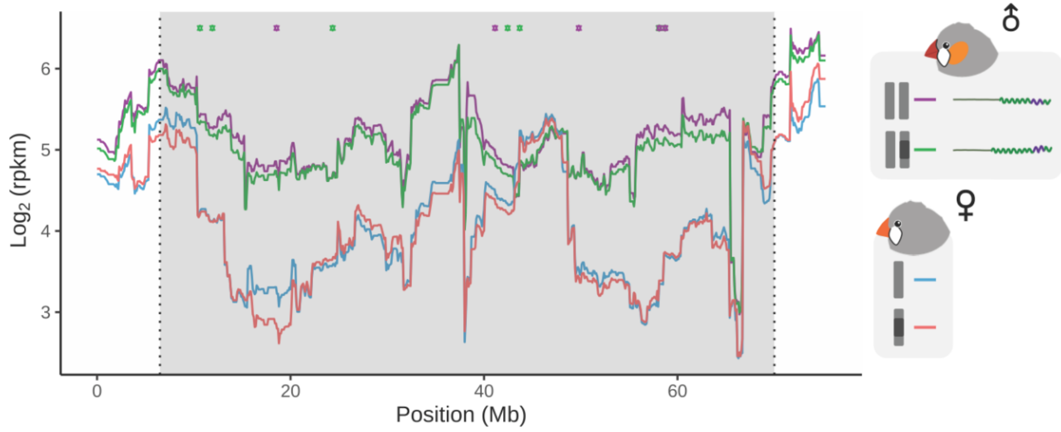


Figure S4.4. Moving average of gene expression across the Z chromosome for males and females of different genotypes*. Genes were only included if they were expressed at >2rpkm in at least one individual of at least one genotype. Average expression was recorded in 5000kbp sliding windows with 100kbp steps. Black dotted lines indicate boundaries of the inverted region and grey bars indicate boundaries of different strata. Asterisks show genes that are significantly differentially expressed between AB and AA males (pfdr<0.05, log2 fold change > 1 or < -1). Asterisks and fitted lines are coloured by genotype and sex (AA: purple, AB: green, A: blue, B: red) with AB biased genes coloured green and AA biased genes coloured purple.*

Chapter 4 supplementary tables

Table S4.1. Online material: ***List of diagnostic SNPs to karyotype the Z-linked inversion.***

Table S4.2. Inversion haplotypes of samples from Singhal et al., 2015.

Sample Alias	Sample Accession	Sample Source	Sex	A support vs B (%)	A support vs C (%)	B support vs C (%)	Inferred Karyotype	Mean Coverage
26462	SAMEA3532857	W	F	**89.5**	**93.0**	20.0	A	27.9
28456	SAMEA3532861	W	F	**90.6**	**93.2**	15.8	A	27.7
26881	SAMEA3532868	W	F	**85.1**	**88.8**	10.5	A	50.2
26896	SAMEA3532870	W	F	**88.5**	**92.6**	13.8	A	18.7
MP2	SAMEA3532853	D	M	**88.3**	**89.8**	17.9	AA	29.8
28402	SAMEA3532862	W	M	**85.6**	**77.9**	12.5	AA	16.7
26733	SAMEA3532866	W	M	**85.1**	**88.3**	19.5	AA	26.1
28481	SAMEA3532867	W	M	**90.3**	**82.0**	13.3	AA	33.4
28404	SAMEA3532864	W	M	37.0	40.6	68.5	AB	20.9
28353	SAMEA3532859	W	M	35.4	44.8	70.2	AB	32.7
26792	SAMEA3532871	W	M	37.7	39.2	66.3	AB	20.7
28313	SAMEA3532875	W	M	30.9	19.3	71.0	AB	28.3
28339	SAMEA3532858	W	F	**7.2**	8.5	**99.1**	B	19.9
26721	SAMEA3532860	W	F	**6.9**	9.5	**99.8**	B	20.7
26820	SAMEA3532865	W	F	**7.4**	14.6	**99.3**	B	29.2
28016	SAMEA3532872	W	F	**8.3**	15.5	**98.8**	B	18.7
26795	SAMEA3532873	W	F	**7.5**	8.9	**99.1**	B	34.2
28078	SAMEA3532874	W	F	**7.7**	16.0	**99.3**	B	25.5
26516	SAMEA3532863	W	M	**7.3**	11.6	**99.4**	BB	22.3
MP3	SAMEA3532854	D	M	77.5	43.8	17.0	AC	23.2
MP4	SAMEA3532855	D	M	77.0	44.3	16.4	AC	19.7
MP5	SAMEA3532856	D	M	75.5	41.9	16.5	AC	28.9
26781	SAMEA3532869	W	M	77.4	61.8	26.3	AC	24.2
MP1	SAMEA3532852	D	F	76.5	**16.7**	**5.8**	C	27.3

Percentage support for each inversion haplotype (A, B & C) in three pairwise comparisons (A vs B, A vs C & B vs C) for each of 24 zebra finches from Singhal et al., 2015. Percentages were calculated using genotypes at diagnostic SNP positions (2123 A vs B, 1085 B vs C & 293 A vs C positions) across the Z chromosome. Karyotypes were determined by percentage support, with a >75% support for a haplotype in both comparisons in which it appears used to infer the haplotype. E.g., >75% support for A in A vs B and in A vs C comparisons suggests A/AA karyotype. All other male individuals were assumed to be hetero-karyotypes. For sample source, D = domesticated, W = wild. Wild samples were collected from Fowlers Gap, New South Wales, Australia, and the domesticated were a captive bred family from East Carolina University. Mean coverage as reported by Singhal et al., 2015.

Table S4.3. Inversion haplotypes of samples with RNA-seq data.

Sample	Sex	No. of raw read pairs	No. of trimmed reads pairs	No. of mapped read pairs	Proportion of diagnostic SNPs			Inferred karyotype
					Heterozygous AB	Homozygous A	Homozygous B	
TGU13	F	71401584	67267309	60655318	0	0	1	B
TGU14	M	76540489	72047716	64437435	0.95	0.017	0.033	AB
TGU17	M	71770103	67576630	60386964	0.929	0.018	0.054	AB
TGU22	F	73578953	69473397	62834678	0	0.941	0.059	A
TGU23	F	72504308	68507902	62285823	0	0.871	0.129	A
TGU25	F	62162745	58552856	52934599	0	0.028	0.972	B
TGU26	M	94182347	88262097	79001666	0.914	0.043	0.043	AB
TGU27	M	73793757	69275918	61968965	0.096	0.904	0	AA
TGU28	F	85171479	79914253	71768906	0	0	1	B
TGU29	M	62747602	58802101	52587371	0.116	0.884	0	AA

Table S4.4. Online material: ***Fixed differences between A and B haplotypes within the coding region.***

Table S4.5. Differential expression between genotypes. *Genes were only included in the analysis if they were expressed > 2rpkm in at least 50% of individuals in either genotype in the pairwise comparison. Differential expression:* p_{fdr}*<0.05 and |log fold change| > 1.*

	AA versus AB males			A versus B females		
	Differentially expressed genes	AB biased	AA biased	Differentially expressed genes	A biased	B biased
Autosomes	146 (1.16%)	53	93	69 (0.57%)	41	28
Z-linked inversion	13 (2.51%)	8	5	6 (1.35%)	5	1

Table S4.6. Number of shared genes with and without allele-specific expression.

Genotype	Chromosome	ASE	Non-ASE	P_{MCMC}	χ^2	N
AB	Autosomes (1-10)	21	1875	$4x10^{-3}$	11.19	2095
AB	Z	8	191			
AA	Autosomes (1-10)	83	3224	$3.4x10^{-5}$	24.88	3505
AA	Z	17	181			

Table S4.7. Number of genes exhibiting allele-specific expression on the Z and autosomes in each male individual.

Sample	Genotype	Chromosome	ASE	Non-ASE	Expected ASE	Expected Non-ASE	P_{MCMC}	χ^2	N
TGU14	AB	Autosomes (1-10)	382	6186	414	6154	1×10^{-6}	44.800	6980
		Z	58	354	26	386			
TGU17	AB	Autosomes (1-10)	390	6070	407	6053	4.7×10^{-4}	13.41	6866
		Z	43	363	26	380			
TGU26	AB	Autosomes (1-10)	421	5590	443	5568	5.5×10^{-5}	18.59	6428
		Z	53	364	31	386			
TGU27	AA	Autosomes (1-10)	410	6044	431	6023	8×10^{-6}	24.39	6755
		Z	41	260	20	281			
TGU29	AA	Autosomes (1-10)	392	6030	428	5994	1×10^{-6}	70.37	6727
		Z	56	249	20	285			

Table S4.8. Allele-specific expression in the Z-linked inversion between A and B haplotypes. *Genes were only included in the analysis if they were expressed > 2rpkm in at least 1 individual. Allele-specific expression was only called for a gene if all three individuals had a significant difference in expression ($p_{fdr} < 0.05$) between the A and B haplotype. Gene function information from Gene Ontology and Gene Cards.*

Gene ID	Position on Z	Chicken Ortholog	Gene Function	Average absolute log2 fold change
LOC100224449	18081886 to 18083539	MRPL50	Mitochondrial	2.486
LOC115491093	11270346 to 11274103	N/A (ITPRIPL1 in mammalia)	ITPRIPL1: protein binding	2.194
LOC115491113	28015157 to 28020234	N/A	N/A	2.386
LOC115491209	49071125 to 49075335	N/A	N/A	4.782
LOC116806769	18512447 to 18518289	N/A (YTHDC2-like)	YTHDC2: Nucleotide/nucleic acid binding, RNA helicase activity	2.604
LOC116808147	17982405 to 17995747	HSF5	DNA binding	1.807
MCCC2	34897755 to 34935512	MCCC2	Nucleotide binding, ligase activity, enables protein/ATP binding	6.551
RPS6	69280757 to 69285251	RPS6	G1/S transition, mitosis, placental development, cytoplasmic translation	9.439

Appendix C: Chapter 5 supplementary materials

Supplementary materials from Chapter 5. The single-cell consequences of an X-linked meiotic driver in stalk-eyed flies.

Chapter 5 supplementary results

Expression of marker genes across cell types in *Teleopsis dalmanni* testes

First, we used a series of known *Drosophila melanogaster* cell-type-specific testes markers (Witt et al., 2019; Mahadevaraju et al., 2021; Li et al., 2022) to identify *T. dalmanni* cell types (Figures 5.1 & S5.1c, Table S5.2). We used the germ-cell-specific marker *vasa* to separate the germline from somatic tissues (Ohlstein & McKearin, 1997). Somatic tissues were split using *Mhc* to identify muscle cells (Hess et al., 2007), *eya* expression to identify cyst cells associated with post-mitotic germline cells (Zoller & Schulz, 2012), and *Impl2*, *fng*, *tj* and *Nrt* to identify cyst cells associated with the mitotic germline (Terry et al., 2006). To then separate the germline stem cells (GSC) and spermatogonia from meiotic phase cells, *bb8* was used, a key gene in mitochondrial derivative development, which is expressed from spermatocytes onwards (Vedelek et al., 2016). The lack of *twe* expression was used to split spermatocytes from spermatids, with *twe* expression peaking in spermatocytes and not detected in spermatids (Courtot et al., 1992). The presence of *cup* genes could also validate spermatids as whilst the post-meiotic germline of *Drosophila* is transcriptomically nearly inactive, *cup* and *comet* genes are expressed (Barreau et al., 2008). To then separate primary from secondary spermatocytes, we used expression of *CycB*, a core G2/M cell cycle component, whose expression peaks at the meiosis I transition between primary and secondary spermatocytes, and *Fest*, a regulator of *CycB*, whose expression begins in primary spermatocytes and extends into late-stage germ cells (Baker et al., 2015). These key markers, used to define cell types, were highly expressed in our dataset and shown in 5.1c. However, there is a wider set of markers in the literature that we used to corroborate our findings, shown in Figure S5.1c and Table S5.2, but some of these genes were less expressed in our dataset and so less reliable in defining cell types.

We then used a large set of eukaryotic mitotic cycle stage classifiers, from the Harvard Chan School: Bioinformatics Core (https://github.com/hbc/tinyatlas/blob/master/cell_cycle/Drosophila_melanogaster.csv), (Figures S5.1a, Table S5.3) to refine the stages of spermatogenesis with expression of G2M markers enriched in the spermatogonia through to primary spermatocytes (Robben et al., 2024).

Using ploidy to distinguish cell types across spermatogenesis

A recent study proposed using SNP-based haploid/diploid phasing to distinguish pre- from post-meiotic cell types (Robben et al., 2024). We expect the GSC/spermatogonia and primary spermatocytes to be diploid and secondary spermatocytes and spermatids to be haploid. Therefore, we followed the approach of Robben et al (2024) to call SNPs across our cells in standard (ST) males. Briefly, using the cellranger generated BAM files for each sample, duplicate reads were marked and removed using GATK's MarkDuplicates, reads with Ns in their cigar string were split with SplitNCigarReads and finally, SNPs were called using HaplotypeCaller. Variants were then filtered for a minGQ of 20, minDP of 4 and minQ of 30. Using this set of SNPs, we identified single-cell level variants using scAlleleCount (https://github.com/barkasn/scAlleleCount). We then calculated the level of heterozygosity for each cell as an estimate of ploidy, classifying cells as haploid if heterozygous at < 95% of SNPs.

As we observed a decrease in transcriptional activity over spermatogenesis, we first checked for any confounding relationship between coverage depth and heterozygosity level that might bias our ability to robustly distinguish haploid and diploid cell types. We focused on diploid cell types (somatic tissue, cyst, GSC/spermatogonia and primary spermatocytes) for this test. We found a negative and significant relationship ($p < 0.0001$, $\rho = -0.53$) where cells with less coverage exhibited lower heterozygosity (Figure S5.2a, S5.2b). This is likely due to reduced power to confidently identify the minor allele when coverage is low. Therefore, to mitigate this bias we significantly increased our filtering for SNP calling to require at <= 2 reads for homozygotes to be called, and 4 for heterozygotes (2 reads mapping to both ref and alt) and a cell requiring variant information at >= 10 SNPs. However, even with this strict filtering, whilst the significant relationship between coverage and heterozygosity was weaker ($p < 0.0001$, $\rho = -0.44$), we still failed to identify differences in ploidy across cell types where expected (i.e., between somatic cells and spermatids) (5.S2c & S5.2d). We hypothesise several reasons for our inability to detect ploidy differences between cell types, discussed below.

First, levels of heterozygosity in our population might be generally low as a result of inbreeding. This would lead to the incorrect assignment of diploid cells as haploid. We investigated this using DNA-sequencing data from five standard males collected from the same population (Appendix C. Chapter 5 supplementary methods). However, levels of heterozygosity are high and there is no evidence of inbreeding (Table S5.5).

Second, the stringent filtering we used to mitigate our reduced power to identify the minor allele when coverage is low means we had very few cells and sites remaining from which to assign ploidy. Specifically, after filtering, there were 2,359 cells remaining from the 4,469 genotyped cells with relaxed filtering. This will significantly reduce our power to robustly identify haploid and diploid cells.

Finally, across much of eukaryotic life, from *Drosophila* to humans, the post-meiotic germline becomes close to transcriptionally inert (Erickson, 1990) after a huge peak in activity in the primary spermatocytes (Witt et al., 2019; Xia et al., 2020). Transcripts that are transcribed in primary spermatocytes can remain dormant in the cytoplasm of the cell to be transcribed later in spermatogenesis (Barreau et al., 2008; Raz et al., 2023). Furthermore, in many species (e.g., rats (Ventelä et al., 2003) and *Drosophila* (Greenbaum et al., 2011) the spermatids within a bundle are joined by large cytoplasmic bridges that enable the movement of mRNA (Braun et al., 1989), proteins (Kaufman et al., 2020) and structures such as large as organelles (Ventelä et al., 2003). This means the cytoplasmic content does not just reflect the transcriptomic activity of one particular cell but also, to an extent, that of those neighbouring it (Braun et al., 1989; Ventelä et al., 2003; Greenbaum et al., 2011). Thus, if SNP calling for ploidy determination is reliant on RNA-seq approaches, reads sequenced from these cells were transcribed not just from themselves but also from earlier diploid cell states and from their neighbours. This could therefore give the appearance of diploidy even if the underlying genomic ploidy is haploid.

In summary, whilst assigning ploidy offers a valuable and alternate route to defining cell types in principle, particularly in non-model organisms where marker genes are either not reliable or absent, we urge caution when interpreting these results, and if necessary propose the use of joint sequencing (Vandereyken et al., 2023) or single-nucleus approaches instead.

Chapter 5 supplementary methods

Alignment, SNP calling and heterozygosity calculations

DNA for nine *T. dalmanni* individuals was extracted using standard approaches. Illumina DNA-seq libraries were generated at the NERC Environmental Omics Facility (NEOF) Liverpool before sequencing with Illumina NovaSeq using S4 chemistry, aiming for a coverage of 30X. FASTQ files were quality trimmed and aligned to the indexed *T. dalmanni* reference genome (van Rensburg et al., 2024) using the BWA-MEM algorithm implemented in BWA v0.7.17 (H. Li & Durbin, 2009). BAM files were sorted and indexed using Samtools v1.11 (Danecek et al., 2011). Read group information was added to the BAM files using Picard tools v2.27.5 (http://broadinstitute.github.io/picard) and PCR duplicates removed using Picard tools MarkDuplicates. Next, variant calling was performed using GATK HaplotypeCaller v4.3.0 (van der Auwera & O'Connor, 2020) to generate GVCF files and remove reads with a mapping quality < 20 and base quality score < 20. GVCFs were merged into a single GVCF using GATK CombineGVCFs and genotypes called using GATK GenotypeGVCFs. Variant filtering was performed using BCFtools v1.11 (Danecek et al., 2021). Genotypes of sites were set to missing (.) if they met one of the following parameters: Depth (DP) < 5 or > 98, genotype quality (GQ) < 50, or SNP quality (QUAL) <= 50. Additionally, only biallelic SNPs were kept. Then, BCFtools view was used to extract only the autosomes (Chr1; Chr2) and the extracted vcf file was indexed using Tabix (htslib) v1.13 (Bonfield et al., 2021). Levels of heterozygosity and the inbreeding coefficient (F) per individual were calculated using VCFtools (-het) v0.1.17 (Danecek et al., 2011).

Chapter 5 supplementary figures

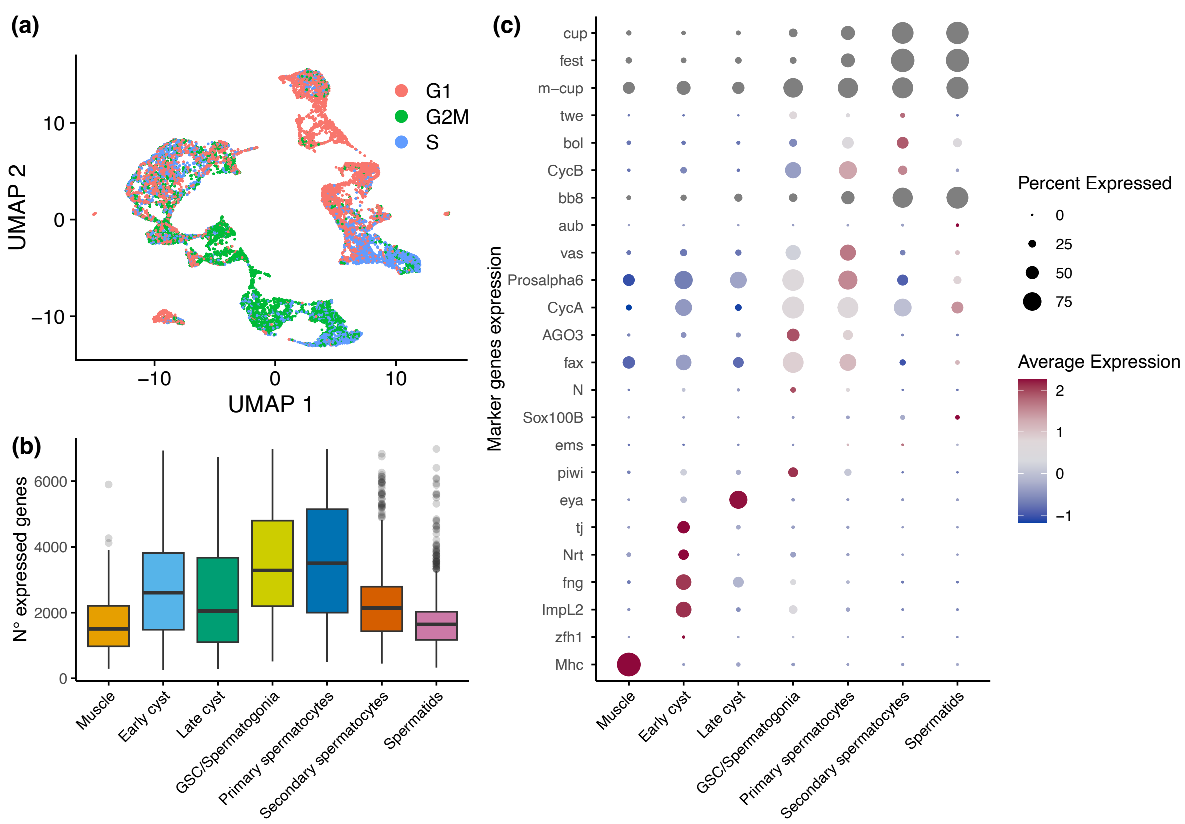


Figure S5.1. Additional cell type assignment approaches.

(a) *Uniform Manifold Approximation and Projection (UMAP) of cells classified by mitotic stage marker expression as a proxy for meiotic stage.* ***(b)*** *Boxplots of number of autosomal genes expressed across all cell types (gene classified as expressed if counts > 1).* ***(c)*** *Dot plot of relative expression of orthologs of Drosophila melanogaster cell-type-specific testis markers. Size of dots indicates the relative number of cells expressing the marker in a cluster and colour indicates the level of expression (blue lowest and red highest).*

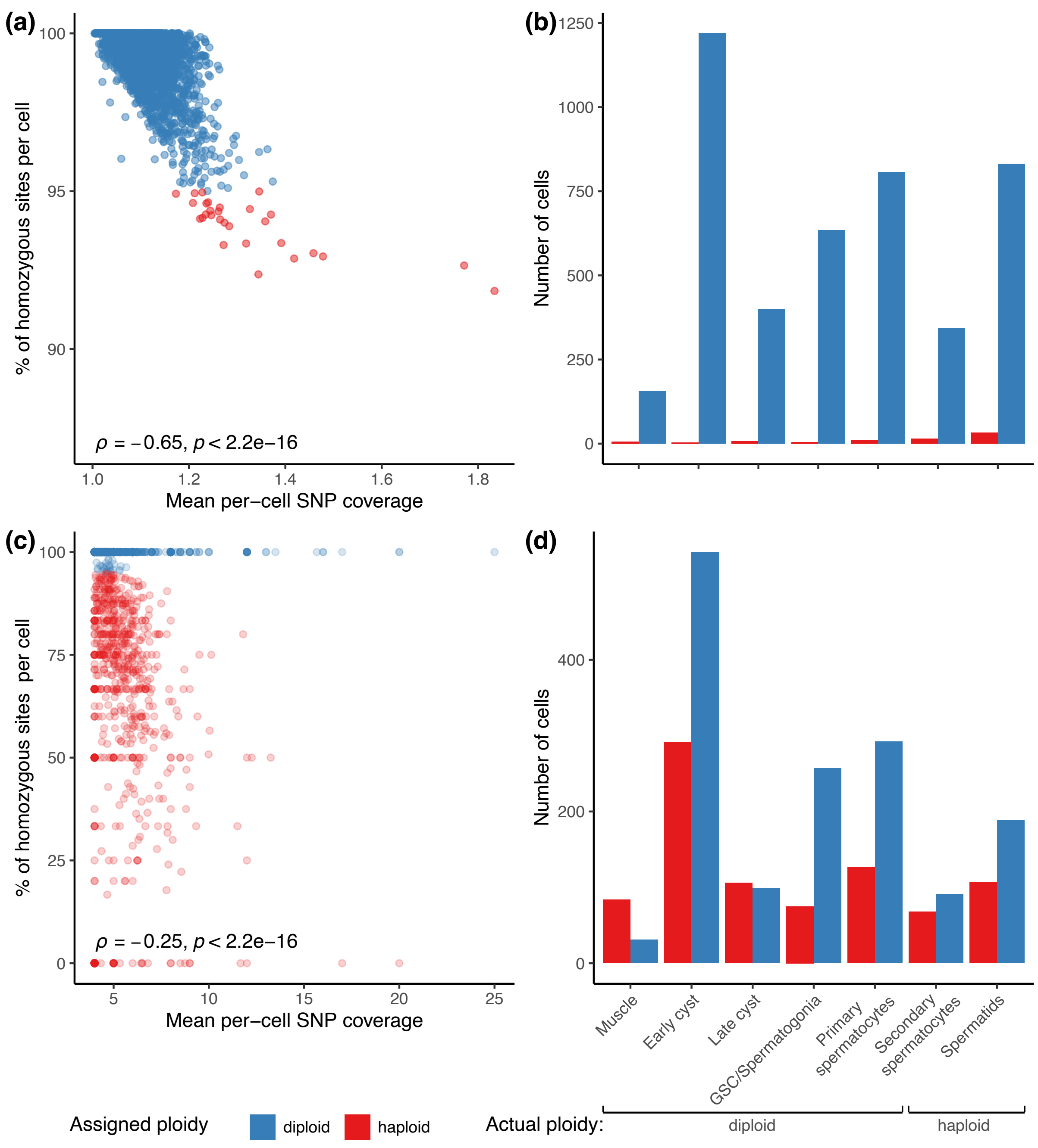


Figure S5.2. Using ploidy to distinguish cell types in ST samples.

__(a)__ Percentage of sites per cell that are homozygous (all reads matching either reference or alternate at a site). Data only shown for standard (ST) male diploid cell types (cyst, muscle, GSC/spermatogonia, and primary spermatocytes) No depth or minimum number of site thresholds were set for calling the ploidy of each cell. __(b)__ Number of cells classified as haploid or diploid for each cell type following the filtering in __(a)__. __(c)__ and __(d)__ are the same as __(a)__ and __(b)__ however a threshold of being genotyped at >= 10 sites per cell with depth of >= two for calling homozygous or four for heterozygous (two mapping to both ref and alt)

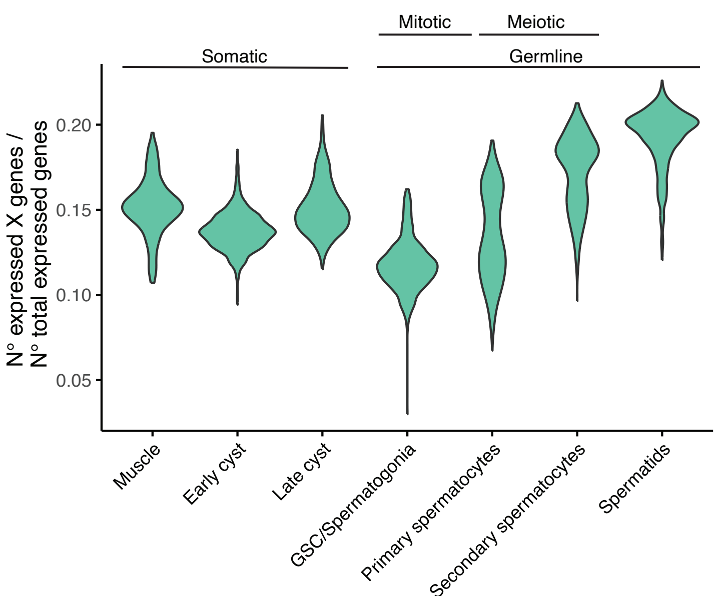


Figure S5.3. X chromosome activity across T. dalmanni spermatogenesis.

Violin plot showing the relative number of X-linked genes expressed across T. dalmanni cell types in standard (ST) males. Relative number calculated as the number of expressed X-linked genes divided by the total number of expressed genes.

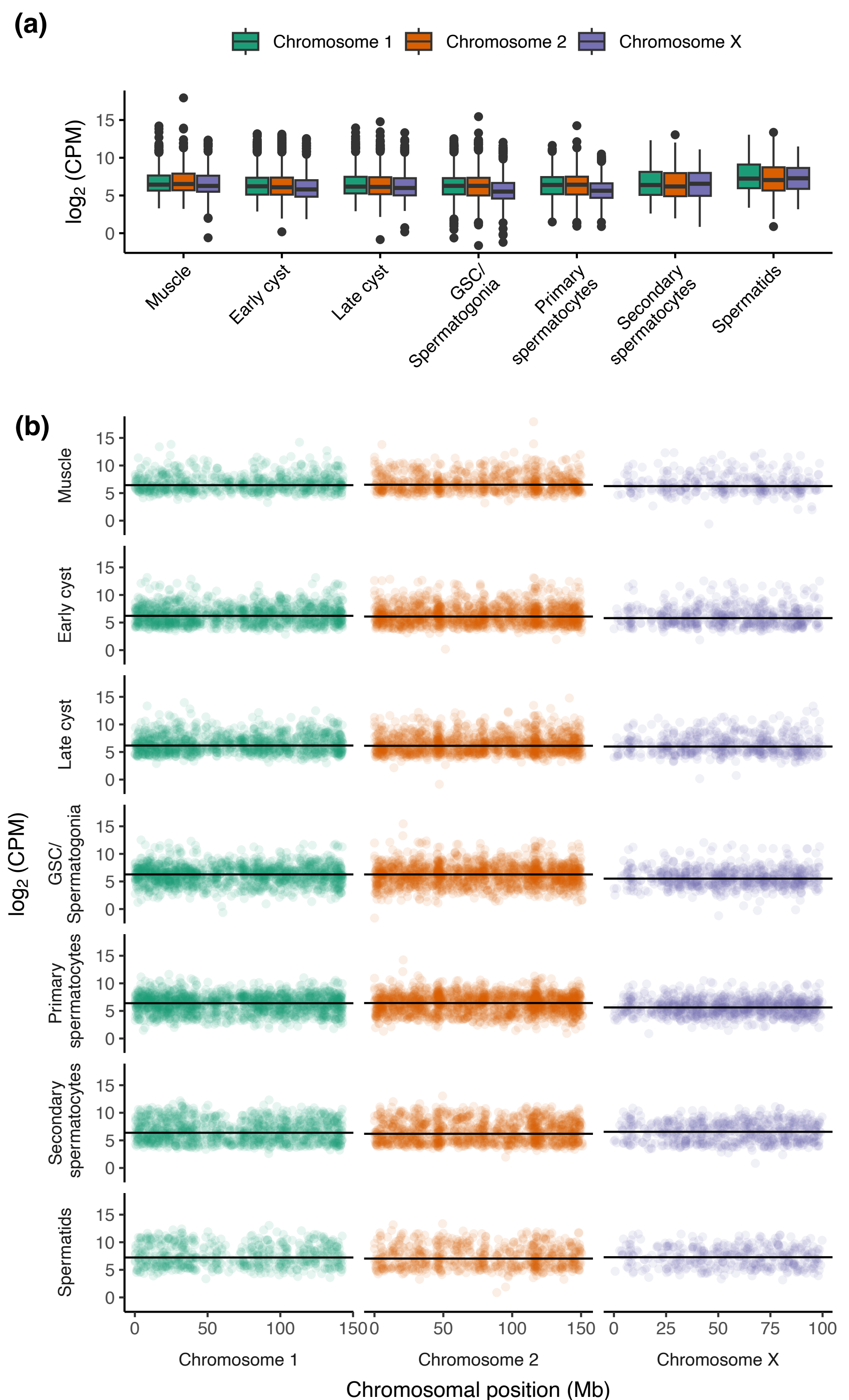


Figure S5.4. Expression of the X chromosome and autosomes across cell types.

(a) *Expression values of genes in ST males from chromosome 1, chromosome 2 and the X-chromosome across cell types.* ***(b)*** *Spatial distribution of gene expression across the genome in each cell type. Expression is measured as the average log_2(CPM) across ST males for each cell type, utilising a pseudobulk approach. Black line is the median expression across a chromosome for each cell type.*

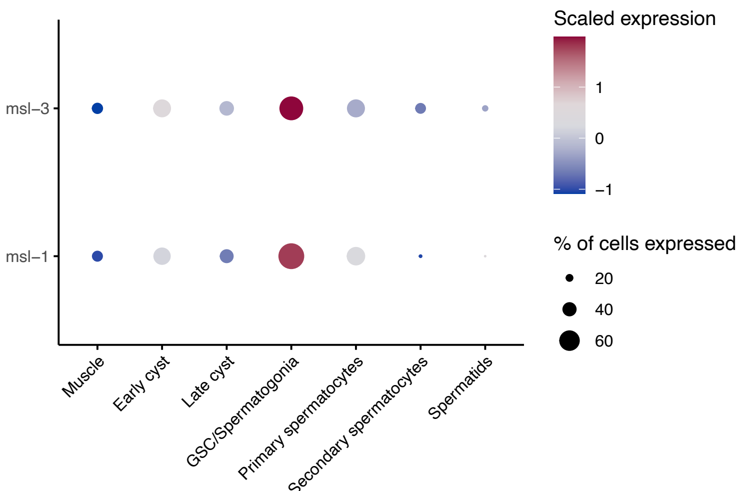


Figure S5.5. Expression of orthologs of the Drosophila dosage compensation complex.

Dot plot of scaled expression of T. dalmanni orthologs of male-specific-lethal 1 and 3 (msl-1/3), core components of the dosage compensation complex in Drosophila. Colour signifies scaled expression and dot size represents the percentage of cells expressing the specific gene.

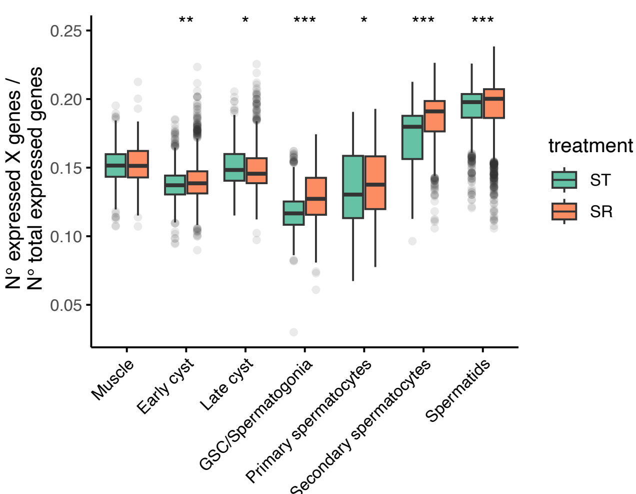


Figure S5.6. Comparing expression of the X chromosome across cell types in standard (ST) and drive (SR) males.

*Box plot showing the relative number of X-linked genes expressed across cell types in standard (ST) and drive (SR) males. Relative number calculated as the number of expressed X-linked genes divided by the total number of expressed genes. A two-sided Wilcox test was used to determine if the proportion of expressed X-linked genes for each cell type deviated from 0. $p < 0.00001$ = ***, $p < 0.001$ = **, $p < 0.05$ = *.*

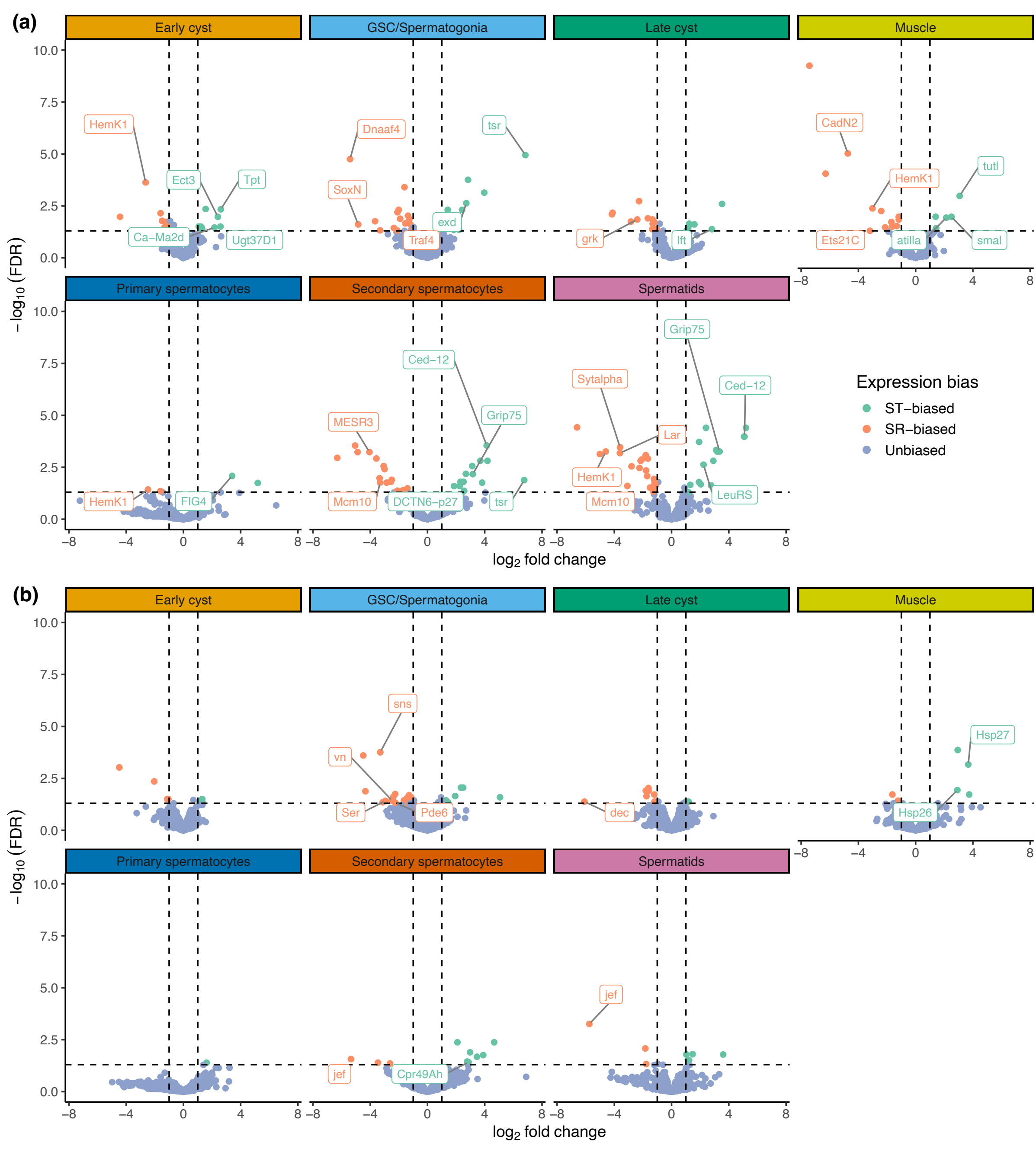


Figure S5.7: Differential gene expression across cell types.

*Volcano plots of differentially expressed genes in each cell type for **(a)** the X chromosome and **(b)** the autosomes with $\log_2$ fold-change on X axis and False Discovery Rate (FDR) adjusted p-value on the Y axis. Blue points are unbiased genes and green and orange dots are drive (SR-) and standard (ST-) biased genes respectively (FDR < 0.05 and absolute $\log_2$(fold-change) > 1). Labelled points are the top 8 significant genes per cell type (ordered by FDR) with Drosophila orthologs (Note that not all cell types have 8 genes matching this criteria).*

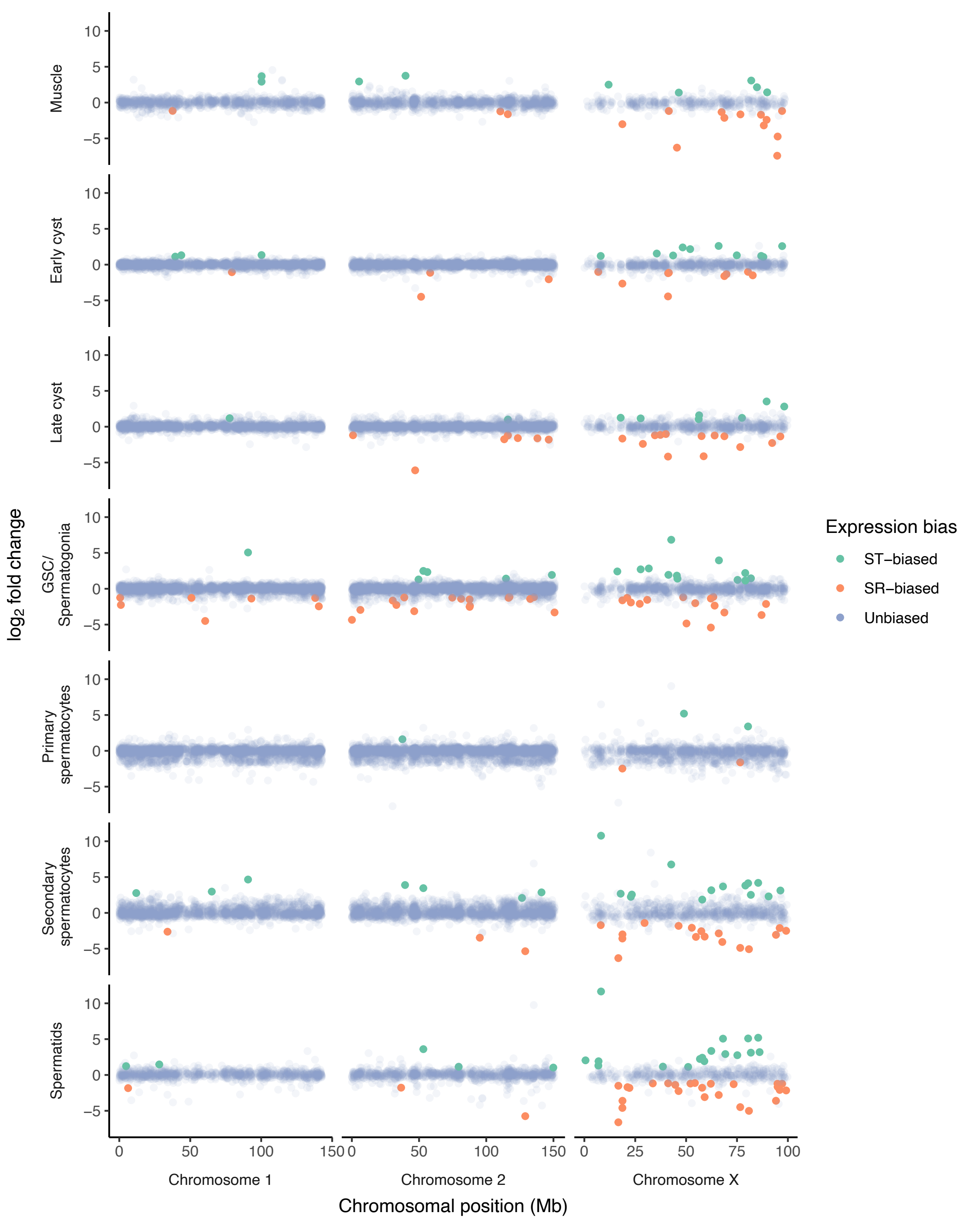


Figure S5.8. Spatial patterns of differential gene expression across cell types.

Dot plots of differentially expressed genes in each cell type across the genome with chromosomal position on X axis and log_2 fold-change on Y axis. Blue points are unbiased genes and green and orange dots are drive (SR-) and standard (ST-) biased genes respectively (FDR < 0.05 and absolute log_2(fold-change) > 1).

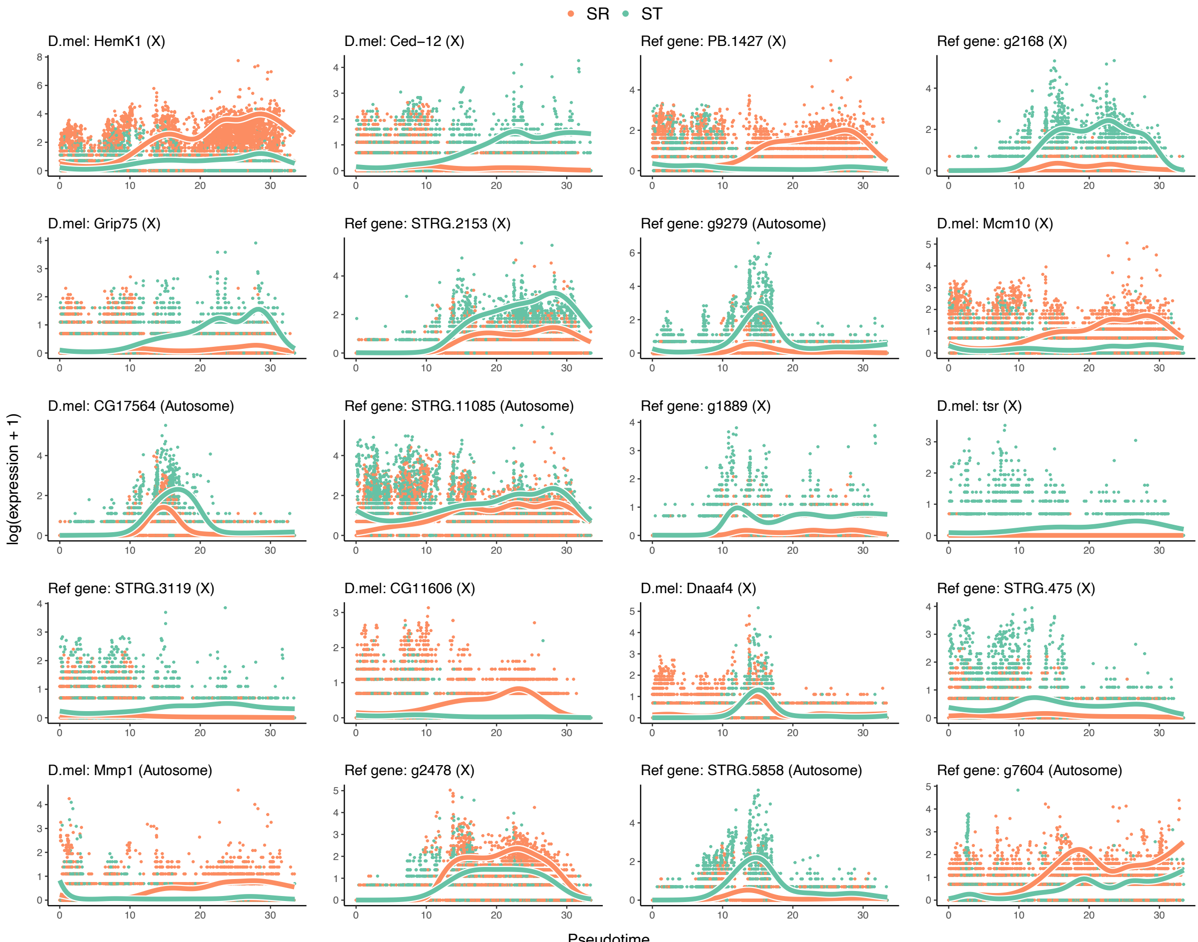


Figure S5.9. Differential trajectories of selected genes across spermatogenesis.

Differential trajectories for cells assigned to germline cell types (GSC/spermatogonia, primary and secondary spermatocytes, and spermatids). Plotted are genes identified as both differentially expressed and with differential trajectories between standard (ST) and drive (SR) individuals (p-value < 0.05 & $\log_2$ fold-change > 2). Genes are then ordered by descending Wald stat from the condition test for identifying differential trajectories, with the top 20 shown. D.mel refers to genes with Drosophila melanogaster orthologs, and Ref gene are those without.

Chapter 5 supplementary tables

Table S5.1a. Cell Numbers

Sample	Type	Number of cells	Median no of features	Median no of UMIs
sr1	SR	1927	1920.3	6088
sr2	SR	1223	2638.9	12123
sr3	SR	2060	2151.1	7431
sr5	SR	2788	2772	8021
st1	ST	1793	2485.8	6362
st2	ST	696	2942.7	9734.5
st3	ST	887	3441.1	14294
st5	ST	1172	2702.5	9149.5

Table S5.1b. Cell Numbers

Sample	Type	Cell type	Number of cells	Median no of features	Median no of UMIs
sr1	SR	Early cyst	288	1419.5	3268
sr1	SR	GSC/Spermatogonia	32	5447.5	32690.5
sr1	SR	Late cyst	409	1795	6504
sr1	SR	Muscle	24	1777	10575
sr1	SR	Primary spermatocytes	82	766.5	1205
sr1	SR	Secondary spermatocytes	194	988.5	2351.5
sr1	SR	Spermatids	898	1697.5	8574
sr2	SR	Early cyst	363	3063	12770
sr2	SR	GSC/Spermatogonia	47	3857	16803
sr2	SR	Late cyst	153	2432	9421
sr2	SR	Muscle	56	1606	9747
sr2	SR	Primary spermatocytes	49	2510	7092
sr2	SR	Secondary spermatocytes	142	2632	17237.5

sr2	SR	Spermatids	413	1939	11461
sr3	SR	Early cyst	435	2928	11399
sr3	SR	GSC/Spermatogonia	115	2737	7485
sr3	SR	Late cyst	233	2076	6949
sr3	SR	Muscle	60	1238.5	5144.5
sr3	SR	Primary spermatocytes	94	3036	9709
sr3	SR	Secondary spermatocytes	244	2090.5	8792.5
sr3	SR	Spermatids	879	1555	6197
sr5	SR	Early cyst	1285	2058	6234
sr5	SR	GSC/Spermatogonia	359	4138	18802
sr5	SR	Late cyst	416	2743	10833
sr5	SR	Muscle	33	1100	3820
sr5	SR	Primary spermatocytes	478	4006	14582
sr5	SR	Secondary spermatocytes	62	2636	7514
sr5	SR	Spermatids	155	1540	3945
st1	ST	Early cyst	334	2043.5	5637.5
st1	ST	GSC/Spermatogonia	293	2951	7792
st1	ST	Late cyst	194	1099.5	2338.5
st1	ST	Muscle	58	1370	3372
st1	ST	Primary spermatocytes	413	3454	9309
st1	ST	Secondary spermatocytes	156	2127	7894.5
st1	ST	Spermatids	345	1398	5184
st2	ST	Early cyst	129	4101	23515
st2	ST	GSC/Spermatogonia	81	3173	9399
st2	ST	Late cyst	48	3432.5	16073.5
st2	ST	Muscle	32	1717.5	7827
st2	ST	Primary spermatocytes	171	2909	9172
st2	ST	Secondary spermatocytes	133	2127	6485
st2	ST	Spermatids	102	1594	7146
st3	ST	Early cyst	351	4092	24308

st3	ST	GSC/Spermatogonia	140	3197	9158.5
st3	ST	Late cyst	87	2468	8911
st3	ST	Muscle	23	2032	9646
st3	ST	Primary spermatocytes	111	4228	15087
st3	ST	Secondary spermatocytes	37	3072	26706
st3	ST	Spermatids	138	2093.5	14161
st5	ST	Early cyst	440	2746.5	11340
st5	ST	GSC/Spermatogonia	128	2744	7467.5
st5	ST	Late cyst	92	1769.5	6131.5
st5	ST	Muscle	54	1568	6452
st5	ST	Primary spermatocytes	128	3447	9299.5
st5	ST	Secondary spermatocytes	40	3668	20361.5
st5	ST	Spermatids	290	1676	8035

Table S5.2. Marker Genes. *List of key marker genes from the literature*

***Drosophila* Marker (FlyBase)**	***Drosophila* Marker (Gene Name)**	***Drosophila* cell type**	**DOI**	**Reference**	***Teleopsis dalmanni* Ortholog**
FBgn0004872	piwi	C	10.1261/rna.744307	Nishida et al. (2007)	PB.1027, STRG.14307
FBgn0000576	ems	E	10.1159/000200079	Nanda et al. (2009)	gene_4133
FBgn0024288	Sox100B	E	10.1159/000200079	Nanda et al. (2009)	PB.2394
FBgn0001224	Hsp23	E	10.1242/jcs.110.17.1989	Michaud et al. (1997)	g9518, g9515
FBgn0004647	N	E	10.1016/j.stem.2013.05.003	Chen et al. (2013)	gene_9147
FBgn0014163	fax	E	10.1016/j.devcel.2005.08.012	Decotto & Spradling (2005)	gene_2995
FBgn0015399	kek1	E	10.1534/genetics.116.196535	Tang et al. (2017)	gene_9472
FBgn0250816	AGO3	G	10.1261/rna.055996.116	Quénerch'du et al. (2016)	gene_9924
FBgn0000404	CycA	G	10.1242/dev.01032	Perezgasga et al. (2004)	PB.8659
FBgn0250843	Prosalpha6	G	10.1186/s12864-018-5085-z	Vedelek et al. (2018)	gene_5828
FBgn0283442	vas	G, PS	10.1101/gad.4.6.905	Lasko & Ashburner (1990)	PB.762
FBgn0038488	m-cup	G, PS	10.1242/dev.021949	Barreau et al. (2008)	PB.2973
FBgn0000146	aub	G, PS	10.1261/rna.055996.116	Quénerch'du et al. (2016)	gene_8914, gene_8916
FBgn0000405	CycB	G,PS	10.1242/dev.01032	Perezgasga et al. (2004)	PB.4546, PB.4560, PB.4544
FBgn0264695	Mhc	M	10.1016/j.modgep.2006.11.007	Hess et al. (2006)	gene_9290
FBgn0000320	eya	LC	10.4161/spmg.21380	Zoller & Schulz (2012)	PB.370

FBgn0000964	tj	EC	10.1038/ncb1058	Li et al. (2003)	g1204
FBgn0004108	Nrt	EC	10.1016/j.ydbio.2006.02.048	Terry et al. (2006)	gene_1650
FBgn0011591	fng	EC	10.1016/j.ydbio.2006.02.048	Terry et al. (2006)	PB.9184
FBgn0001257	ImpL2	EC	10.1016/j.ydbio.2006.02.048	Terry et al. (2006)	g9584
FBgn0011206	bol	PS	10.1038/11091	Maines & Wasserman (1999)	gene_3925
FBgn0002673	twe	PS, SS	10.1016/0092-8674(92)90616-k	Alphey et al. (1992)	PB.1956
FBgn0034435	fest	PS, SS, ST	10.1242/dev.122341	Baker et al. (2015)	STRG.7827
FBgn0039071	bb8	PS, ST	10.1371/journal.pone.0161289	Vedelek et al. (2016)	gene_9679, g10101
FBgn0030840	p-cup	ST	10.1242/dev.021949	Barreau et al. (2008)	STRG.2391
FBgn0031142	r-cup	ST	10.1242/dev.021949	Barreau et al. (2008)	STRG.2391
FBgn0037502	wa-cup	ST	10.1242/dev.021949	Barreau et al. (2008)	STRG.2391
FBgn0028487	f-cup	ST	10.1242/dev.021949	Barreau et al. (2008)	gene_4522
FBgn0004606	zfh1	CYSC	10.1016/j.stem.2008.05.001	Leatherman & DiNardo (2009)	gene-3960

Cell type key. E: Epithelial, PS: Primary spermatocyte, SS: Secondary spermatocyte ,ST: Spermatid G: GSC/Spermatogonia, EC: Early Cyst, LC: Late Cyst, M: Muscle, CYSC: Cyst Stem Ce

Table S5.3. Cell cycle markers and orthologs

Phase	Drosophila Marker (FlyBase)	Drosophila Marker (Gene Name)	*Teleopsis dalmanni* ortholog
G2/M	FBgn0000063	Mps1	PB.3768
G2/M	FBgn0000147	aurA	gene-4755
G2/M	FBgn0001086	fzy	gene-9691
G2/M	FBgn0002673	twe	PB.1956
G2/M	FBgn0003041	pbl	gene-3709
G2/M	FBgn0003525	stg	PB.1963
G2/M	FBgn0003545	sub	gene-6143
G2/M	FBgn0003607	Su(var)205	gene-9288
G2/M	FBgn0003607	Su(var)205	gene-833
G2/M	FBgn0004106	Cdk1	gene-9054
G2/M	FBgn0004106	Cdk1	gene-9940
G2/M	FBgn0004378	Klp61F	gene-3313
G2/M	FBgn0005683	pie	gene-5748
G2/M	FBgn0005683	pie	gene-3792
G2/M	FBgn0011692	pav	gene-3328
G2/M	FBgn0015391	glu	gene-9709
G2/M	FBgn0015391	glu	gene-4429
G2/M	FBgn0027948	msps	gene-4852
G2/M	FBgn0029881	pigs	PB.8250
G2/M	FBgn0029970	dlg1	gene-133
G2/M	FBgn0031091	Phf7	PB.7762
G2/M	FBgn0031696	Bub1	PB.1666
G2/M	FBgn0033845	mars	PB.6246
G2/M	FBgn0034282	Mapmodulin	gene-6041
G2/M	FBgn0034657	Ogg1	gene-7540
G2/M	FBgn0034657	LBR	gene-8586
G2/M	FBgn0035769	CTCF	STRG.14736

G2/M	FBgn0037613	Cks85A	gene-7323
G2/M	FBgn0039019	HP1c	gene-6795
G2/M	FBgn0039680	Cap-D2	gene-8633
G2/M	FBgn0039680	Cap-D2	gene-9469
G2/M	FBgn0040233	Sou	gene-652
G2/M	FBgn0261385	scra	PB.5826
G2/M	FBgn0263855	BubR1	gene-2857
G2/M	FBgn0264291	Det	gene-4537
G2/M	FBgn0264848	vih	gene-3698
G2/M	FBgn0284220	Top2	PB.83
S	FBgn0002906	Blm	PB.4034
S	FBgn0003479	spn-A	PB.2003
S	FBgn0005655	PCNA	gene-9759
S	FBgn0010382	CycE	PB.706
S	FBgn0015929	dpa	gene-7982
S	FBgn0017577	Mcm5	gene-4659
S	FBgn0024920	Ts	STRG.11772
S	FBgn0025815	Mcm6	gene-876
S	FBgn0025832	Fen1	gene-6294
S	FBgn0028476	Usp1	PB.4180
S	FBgn0032635	CG15141	PB.1054
S	FBgn0032698	CG10336	gene-10419
S	FBgn0033526	Caf1-105	gene-7810
S	FBgn0034495	CG11788	PB.6275
S	FBgn0035918	Cdc6	STRG.16179
S	FBgn0041186	Slbp	STRG.4192
S	FBgn0052251	CG3225	gene-916
S	FBgn0259113	PolA1	gene-4864
S	FBgn0260985	RfC4	gene-3318
S	FBgn0261976	Psf2	gene-10243

Table S5.4. Online material: **Novel cell type markers**

Table S5.5. Inbreeding values

Sample	No. heterozygous sites	Total no. sites	Inbreeding coefficient (F)	Observed heterozygosity
12-A12	3394560	8381826	-0.10802	0.4049905
13-A13	2444355	8315710	0.1958	0.2939442
14-A14	3229598	8366825	-0.0536	0.3860004
15-A15	2707557	8391274	0.11847	0.3226634
16-A16	3686362	8348588	-0.20725	0.4415551
17-A17	3150261	8339935	-0.03205	0.3777321
18-A18	3160057	8328859	-0.03656	0.3794106
19-A19	3336962	8308420	-0.09742	0.4016362
20-A20	3387492	8329117	-0.11479	0.4067048

Table S5.6. Differential expression between ST & SR

Cell type	Chromosome	SR-biased	ST-biased	Unbiased
Pre-meiotic cyst	Autosome	4	3	4865
	X	9	10	849
GSC/Spermatogonia	Autosome	23	6	5568
	X	15	12	1074
Post-meiotic cyst	Autosome	8	2	4466
	X	13	7	773
Muscle	Autosome	3	4	2403
	X	12	5	425
Primary Spermatocytes	Autosome	0	1	5863
	X	2	2	1326
Secondary Spermatocytes	Autosome	3	7	3779
	X	17	14	815

Spermatids	Autosome	3	5	2018
	X	25	17	482

Table S5.7. DGE enrichment model results

	Estimate	Std. Error	z value	Pr(>lzl)
(Intercept)	-6.0584756	0.28487496	-21.26714	2.29E-100
celltype:Pre-meiotic cyst	-0.6179285	0.2859022	-2.1613281	0.03067
celltype:Post-meiotic cyst	-0.3821644	0.27687808	-1.3802624	0.16750588
celltype:GSC/Spermatogonia	-0.0449223	0.24709334	-0.181803	0.85573732
celltype:Primary Spermatocytes	-2.6295657	0.49316657	-5.3320032	9.71E-08
celltype:Secondary Spermatocytes	-0.0532835	0.2603911	-0.2046287	0.83786225
celltype:Spermatids	0.69418315	0.25285262	2.7454062	0.0060436
chrChr_2	0.59152208	0.25315085	2.33663873	0.01945798
chrChr_X	2.77314708	0.22391712	12.384703	3.16E-35

Table S5.8. Online material: Differentially expressed genes

Table S5.9. Online material: Genes with differential trajectories

Table S5.10. GO terms

Ontology	ID	Description	GeneRatio	BgRatio	RichFactor	FoldEnrichment	zScore	pvalue	p.adjust	qvalue	Drosophila orthologs	Count
BP	GO:0003341	cilium movement	12/78	36/3107	0.33	13.28	11.89	0.00	0.00	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84D/Tektin-C/CG17450/Rsph4a/kl-3/tous/CG9313/Ccdc39/Dnaaf3	12.00
BP	GO:0044782	cilium organization	13/78	58/3107	0.22	8.93	9.78	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/Rsph4a/kl-3/CG7886/CG9313/Ccdc39/CG17687/Dnaaf3	13.00
BP	GO:0060271	cilium assembly	12/78	52/3107	0.23	9.19	9.56	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/Rsph4a/kl-3/CG9313/Ccdc39/CG17687/Dnaaf3	12.00
BP	GO:0001539	cilium or flagellum-dependent cell motility	9/78	27/3107	0.33	13.28	10.28	0.00	0.00	0.00	hmw/alphaTub84D/Tektin-C/CG17450/Rsph4a/kl-3/tous/Ccdc39/Dnaaf3	9.00
BP	GO:0060285	cilium-dependent cell motility	9/78	27/3107	0.33	13.28	10.28	0.00	0.00	0.00	hmw/alphaTub84D/Tektin-C/CG17450/Rsph4a/kl-3/tous/Ccdc39/Dnaaf3	9.00
BP	GO:0120031	plasma membrane bounded cell projection assembly	13/78	72/3107	0.18	7.19	8.53	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/Rsph4a/kl-3/tsr/CG9313/Ccdc39/CG17687/Dnaaf3	13.00

BP	GO:0030031	cell projection assembly	13/78	75/3107	0.17	6.90	8.31	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/Rsph4a/kl-3/tsr/CG9313/Ccdc39/CG17687/Dnaaf3	13.00
BP	GO:0060294	cilium movement involved in cell motility	8/78	23/3107	0.35	13.86	9.93	0.00	0.00	0.00	hmw/alphaTub84D/Tektin-C/CG17450/Rsph4a/kl-3/tous/Dnaaf3	8.00
BP	GO:0035082	axoneme assembly	9/78	32/3107	0.28	11.20	9.31	0.00	0.00	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84D/Rsph4a/kl-3/CG9313/Ccdc39/Dnaaf3	9.00
BP	GO:0007018	microtubule-based movement	12/78	71/3107	0.17	6.73	7.84	0.00	0.00	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84D/Tektin-C/CG17450/Rsph4a/kl-3/tous/CG9313/Ccdc39/Dnaaf3	12.00
BP	GO:0001578	microtubule bundle formation	9/78	37/3107	0.24	9.69	8.53	0.00	0.00	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84D/Rsph4a/kl-3/CG9313/Ccdc39/Dnaaf3	9.00
BP	GO:0120036	plasma membrane bounded cell projection organization	19/78	219/3107	0.09	3.46	6.05	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/Nost/Rsph4a/Cul3/kl-3/nompA/tsr/beat-Ib/Mmp1/CG7886/CG9313/Ccdc39/CG17687/Dnaaf3	19.00
BP	GO:0070925	organelle assembly	17/78	177/3107	0.10	3.83	6.21	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/alphaTub84D/Tektin-C/CG32371/Rsph4a/kl-3/	17.00

											tn/bru1/mamo/Grip75/CG9313/ Ccdc39/CG17687/Dnaaf3	
BP	GO:0030030	cell projection organization	19/78	223/3107	0.09	3.39	5.95	0.00	0.00	0.00	Dnaaf4/hmw/CG13693/TTLL3B/ alphaTub84D/Tektin-C/ Nost/Rsph4a/Cul3/kl-3/nompA/ tsr/beat-Ib/Mmp1/CG7886/ CG9313/Ccdc39/CG17687/Dnaaf3	19.00
BP	GO:0007600	sensory perception	8/78	38/3107	0.21	8.39	7.35	0.00	0.00	0.00	Dnaaf4/hmw/CG8086/tous/nompA /Gr28b/ppk11/CG9313	8.00
BP	GO:0030317	flagellated sperm motility	6/78	18/3107	0.33	13.28	8.38	0.00	0.00	0.00	hmw/alphaTub84D/CG17450/kl-3/ tous/Dnaaf3	6.00
BP	GO:0097722	sperm motility	6/78	18/3107	0.33	13.28	8.38	0.00	0.00	0.00	hmw/alphaTub84D/CG17450/kl-3/ tous/Dnaaf3	6.00
BP	GO:0007017	microtubule-based process	18/78	214/3107	0.08	3.35	5.72	0.00	0.00	0.00	Dnaaf4/hmw/betaTub60D/TTLL3B /alphaTub84D/Tektin-C/ CG32371/CG17450/Rcd7/ betaTub85D/Rsph4a/kl-3/ tous/tsr/Grip75/CG9313/Ccdc39/ Dnaaf3	18.00
BP	GO:0007605	sensory perception of sound	6/78	20/3107	0.30	11.95	7.88	0.00	0.00	0.00	Dnaaf4/hmw/CG8086/tous/nompA /CG9313	6.00
BP	GO:0007010	cytoskeleton organization	20/78	280/3107	0.07	2.85	5.19	0.00	0.01	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84 D/Rab3-GEF/CG32371/CG17450/ Rcd7/betaTub85D/Rsph4a/kl-3/	20.00

											tn/tsr/Ced-12/bru1/CG33521/ Grip75/CG9313/Ccdc39/Dnaaf3	
BP	GO:0050954	sensory perception of mechanical stimulus	6/78	24/3107	0.25	9.96	7.07	0.00	0.01	0.00	Dnaaf4/hmw/CG8086/tous/nompA /CG9313	6.00
BP	GO:0070286	axonemal dynein complex assembly	5/78	16/3107	0.31	12.45	7.37	0.00	0.02	0.00	Dnaaf4/kl-3/CG9313/ Ccdc39/Dnaaf3	5.00
BP	GO:0000226	microtubule cytoskeleton organization	15/78	184/3107	0.08	3.25	5.04	0.00	0.03	0.00	Dnaaf4/hmw/TTLL3B/alphaTub84D/CG32371/CG17450/Rcd7/ betaTub85D/Rsph4a/kl-3/ tsr/Grip75/CG9313/Ccdc39/ Dnaaf3	15.00
CC	GO:0005930	axoneme	7/76	20/3118	0.35	14.36	9.47	0.00	0.00	0.00	CG13693/TTLL3B/Rsph4a/ kl-3/CG9313/Ccdc39/CG17687	7.00
CC	GO:0097014	ciliary plasm	7/76	20/3118	0.35	14.36	9.47	0.00	0.00	0.00	CG13693/TTLL3B/Rsph4a/ kl-3/CG9313/Ccdc39/CG17687	7.00
CC	GO:0032838	plasma membrane bounded cell projection cytoplasm	7/76	29/3118	0.24	9.90	7.61	0.00	0.00	0.00	CG13693/TTLL3B/Rsph4a/ kl-3/CG9313/Ccdc39/CG17687	7.00

CC	GO:0005856	cytoskeleton	16/76	203/3118	0.08	3.23	5.20	0.00	0.00	0.00	betaTub60D/CG13693/TTLL3B/alphaTub84D/Tektin-C/CG32371/betaTub85D/Rsph4a/kl-3/CG8086/tsr/Grip75/CG7886/CG9313/Ccdc39/CG17687	16.00
CC	GO:0005929	cilium	8/76	50/3118	0.16	6.56	6.27	0.00	0.00	0.00	CG13693/TTLL3B/Rsph4a/kl-3/CG7886/CG9313/Ccdc39/CG17687	8.00
MF	GO:0005200	structural constituent of cytoskeleton	5/71	18/2852	0.28	11.16	6.91	0.00	0.01	0.01	betaTub60D/alphaTub84D/CG17450/betaTub85D/Grip75	5.00

Ontology: Which GO domains the GO terms fall under (BP: Biological Process MF: Molecular Function CC: Cellular Component).

GeneRatio: Number of focal genes in the input associated with the GO term/number of focal genes

BgRatio: Number of genes in the background associated with the GO term/total number of background genes

RichFactor: Number of focal genes in a GO term/Number of background genes in that GO term

FoldEnrichment: Observed frequency of GO term in focal genes/Expected frequency of GO term in background genes

p.adjust: Bonferroni adjusted p.value (cut off < 0.05)

qvalue: estimate of false positives among enriched terms (cut off < 0.05)

Chapter 5 supplementary references

Baker, C. C., Gim, B. S., & Fuller, M. T. (2015). Cell type-specific translational repression of Cyclin B during meiosis in males. *Development* , *142*(19), 3394–3402.

Barreau, C., Benson, E., & White-Cooper, H. (2008). Comet and cup genes in *Drosophila* spermatogenesis: the first demonstration of post-meiotic transcription. *Biochemical Society Transactions*, *36*(Pt 3), 540–542.

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience*, *10*(2), giab007

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L., & Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature*, *337*(6205), 373–376.

Courtot, C., Fankhauser, C., Simanis, V., & Lehner, C. F. (1992). The *Drosophila cdc25* homolog twine is required for meiosis. *Development, 116*(2), 405–416.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008

Erickson, R. P. (1990). Post-meiotic gene expression. *Trends in Genetics*, *6*, 264–268.

Greenbaum, M. P., Iwamori, T., Buchold, G. M., & Matzuk, M. M. (2011). Germ Cell Intercellular Bridges. *Cold Spring Harbor Perspectives in Biology*, *3*(8), a005850

Hess, N. K., Singer, P. A., Trinh, K., Nikkhoy, M., & Bernstein, S. I. (2007). Transcriptional regulation of the *Drosophila melanogaster* muscle myosin heavy-chain gene. *Gene Expression Patterns*, *7*(4), 413–422.

Kaufman, R. S., Price, K. L., Mannix, K. M., Ayers, K. M., Hudson, A. M., & Cooley, L. (2020). Drosophila sperm development and intercellular cytoplasm sharing through ring canals do not require an intact fusome. *Development, 147*(22), dev190140.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754–1760.

Li, H., Janssens, J., De Waegeneer, M., Kolluru, S. S., Davie, K., Gardeux, V., Saelens, W., David, F. P. A., Brbić, M., Spanier, K., Leskovec, J., McLaughlin, C. N., Xie, Q., Jones,

R. C., Brueckner, K., Shim, J., Tattikota, S. G., Schnorrer, F., Rust, K., … Zinzen, R. P. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science*, *375*(6584), eabk2432.

Mahadevaraju, S., Fear, J. M., Akeju, M., Galletta, B. J., Pinheiro, M. M. L. S., Avelino, C. C., Cabral-de-Mello, D. C., Conlon, K., Dell'Orso, S., Demere, Z., Mansuria, K., Mendonça, C. A., Palacios-Gimenez, O. M., Ross, E., Savery, M., Yu, K., Smith, H. E., Sartorelli, V., Yang, H., … Oliver, B. (2021). Dynamic sex chromosome expression in *Drosophila* male germ cells. *Nature Communications*, *12*(1), 892.

Ohlstein, B., & McKearin, D. (1997). Ectopic expression of the *Drosophila* Bam protein eliminates oogenic germline stem cells. *Development, 124*(18), 3651–3662.

Raz, A. A., Vida, G. S., Stern, S. R., Mahadevaraju, S., Fingerhut, J. M., Viveiros, J. M., Pal, S., Grey, J. R., Grace, M. R., Berry, C. W., Li, H., Janssens, J., Saelens, W., Shao, Z., Hu, C., Yamashita, Y. M., Przytycka, T., Oliver, B., Brill, J. A., … Fuller, M. T. (2023). Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of Drosophila testes. *eLife*, *12*, e82201

Robben, M., Ramesh, B., Pau, S., Meletis, D., Luber, J., & Demuth, J. (2024). scRNA-seq Reveals Novel Genetic Pathways and Sex Chromosome Regulation in *Tribolium* Spermatogenesis. *Genome Biology and Evolution*, *16*(3), evae059

Terry, N. A., Tulina, N., Matunis, E., & DiNardo, S. (2006). Novel regulators revealed by profiling *Drosophila* testis stem cells within their niche. *Developmental Biology*, *294*(1), 246–257.

van der Auwera, G., & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated.

Vandereyken, K., Sifrim, A., Thienpont, B., & Voet, T. (2023). Methods and applications for single-cell and spatial multi-omics. *Nature Reviews. Genetics*, *24*(8), 494–515.

van Rensburg, A. J., Hipperson, H., Bradshaw, S. L., Bates, S., Burke, T., & Pomiankowski, A. (2024). Genomics of an X-linked meiotic driver in stalk-eyed flies. *In Prep*.

Vedelek, V., Laurinyecz, B., Kovács, A. L., Juhász, G., & Sinka, R. (2016). Testis-Specific Bb8 Is Essential in the Development of Spermatid Mitochondria. *PloS One*, *11*(8), e0161289.

Ventelä, S., Toppari, J., & Parvinen, M. (2003). Intercellular Organelle Traffic through Cytoplasmic Bridges in Early Spermatids of the Rat: Mechanisms of Haploid Gene Product Sharing. *Molecular Biology of the Cell*, *14*(7), 2768–2780.

Witt, E., Benjamin, S., Svetec, N., & Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila*.

eLife, *8*, e47138

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S. Y., Keefe, D. L., Alukal, J. P., Boeke, J. D., & Yanai, I. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell*, *180*(2), 248–262.e21.

Zoller, R., & Schulz, C. (2012). The *Drosophila* cyst stem cell lineage: Partners behind the scenes? *Spermatogenesis*, *2*(3), 145–157.

Appendix D. The Avian W Chromosome: Simply a Y with a Different Name?

Peter D Price[1], Thea F Rogers[1], Alison E Wright[1]

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, United Kingdom

This appendix was an invited piece for the 'Fourth Report on Chicken Genes and Chromosomes 2022'

Smith J, Alfieri JM, Anthony N, Arensburger P, Athrey GN, Balacco J, Balic A, Bardou P, Barela P, Bigot Y, Blackmon H, Borodin PM, Carroll R, Casono MC, Charles M, Cheng H, Chiodi M, Cigan L, Coghill LM, Crooijmans R, Das N, Davey S, Davidian A, Degalez F, Dekkers JM, Derks M, Diack AB, Djikeng A, Drechsler Y, Dyomin A, Fedrigo O, Fiddaman SR, Formenti G, Frantz LAF, Fulton JE, Gaginskaya E, Galkina S, Gallardo RA, Geibel J, Gheyas A, Godinez CJP, Goodell A, Graves JAM, Griffin DK, Haase B, Han J-L, Hanotte O, Henderson LJ, Hou Z-C, Howe K, Huynh L, Ilatsia E, Jarvis E, Johnson SM, Kaufman J, Kelly T, Kemp S, Kern C, Keroack JH, Klopp C, Lagarrigue S, Lamont SJ, Lange M, Lanke A, Larkin DM, Larson G, Layos JKN, Lebrasseur O, Malinovskaya LP, Martin RJ, Martin Cerezo ML, Mason AS, McCarthy FM, McGrew MJ, Mountcastle J, Muhonja CK, Muir W, Muret K, Murphy T, Ng'ang'a I, Nishibori M, O'Connor RE, Ogugo M, Okimoto R, Ouko O, Patel HR, Perini F, Pigozzi MI, Potter KC, Price PD, Reimer C, Rice ES, Rocos N, Rogers TF, Saelao P, Schauer J, Schnabel R, Schneider V, Simianer H, Smith A, Stevens MP, Stiers K, Tiambo CK, Tixier-Boichard M, Torgasheva AA, Tracey A, Tregaskes CA, Vervelde L, Wang Y, Warren WC, Waters PD, Webb D, Weigend S, Wolc A, Wright AE, Wright D, Wu Z, Yamagata M, Yang C, Yin Z-T, Young M, Zhang G, Zhao B & Zhou H (2023) Fourth Report on Chicken Genes and Chromosomes 2022. Cytogenetic and Genome Research 162:405–527

All microchromosomes are represented in the most recent assembly. Therefore, whole-genome alignments to it can be used to assess assembly errors and/or real rearrangements in new bird assemblies for which karyotype information is available. This presents a more robust comparison than another popular bird model, the zebra finch, which has undergone macrochromosome fission and internal rearrangement (Fig. 9). It is also more useful for aligning the genomes of birds with very rearranged chromosomes, since the multiple rearrangements are independent in different clades (e.g., golden eagle and falcon/parrot, Fig. 9).

Conflict of Interest Statement
There are no conflicts of interest to declare.

Funding Sources
P.D.W. is supported by Australian Research Council Discovery Projects (DP180100931, DP210103512, and DP220101429). J.M.G. is supported by Australian Research Council Discovery Projects (DP210103512 and DP220101429). H.R.P. is supported by the Australian National University and NHMRC Synergy Grant (APP2011277). The authors acknowledge the provision of computing resources provided by the Australian BioCommons Leadership Share (ABLeS) program. This program is co-funded by Bioplatforms Australia (enabled by NCRIS), the National Computational Infrastructure and Pawsey Supercomputing Centre.

The Avian W Chromosome: Simply a Y with a Different Name?

(Prepared by P.D. Price, T.F. Rogers, and A.E. Wright)

Sex chromosomes have long fascinated biologists due to their unique gene content and evolutionary trajectories relative to the rest of the genome [Furman et al., 2020]. In particular, the halting of recombination between sex chromosome pairs has resulted in the evolution of highly degenerate sex-limited W and Y chromosomes in many species [Charlesworth, 1991]. Identifying the function of these chromosomes and understanding if and how they can resist the degenerative forces arising from reduced recombination has been the focus of numerous studies [Bachtrog et al., 2011].

We now know a considerable amount about Y chromosomes, despite the difficulties in sequencing highly heterochromatic and repetitive genomic regions [Tomaszkiewicz et al., 2017]. Their evolution is typically characterised by the accumulation of genes with male-specific functions, large-scale gene amplification, and rapid turnover of gene content across lineages [Bachtrog, 2013; Subrini and Turner, 2021]. In contrast, our understanding of the W chromosome has lagged. However, the last decade has seen an explosion in the number of W-linked genes sequenced across birds [Zhou Q et al., 2014; Bravo et al., 2021], ranging from songbirds [Smeds et al., 2015; Xu et al., 2019; Sigeman et al., 2021; Huang et al., 2022; Warmuth et al., 2022] to fowl [Moghadam et al., 2012; Ayers et al., 2013; Wright et al., 2014] to paleognaths [Xu and Zhou, 2020; Liu J et al., 2021], and a reference assembly of the chicken W chromosome [Warren et al., 2017]. In theory, W chromosomes are in many ways comparable to Y chromosomes, as both are sex-limited and often don't recombine, and so they might be expected to share similar evolutionary fates. However, there are key differences, most notably that the W chromosome is limited to females whereas the Y chromosome is only present in males [Bachtrog et al., 2011; Mank, 2012]. Below, we outline new insights into avian W chromosomal evolution and ask whether W and Y chromosomes are really that different.

What Are the Evolutionary Dynamics of W Chromosomes across Birds?

It has been known for decades that the chicken W chromosome is a degenerated version of the Z, with the most recent build of the W reference (GRCg7b) identifying only ~80 protein-coding genes across ~9 Mb [Warren et al., 2017]. However, establishing whether the chicken W chromosome is representative of the avian W more generally has only recently been possible due to the plethora of W-linked sequences now available across the avian phylogeny [Wang Z et al., 2014; Zhou Q et al., 2014; Xu et al., 2019; Sigeman et al., 2020].

Sex chromosomes diverge as recombination is suppressed between them, typically assumed to occur in a stepwise process through sequential inversions [Charlesworth et al., 2005]. Consistent with this, 'strata' of different ages can be detected on avian Z and W chromosomes [Wang Z et al., 2014; Wright et al., 2014]. These strata are thought to reflect the halting of recombination through large-scale chromosomal rearrangements, such as inversions [Wright et al., 2016]. However, recent evidence suggests that recombination suppression at the earliest stages of avian sex chromosome divergence is a more mosaic and gradual process [Sigeman et al., 2021]. This cessation of recombination has occurred independently in different avian lineages [Zhou Q et al., 2014] and many species exhibit a heavily degraded W chromosome, similar to the chicken.

Despite degeneration proceeding independently across birds, the set of ancestral genes retained on the W chromosome is remarkably conserved, suggesting that

Downloaded from http://karger.com/cgr/article-pdf/162/8-9/405/3975130/000529376.pdf by guest on 03 October 2024

decay is non-random [Xu and Zhou, 2020]. For instance, over 80% of W-linked genes in the oldest stratum are conserved across chicken, songbirds and tinamous [Xu and Zhou, 2020]. This is in stark contrast to the Y chromosome, where frequent gene movement onto and off the chromosome is common [Hughes et al., 2015; Mahajan and Bachtrog, 2017]. This has been interpreted as a product of differing selective pressures acting on Y versus W chromosomes, with the W chromosome subject to stronger purifying selection compared to the Y due to its higher effective population size relative to the autosomes [Wright and Mank, 2013].

However, there is still remarkable variation in the extent of Z-W divergence across birds [Zhou Q et al., 2014]. In contrast to the chicken, the paleognath W chromosome recombines along a large proportion of its length and so has experienced limited decay, although this varies across species [Zhou Q et al., 2014; Yazdi and Ellegren, 2018; Liu J et al., 2021] with greater recombination suppression in ostrich and emu than tinamou. The growing amount of long-read sequencing data for birds has also revealed that fusions between sex chromosomes and autosomes to create neo-sex chromosomes are not uncommon, with two independent origins across Psittaciformes [Huang et al., 2022], four across Sylvioidea [Pala et al., 2012; Sigeman et al., 2020, 2021], one in the eastern yellow robin (*Eopsaltria australis*) [Gan et al., 2019], cuckoo (*Crotophaga ani*) [Kretschmer et al., 2020] and Raso lark (*Alauda razae*) [Dierickx et al., 2020] identified to date. No doubt this number will increase as more bird genomes are probed for sex chromosomes, making it possible to test the evolutionary pressures responsible for driving these fusions. Together, this challenges the traditional view that the avian W chromosome is genetically inert and highly conserved across species.

Is the Avian W Chromosome Selected for Female-Specific Functions?

Given the sex-limited inheritance pattern of Y and W chromosomes, theory predicts that they should be subject to sex-specific selection and accumulate genes with sex-specific functions [Rice, 1984]. Indeed, the Y chromosome in many species is enriched with genes predominantly expressed in testes that function in spermatogenesis [Bachtrog, 2013; Subrini and Turner, 2021], although there is a growing awareness of its role in non-reproductive traits [Cīrulis et al., 2022]. It follows that we might expect the W to be subject to female-specific selection to retain genes with female fitness benefits.

The avian W chromosome lacks a candidate sex-determining gene [Schmid et al., 2015]. Instead, sex in birds is determined by dosage of the Z-linked gene, *DMRT1* [Hirst et al., 2017]. However, there are several lines of evidence implicating the avian W chromosome in female fertility, although the precise functions of genes on the W have yet to be defined. First, W-linked genes are highly expressed in developing chicken ovaries [Moghadam et al., 2012; Ayers et al., 2013; Xu and Zhou, 2020]. This is consistent with feminization of the W [Mank et al., 2010], as key stages of oogenesis are restricted to embryogenesis unlike spermatogenesis, which is a continuous process throughout adult life. Second, W-linked genes expressed during late female development are convergently upregulated in the ovaries of chicken layer breeds subject to artificial selection for fecundity relative to their modern ancestor, the Red Jungle Fowl, and other chicken breeds [Moghadam et al., 2012].

However, unlike most Y-linked genes, which typically exhibit testes-specific expression [Subrini and Turner, 2021], expression of genes on the avian W chromosome is not limited to the ovary. Instead, studies from chicken and collared flycatcher show that many W genes are active in both somatic and reproductive tissue [Smeds et al., 2015; Bellott et al., 2017; Xu and Zhou, 2020]. While this does not preclude a specific role of the W chromosome in oogenesis, it has led to suggestions that this chromosome has instead been selected to maintain gene dosage and ancestral expression levels of essential genes. Consistent with this, many avian W-linked genes are subject to purifying selection [Wright et al., 2014; Sigeman et al., 2021], exhibit a high degree of sequence conservation as well as similar expression patterns to their Z-linked partner [Ayers et al., 2013; Smeds et al., 2015; Xu and Zhou, 2020], and have human orthologs that exhibit detrimental effects when haploid [Bellott et al., 2017; Xu et al., 2019; Bellott and Page, 2021; Sigeman et al., 2021].

It is plausible that apparent differences in the function of Y and W chromosomes could arise from their contrasting inheritance patterns. For instance, W-linked genes, which only pass through the female germ line, are not exposed to sperm competition and so might be subject to weaker sex-specific selection than genes on the Y chromosome. However, it is worth noting that our understanding of the function of the avian W is based on expression data from a limited number of species (chicken and flycatcher) taken across whole, heterogeneous, adult tissue. This precludes accurate contrasts of expression between Z and W orthologs [Price et al., 2022] and so could lead to false inferences of selection to maintain gene dos-

Downloaded from http://karger.com/cgr/article-pdf/162/8-9/405/3975130/000529376.pdf by guest on 03 October 2024

age between gametologs. Further expression analyses, incorporating a broader taxonomic range and data for individual cell types throughout female development, are essential to ascertain why specific genes have been retained on the avian W chromosome.

How Do Multi-Copy Gene Families Evolve on the W Chromosome?

Y chromosome degeneration is frequently characterised by massive gene amplification where many remaining Y-linked genes persist as members of multi-copy gene families [Skaletsky et al., 2003; Soh et al., 2014; Bachtrog et al., 2019; Vegesna et al., 2020]. However, until recently, gene amplification on the W chromosome has received comparatively less attention and it remained unclear whether large-scale gene amplification is a general feature of sex chromosome evolution or a peculiar quirk of the Y.

A handful of W-linked multi-copy gene families have been identified in a limited number of avian species [Hori et al., 2000; Backström et al., 2005; Davis et al., 2010; Smeds et al., 2015; Rogers et al., 2021]. The most comprehensively studied of these is histidine triad nucleotide-binding protein (HINTW), an ampliconic gene family that is hypothesized to play a role in female reproduction and oogenesis [O'Neill et al., 2000; Ceplitis and Ellegren, 2004]. At present, approximately 10 different copies of *HINTW* are annotated in the most recent chicken W chromosome assembly (GRCg7b), however, this is likely an underestimation with previous studies estimating over 40 copies [Hori et al., 2000; Backström et al., 2005]. Furthermore, large-scale amplification of *HINTW* is conserved across avian non-ratites [Hori et al., 2000]. Currently, evidence for the functionality of HINTW is lacking. However, it is known that HINT can form a heterodimer and the amino acid residues that form the dimer binding site are conserved in many HINTW copies, although many copies are nonfunctional [Hori et al., 2000; O'Neill et al., 2000]. Therefore, HINTW might act to disrupt the function of its Z-linked ortholog (*HINTZ*) by forming a heterodimer. Interestingly, the size of the *HINTW* gene family varies between chicken layer breeds subject to artificial selection for fecundity relative to other chicken breeds [Rogers et al., 2021], potentially suggesting a role of female-specific selection in driving gene amplification, although this relationship was absent across duck breeds.

The paucity of multi-copy gene families on the avian W chromosome is in stark contrast to the abundance of ampliconic genes often present on the Y chromosome. Several mechanisms have been proposed to drive the evolution of multi-copy gene families on the Y, including meiotic drive, sperm competition, genetic drift, and gene conversion [Skaletsky et al., 2003; Ellis et al., 2011; Cocquet et al., 2012; Larson et al., 2018; Bachtrog et al., 2019; Vegesna et al., 2020]. In theory, the strength of these processes might differ between the Y and W due to their contrasting inheritance patterns [Wright and Mank, 2013]. Notably, the Y chromosome is exposed to spermatogenesis, whereas the W is subject to oogenesis, and this likely leads to differences in the potential for antagonistic co-evolution between the X and Y versus the Z and W. Antagonistic co-evolution is predicted to drive the co-amplification of genes on sex chromosomes but should be weaker during oogenesis than spermatogenesis, potentially explaining the limited number of W-linked multi-copy gene families [Bachtrog, 2020]. Targeted avian gene knockouts [Ioannidis et al., 2021] provide an exciting opportunity to elucidate the functionality of HINTW copies, whether this varies across avian species, and the potential for antagonism between W and Z orthologs.

Is There a "Toxic W" Effect?

There appears to be a cost for males to carrying a degenerated Y chromosome [Brown et al., 2020; Xirocostas et al., 2020; Nguyen and Bachtrog, 2021; Connallon et al., 2022], where males in species with XY chromosomes tend to die earlier [Xirocostas et al., 2020]. Several hypotheses have been put forward to explain this phenomenon, including the presence of deleterious recessive mutations on the single X in males that would otherwise be shielded in females ("unguarded X") or the accumulation of mutations and repetitive elements on the Y chromosome ("toxic Y"). There is also evidence that the Y chromosome acts as a heterochromatin sink, reducing the efficiency of heterochromatin maintenance across the rest of the male genome [Francisco and Lemos, 2014; Brown et al., 2020].

Similar processes may operate on the W chromosome, where females exhibit a shorter lifespan than males across a range of species [Xirocostas et al., 2020]. Consistent with a "toxic W" effect, the avian W chromosome is a haven for repetitive material and transposable elements in several species. For instance, females in species with a degenerate W carry between 20 and 90% more endogenous retroviruses than males [Peona et al., 2021]. Furthermore, transposable element suppression is less effective on the crow W chromosome than the rest of the genome, leading to higher expression of transposable elements in females [Warmuth et al., 2022]. Although transposable elements can facilitate adaptive evolution, they also have the potential to reduce fitness through the disruption of gene activ-

Downloaded from http://karger.com/cgr/article-pdf/162/8-9/405/3975130/000529376.pdf by guest on 03 October 2024

ity and the promotion of chromosomal rearrangements [McDonald, 1993]. In theory, they may also contribute to an increased chance of female sterility in hybrids, where mechanistic mismatches between transposable repressor mechanisms and the W chromosome lead to reduced female fertility. This would provide further support for Haldane's rule where the heterogametic sex is more likely sterile in hybrids [Haldane, 1922].

Final Remarks

Recent studies have provided new insight into avian W chromosome evolution, challenging the traditional view that the avian W is genetically inert and highly conserved across species. There are clear parallels with Y chromosomal evolution but also key differences, primarily regarding the relative importance of the W in reproduction and fertility traits. Recent technological advances offer new potential to resolve uncertainty over the functionality of the W, for instance by using single-cell RNA-seq to establish fine-scale expression patterns of Z- and W-linked genes through development and across species [e.g., Estermann et al., 2020] and targeted gene knockouts to test gene function [e.g., Ioannidis et al., 2021]. Therefore, the next couple of years hold much promise for disentangling the function and evolution of the W chromosome in birds.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Sources
This work was funded by a NERC Independent Research Fellowship to A.E.W. (NE/N013948/1) and a NERC ACCE DTP to P.D.P.

Recovering the "Missing" Avian Genes Using Multi-Omics Data

(Prepared by Z.-T. Yin, J. Smith, and Z.-C. Hou)

Gene gain and loss are common events in the evolution of species, especially of birds, which have evolved many unique characteristics such as feathers, wings and flight capabilities, strong and lightweight skeletons, toothless beaks, high metabolic rates and heat absorption, sex, and unique respiratory and excretory systems [Kennedy and Vevers, 1976; Blomme et al., 2006]. The release of the first chicken genome provided the basis for systematic analysis of the similarities and differences between vertebrate and avian genomes [International Chicken Genome Sequencing Consortium, 2004]. In comparison with other amniotes, bird genomes are more compact, and this difference may be related to the overall smaller cell size [Hughes and Hughes, 1995; Hughes and Friedman, 2008]. The reductions in genome size may be the result of the loss of noncoding DNA sequences, with bird genomes having less repetitive DNA, fewer pseudogenes, and shorter introns than mammalian genomes [International Chicken Genome Sequencing Consortium, 2004; Hughes and Piontkivska, 2005]. Importantly, the evolution of avian genomes also appears to involve the loss of protein-coding genes, as the total number of uniquely identified avian coding genes is much smaller than in other tetrapods (i.e., 23,294 in humans, GRCh38.p14; 19,404 in lizards, AnoCar2.0; 17,007 in chickens, GRCg7b). Paralog analysis revealed a higher overall incidence of gene families with fewer members in birds compared to other vertebrates [Hughes and Friedman, 2008]. Likewise, birds have a high rate of chromosomal rearrangements compared to other organisms, all of which may result in the deletion of protein-coding genes [Backström et al., 2010]. In recent years, the genomes of a large number of birds and lizards have been assembled and annotated, including zebra finches [Warren et al., 2010], chickens [International Chicken Genome Sequencing Consortium, 2004], turkeys [Dalloul et al., 2010], and duck [Zhu et al., 2021]. Moreover, large-scale bird genome projects [Jarvis et al., 2014; Zhang G et al., 2014], and chicken pangenomes [Wang K et al., 2021; Li M et al., 2022] have also generated considerable genomic data. These large comparative genomic datasets identified hundreds of lost genomic-blocks in the bird genomes, and also suggested that hundreds of genes are missing in birds [Lovell et al., 2014; Zhang G et al., 2014].

The missing genes seem to be directly related to the unique physiological phenomena of birds. Several functionally important genes in mammals are supposed "missing" in chickens and have caused long-debated questions in bird biology. Spurious discovery of the missing/hidden genes in the bird genome has continued for decades. Previously, *BGN* [Blaschke et al., 1996], *CORO1A* [Xavier et al., 2008], *MAPK3* [Lemoine et al., 2009], *MMP14* [Simsa et al., 2007], *TBX6* [Lardelli, 2003; Ahn et al., 2012], *TSSK4* [Shang et al., 2013], and five adipokine genes [Dakovic et al., 2014] were reported to be missing in birds, however, several long-debated genes including *TNF-α*, and *leptin* have been cloned in birds [Prokop et al., 2014; Seroussi et al., 2016; Rohde et al., 2018]. This hide-and-seek game still continues, and does not appear to be ending anytime soon [Elleder and Kaspers, 2019].

Cytogenet Genome Res 2022;162:405–527
DOI: 10.1159/000529376

Downloaded from http://karger.com/cgr/article-pdf/162/8-9/405/3975130/000529376.pdf by guest on 03 October 2024