2022

SPERMATOGENESIS MOLECULAR EVOLUTION IN MURINE RODENTS

Emily Emiko Konishi Kopania

SPERMATOGENESIS MOLECULAR EVOLUTION IN MURINE RODENTS

By

EMILY EMIKO KONISHI KOPANIA

BS Biological Sciences, University of Southern California, Los Angeles, CA, 2016
MS Molecular Genetics and Biochemistry, University of Southern California, Los Angeles, CA, 2016

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Ecology and Evolution

The University of Montana
Missoula, MT

May 2022

Approved by:

Scott Whittenburg,
Graduate School Dean

Dr. Jeffrey M. Good, Chair
Division of Biological Sciences

Dr. Douglas J. Emlen
Division of Biological Sciences

Dr. Lila Fishman
Division of Biological Sciences

Dr. Travis Wheeler
Department of Computer Science

Dr. Erica L. Larson
Department of Biological Sciences, University of Denver

Kopania, Emily, Ph.D., January 2022 Ecology and Evolution

Spermatogenesis Molecular Evolution in Murine Rodents

Chairperson: Dr. Jeffrey Good

Reproductive traits are fascinating from an evolutionary perspective because they are necessary for individuals to produce offspring and increase their evolutionary fitness. Given the essentiality of reproduction to fitness, genes involved in reproduction may be expected to be highly conserved. However, some genes involved in reproduction evolve very rapidly, including many spermatogenesis genes. This rapid evolution may result from intense sexual selection acting on reproductive traits, particularly in species where females mate multiply thus creating the potential for sperm competition. In addition to sexual selection, other evolutionary forces may shape rapid spermatogenesis evolution, including genomic conflict and relaxed pleiotropic constraint due to the high specificity of genes involved in spermatogenesis. It is unclear how these forces may interact, their relative importance in spermatogenesis molecular evolution, and how the intensity of these forces changes across spermatogenesis developmental stages. Rapid spermatogenesis evolution is thought to have important downstream consequences, including rapid phenotypic evolution of male reproductive traits and reproductive barriers that contribute to speciation. However, direct connections between molecular evolution, phenotypic evolution, and speciation have rarely been made for male reproductive traits. Thus, my dissertation seeks to understand what are the causes and consequences of rapid spermatogenesis molecular evolution?

House mice (*Mus musculus*) and closely related species are an ideal system in which to address this question because they experience sperm competition, form natural hybrid zones and produce sterile hybrid males, readily breed and hybridize in the laboratory, and have extensive genomic resources available. Furthermore, house mice are part of the massive *Murinae* subfamily of rodents, which comprise over 10% of all mammal species and show remarkable variation in reproductive traits, including sperm morphology. Spermatogenesis is a complex developmental process, so understanding variation in the intensity of different evolutionary forces across spermatogenesis stages is critical to understanding spermatogenesis evolution. Fluorescence-activated cell sorting is one way to generate enriched cell populations representing different spermatogenesis stages. In this dissertation, I use gene expression data from sorted cell populations in house mice, as well as genomic and phenotypic data from mice and other murine rodents to study mammalian spermatogenesis evolution.

In Chapter 1, I use data from enriched cell populations representing two different spermatogenesis stages and four different species of mice to investigate the relative rates of molecular evolution across spermatogenesis and the types of mutations underlying gene expression evolution in different spermatogenesis stages. I show that lineage-specificity of genes expressed, gene expression level divergence, and protein sequence divergence all increase during the late stages of spermatogenesis. I also show that protein coding divergence, but not gene expression divergence, is higher on the X chromosome than the autosomes across spermatogenesis cell types. Lastly, I use published data from F1 mouse crosses to do allele-specific expression analyses and show that the types of regulatory mutations underlieing expression divergence are strikingly different between early and late spermatogenesis. This study provides insight into mammalian spermatogenesis molecular evolution and shows the importance of developmental context in molecular evolutionary studies. In Chapter 2, I perform two genetic

experiments involving advanced-generation hybrid mouse crosses to explore hybrid incompatibilities on the sex chromosomes and their effects on hybrid male spermatogenesis expression and reproductive phenotypes. My results refute the hypothesis that genomic conflict between the sex chromosomes contributes to sex chromosome overexpression during late spermatogenesis in sterile mouse hybrids. However, they do show that incompatibilities between the X and Y chromosomes, between the Y chromosome and autosomes, or both likely contribute to male hybrid sterility in house mice. These findings advance our understanding of genetic incompatibilities contributing to male hybrid sterility, a common barrier to reproduction between species. In Chapter 3, I expand my research on spermatogenesis evolution to the *Murinae* subfamily, using exome capture and phenotype data to investigate the role of sexual selection in sperm morphological evolution and test for positive selection acting on male reproductive genes. My analyses indicate that relative testes mass is evolving indepently of phylogeny, and therefore may be evolving in response to sperm competition. Most *Murinae* sperm have a hook on the sperm head, and I show that hook length and angle are correlated with relative testes mass suggesting that these traits may also be selected on by sperm competition. Lastly, I find that genes expressed in rapidly evolving male reproductive tissues and spermatogenesis cell types, specifically seminal vesicles and postmeiotic spermatids, tend to experience more positive selection than other male reproductive genes, so their rapid evolution is likely due in part to positive selection. These findings contribute to our understanding of the underlieing causes of the rapid evolution of reproduction at both the phenotypic and molecular levels.

In addition to these three chapters, I contributed to several related projects that address the overarching questions of my dissertation: a review on sex chromosome evolution in mammals in the context of spermatogenesis (Larson, et al. 2018), two methodological papers on quantifying sperm morphology (Skinner, et al. 2019a; Skinner, et al. 2019b), a peer-reviewed research article on disrupted X chromosome expression at different spermatogenesis stages in sterile house mouse hybrids (Larson, et al. 2021), and a study on X chromosome evolution in dwarf hamsters (Moore, et al. 2022). Collectively, my dissertation and related projects contribute to our understanding of reproduction and molecular evolution in mammals.

References

Larson EL, Kopania EEK, Good JM. 2018. Spermatogenesis and the Evolution of Mammalian Sex Chromosomes. *Trends Genet*. 34:722-732.

Larson EL, Kopania EEK, Hunnicutt KE, Vanderpool D, Keeble S, Good JM. 2021. Stage-specific disruption of X chromosome expression during spermatogenesis in sterile house mouse hybrids. *G3 Genes|Genomes|Genetics*.

Moore EC, Thomas GWC, Mortimer S, Kopania EEK, Hunnicutt KE, Clare-Salzler ZJ, Larson EE, Good JM. 2022. The evolution of widespread recombination suppression on the Dwarf Hamster (Phodopus) X chromosome. *bioRxiv*.2021.2011.2015.468705.

Skinner BM, Bacon J, Rathje CC, Larson EL, Kopania EEK, Good JM, Affara NA, Ellis PJI. 2019a. Automated Nuclear Cartography Reveals Conserved Sperm Chromosome Territory Localization across 2 Million Years of Mouse Evolution. *Genes*. 10:109.

Skinner BM, Rathje CC, Bacon J, Johnson EEP, Larson EL, Kopania EEK, Good JM, Yousafzai G, Affara NA, Ellis PJI. 2019b. A high-throughput method for unbiased quantitation and categorization of nuclear morphology. *Biol Reprod*. 100:1250-1260.

Acknowledgements

I am extremely grateful for the scientific and personal support I have received while working on this dissertation and throughout my life.

Thank you to my co-authors and collaborators. I have had the privilege of working with many amazing scientists who are also great colleagues. Thank you to University of Montana core facilities staff: Lab Animal Resources, the Genomics Core, and the Flow Cytometry Core, especially Pam Shaw.

Matt Dean, my undergraduate and master's advisor, introduced me to the field of evolutionary biology and encouraged me to apply to Jeff Good's lab for my Ph.D. Matt taught me to believe in myself as a scientist and I am grateful for his continued support over the years.

I have been fortunate to have a fantastic committee. Thank you to Travis Wheeler for your vast computational knowledge and for always having a kind word. Doug Emlen, your excitement for science was always inspiring, and I am honored to have had the opportunity to discuss research on sexual selection from someone who is so knowledgeable in this field. Thank you, Lila Fishman, for sharing your incredible evolutionary genomics expertise, especially on speciation and genomic conflict. I am also grateful for the many letters of recommendation you wrote me over the course of my Ph.D. and for the advice on grant writing I received from you. I especially want to thank Erica Larson, who has been a close collaborator and encouraging mentor over the course of my Ph.D. Much of my dissertation research builds from work that Erica did, and she always has great advice about everything from career decisions to big-picture research questions to the details of experiments and analyses.

I feel extremely fortunate to have had Jeff Good as my Ph.D. advisor. Jeff is a brilliant scientist and inspiring person to work with, always coming up with new ideas and creative approaches. Thank you, Jeff, for sharing your expertise with me, teaching me how to think broadly about evolutionary biology, and teaching me how to be a scientist. Thank you, also, for your career advice, encouragement, and supportive mentorship. The Good Lab has been an intellectually enriching, supportive, and fun environment in which to do my Ph.D. I would like to thank all the students, postdocs, and technicians I worked with in the lab and in the broader UM evolutionary genomics community for technical help with lab work and analyses, and for many exciting conversations about science. I would particularly like to thank Emily Moore for being an excellent person to talk science with and an incredible friend.

The Ecology and Evolution department at the University of Montana was a fantastic place to do my Ph.D. and I would like to thank all the faculty, staff, postdocs, and students in the program. Many people in this department have become dear friends. I genuinely enjoyed my time in graduate school, mostly because of these friends. I would also like to thank my hometown and college friends who were very encouraging as I was working on my Ph.D.

Peter Williams, thank you for your love and support throughout our time in graduate school. No matter the challenges you were facing while working on your own Ph.D., you always found the time and energy to be there for me. Thank you for listening to my practice talks, encouraging me through difficult times, and making grad school fun.

And most importantly, thank you to my family: Mom, Dad, Ben, Jib, and Keanu. Jib was a wonderful pup, and Keanu continues to find new ways to be adorable and keep me smiling. Ben, you are the best brother ever. Thank you for tolerating the Montana cold to visit me. Mom and Dad, you have always been my biggest supporters and I could not have done this without you. Thank you for fostering my love of science ever since I was a little kid, and for all the time and effort you put into raising me. Working on a Ph.D. was a challenging undertaking, and I was able to face that challenge thanks to the confidence and strong work ethic you instilled in me.

Table of Contents

Molecular Evolution across Mouse Spermatogenesis

Emily E.K. Kopania [1],*[,1] Erica L. Larson [2] Colin Callahan,[1] Sara Keeble [1] and Jeffrey M. Good *[,1]

[1]Division of Biological Sciences, University of Montana, Missoula, MT, USA

[2]Department of Biological Sciences, University of Denver, Denver, CO, USA

*Corresponding authors: E-mails: emily.kopania@umconnect.umt.edu; jeffrey.good@umontana.edu.

Associate editor: Patricia Wittkopp

Abstract

Genes involved in spermatogenesis tend to evolve rapidly, but we lack a clear understanding of how protein sequences and patterns of gene expression evolve across this complex developmental process. We used fluorescence-activated cell sorting (FACS) to generate expression data for early (meiotic) and late (postmeiotic) cell types across 13 inbred strains of mice (*Mus*) spanning ~7 My of evolution. We used these comparative developmental data to investigate the evolution of lineage-specific expression, protein-coding sequences, and expression levels. We found increased lineage specificity and more rapid protein-coding and expression divergence during late spermatogenesis, suggesting that signatures of rapid testis molecular evolution are punctuated across sperm development. Despite strong overall developmental parallels in these components of molecular evolution, protein and expression divergences were only weakly correlated across genes. We detected more rapid protein evolution on the X chromosome relative to the autosomes, whereas X-linked gene expression tended to be relatively more conserved likely reflecting chromosome-specific regulatory constraints. Using allele-specific FACS expression data from crosses between four strains, we found that the relative contributions of different regulatory mechanisms also differed between cell types. Genes showing *cis*-regulatory changes were more common late in spermatogenesis, and tended to be associated with larger differences in expression levels and greater expression divergence between species. In contrast, genes with *trans*-acting changes were more common early and tended to be more conserved across species. Our findings advance understanding of gene evolution across spermatogenesis and underscore the fundamental importance of developmental context in molecular evolutionary studies.

Key words: **gene expression, allele-specific expression, faster-X evolution, fluorescence activated cell sorting (FACS), phylogenetic contrasts.**

Introduction

Mature sperm are the most morphologically diverse animal cell type, likely as a consequence of intense selection on sperm form and function (Pitnick et al. 2009). Genes involved in spermatogenesis also tend to evolve rapidly (Swanson et al. 2003; Good and Nachman 2005; Turner et al. 2008; Larson et al. 2016; Finseth and Harrison 2018), suggesting that pervasive sexual selection also shapes molecular evolution (Swanson and Vacquier 2002; Harrison et al. 2015). However, direct genotype-to-phenotype connections remain elusive for primary sexually selected traits, and there are additional evolutionary forces acting during spermatogenesis that shape overall patterns of molecular evolution (Good and Nachman 2005; Burgoyne et al. 2009; Dean et al. 2009; Larson et al. 2016; Schumacher and Herlyn 2018). For example, many spermatogenesis genes are highly specialized (Eddy 2002; Chalmel et al. 2007; Green et al. 2018), which can relax pleiotropic constraint and contribute to rapid evolution even in the absence of positive directional selection (Winter et al. 2004; Larracuente et al. 2008; Meisel 2011). Other components of spermatogenesis are highly conserved because small disruptions can lead to infertility (Burgoyne et al. 2009). Thus, spermatogenesis genes are likely to experience strong and sometimes contradictory evolutionary pressures. Understanding how these processes interact to shape molecular evolution across spermatogenesis is essential to understanding how natural selection shapes the genetic determinants of male fertility.

There are many components or levels of molecular evolution, spanning from protein sequence changes to differences in gene expression level, timing, and developmental specificity (King and Wilson 1975; Wray et al. 2003; Larracuente et al. 2008; Kaessmann 2010; Piasecka et al. 2013; Cridland et al. 2020). Many of these components have been shown to evolve relatively rapidly during spermatogenesis (Meiklejohn et al. 2003; Khaitovich et al. 2005; Voolstra et al. 2007; Brawand et al. 2011; Harrison et al. 2015; Vicens et al. 2017; Cridland et al. 2020; Sánchez-Ramírez et al. 2021), and generally trend toward increased divergence during the later stages of development (Good and Nachman 2005; Piasecka et al. 2013; Larson et al. 2016). Novel genes disproportionately arise with testis-specific expression (Levine et al. 2006; Zhao et al. 2014; Cridland et al. 2020; Schroeder et al. 2020; Lange et al. 2021), likely as a consequence of the more permissive regulatory environment of the later stages of sperm development (Kaessmann 2010; Soumillon et al. 2013). Likewise, the later stages of spermatogenesis tend to be enriched for novel

Article

testis-specific genes (Eddy 2002; Chalmel et al. 2007; Green et al. 2018). These developmental signatures of novelty and specialization are further reflected in patterns of increased divergence of protein sequences (Good and Nachman 2005; Kousathanas et al. 2014) and expression levels (Larson et al. 2016) between species during the later stages of sperm development. Parallel signatures of rapid molecular evolution likely reflect both relaxed constraints during the late stages of spermatogenesis, and enhanced positive selection on late-developing sperm phenotypes (Eddy 2002; Good and Nachman 2005; Larracuente et al. 2008; Larson et al. 2016; Cutter and Bundus 2020). However, it remains unclear how strongly different forms of molecular evolution are correlated. For example, changes in gene expression may often be cell or stage-specific and therefore may be less pleiotropic than protein-coding changes. This pleiotropic constraint hypothesis primarily applies to *cis*-regulatory changes, which likely affect one gene, whereas *trans*-regulatory changes can affect many genes across multiple cell types (Wray et al. 2003; Carroll 2008; Cutter and Bundus 2020).

The X chromosome provides a compelling example of how the conflicting selective pressures acting on spermatogenesis may shape different components of molecular evolution. Theory predicts that the X chromosome should evolve more rapidly than the autosomes, particularly if most beneficial mutations are recessive, because X-linked recessive beneficial mutations will always be exposed to selection in males (Charlesworth et al. 1987; Vicoso and Charlesworth 2009). Differences in effective population size (N_e) on the X chromosome may also affect relative rates of fixation on the X chromosome and autosomes due to genetic drift, but the relative differences in N_e depend on the relative reproductive success of different sexes in a population (Vicoso and Charlesworth 2009). Consistent with more efficient X-linked selection, protein-coding evolution tends to be faster on the X chromosome compared with the autosomes in several taxa, and this effect is often strongest for genes with male-biased expression (Khaitovich et al. 2005; Baines and Harr 2007; Baines et al. 2008; Meisel and Connallon 2013; Parsch and Ellegren 2013; Larson et al. 2016). Novel genes tend to arise more often on the X chromosome, and these are often expressed during spermatogenesis (Levine et al. 2006; Kaessmann 2010). There is also some evidence for rapid expression evolution on the X chromosome in flies and mammals (Khaitovich et al. 2005; Brawand et al. 2011; Meisel et al. 2012; Coolon et al. 2015), but X-linked expression in mice appears conserved relative to autosomal genes expressed during the later stages of spermatogenesis (Larson et al. 2016). Stage-specific differences in relative rates of expression evolution on the X chromosome may result from the unique regulatory pattern that the sex chromosomes undergo during mammalian spermatogenesis. In males, the X chromosome is inactivated early in meiosis (i.e., meiotic sex chromosome inactivation, MSCI; McKee and Handel 1993) and remains partially repressed during the post-meiotic haploid stages of sperm development (i.e., postmeiotic sex chromosome repression, PSCR; Namekawa et al. 2006). The theory underlying faster-X protein-coding evolution may also apply to *cis*-regulatory gene expression evolution, but X chromosome expression divergence is likely also affected by *trans*-regulatory changes on other chromosomes and regulatory constraints unique to the X chromosome (e.g., MSCI and PSCR, Meisel et al. 2012). Thus, comparing relative expression divergence on the X chromosome compared with the autosomes can give insight into the types of mutations and selective forces affecting X chromosome expression.

These stage-specific patterns highlight the importance of studying specific components of molecular evolution in a developmental framework (fig. 1A; Larson, Kopania, et al. 2018; Cutter and Bundus 2020). However, studies of molecular evolution have primarily focused on pairwise contrasts across nuanced aspects of tissue development (Good and Nachman 2005; Larson et al. 2016), or examined protein-coding versus regulatory evolution in whole tissues (Khaitovich et al. 2005; Voolstra et al. 2007; Mack et al. 2016; Vicens et al. 2017; Cridland et al. 2020), without combining both in a phylogenetic framework (but see Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mößinger K, Brüning T, Lamanna F, Belles MR, Conrad C, Kondova I, Bontrop R, Behr R, Khaitovich P, Pääbo S, Marques-Bonet T, Grützner F, Almstrup K, Schierup MH, Kaessmann H, 2021, unpublished data, https://www.biorxiv.org/content/10.1101/2021.11.08.467712v1, last accessed November 30, 2021). Relying on whole tissue expression comparisons may be particularly problematic for spermatogenesis, because differences in testis composition are expected to evolve rapidly between species (Ramm and Schärer 2014; Yapar E, Saglican E, Dönertaş HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfs RV, Khaitovich P, Somel M, 2021, unpublished data, https://www.biorxiv.org/content/10.1101/010553v2, last accessed July 12, 2021) and may confound patterns of expression level divergence (Good et al. 2010; Larson et al. 2016; Hunnicutt et al. 2021). Nonetheless, collection of stage or cell-specific expression data remains technically demanding (da Cruz et al. 2016; Green et al. 2018), likely limiting widespread use in comparative studies. As a consequence, most evolutionary studies of gene expression have relied on whole tissue comparisons between closely related species pairs, instead of using more powerful phylogenetic approaches (Rohlfs and Nielsen 2015; Dunn et al. 2018).

In this study, we use a comparative developmental approach to gain a more comprehensive understanding of molecular evolution across spermatogenesis in house mice (*Mus*). Mice are the predominant laboratory model for mammalian reproduction (Phifer-Rixey and Nachman 2015; Firman 2020), with abundant genomic resources (Keane et al. 2011; Thybert et al. 2018), and established wild-derived inbred strains that can be crossed to resolve mechanisms underlying expression divergence (i.e., *cis*- vs. *trans*-regulatory changes; Mack et al. 2016). Mice also show divergence in sperm head morphologies across closely related species (Skinner et al. 2019) and experience sperm competition in the wild (Dean et al. 2006), providing a compelling system for understanding the evolution of spermatogenesis.

We used fluorescence-activated cell sorting (FACS) to resolve patterns of gene expression in two enriched spermatogenic cell populations across several mouse strains, species,

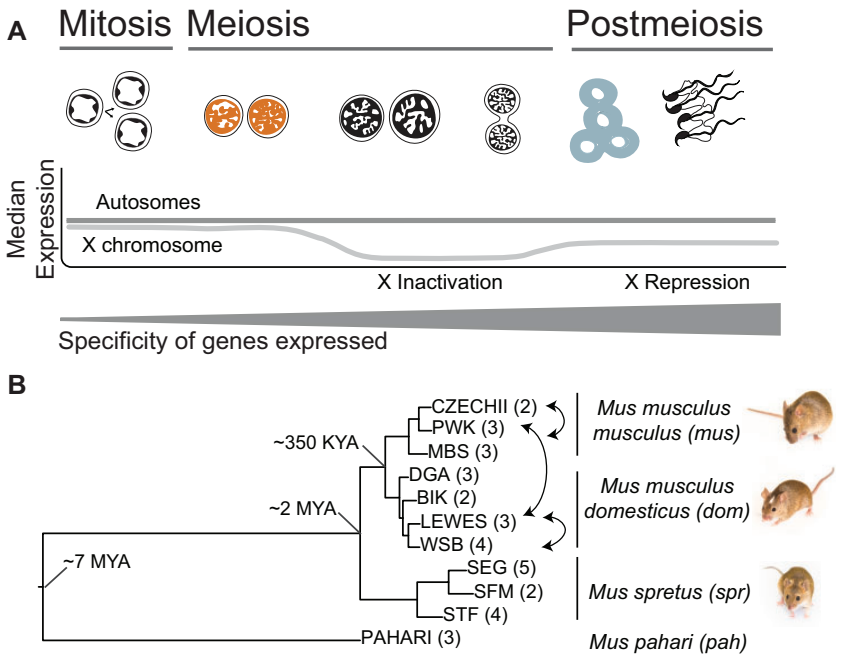

Fig. 1. (*A*) Predictive framework depicting the major stages of spermatogenesis and expected relative expression levels of the X chromosome and autosomes at each stage (Namekawa et al. 2006). The two cell populations used in this study are leptotene-zygotene ("early," second from left, orange) and round spermatids ("late," second from right, blue). The relative thickness of the gray bar represents the predicted cell type specificity at each stage (Eddy 2002; Chalmel et al. 2007; Larson et al. 2016; Green et al. 2018). (*B*) Maximum likelihood tree of concatenated exome data from the four *Mus* species or subspecies used in this study: *Mus musculus musculus* (*mus*), *Mus musculus domesticus* (*dom*), *Mus spretus* (*spr*), *Mus pahari* (*pah*). Tips are labeled with the inbred strains from each lineage, with select crosses used to generate F1 hybrids indicated with arrows. Number of individuals sampled for each strain indicated in parentheses. Approximate divergence times are placed at each major node (Chevret et al. 2005). All nodes had 100% bootstrap support.

and cross types (fig. 1*A*). Our study used two main comparisons. First, we evaluated divergence in spermatogenic protein sequences and gene expression levels across thirteen inbred strains of mice, including two subspecies of the house mouse (*Mus musculus*) and two other *Mus* species spanning 7 My of evolution (fig. 1*B*; Chevret et al. 2005). Second, we used published data from reciprocal crosses between a subset of these inbred strains to resolve the relative contribution of *cis*- versus *trans*-regulatory changes to expression divergence. We used these data to address five main questions: 1) Is gene expression more lineage-specific during late spermatogenesis? 2) Do protein-coding sequences and gene expression levels evolve faster during the later stages of spermatogenesis? 3) Is the rate of molecular evolution elevated on the X chromosome compared with the autosomes, and does this relationship change across spermatogenesis? 4) To what extent are protein-coding and gene expression divergence correlated, and does this relationship change across developmental stages? 5) Are there differences in the relative contributions of regulatory mechanisms (*cis*- vs. *trans*-regulatory changes) across spermatogenesis?

Results

Spermatogenesis Gene Expression by Cell Type and Lineage

We collected spermatogenesis expression data from 34 mice representing four different species or subspecies: *Mus musculus musculus*, *Mus musculus domesticus*, *Mus spretus*, and *Mus pahari*. We will use the abbreviations *mus*, *dom*, *spr*, and *pah* to reference the four major groups, and refer to all taxa as "lineages" for concision (fig. 1*B*). For each sample, we generated expression data for two spermatogenic cell types, an early meiotic cell type (leptotene-zygotene cells from early prophase of meiosis I, hereafter "early") and a postmeiotic cell type (round spermatids, hereafter "late"). We identified 23,164 one-to-one orthologs, including both protein-coding and nonprotein-coding genes, that were annotated in all four mouse lineages and the mouse reference (GRCm38). From this set, we defined expressed genes as those with an FPKM $>$ 1 in all samples of a given cell type. Expression variance cleanly separated samples by cell type and lineage (supplementary fig. S1, Supplementary Material online), indicating successful enrichment of different cell types. Most expressed genes were detected in both cell types (table 1). However, approximately one third of the detected genes were preferentially expressed or "induced" in a given cell type (transcripts with $> 2\times$ median expression level in one cell type across all lineages; table 1). We also identified expressed genes that show testis-specific expression based on published multi-tissue expression data (Chalmel et al. 2007). We found that 493 testis-specific genes were induced late, whereas only 65 testis-specific genes were induced early (table 1), consistent with increased specificity late in spermatogenesis (Eddy 2002; Larson et al. 2016; Green et al. 2018). To distinguish experimental noise from biologically meaningful expression, we also

Table 1. Counts of Genes in Each Data Set and Cell Type across Spermatogenesis.

	Early	Late	Both Early and Late
Expressed	9,570	8,986	7,670
Induced	3,375	2,769	0
Testis-specific (TS)[a]	544	655	524
Induced and TS	65	493	0
Active (*dom*)	8,206 (98.2%)	8,581 (90.4%)	6,355
Active (*mus*)	8,782 (97.5%)	10,098 (83.4%)	7,289
Active (*spr*)	8,728 (97.1%)	9,509 (86.0%)	7,227
Active (*pah*)	8,124 (97.6%)	9,563 (83.9%)	6,682

NOTE.—Numbers in parentheses represent the percent of genes in the "active" data sets that were also in the "expressed" data set. Early, spermatocytes (leptotene/zygotene); Late, round spermatids.
[a]Testis-specific inferred from Chalmel et al. (2007).

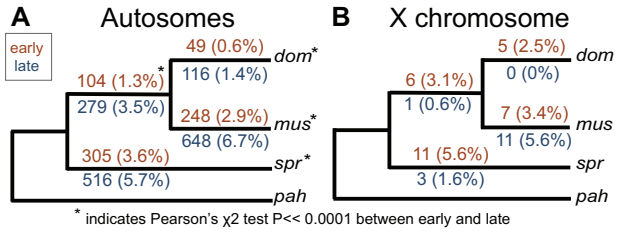

FIG. 2. Number of genes that were lineage-specific on each internal branch of the mouse phylogeny used in this study. Numbers in parentheses are the percent of active genes that were lineage-specific. Results are presented separately for the autosomes (*A*) and X chromosome (*B*). Orange values above each branch represent the early cell type and blue values below represent the late cell type. Asterisks indicate a significant difference between early and late on that branch based on a Pearson's χ^2 test.

used a Bayesian approach to determine if a gene was "active" in a tissue or cell type (Thompson et al. 2020) and found broad overlap with genes in the expressed data set (table 1). Using the same framework, we identified genes showing evidence for lineage-specific expression ("active" in a single lineage or subset of lineages). We tested for lineage-specificity in each cell type separately, so a gene that we considered lineage-specific in one cell type may be expressed in other lineages during other spermatogenesis stages.

We found that lineage-specificity was rare overall, but more common for autosomal genes active during late spermatogenesis (Pearson's χ^2 test; *dom*: $P \ll 0.0001$, *mus*: $P \ll 0.0001$, *spr*: $P \ll 0.0001$, *dom-mus* common ancestor: $P \ll 0.0001$; fig. 2*A*). X-linked genes showed no significant differences in lineage-specificity between early and late cell types (fig. 2*B*), which could reflect a lack of specialization on the sex chromosomes, or reduced power to detect differences between cell types given small sample sizes. Few genes were lineage-specific in both cell types, and all were autosomal (*dom*: 9 genes, *mus*: 24 genes, *spr*: 24 genes, *dom-mus*: 21 genes). We found similar results using a log fold-change (logFC) approach with different logFC cutoff values to identify lineage-specific genes (supplementary fig. S2 and table S1, Supplementary Material online). Lineage-specific genes were not enriched for any processes specifically related to male reproduction. We also tested if lineage-specific genes tended to have higher or lower associations with coexpression networks using weighted gene coexpression network analysis (WGCNA, Langfelder and Horvath 2008). We did not see a general pattern across all lineage-specific genes, but genes specific to a given lineage tended to have higher association with coexpression modules associated with that lineage (supplementary fig. S3A, Supplementary Material online). Our results suggest that lineage-specific expression of spermatogenic genes is relatively uncommon at these shallow phylogenetic scales, but more likely to arise later in spermatogenesis.

Greater Protein-Coding and Gene Expression Divergence during Late Spermatogenesis

Having detected subtle increases in lineage specificity late in spermatogenesis, we next tested if rates of protein sequence evolution (dN/dS) and expression level divergence were also elevated during the postmeiotic stage, as has been reported previously (Larson et al. 2016). Genes induced late in spermatogenesis showed significantly higher rates of protein-coding divergence on both the autosomes ($n = 2{,}046$ genes induced early, median $dN/dS = 0.11$; $n = 1{,}711$ genes induced late, median $dN/dS = 0.20$; Wilcoxon rank sum test $P \ll 0.0001$) and the X chromosome ($n = 54$ genes induced early, median $dN/dS = 0.25$; $n = 61$ genes induced late, median $dN/dS = 0.41$; Wilcoxon rank sum test $P = 0.049$; fig. 3*A*, supplementary tables S2 and S3, Supplementary Material online). The 489 testis-specific genes showed elevated dN/dS overall, but most testis-specific genes were expressed in both cell types and there was no significant difference between genes expressed early and late for the autosomes ($n = 350$ genes expressed early, median $dN/dS = 0.28$; $n = 424$ genes expressed late, median $dN/dS = 0.30$; Wilcoxon rank sum test $P = 1$) or the X chromosome ($n = 16$ genes expressed early; median $dN/dS = 0.59$; $n = 24$ genes expressed late, median $dN/dS = 0.58$; Wilcoxon rank sum test $P = 1$). However, 348 testis-specific genes were preferentially expressed in the late cell type, representing ~20% of all genes induced late for which we were able to calculate dN/dS. Taken together, these results confirm that tissue specificity plays an important role in the rapid protein-coding divergence of spermatogenic genes, and that most of this signature involves genes induced during postmeiotic spermatogenesis.

We used a phylogenetic ANOVA to estimate expression divergence while controlling for phylogenetic relatedness and variance within lineages (i.e., the expression variance and

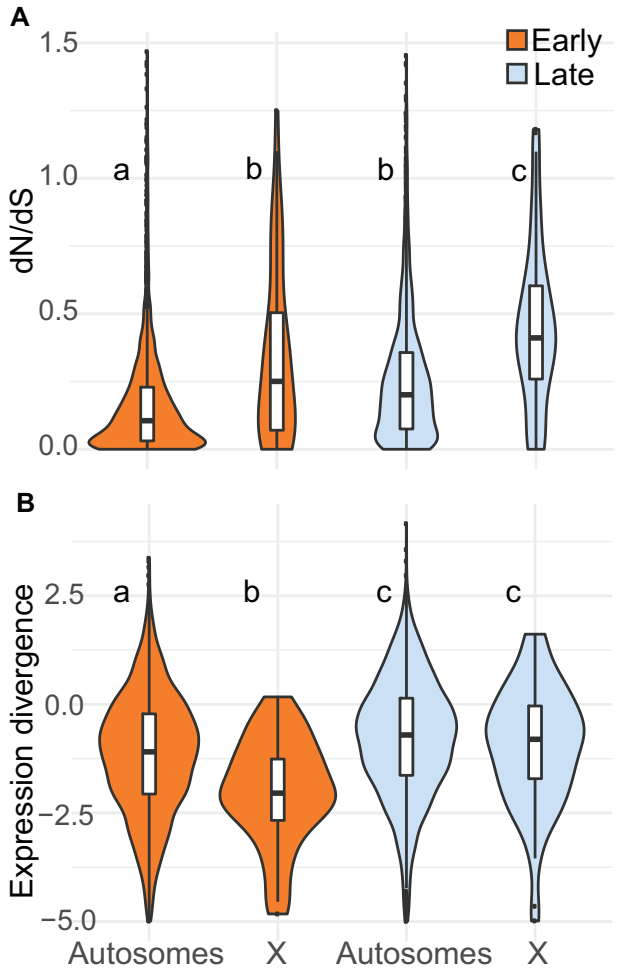

FIG. 3. (*A*) Protein-coding and (*B*) expression divergence on the autosomes and X chromosome for genes induced in each cell type. Expression divergence values on the y-axis are $-\log(\text{beta}_i)$, where $beta_i$ is the measure of expression divergence from EVE. Higher values on the y-axis represent higher divergence. The center of each violin plot is a standard boxplot, with the center horizontal line representing the median divergence value. The violins show the probability density of divergence values for each group. A wider part of the violin at a given value means genes expressed in that group are more likely to have that divergence value. The letters above each violin indicate significant differences between the cell types and chromosome types based on a Wilcoxon rank sum test.

evolution [EVE] model; Rohlfs and Nielsen 2015). We report expression divergence from EVE as $-\log(beta_i)$, where $beta_i$ is a metric from EVE that represents the ratio of within-lineage variance to between-lineage evolutionary divergence, and higher positive $-\log(beta_i)$ values correspond to greater divergence between lineages. Expression divergence was higher for genes induced late in spermatogenesis on both the autosomes ($n = 2{,}461$ genes induced early, median EVE divergence $= -1.09$; $n = 2{,}305$ genes induced late, median EVE divergence $= -0.70$; Wilcoxon rank sum test $P \ll 0.0001$) and the X chromosome ($n = 44$ genes induced early, median EVE divergence $= -2.04$; $n = 68$ genes induced late, median EVE divergence $= -0.80$; Wilcoxon rank sum test $P = 0.00019$; fig. 3*B*). This pattern held for all expressed genes, testis-specific genes, and different threshold cutoffs for considering genes induced (supplementary tables S4 and S5, Supplementary Material online). We also found higher divergence late for expressed and induced autosomal genes (supplementary table S5, Supplementary Material online) based on pairwise expression divergences using logFC and the metric from (Meisel et al. 2012); however, the pairwise framework did not give a consistent pattern on the X chromosome. When looking at all genes, most pairwise comparisons showed higher divergence late, but induced genes showed no difference between early and late spermatogenesis for most comparisons. However, the *dom* versus *spr* comparison had lower divergence late for all expressed genes and induced genes (supplementary table S5, Supplementary Material online).

Next, we tested if pleiotropic constraint imposed by protein–protein interactions contributed to less divergence during early spermatogenesis. We compared EVE expression divergence and dN/dS protein sequence divergence to the number of protein–protein interactions for genes in the mouse interactome database (MIPPIE, Alanis-Lobato et al. 2020). We found that genes induced early had fewer high-scoring protein–protein interactions (FDR-corrected Wilcoxon rank sum $P \ll 0.0001$, supplementary fig. S4, Supplementary Material online), suggesting that these genes may actually be less constrained by protein–protein interactions. However, this difference was subtle, and protein-protein interactions are only one measure of potential pleiotropy, so genes induced early may still be constrained by their roles in other tissues or cell types. For both cell types, the number of protein–protein interactions was significantly negatively correlated with dN/dS (early: $\rho = -0.122$, Spearman's rank correlation $P \ll 0.001$; late: $\rho = -0.143$, Spearman's rank correlation $P \ll 0.001$), but not EVE divergence (early: $\rho = -0.032$, Spearman's rank correlation $P = 0.5$; late: $\rho = -0.060$, Spearman's rank correlation $P = 0.5$), consistent with hypotheses that protein sequence evolution is more constrained by pleiotropy and protein–protein interactions compared with gene expression evolution (Carroll 2008).

Collectively, we found strong evidence for more rapid protein-coding and gene expression level divergence during post-meiotic spermatogenesis, suggesting that these general patterns hold after controlling for phylogeny and at deeper divergence levels than had previously been shown in mice (Larson et al. 2016). Despite our expanded phylogenetic sample, we still lacked the power to determine if more rapid expression and protein-coding divergence is due to positive directional selection (supplementary fig. S5, Supplementary Material online).

Weak Positive Correlation between Gene Expression and Protein-Coding Divergence

We next tested for more general relationships between protein-coding and expression divergence across sets of genes expressed or induced during spermatogenesis (supplementary fig. S6 and table S6, Supplementary Material online). Across all autosomal genes expressed early, there was a weak positive correlation between dN/dS and pairwise expression divergence ($\rho = 0.13$–0.17, Spearman's rank

Table 2. Correlation between Pairwise Expression Divergence Values for All Possible Pairwise Comparisons.

		dom versus *mus*	*dom* versus *spr*	*mus* versus *spr*	*dom* versus *pah*	*mus* versus *pah*
Early, X-linked	***dom* versus *spr***	**0.34**				
	mus* versus *spr	***0.07***	***0.28***			
	dom* versus *pah	***0.07***	***0.14***	***0.19***		
	mus* versus *pah	***0.16***	***0.10***	***0.03***	**0.62**	
	spr* versus *pah	***0.14***	**0.27**	***0.16***	**0.58**	**0.67**
Early, autosomal	***dom* versus *spr***	***0.32***				
	mus* versus *spr	**0.32**	**0.61**			
	dom* versus *pah	**0.28**	**0.28**	**0.27**		
	mus* versus *pah	**0.29**	**0.26**	**0.30**	**0.74**	
	spr* versus *pah	**0.24**	**0.32**	**0.34**	***0.55***	***0.57***
Late, X-linked	***dom* versus *spr***	**0.36**				
	mus* versus *spr	**0.50**	**0.45**			
	dom* versus *pah	**0.20**	**0.23**	**0.22**		
	mus* versus *pah	**0.28**	**0.28**	**0.36**	**0.74**	
	spr* versus *pah	**0.15**	***0.20***	**0.20**	**0.73**	**0.72**
Late, autosomal	***dom* versus *spr***	**0.35**				
	mus* versus *spr	**0.37**	**0.59**			
	dom* versus *pah	**0.30**	**0.33**	**0.30**		
	mus* versus *pah	**0.30**	**0.30**	**0.33**	**0.76**	
	spr* versus *pah	**0.25**	**0.32**	**0.33**	**0.64**	**0.63**

NOTE.—Numbers presented are ρ values from a Spearman's rank correlation test. We tested for correlations in pairwise expression divergence value among induced genes in each stage and chromosome group (early X, early autosomal, late X, and late autosomal). Gray boxes indicate no significant correlation between pairwise divergence values after FDR correction (Spearman's rank correlation $P > 0.05$). Italic values indicate the lowest Spearman's ρ value for each pairwise comparison across the four stages and chromosome groups.

correlation $P \ll 0.0001$). For induced genes, this correlation was weaker but still significant ($\rho = 0.07$–0.11, Spearman's rank correlation $P < 0.05$). For the late cell type, there was also a weak positive correlation between pairwise expression divergence and dN/dS on the autosomes, but the correlation was weaker than that seen in the early cell type ($\rho = 0.03$–0.05, Spearman's rank correlation $P < 0.05$). There was no correlation for the set of genes induced late. When looking only at genes with evidence for positive directional selection at the protein-coding level after correction for multiple tests (366 genes), the correlation was stronger on the autosomes late for the *dom* versus *spr* ($n = 250$ genes, $\rho = 0.17$, Spearman's rank correlation $P = 0.02$) and *mus* versus *spr* comparisons ($n = 249$ genes, $\rho = 0.18$, Spearman's rank correlation $P \ll 0.0001$). When comparing dN/dS to EVE expression divergence, we only saw a significant positive correlation for genes expressed late that were also under positive selection at the protein-coding level ($n = 160$ genes, $\rho = 0.18$, Spearman's rank correlation $P = 0.04$). We also tested if dN/dS was correlated with module eigengene values in our WGCNA. There was a weak positive correlation for eigengene values in the late cell type module ($\rho = 0.033$, FDR-corrected $P = 0.03$, supplementary fig. S3C, Supplementary Material online), but not the early cell type module ($\rho = 0.026$, FDR-corrected $P = 0.07$). In summary, we tended to observe a positive relationship between protein-coding and expression level divergence, but the strength of this relationship was weak and varied by gene set and divergence metric.

Faster-X Protein-Coding but Not Gene Expression Evolution

In addition to comparisons between spermatogenesis cell types, we compared relative rates of molecular evolution between X-linked and autosomal genes within a cell type. We found that protein-coding divergence was higher on the X chromosome, both early and late, across all gene sets (fig. 3A, supplementary tables S3 and S4, Supplementary Material online) consistent with several previous studies (Khaitovich et al. 2005; Baines et al. 2008; Meisel and Connallon 2013; Kousathanas et al. 2014; Larson et al. 2016). For expression evolution, we found lower divergence on the X chromosome early using EVE ($n = 2{,}461$ autosomal genes, median EVE divergence $= -1.09$; $n = 44$ X-linked genes, median EVE divergence $= -2.04$; Wilcoxon rank sum test $P = 0.00015$; fig. 3B), but higher X-linked divergence when using pairwise comparisons (supplementary table S5, Supplementary Material online). A major difference between these approaches was that EVE calculates divergence across a phylogeny, so genes that show divergent expression levels in one lineage may still be conserved across the entire phylogeny. We detected significant correlations between pairwise divergence values for different pairwise comparisons on the autosomes, and during late spermatogenesis, but lower or nonsignificant correlations on the X early (table 2). Thus, many genes on the X chromosome expressed early showed relatively high divergence between two particular lineages, but lower divergence across other pairwise comparisons and across the phylogeny as a whole. This lineage-specific variance underscores the importance of evaluating gene expression divergence in a phylogenetic framework (Rohlfs and Nielsen 2015; Dunn et al. 2018).

In late spermatogenic cells (i.e., round spermatids), X-linked expression divergence was similar to or lower than on the autosomes depending on the contrast and approach. Using EVE, we found similar divergence on the X chromosome and autosomes late ($n = 2{,}305$ autosomal genes, median EVE divergence $= -0.70$; $n = 68$ X-linked genes, median EVE divergence $= -0.80$; Wilcoxon rank sum test $P = 0.34$;

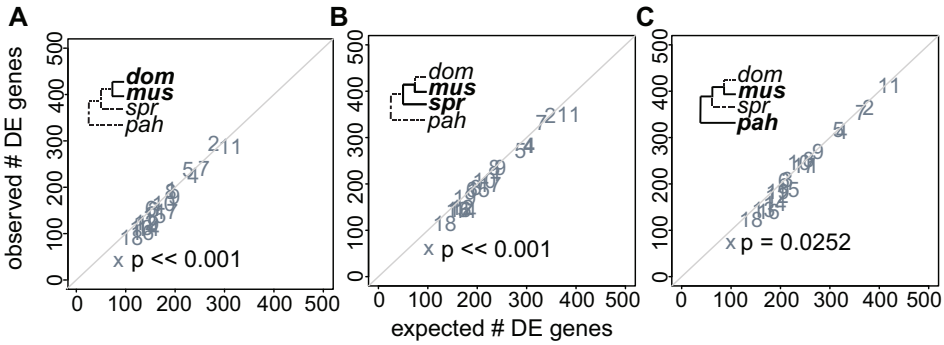

FIG. 4. Observed versus expected number of genes differentially expressed (DE) in late spermatogenesis for three pairwise comparisons at different levels of evolutionary divergence: (*A*) *dom* versus *mus*, (*B*) *spr* versus *mus*, and (*C*) *pah* versus *mus*. Each point represents a different chromosome. The diagonal line is the one-to-one line at which the observed number of DE genes equals the expected number. *P* values are shown for the X chromosome only. They are based on a hypergeometric test for enrichment and corrected for multiple tests using a false discovery rate correction. A significant *P* value indicates that the observed number of DE genes is different from the expected number.

fig. 3*B*), whereas pairwise comparisons gave mixed results, depending on which two lineages were compared (supplementary table S5, Supplementary Material online). There were proportionally fewer differentially expressed genes on the X chromosome (fig. 4, supplementary fig. S7, Supplementary Material online), and this pattern was strongest for the more closely related comparisons (hypergeometric test; *mus* vs. *dom* $P \ll 0.0001$, *spr* vs. *dom* $P \ll 0.0001$, *spr* vs. *mus* $P \ll 0.0001$). Across all metrics of expression divergence and both developmental stages, there was no evidence for pervasive faster-X gene expression level evolution. We also asked if there were differences in the degree of module association for X chromosome and autosomal genes based on WGCNA. X-linked genes tended to have higher eigengene values for the early cell type module (Wilcoxon rank sum test $P \ll 0.001$), but lower values for the late cell type module (Wilcoxon rank sum test $P \ll 0.001$, supplementary fig. S3*B*, Supplementary Material online). Because the X chromosome is repressed during late spermatogenesis, these differences in module association are likely a consequence of overall differences in expression level.

Relative Contributions of *cis*- and *trans*-Regulatory Evolution Vary across Spermatogenesis

Having shown differences in expression divergence between cell types, we next asked if there were differences in the types of regulatory mutations (e.g., *cis*- vs. *trans*-regulatory changes) underlying expression divergence of autosomal genes in each cell type. Note that allele-specific expression cannot be examined for X-linked genes in hemizygous males. We used whole testis (Mack et al. 2016) and FACS-sorted (Larson et al. 2017) data from reciprocal crosses between house mouse subspecies (*dom* × *mus*) to estimate allele-specific expression (ASE) and assign genes to eight different regulatory categories: *cis*, *trans*, *cis* × *trans*, compensatory, *cis* + *trans* opposite, *cis* + *trans* same, other, and conserved (Coolon et al. 2014; Mack et al. 2016).

Across all cell types and genotypes, 50–90% of genes were conserved. Comparing the two spermatogenic stages, we saw striking differences in the proportions of nonconserved genes within each regulatory category (fig. 5, supplementary table S7, Supplementary Material online). *Trans* was more common than *cis* early, whereas *trans* and *cis* made up a similar proportion of regulatory changes late (fig. 5, supplementary table S7, Supplementary Material online). Compensatory changes (compensatory and *cis*+*trans* opposite) were more common than reinforcing (*cis*+*trans* same) in both cell types, but there was a higher relative proportion of reinforcing late (fig. 5, supplementary table S7, Supplementary Material online). Correlated error can lead to an overestimation of compensatory effects in some instances; therefore we verified our result showing a bias towards compensatory changes using a subtraction approach with cross-replicate analysis (Fraser 2019; see supplementary methods for details, Supplementary Material online). We found significant negative correlations between *cis* and *trans* effects, with a trend towards more negative correlations early (early: $r = -0.13$ to -0.16, $P \ll 0.0001$; late: $r = -0.12$ to -0.15, $P \ll 0.0001$). We also asked if genes tended to be assigned to the same regulatory category or switch categories between the two cell types. Overall, most genes assigned to a given regulatory category in one cell type were either not expressed or conserved in the other cell type (supplementary table S8, Supplementary Material online). Of the 1,052 genes that were assigned to a regulatory category in both cell types, 501 remained in the same category and 551 switched categories, indicating that different types of mutations may shift the regulation of the same genes in different cell types.

We focused on results for the *dom* (LEWES)$^{\♀}$ × *mus* (PWK)$^{\♂}$ cross (fig. 5) because these F1 hybrids are more fertile and therefore less likely to have misexpressed genes due to hybrid incompatibilities (Good et al. 2010). However, the subfertile reciprocal hybrids also showed similar overall proportions of genes in each regulatory category. The proportions of different regulatory mechanisms in whole testes were more similar to the late cell type (supplementary table S7, Supplementary Material online), consistent with previous studies showing high overlap in expression profiles between whole testes and spermatid stage cells (Soumillon et al. 2013). We further verified our results using pure strain (LEWES and PWK) expression data from our phylogenetic expression data set to determine differences in parental strain expression

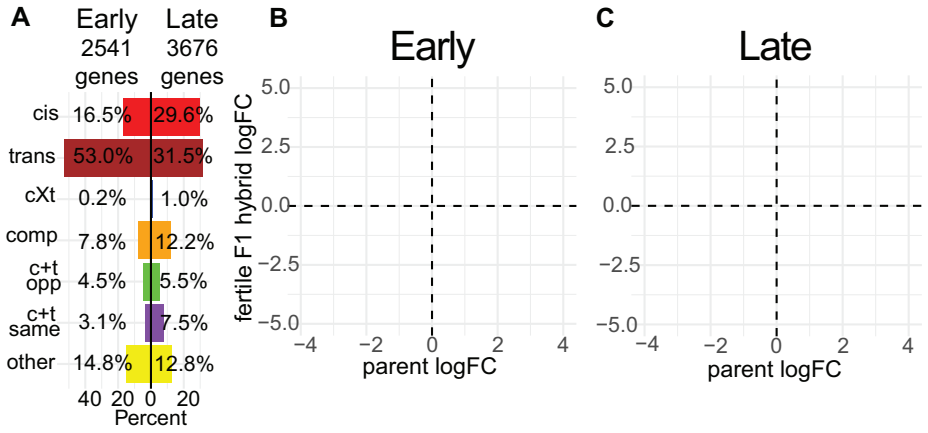

FIG. 5. Regulatory category results for the fertile F1 hybrid (LEWES♀ X PWK♂). (*A*) Percent of nonconserved genes in each regulatory category both early and late. (*B* and *C*) Expression logFC between alleles within the fertile F1 (*y*-axis) plotted against the expression logFC between the parental subspecies (*x*-axis). Each point represents a single gene. Colors correspond to (*A*) and indicate the regulatory category to which that gene was assigned. cXt = *cis* X *trans*; comp = compensatory; c + t opp = *cis* + *trans* opposite; c + t same = *cis* + *trans* same.

levels (supplementary table S7, Supplementary Material online). Finally, we evaluated the relative contributions of regulatory mechanisms contributing to expression differences between strains within each *M. musculus* subspecies using expression data from within-subspecies F1s (WSB X LEWES and CZECHII X PWK) and from the respective parental inbred strains. Consistent with results from the more divergent F1 hybrids, there was more *trans* than *cis* early but some variation depending on subspecies and cross-type (*cis* early: 8–14%, *trans* early: 46–59%, *cis* late: 12–22%, *trans* late: 28–29%; supplementary table S7, Supplementary Material online). In summary, early and late spermatogenesis differed in the types of regulatory mutations contributing to expression divergence, with a proportionally higher contribution of *trans*-regulatory changes early. This pattern was consistent across different degrees of evolutionary divergence and between reciprocal crosses.

cis-Regulatory Changes Tended to Have Larger Effects on Expression Level Divergence

Given that *trans*-regulatory changes were proportionally more common during early spermatogenesis (fig. 5), and that expression levels tended to be more conserved early (fig. 3), we hypothesized that *trans*-regulatory changes would have smaller effect sizes (Coolon et al. 2014; Hill et al. 2021). Consistent with this, genes with *trans* changes showed lower median divergence than those with *cis* changes (fig. 6). We saw higher divergence for reinforcing mutations based on logFC, but not EVE (fig. 6), suggesting that genes with reinforcing changes specific to the *dom* and *mus* comparison may not accumulate more divergence at deeper phylogenetic levels. For the early cell type, 26% of genes in the reinforcing category overlapped with genes that had high pairwise divergence between *dom* and *mus*, whereas only 10–16% of genes in this category overlapped with high divergence genes in other pairwise comparisons (supplementary table S9, Supplementary Material online). Similar patterns were observed for late cell type genes, with 22% of genes in the reinforcing category overlapping those with high divergence between *dom* and *mus* but only 10–14% overlapping with genes showing high divergence in other pairwise comparisons (supplementary table S9, Supplementary Material online). Collectively, *cis*-regulatory changes tended to have larger effects on expression divergence than *trans*-regulatory changes, and reinforcing mutations tended to have large effects on expression divergence between *mus* and *dom*, but not at deeper levels of evolutionary divergence.

Discussion

Developmental stage and context play an important role in shaping the molecular evolution of reproductive genes (Dean et al. 2009; Larson et al. 2016; Finseth and Harrison 2018; Schumacher and Herlyn 2018), with genes expressed in later developmental stages evolving more rapidly (Good and Nachman 2005; Larson et al. 2016). However, comparing gene expression and protein divergence across developmental stages has rarely been done in a phylogenetic framework. In this study, we combined comparative genomics with cell sorting in four species to understand mouse spermatogenesis evolution across a common developmental framework. Our results give insight into how evolution proceeds at different stages of sperm development, at different molecular levels, and on different chromosome types.

Molecular Divergence across Development

There is a long-standing prediction that early developmental stages should be more constrained, with evolutionary divergence gradually increasing across development (Abzhanov 2013), which likely contributes to more rapid molecular evolution during the later stages of sperm development. In addition, the postmeiotic stages are enriched for genes with narrower expression profiles or highly specific biological functions and are therefore expected to experience relaxed pleiotropic constraint (Eddy 2002; Good and Nachman 2005; Green et al. 2018), also motivating our general hypothesis that the postmeiotic round spermatid stage would diverge

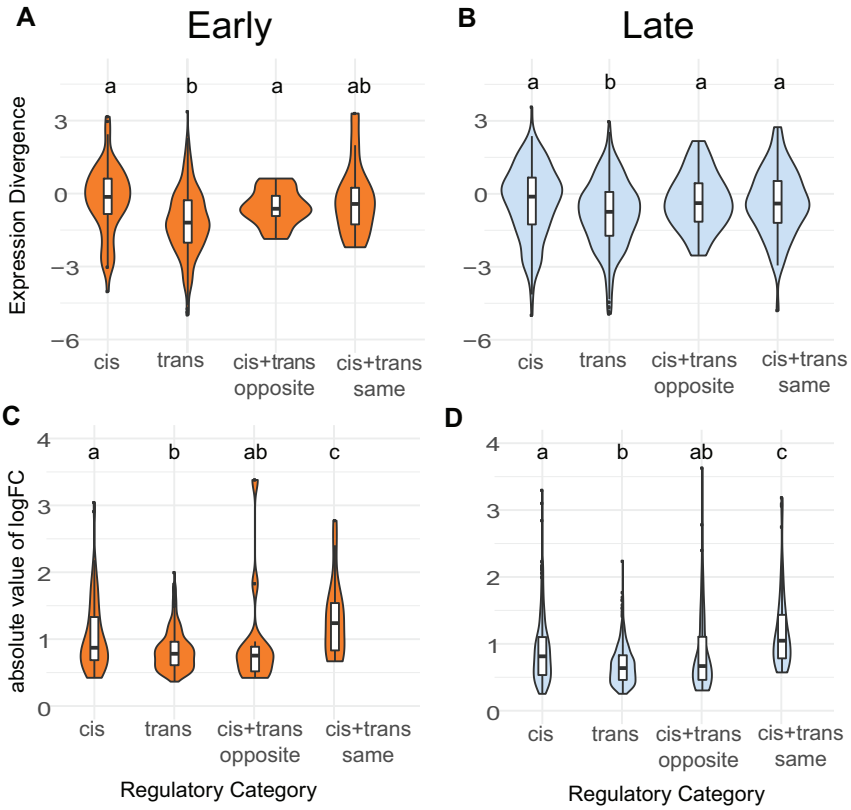

Fig. 6. Expression divergence violin plots by regulatory category for the fertile hybrid. Expression divergence is calculated using the value from EVE (*A* and *B*) and as the absolute value of the logFC in expression between parental subspecies (*C* and *D*). Plots (*A*) and (*C*) correspond to the early cell type and plots (*B*) and (*D*) correspond to the late cell type. Letters indicate significant differences between categories based on a pairwise Wilcoxon rank sum test.

more rapidly. Sexual selection is also likely to be a primary determinant of spermatogenic evolution, but variation in the intensity of sexual selection across spermatogenesis is not well understood (White-Cooper et al. 2009). Sperm competition and cryptic female choice can select for changes in sperm production rate, form, or function, and many aspects of sperm morphology correlate with the intensity of postmating sexual selection (Lüpold et al. 2016; McLennan et al. 2017; Pahl et al. 2018). Rates of mitotic and initial meiotic divisions during early spermatogenesis can control the overall rate of sperm production (Ramm and Schärer 2014). Therefore, selection for increased sperm production likely acts during the development of spermatogonia (diploid mitotic cells; White-Cooper et al. 2009). In contrast, sexual selection shaping the form and function of mature sperm (e.g., sperm swimming speed and fertilization ability) likely acts on later developmental stages such as haploid spermatids (Alavioon et al. 2017). However, many genes involved in mature spermatozoa functions are also highly expressed during early meiosis (da Cruz et al. 2016), suggesting that spermatozoa may be shaped by regulatory networks operating throughout spermatogenesis.

All aspects of molecular evolution that we considered showed more divergence when considering genes induced in late spermatogenesis: lineage-specific expression (fig. 2), protein-coding divergence, and expression level divergence (fig. 3). On first principles, these likely result from a combination of positive selection and relaxed developmental and pleiotropic constraint (Eddy 2002; Swanson and Vacquier 2002; Winter et al. 2004; Good and Nachman 2005; Abzhanov 2013; Green et al. 2018). However, our study was underpowered to formally test for positive selection using likelihood ratio test approaches (Anisimova et al. 2001; Rohlfs and Nielsen 2015). Thus, the relative contributions of positive selection and relaxed constraint to rapid spermatogenesis evolution remain unclear, especially for gene expression phenotypes.

Induced genes provided strong evidence for rapid evolution late, but results were less clear when looking at other genes. Spermatogenesis is a transcriptionally complex process, with most genes in the genome expressed in the testes (Soumillon et al. 2013) and high overlap between genes expressed early and late in our data set (table 1). For protein-coding divergence, we saw more rapid evolution late only when looking at the induced data set, but not when looking at all expressed genes, likely because most genes in our data set were expressed in both cell types. For expression divergence, there was more rapid evolution late even when looking at all expressed genes. This suggests that even genes with broader (i.e., noninduced) expression patterns tended to show more conserved expression early in spermatogenesis.

Testis-specific genes tended to be both induced late and rapidly evolving at the protein-coding level. Testis-specific and male-biased gene sequences often evolve rapidly, which

could be the result of positive selection on genes with specific spermatogenesis functions as well as relaxed constraint because these genes tend to have highly specific functions (Meiklejohn et al. 2003; Baines et al. 2008; Meisel 2011; Parsch and Ellegren 2013). However, we did not see a significant faster late pattern for protein-coding or pairwise expression divergence when looking only at testis-specific genes. Although there were relatively few testis-specific genes, it appears that they tended to be rapidly evolving regardless of which spermatogenesis stage they were expressed in. If generally true, more rapid divergence late in spermatogenesis may partially reflect a higher proportion of testis-specific genes induced in the late cell type (table 1).

In addition to these broad patterns of molecular evolution, we explored the potential functional relevance of rapid divergence for specific genes (supplementary table S10, Supplementary Material online). We detected 20 genes with high (>2.5) EVE divergence in either cell type, and of these 15 were broadly expressed, but five may have specific roles in spermatogenesis (The UniProt Consortium 2020). For example, *Rnf19a* had an EVE value of 4.2 in the late cell type and has a known role in the formation of the sex body, which isolates the sex chromosomes in the nucleus during meiosis, a process that is required for proper spermatogenesis (Párraga and del Mazo 2000) and appears to be disrupted in sterile hybrid mice (Bhattacharyya et al. 2013).

Gene Expression versus Protein-Coding Divergence

Protein-coding changes alter a gene in every tissue and developmental stage in which it is expressed, whereas expression changes have the potential to be more specific (Wray et al. 2003; Carroll 2008). Expression changes, specifically *cis*-regulatory changes, should be less constrained by pleiotropy and may underlie evolutionary changes when purifying selection acts more strongly against protein-coding divergence (Wray et al. 2003; Carroll 2008). Under this model, we might expect to see less pronounced differences in relative expression levels when comparing early versus late stages. However, more recent work has shown that *cis*-regulatory elements such as enhancers can be highly pleiotropic, so *cis*-regulatory changes may be more constrained than once thought (Sabarís et al. 2019; Hill et al. 2021). If gene expression and protein-coding are subject to similar constraints, we would expect them to show similar evolutionary patterns across spermatogenesis, as we observed for autosomal genes (fig. 3).

Interestingly, despite parallel trends in relative divergence across spermatogenesis, expression level divergence and protein-coding divergence were not strongly correlated across genes, suggesting that these two types of molecular changes mostly evolve independently (Khaitovich et al. 2005). Perhaps surprisingly, there was no overlap between genes with very rapid protein-coding divergence ($dN/dS > 1.5$) and high expression divergence (EVE divergence > 2.5). Likewise, only 26 genes with high pairwise expression divergence in at least one comparison (pairwise divergence metric > 1) also had high protein-coding divergence ($dN/dS > 1.5$; supplementary table S10, Supplementary Material online). Whether expression or protein-coding is more rapid for a particular gene may depend on factors such as expression breadth and protein function, but rarely did spermatogenic genes appear to be rapidly evolving for both gene expression and protein sequences.

We also investigated the evolution of lineage-specificity. Testes and sperm tend to be enriched for lineage-specific genes (Brawand et al. 2011) and novel genes (Cridland et al. 2020; Schroeder et al. 2020; Lange et al. 2021). Lineage-specific and novel genes may be common in spermatogenesis because testes are highly transcriptionally active and have a high tissue-specific expression profile, which may allow new genes to arise without disrupting other processes (Levine et al. 2006; Kaessmann 2010; Soumillon et al. 2013; Zhao et al. 2014). We found that late spermatogenesis also had proportionally more lineage-specific genes (fig. 2). Increased lineage-specificity late is consistent with and likely contributed to higher protein and expression level divergence late, as all results suggest that spermatogenesis can tolerate more genetic changes during the late stages without impacting fertility.

X Chromosome Evolution

The X chromosome is predicted to evolve faster than the autosomes because it is hemizygous in males so beneficial recessive mutations will fix more quickly (Charlesworth et al. 1987; Vicoso and Charlesworth 2009). Empirical studies show evidence for a faster-X effect at the protein-coding level in many taxa, particularly for male reproductive genes (Khaitovich et al. 2005; Baines et al. 2008; Meisel and Connallon 2013; Parsch and Ellegren 2013; Larson et al. 2016; but see Whittle et al. 2020). Our data provide strong evidence for faster-X protein-coding evolution for both early and late spermatogenesis, demonstrating that the faster-X effect applies across genes involved in different spermatogenesis stages in mice.

Our results were more complex for expression evolution, with phylogenetic (Rohlfs and Nielsen 2015) and pairwise approaches (Meisel et al. 2012) sometimes yielding contrasting results. In the early cell type, pairwise comparisons supported a faster-X effect, whereas the phylogenetic model did not (fig. 3*B*, supplementary table S5, Supplementary Material online). Correlations between different pairwise divergence values were relatively low on the X chromosome early, suggesting that X-linked genes with high expression level divergence in one pairwise comparison did not tend to have high divergence in other comparisons (table 2). In the late cell type, both phylogenetic and pairwise divergence metrics supported a similar rate of X-linked and autosomal expression evolution (fig. 3*B*, supplementary table S5, Supplementary Material online). It is well-established that lineage-specific changes can create false signatures of rapid divergence in pairwise comparisons (Felsenstein 1985), including in studies of gene expression evolution (Dunn et al. 2018). Thus, our results highlight the importance of accounting for shared evolutionary history when inferring general evolutionary trends (Rohlfs and Nielsen 2015; Dunn et al. 2018).

Overall, our results did not support a faster-X effect for testis gene expression evolution, in contrast to several

previous studies (Khaitovich et al. 2005; Brawand et al. 2011; Meisel et al. 2012). These studies were in other systems and used whole testes samples, which are made up of different cell types, so signals of expression divergence may partially reflect differences in cell type composition rather than true per cell changes in expression levels (Good et al. 2010; Hunnicutt et al. 2021; Yapar E, Saglican E, Dönertaş HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfs RV, Khaitovich P, Somel M, 2021, unpublished data, https://www.biorxiv.org/content/10.1101/010553v2, last accessed July 12, 2021). One previous study used cell type-specific data and found that the X chromosome showed fewer differentially expressed genes during late spermatogenesis between *mus* and *dom* (Larson et al. 2016), and our phylogenetic sampling demonstrates that this result likely applies across mouse species.

Theoretical predictions for the faster-X effect on protein-coding evolution may also apply to gene expression changes, but only for *cis*-regulatory changes or *trans*-regulatory changes where both the causative mutations and affected loci are on the X chromosome (Meisel and Connallon 2013; Larson et al. 2016). The lack of faster-X effect for gene expression could indicate that *trans*-regulatory changes on other chromosomes play an important role in X chromosome spermatogenesis expression evolution. Unfortunately, we are unable to differentiate allele-specific testis expression for X-linked genes in hemizygous males and thus the contribution of *cis*- versus *trans*-regulatory changes remain speculative. Nonetheless, it is plausible that contrasting patterns of expression level and protein sequence divergence on the X chromosome could also reflect the fact that X-linked regulatory phenotypes experience additional constraints during spermatogenesis (Larson et al. 2016). For example, the sex chromosomes undergo MSCI and PSCR, which likely imposes an overall repressive regulatory environment that constrains gene expression levels but not protein-coding changes. Disruption of MSCI and PSCR strongly impairs male fertility, so evolutionary constraints on X chromosome expression during spermatogenesis are expected to be strong (Burgoyne et al. 2009; Good et al. 2010; Larson et al. 2017). These stage-specific mechanisms would not explain lower regulatory divergence early, which we also observed (fig. 3*B*). Overall, our results support the hypothesis that regulatory constraints reduce X-linked expression level divergence during at least some stages of spermatogenesis, while still allowing rapid protein-coding divergence (Larson et al. 2016; Larson, Kopania, et al. 2018). This finding underscores how different components of molecular evolution may experience unique evolutionary pressures that result in distinct patterns of divergence (Brawand et al. 2011; Halligan et al. 2013; Larson et al. 2016).

Regulatory Mechanisms Underlying Expression Divergence

Resolving the relative contributions of *cis*- versus *trans*-acting mutations underlying expression divergence is an important step toward understanding the genetic architecture of expression phenotypes and how different evolutionary forces may act on gene expression (Benowitz et al. 2020; Hill et al. 2021). Although considerable progress has been made in a few key model systems on this important question (Goncalves et al. 2012; Coolon et al. 2014; Mack et al. 2016; Benowitz et al. 2020; Cridland et al. 2020; Sánchez-Ramírez et al. 2021), available data mostly come from whole tissues or organisms. Our results showed that the relative contribution of underlying regulatory mechanisms can differ dramatically between two cell types within a single complex tissue. Genes assigned to a regulatory category in one cell type were often conserved, not expressed, or assigned to a different category in the other cell type, suggesting that most regulatory mutations were cell type-specific in our experiments. This finding supports the hypothesis that regulatory changes may experience less pleiotropic constraint than protein-coding changes, even for genes that are expressed in multiple cell types (Carroll 2008). Although these striking differences are perhaps an expected consequence of different selective pressures acting on cellular function and developmental stage, they also underscore how difficult it is to resolve regulatory phenotypes from complex tissues.

Trans-regulatory changes acting during early development are more likely to cause wide-ranging disruptions to regulatory networks, which are more likely to have detrimental effects on downstream developmental stages. Thus, *trans*-regulatory changes altering expression during early development are predicted to be removed by purifying selection, whereas *cis*-regulatory changes are generally thought to be less pleiotropic and therefore more common in early stages (Carroll 2008; Hill et al. 2021). Based on this simple logic, we predicted that *cis*-regulatory mutations may be proportionally more common in early spermatogenesis, but we found the opposite pattern (fig. 5, supplementary table S7, Supplementary Material online). The relative contributions of *cis*- and *trans*-regulatory changes to expression divergence likely depend on other factors, including a tendency of *cis* mutations to have larger individual effect sizes (Coolon et al. 2014; Hill et al. 2021). We did observe proportionally more *cis*-regulatory changes of large effect during late spermatogenesis (fig. 6*D*) underlying higher overall expression divergence at this stage (fig. 3). Thus, differences in individual effect sizes of *cis*- versus *trans*-acting changes likely play a central role in shaping regulatory evolution across mouse spermatogenesis.

Cis- and *trans*-regulatory mutations can combine to affect the expression of a single gene, either in the same direction (reinforcing) or in opposite directions (compensatory; Goncalves et al. 2012; Coolon et al. 2014; Mack et al. 2016). We observed a higher proportion of compensatory mutations than reinforcing mutations across both spermatogenesis cell types and in whole testes. Even after controlling for correlated error (Fraser 2019), we observed a negative correlation between *cis*- and *trans*-regulatory effects, supporting our result that compensatory mutations were more common than reinforcing mutations. This was expected given that gene expression tends to evolve under stabilizing selection (Rohlfs and Nielsen 2015), and it is consistent with previous studies across many tissue types in mice (Goncalves et al. 2012; Mack et al. 2016), flies (Coolon et al. 2014; Benowitz et al. 2020), and roundworms (Sánchez-Ramírez et al. 2021). We also saw

relatively more reinforcing mutations during postmeiotic spermatogenesis. Reinforcing mutations tended to have a larger effect size based on expression differences (logFC) between *mus* and *dom* (fig. 6*D*), thus large-effect reinforcing changes also likely contribute to higher expression level divergence in late spermatogenesis.

Given the striking differences that we saw between just two cell types, it is likely that complex tissues composed of many cell types may often give different results than isolated cell populations. Consistent with this prediction, our observed proportions of genes in each regulatory category differ from some other published results in house mouse whole tissues (i.e., liver, Goncalves et al. 2012; whole testes, Mack et al. 2016), primarily in that we saw a higher proportion of genes in the *trans* category. We also found some different patterns when reanalyzing whole testes expression data from (Mack et al. 2016) that likely reflect technical differences in the analytical pipelines used between studies (supplementary table S7, see supplementary methods for details, Supplementary Material online). In general, our analysis used more conservative approaches to test for significant DE or ASE. Thus, only genes showing relatively pronounced differences in expression levels between genotypes or alleles were assigned to regulatory mechanisms in our study.

We also found that the relative proportion of *cis*- and *trans*-regulatory changes were similar between whole testes and the late cell type in the fertile F1 hybrid (supplementary table S7, Supplementary Material online), consistent with the observation that postmeiotic spermatids have a disproportionately large contribution to mouse whole testes expression patterns (Hunnicutt et al. 2021). These results suggest that changes in the relative intensities of different selective pressures acting across spermatogenesis not only change the extent of expression level divergence, but also select for different mechanisms of regulatory evolution underlying these expression changes. Given this, analyzing such patterns at the level of whole organisms or tissues seems unlikely to provide a clear understanding of how mechanisms of regulatory evolution proceed in underlying cells. Indeed, even enriched cell populations as we have generated may be limited by relative purities.

By considering both expression divergence across the *Mus* phylogeny and underlying mechanisms of regulatory divergence between two lineages (*mus* and *dom*), our study also provided a novel opportunity to connect different types of regulatory changes to patterns of expression divergence at a deeper phylogenetic scale. Although *trans*-acting changes were relatively common (fig. 5), genes with *cis*-regulatory changes between *mus* and *dom* tended to have higher phylogeny-wide expression divergence than those with *trans*-regulatory changes for both cell types (fig. 6*A*, 6*B*). This suggests that genes showing *cis*-regulatory changes were also more likely to accumulate regulatory differences over time, resulting in phylogeny-wide expression divergence, whereas genes showing *trans*-regulatory changes at relatively shallow evolutionary scales tended to be relatively conserved across the *Mus* phylogeny. Genes with reinforcing changes also had relatively low phylogeny-wide expression level divergence (fig. 6*A* and *B*), in contrast to their high pairwise divergence between *mus* and *dom* (fig. 6*C* and *D*). Genes in this category likely have large-effect, lineage-specific changes in expression that may be under purifying selection over deeper phylogenetic levels. Finally, our phylogenetic contrast revealed rapid expression level divergence late in spermatogenesis. By combining these data with allele-specific expression data, we further showed that *cis*-regulatory changes are likely to underlie this rapid phylogeny-wide expression divergence in late spermatogenesis.

Materials and Methods

Mouse Resources

We investigated gene expression and protein-coding evolution in 12 *Mus musculus domesticus* (*dom*) individuals from four inbred strains (2 BIK/g, 3 DGA, 3 LEWES/EiJ, 4 WSB/EiJ), 8 *M. m. musculus* (*mus*) individuals from three inbred strains (2 CZECHII/EiJ, 3 MBS, 3 PWK/PhJ), 11 *M. spretus* (*spr*) individuals from three inbred strains (5 SEG, 2 SFM, 4 STF), and 3 *M. pahari* (*pah*) individuals from one inbred strain (3 PAHARI/EiJ; fig. 1*B*). By using multiple wild-derived inbred strains of *dom*, *mus*, and *spr*, we sampled natural within-species variation while also having biological replicates of genetically similar individuals. These mice were maintained in breeding colonies at the University of Montana (UM) Department of Laboratory Animal Resources (IACUC protocol 002-13). These colonies were initially established from mice purchased from The Jackson Laboratory, Bar Harbor, ME (CZECHII/EiJ, PWK/PhJ, WSB/EiJ, LEWES/EiJ, PAHARI/EiJ) or acquired from Matthew Dean's colonies at the University of Southern California which were derived from François Bonhomme's stocks at the University of Montpellier, Montpellier, France (MBS, BIK, DGA, STF, SFM, SEG). We weaned males at ~21 days postpartum (dpp) into same sex sibling groups and caged males individually at least 15 days prior to euthanization to avoid dominance effects on testes expression. We euthanized mice at 60–160 dpp by CO_2 followed by cervical dislocation.

For expression data from reciprocal F1 males, we used FACS enriched expression data from (Larson et al. 2017). These data include males from reciprocal F1 crosses between different inbred strains within each *M. musculus* subspecies (*mus*: CZECHII females X PWK males, *dom*: WSB females X LEWES males), as well as reciprocal *mus* and *dom* F1 hybrids (LEWES females X PWK males and PWK females X LEWES males), allowing us to compare results at two different levels of divergence (i.e., within and between lineages). We also analyzed whole testes expression data from (Mack et al. 2016) to compare FACS-enriched cell types to whole testes, including crosses between different strains within each *M. musculus* subspecies (LEWES females X WSB males and PWK females X CZECHII males) and the same reciprocal F1 hybrid crosses to those in (Larson et al. 2017).

Testis Cell Sorting and RNAseq

We collected testes from mice immediately following euthanization and isolated cells at different stages of

spermatogenesis using FACS (Getun et al. 2011). The full FACS protocol is available on GitHub (https://github.com/goodest-goodlab/good-protocols/tree/main/protocols/FACS, last accessed June 16, 2021). Briefly, we decapsulated testes and washed them twice with 1 mg/ml collagenase (Worthington Biochemical), 0.004 mg/ml DNase I (Qiagen), and GBSS (Sigma), followed by disassociation with 1 mg/ml trypsin (Worthington Biochemical) and 0.004 mg/ml DNase I. We then inactivated trypsin with 0.16 mg/ml fetal calf serum (Sigma). For each wash and disassociation step, we incubated and agitated samples at 33 °C for 15 min on a VWR minishaker at 120 rpm. We stained cells with 0.36 mg/ml Hoechst 33324 (Invitrogen) and 0.002 mg/ml propidium iodide, filtered with a 40 μm cell filter, and sorted using a FACSAria IIu cell sorter (BD Biosciences) at the UM Center for Environmental Health Sciences Fluorescence Cytometry Core. We periodically added 0.004 mg/ml DNase I as needed during sorting to prevent DNA clumps from clogging the sorter. We sorted cells into 15 μl beta-mercaptoethanol (Sigma) per 1 ml of RLT lysis buffer (Qiagen) and kept samples on ice whenever they were not in the incubator or the cell sorter. For this study, we focused on two cell populations: early meiotic spermatocytes (leptotene/zygotene) and post-meiotic round spermatids. We extracted RNA using the Qiagen RNeasy Blood and Tissue Kit and checked RNA integrity with a Bioanalyzer 2000 (Agilent) or TapeStation 2200 (Agilent). All samples except one had RIN $\geq$ 7 (supplementary table S11, Supplementary Material online). We prepared RNAseq libraries using the Agilent SureSelect protocol and sequenced samples at the Hudson Alpha Institute for Biotechnology using Illumina NextSeq (75 bp single end). All sample libraries were prepared and sequenced together to minimize batch effects.

Mus Strain Phylogeny

We generated the phylogeny in figure 1*B* using available exome (Chang et al. 2017; Sarver et al. 2017) and whole genome (Keane et al. 2011; Thybert et al. 2018) sequence data (PRJNA326865, PRJNA323493, PRJEB2003, PRJEB14896). Genotypes were based on iterative mapping assemblies relative to the house mouse reference genome (mm10) conducted using *pseudo-it* v3.0 (Sarver et al. 2017) that restricts genotyping to targeted exons. We ran *pseudo-it* with one iteration to generate consensus fasta files for each sample. We then extracted exons, aligned these regions using MAFFT v7.271 (Katoh and Standley 2013), converted to PHYLIP format using AMAS (Borowiec 2016), and inferred a maximum likelihood concatenated tree using IQ-TREE v2.1.4-beta (Nguyen et al. 2015).

Processing of Gene Expression Data

We used R version 3.6.3 and Bioconductor version 3.10 for all analyses. We trimmed raw reads for adaptors and low-quality bases using expHTS (Streett et al. 2015) and mapped trimmed reads with TopHat version 2.1.0 (Kim et al. 2013). Genome assemblies were previously published for all four lineages (Keane et al. 2011; Thybert et al. 2018), allowing us to map reads to the correct assembly and reduce reference bias (Sarver et al. 2017). Mapping rates were consistent across lineages (supplementary table S11, Supplementary Material online). To select orthologous genes among the four lineages, we used BiomaRt (Durinck et al. 2005, 2009) to identify one-to-one Ensembl orthologs and retained only those that were present in all genome assemblies and the mouse reference build GRCm38.

We counted reads using featureCounts and included multiply-mapping reads (Liao et al. 2014). We used edgeR 3.28.1 (Robinson et al. 2010) to normalize expression data, calculate fragments per kilobase per million reads (FPKM), and perform differential expression (DE) analyses. A gene was defined as "expressed" in our data set if it had an FPKM $>$ 1 in at least eight samples. We tested different FPKM cutoffs for considering a gene "expressed" as well as different ways of handling multiply-mapped reads, and our results were consistent across these approaches (supplementary table S4, Supplementary Material online). A gene was expressed in a particular lineage and cell type if it had an FPKM $>$ 1 in all samples of that lineage and cell type. A gene was considered induced in a particular cell type if its median FPKM in that cell type across all lineages was greater than two times its median FPKM in the other cell type across all lineages. We also tested different threshold cutoffs for considering a gene induced. Testis-specific genes were those only expressed in testis based on the mouse tissue expression data from (Chalmel et al. 2007).

We defined lineage-specific genes in two ways. First, we used a log fold-change (logFC) method in which a gene was considered lineage-specific if its median expression level in a lineage was greater than two times its median expression level in any of the other three lineages. We tested different logFC threshold cutoffs ranging from 1.5 to 10 and saw similar results as the logFC $>$ 2 cutoff (supplementary table S1, Supplementary Material online). Second, we used a Bayesian approach to determine if a gene was active or inactive in an expression data set based on transcript levels as implemented with the program Zigzag (Thompson et al. 2020). Genes identified as being active (posterior $P > 0.5$) in one lineage and inactive (posterior $P < 0.5$) in the other lineages were considered lineage-specific. We ran Zigzag twice and only included genes with consistent active or inactive assignments between the two runs. Both the logFC and Zigzag analyses were performed for each cell type, so a gene could be lineage-specific in one cell type but not the other. For each lineage, we determined the proportion of expressed (logFC) or active (Zigzag) genes that were lineage-specific and used a Pearson's χ^2 test to determine if one cell type had greater lineage-specificity than the other. We used the R package topGO with the default algorithm and Fisher's Exact Test to do a gene ontology (GO) enrichment test on lineage-specific genes.

Protein-Coding Divergence

We used the "iqtree-omp" command in IQTree version 1.5.5 (Nguyen et al. 2015) to infer a mouse species tree based on gene trees estimated from the reference sequences for all four mouse lineages (Keane et al. 2011; Thybert et al. 2018). We

took the longest transcript for all one-to-one orthologs and aligned these using MAFFT v7.271 (Katoh and Standley 2013) and converted to PHYLIP format using AMAS (Borowiec 2016). We used a custom script to exclude genes that did not begin with a start codon, had early stop codons, or had sequence lengths that were not multiples of three. We then used the Codeml program in the PAML package to calculate protein-coding divergence and test for positive selection on protein-coding genes (Yang 2007). We used the M0 model to calculate phylogeny-wide dN/dS for each gene, which we report as the overall protein-coding divergence values. We also performed a likelihood ratio test between the M8 and M8a site-based models to test for positive directional selection on each gene (Swanson et al. 2003).

Differential Expression

We performed all analyses of expression level divergence for three different gene sets: expressed genes, induced genes, and testis-specific genes. To calculate expression divergence in a phylogenetic framework, we used the EVE model (Rohlfs and Nielsen 2015), which performs a phylogenetic ANOVA using an Ornstein-Uhlenbeck model to evaluate divergence while controlling for evolutionary relatedness. We report expression divergence from EVE as $-\log(\text{beta}_i)$, where *beta*$_i$ is a metric from EVE that represents the ratio of within-lineage variance to between-lineage evolutionary divergence. By taking the negative log, higher positive numbers correspond to greater evolutionary divergence. We excluded genes with extremely low divergence values [$-\log(\text{beta}_i) < -5$] because this subset did not show a linear relationship between evolutionary divergence and population variance and therefore violated underlying assumptions of the EVE model (supplementary fig. S8, Supplementary Material online).

We also calculated expression divergence in a pairwise framework (Meisel et al. 2012). This method takes the difference in expression level between two lineages and normalizes based on the average expression of the gene in both lineages:

$$D_{a,ij} = \frac{S_{a,i} - S_{a,j}}{(S_{a,i} + S_{a,j})/2}. \quad (1)$$

$D_{a,\,ij}$ is the divergence of gene *a* between lineages *i* and *j*. $S_{a,\,i}$ is the median FPKM of gene *a* in lineage *i*, and $S_{a,\,j}$ is the median FPKM of gene *a* in lineage *j*. We also calculated the logFC in expression between every pairwise comparison of lineages as an additional pairwise divergence metric (Robinson et al. 2010). For the EVE, pairwise divergence, and logFC methods, we compared relative expression divergence between cell types and between the X chromosome and autosomes using a Wilcoxon rank sum test. We tested if certain cell types or chromosome types showed greater correlation among pairwise divergence values using Spearman's rank correlation.

To compare rates of divergence with number of protein–protein interactions, we downloaded publicly available data from the mouse integrated protein–protein interaction reference (MIPPIE, Alanis-Lobato et al. 2020). We used scripts provided by MIPPIE to calculate the number of protein–protein interactions among genes induced early and among genes induced late based on MIPPIE data, only counting interactions with high (> 0.6) MIPPIE scores. We then compared the median number of interactions between early and late genes using a Wilcoxon rank sum test and tested if the number of interactions was correlated with EVE expression divergence or dN/dS protein sequence divergence using Spearman's rank correlation tests. We also tested if groups of genes had higher coexpression network association using a coexpression network analysis implemented in the R package WGCNA (Langfelder and Horvath 2008). We tested if WGCNA modules were associated with cell types or lineages using linear models with posthoc Tukey tests implemented in the R package multcomp. We then used Wilcoxon rank sum tests with FDR-correction for multiple tests to compare gene eigenvalues between the X chromosome and autosomes, and between lineage-specific and nonlineage-specific genes to test if certain groups of genes had higher module associations.

We also compared relative expression divergence on the X chromosome versus the autosomes using the proportion of DE genes on each chromosome (Good et al. 2010; Larson et al. 2016). First, we calculated the proportion of expressed genes that are DE across all autosomes. We then multiplied this proportion by the number of genes expressed on each chromosome to calculate the expected number of DE genes for each chromosome. We plotted the observed number of DE genes against the expected number and used a hypergeometric test to evaluate if each chromosome was over- or under-enriched for DE genes.

Allele-Specific Expression and Regulatory Divergence

We used the modtools and lapels-suspenders pipelines (Huang et al. 2014) to reduce mapping bias and to assign the parental origin of reads in F1 individuals (see supplementary methods for details, Supplementary Material online). This approach requires mapping to pseudogenomes generated using modtools to resolve differences in genome coordinates between different references. We used published pseudogenomes for WSB and PWK, which incorporate single nucleotide variants (SNVs) and indels from these strains into the GRCm38 mouse reference build (Huang et al. 2014). For LEWES and CZECHII, we generated our own pseudogenomes with modtools version 1.0.2 using published VCF files (Morgan et al. 2016; Larson, Vanderpool, et al. 2018). We developed a custom pipeline (see supplementary methods for details, Supplementary Material online) to assign autosomal genes to regulatory categories following previous recommendations (Coolon et al. 2014; Mack et al. 2016; Combs and Fraser 2018; Benowitz et al. 2020). To determine significant differences between cell types, we performed a Pearson's χ^2 test followed by false discovery rate correction for multiple tests.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Michael Nachman and three anonymous reviewers for their helpful comments on an earlier version of this manuscript. We would like to thank Pamela K. Shaw and the UM Fluorescence Cytometry Core supported by an Institutional Development Award from the NIGMS (P30GM103338), the UM Genomics Core supported by the M.J. Murdock Charitable Trust, the UM Lab Animal Resources staff, Gregg Thomas for assistance generating the phylogeny in figure 1B, Nathanael Herrera for mouse photos, and Frank Albert and members of the Good Lab for helpful advice. This work was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (R01-HD073439, R01-HD094787 to JMG). E.E.K.K. was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1313190). E.L.L. was supported by the National Science Foundation (DEB-2012041). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

Author Contributions

J.M.G conceived and funded the project. E.E.K.K., E.L.L., and J.M.G. designed the experiments. E.L.L. and S.K. performed the mouse husbandry and breeding experiments. C.C., S.K., and E.L.L. performed mouse dissections and cell sorts. C.C. and E.L.L. prepared the sequencing libraries. E.E.K.K. analyzed the data. E.E.K.K., E.L.L., and J.M.G. wrote the manuscript with input from all authors.

Data Availability

RNAseq data generated for this project are available through the National Center for Biotechnology Information under accession PRJNA735780. Individual sample accessions are in supplementary table S11, Supplementary Material online. A table of genes in our analyses and whether they were considered expressed, induced, or active in each cell type is available in supplementary table S12, Supplementary Material online. Scripts used for expression divergence and allele-specific expression analyses are available on GitHub: https://github.com/ekopania/mus_spermatogenesis_analyses (last accessed November 18, 2021) and https://github.com/ekopania/cis-trans-pipeline (last accessed September 15, 2021).

References

Abzhanov A. 2013. von Baer's law for the ages: lost and found principles of developmental evolution. *Trends Genet*. 29(12):712–722.

Alanis-Lobato G, Möllmann JS, Schaefer MH, Andrade-Navarro MA. 2020. MIPPIE: the mouse integrated protein–protein interaction reference. *Database*. 2020:baaa035.

Alavioon G, Hotzy C, Nakhro K, Rudolf S, Scofield DG, Zajitschek S, Maklakov AA, Immler S. 2017. Haploid selection within a single ejaculate increases offspring fitness. *Proc Natl Acad Sci U S A*. 114(30):8053–8058.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*. 18(8):1585–1592.

Baines JF, Harr B. 2007. Reduced X-linked diversity in derived populations of house mice. *Genetics* 175(4):1911–1921.

Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008. Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. *Mol Biol Evol*. 25(8):1639–1650.

Benowitz KM, Coleman JM, Allan CW, Matzkin LM. 2020. Contributions of *cis*- and *trans*-regulatory evolution to transcriptomic divergence across populations in the *Drosophila mojavensis* larval brain. *Genome Biol Evol*. 12(8):1407–1418.

Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, Simecek P, Forejt J. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc Natl Acad Sci U S A*. 110(6):E468–E477.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660.

Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478(7369):343–348.

Burgoyne PS, Mahadevaiah SK, Turner JMA. 2009. The consequences of asynapsis for mammalian meiosis. *Nat Rev Genet*. 10(3):207–216.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J-J, Wolgemuth DJ, Jégou B et al. 2007. The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A*. 104(20):8346–8351.

Chang PL, Kopania E, Keeble S, Sarver BAJ, Larson E, Orth A, Belkhir K, Boursot P, Bonhomme F, Good JM et al. 2017. Whole exome sequencing of wild-derived inbred strains of mice improves power to link phenotype and genotype. *Mamm Genome*. 28(9–10):416–425.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat*. 130(1):113–146.

Chevret P, Veyrunes F, Britton-Davidian J. 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biol J Linn Soc*. 84(3):417–427.

Combs PA, Fraser HB. 2018. Spatially varying cis-regulatory divergence in *Drosophila* embryos elucidates cis-regulatory logic. *PLoS Genet*. 14(11):e1007631.

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res*. 24(5):797–808.

Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. 2015. Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the *Drosophila* X chromosome. *Mol Biol Evol*. 32(10):2605–2615.

Cridland JM, Majane AC, Sheehy HK, Begun DJ. 2020. Polymorphism and divergence of novel gene expression patterns in *Drosophila melanogaster*. *Genetics* 216(1):79–93.

Cutter AD, Bundus JD. 2020. Speciation and the developmental alarm clock. *eLife* 9:e56276.

da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, Farías J, Curti G, Capoano CA, Folle GA, Benavente R, Sotelo-Silveira JR, Geisinger A. 2016. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics*. 17:294.

Dean MD, Ardlie KG, Nachman MW. 2006. The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). *Mol Ecol*. 15(13):4141–4151.

Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW. 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol Biol Evol*. 26(8):1733–1743.

Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A. 2018. Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proc Natl Acad Sci U S A*. 115(3): E409–E417.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21(16):3439–3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 4(8):1184–1191.

Eddy EM. 2002. Male germ cell gene expression. *Recent Prog Horm Res*. 57:103–128.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125(1):1–15.

Finseth FR, Harrison RG. 2018. Genes integral to the reproductive function of male reproductive tissues drive heterogeneity in evolutionary rates in Japanese Quail. *Genes Genom Genet*. 8:39–51.

Firman RC. 2020. Of mice and women: advances in mammalian sperm competition with a focus on the female perspective. *Philos Trans R Soc Lond B Biol Sci*. 375(1813):20200082.

Fraser HB. 2019. Improving estimates of compensatory cis–trans regulatory divergence. *Trends Genet*. 35(1):3–5.

Getun IV, Torres B, Bois PRJ. 2011. Flow cytometry purification of mouse meiotic cells. *J Vis Exp*. (50):2602.

Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. 2012. Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Res*. 22(12):2376–2384.

Good JM, Giger T, Dean MD, Nachman MW. 2010. Widespread overexpression of the X chromosome in sterile F_1 hybrid mice. *PLoS Genet*. 6(9):e1001148.

Good JM, Nachman MW. 2005. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol Biol Evol*. 22(4):1044–1052.

Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB et al. 2018. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev Cell*. 46(5):651–667.e10.

Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet*. 9(12):e1003995.

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci U S A*. 112(14):4393–4398.

Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet*. 22(4):203–215.

Huang S, Holt J, Kao C-Y, McMillan L, Wang W. 2014. A novel multi-alignment pipeline for high-throughput sequencing data. *Database* 2014(0):bau057.

Hunnicutt KE, Good JM, Larson EL. 2021. Unraveling patterns of disrupted gene expression across a complex tissue. *Evolution*. Available from: https://doi.org/10.1111/evo.14420.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 20(10):1313–1326.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309(5742):1850–1854.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14(4):R36.

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

Kousathanas A, Halligan DL, Keightley PD. 2014. Faster-X adaptive protein evolution in house mice. *Genetics* 196(4):1131–1143.

Lange A, Patel PH, Heames B, Damry AM, Saenger T, Jackson CJ, Findlay GD, Bornberg-Bauer E. 2021. Structural and functional characterization of a putative de novo gene in *Drosophila*. *Nat Commun*. 12(1):1667.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 9(1):559.

Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in *Drosophila*. *Trends Genet*. 24(3):114–123.

Larson EL, Keeble S, Vanderpool D, Dean MD, Good JM. 2017. The composite regulatory basis of the large X-effect in mouse speciation. *Mol Biol Evol*. 34(2):282–295.

Larson EL, Kopania EEK, Good JM. 2018. Spermatogenesis and the evolution of mammalian sex chromosomes. *Trends Genet*. 34(9):722–732.

Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BAJ, Smith AD, Dean MD, Good JM. 2016. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics* 203(4):1841–1857.

Larson EL, Vanderpool D, Sarver BAJ, Callahan C, Keeble S, Provencio LP, Kessler MD, Stewart V, Nordquist E, Dean MD et al. 2018. The evolution of polymorphic hybrid incompatibilities in house mice. *Genetics* 209(3):845–859.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A*. 103(26):9935–9939.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930.

Lüpold S, Manier MK, Puniamoorthy N, Schoff C, Starmer WT, Luepold SHB, Belote JM, Pitnick S. 2016. How sexual selection can drive the evolution of costly sperm ornamentation. *Nature* 533(7604):535–538.

Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res*. 26(4):451–461.

McKee BD, Handel MA. 1993. Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* 102(2):71–80.

McLennan HJ, Lüpold S, Smissen P, Rowe KC, Breed WG. 2017. Greater sperm complexity in the Australasian old endemic rodents (Tribe: hydromyini) is associated with increased levels of inter-male sperm competition. *Reprod Fertil Dev*. 29(5):921–930.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*. 100(17):9894–9899.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Mol Biol Evol*. 28(6):1893–1900.

Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. *Trends Genet*. 29(9):537–544.

Meisel RP, Malone JH, Clark AG. 2012. Faster-X evolution of gene expression in *Drosophila*. *PLoS Genet*. 8(10):e1003013.

Morgan AP, Didion JP, Doran AG, Holt JM, McMillan L, Keane TM, Pardo-Manuel de Villena F. 2016. Genome report: whole genome sequence of two wild-derived *Mus musculus domesticus* inbred strains, LEWES/EiJ and ZALENDE/EiJ, with different diploid numbers. *G3*. 6(12):4211–4216.

Namekawa SH, Park PJ, Zhang L-F, Shima JE, McCarrey JR, Griswold MD, Lee JT. 2006. Postmeiotic sex chromatin in the male germline of mice. *Curr Biol*. 16(7):660–667.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.

Pahl T, McLennan HJ, Wang Y, Achmadi AS, Rowe KC, Aplin K, Breed WG. 2018. Sperm morphology of the Rattini - are the interspecific differences due to variation in intensity of intermale sperm competition? *Reprod Fertil Dev*. 30(11):1434–1442.

Párraga M, del Mazo J. 2000. XYbp, a novel RING-finger protein, is a component of the XY body of spermatocytes and centrosomes. *Mech Dev*. 90(1):95–101.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet*. 14(2):83–87.

Phifer-Rixey M, Nachman MW. 2015. Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife*. 4:e05959.

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 Genet*. 9(4):e1003476.

Pitnick S, Hosken DJ, Birkhead TR. 2009. Sperm morphological diversity. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm biology. London: Academic Press. p. 69–149.

Ramm SA, Schärer L. 2014. The evolutionary ecology of testicular function: size isn't everything. *Biol Rev*. 89(4):874–888.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

Rohlfs RV, Nielsen R. 2015. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst Biol*. 64(5):695–708.

Sabarís G, Laiker I, Preger-Ben Noon E, Frankel N. 2019. Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet*. 35(6):423–433.

Sánchez-Ramírez S, Weiss JG, Thomas CG, Cutter AD. 2021. Widespread misregulation of inter-species hybrid transcriptomes due to sex-specific and sex-chromosome regulatory evolution. *PLoS Genet*. 17(3):e1009409.

Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. 2017. Phylogenomic insights into mouse evolution using a pseudoreference approach. *Genome Biol Evol*. 9(3):726–739.

Schroeder CM, Valenzuela JR, Mejia Natividad I, Hocky GM, Malik HS. 2020. A burst of genetic innovation in *Drosophila* actin-related proteins for testis-specific function. *Mol Biol Evol*. 37(3):757–772.

Schumacher J, Herlyn H. 2018. Correlates of evolutionary rates in the murine sperm proteome. *BMC Evol Biol*. 18(1):35.

Skinner BM, Johnson EEP, Bacon J, Affara NA, Rathje CC, Yousafzai G, Ellis PJI, Larson EL, Kopania EEK, Good JM. 2019. A high-throughput method for unbiased quantitation and categorisation of nuclear morphology. *Biol Reprod*. 100(5):1250–1260.

Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*. 3(6):2179–2190.

Streett DA, Petersen KR, Gerritsen AT, Hunter SS, Settles ML. 2015. expHTS: analysis of high throughput sequence data in an experimental framework. In: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics. Atlanta (GA): Association for Computing Machinery. p. 523–524. Available from: https://doi.org/10.1145/2808719.2811442.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol*. 20(1):18–20.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet*. 3(2):137–144.

The UniProt Consortium. 2020. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 49:D480–D489.

Thompson A, May MR, Moore BR, Kopp A. 2020. A hierarchical Bayesian mixture model for inferring the expression state of genes in transcriptomes. *Proc Natl Acad Sci U S A*. 117(32):19339–19346.

Thybert D, Roller M, Navarro FCP, Fiddes I, Streeter I, Feig C, Martin-Galvez D, Kolmogorov M, Janoušek V, Akanni W et al. 2018. Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res*. 28(4):448–459.

Turner LM, Chuong EB, Hoekstra HE. 2008. Comparative analysis of testis protein evolution in rodents. *Genetics* 179(4):2075–2089.

Vicens A, Borziak K, Karr TL, Roldan ERS, Dorus S. 2017. Comparative sperm proteomics in mouse species with divergent mating systems. *Mol Biol Evol*. 34(6):1403–1416.

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

Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B. 2007. Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Res*. 17(1):42–49.

White-Cooper H, Doggett K, Ellis RE. 2009. The evolution of spermatogenesis. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm biology. London: Academic Press. p. 151–183.

Whittle CA, Kulkarni A, Extavour CG. 2020. Absence of a faster-X effect in beetles (*Tribolium*, Coleoptera). *G3 (Bethesda)* 10:1125–1136.

Winter EE, Goodstadt L, Ponting CP. 2004. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res*. 14(1):54–61.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 20(9):1377–1419.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.

Zhao L, Saelao P, Jones CD, Begun DJ. 2014. Origin and spread of de novo genes in *Drosophila melanogaster* populations. *Science* 343(6172):769–772.

Supplementary Methods

Regulatory categories: Because the publicly available PWK pseudogenome is based on a VCF that is much older than the CZECHII VCF, the median quality scores are much higher in the CZECHII VCF which causes mapping bias when mapping CZECHII X PWK individuals. To address this, we called variants from publicly available PWK sequence reads using the Genome Analysis Tool Kit (GATK) version 4.1.7.0 program HaplotypeCaller. We also downsampled from the CZECHII sequence reads to match the read counts available for PWK and generated a new CZECHII VCF using HaplotypeCaller. We used modtools to generate new pseudogenomes for PWK and CZECHII based on these modified VCF files and only used these pseudogenomes for mapping CZECHII X PWK individuals.

We trimmed raw reads using trimmomatic version 0.35 (Bolger, et al. 2014) and mapped reads using TopHat v2.1.1 (Kim, et al. 2013), consistent with (Mack, et al. 2016). After mapping reads to both parent pseudogenomes, we converted coordinates to match the mouse reference build GRCm38 using lapels (pylapels version 0.2.0 in Lapels 1.1.1). We then used suspenders (pysuspenders version 0.2.5 in Suspenders 0.2.5), which merges lapels outputs from mappings to the two different parents and assigns reads to either parent based on sequence variants and mapping quality scores. The *mus* and *dom* subspecies are closely related and have a D_{xy} of about 0.5% (Geraldes, et al. 2008), which means most reads will map equally well to both parents. For F1 hybrids, we excluded reads that mapped equally well to both parents. For parents, we mapped reads to both pseudogenomes and only kept reads that were assigned to the correct parent. For example, we would map a PWK sample to both

PWK and LEWES, run it through the full lapels-suspenders pipeline, and only keep reads that were assigned to PWK. This ensured that both F1 and parent data were treated the same and removed genes that were DE between the parents but could not be evaluated for ASE due to a lack of variants. Because these data included a combination of paired-end and single-end data, we removed the second read from all pairs in which both reads mapped, and then converted all SAM flags to single end flags. We then downsampled reads from these parent bam files such that both parents had a similar number of reads to the mean number of reads assigned to each F1 allele (Coolon, et al. 2014; Mack, et al. 2016). This gave us similar power to detect DE and ASE. Hybrids were never downsampled as this could bias *cis* versus *trans* results for a given gene by randomly keeping more reads from one allele or the other. Read counts were similar between the two cell types, so differences in power between the two cell types are unlikely to affect our results. We counted the number of reads mapping to each gene with HTSeq-count (Anders, et al. 2014). (supplementary fig. S9)

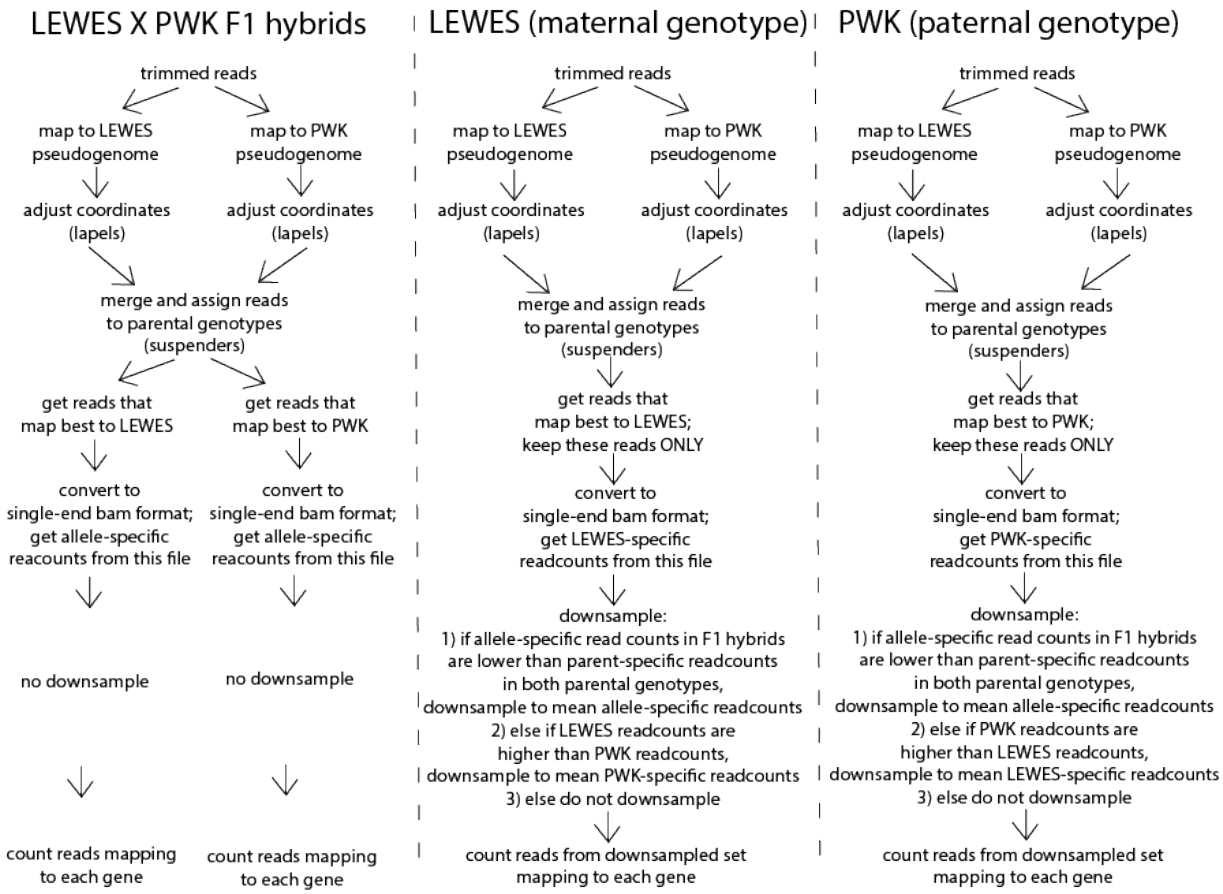

Fig. S9. An example of the read mapping, lapels-suspenders, and downsampling pipeline used to assess regulatory categories.

Our custom pipeline to assign genes to general regulatory categories as defined in previous studies (Coolon, et al. 2014; Mack, et al. 2016; Combs and Fraser 2018; Benowitz, et al. 2020) was as follows. First, we used a negative binomial test implemented in edgeR to determine if a gene is DE between the parental genotypes. We then also used the edgeR negative binomial model to test if the gene showed ASE within the F1 hybrids, similar to previous studies (Combs and Fraser 2018; Benowitz, et al. 2020). We ran edgeR with the calcNormFactors() function, which is generally recommended to account for overrepresentation of highly expressed genes in RNAseq

datasets. However, it is possible that this normalization can falsely remove signatures of ASE from data, so we re-ran our analyses without the calcNormFactors() function and saw nearly identical proportions of genes assigned to each regulatory category (supplementary Table S13). We tested for ASE without blocking by individual because we used crosses from inbred strains that should be nearly genetically identical and therefore should have very little variation among individuals (edgeR model: ~0+allele). We also wanted to treat our DE analyses between parents and ASE analyses in F1s as similarly as possible, and blocking by individual is not possible for parents because each parent represents only one allele. However, we did observe some individual variation in expression levels, so we repeated our analyses with blocking by individual to determine ASE (edgeR model: ~individual+group). With blocking, we observed some changes in the proportions of genes assigned to each regulatory category, but our major result of proportionally more *trans*-regulatory changes in early spermatogenesis held even though it was less pronounced (supplementary Table S13).

We used a Fisher's Exact Test (FET) to determine if the expression difference between the parents was significantly different from the allelic expression difference within the F1 hybrids (Coolon, et al. 2014; Mack, et al. 2016; Benowitz, et al. 2020). If a gene was not DE between the parental genotypes and showed no evidence for ASE in the F1 hybrids, it was treated as conserved. If a gene was not DE, but had ASE with a significant FET, it was assigned to the compensatory category. Genes that were DE but had no ASE with a significant FET were assigned to the *trans* category. A gene with both DE and ASE and a non-significant FET was considered regulated in *cis*. Genes with both DE and ASE plus a significant FET presumably had some combination of *cis* and *trans*

mutations acting on gene expression. We further broke this category down in the following ways. Genes were assigned to the *cis* X *trans* category if DE and ASE were in opposite directions (e.g., the gene was more highly expressed in the maternal parent, but the paternal allele was more highly expressed in the F1 hybrid). Genes were assigned to the *cis* + *trans* same category if DE and ASE were in the same direction and the expression difference was greater between the parental genotypes. This category represents reinforcing mutations, or *cis* and *trans* acting in the same direction. Genes were assigned to the *cis* + *trans* opposite category if DE and ASE were in the same direction and the expression difference was greater in the F1 hybrid. This category represents weak compensatory mutations, where *cis* and *trans* are acting in opposite directions but have not fully compensated each other because the gene is still DE between the parent lineages. Note that we could only assign autosomal genes to regulatory categories, because hemizygosity in males prevented us from determining ASE on the X or Y chromosomes. (supplementary fig. S10)

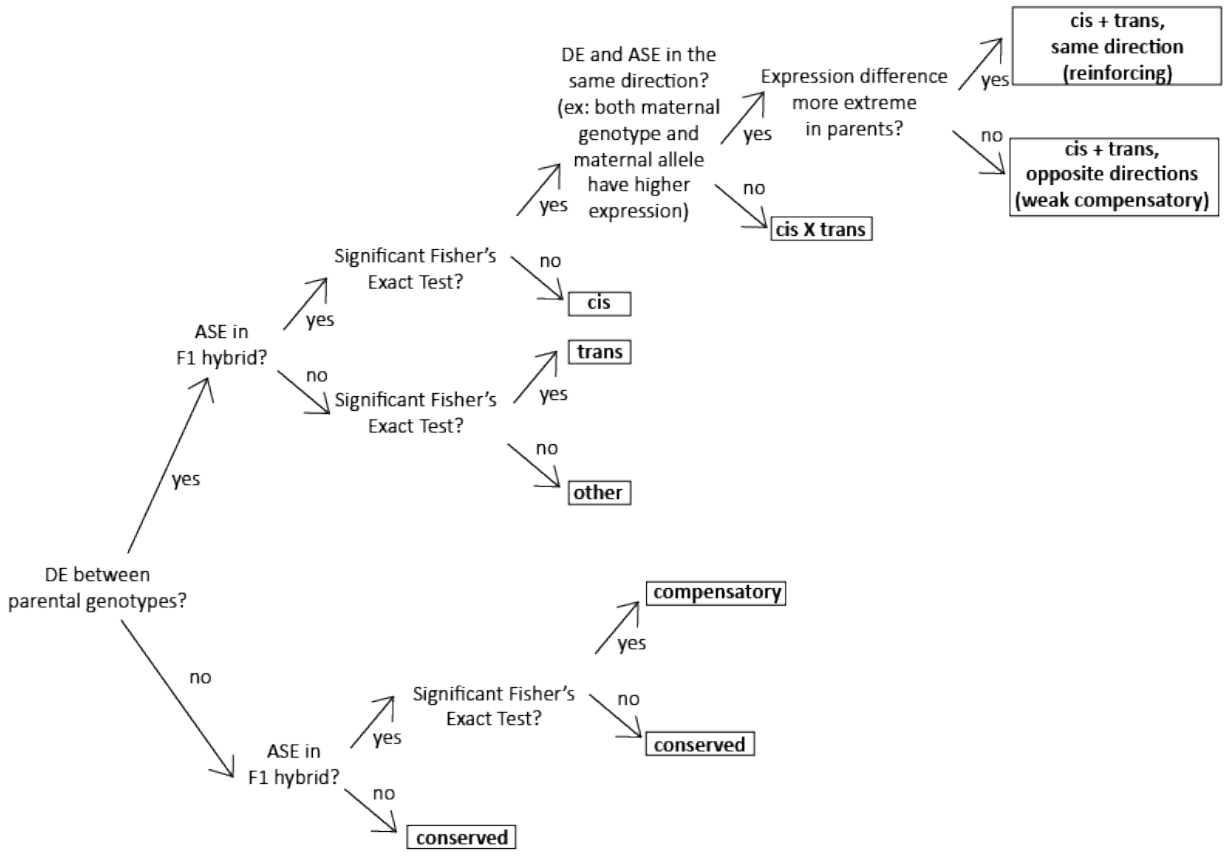

Fig. S10. Flow chart showing the pipeline used to assign genes to regulatory categories.

For each F1 cross, we used expression data from the parental inbred strains as the "parent" in our regulatory category analyses. For the hybrid F1s, we repeated the analyses using data from within-subspecies F1s as the "parent" to compare the effects of having a homozygous parent vs a heterozygous parent. For example, we performed the analyses with LEWES♀ X PWK♂ F1 hybrids using LEWES and PWK as the parents, and then repeated the analysis using data from the same F1 hybrids and using data from WSB♀ X LEWES♂ and CZECHII♀ X PWK♂ F1 mice as the parents. We report

results based on within-subspecies F1 parents in the main text, and report all results in (supplementary table S7).

Cross-replicate analysis (Fraser 2019): Correlated error between *cis* and *trans* regulatory changes can lead to an overestimation of compensatory effects, and one way to address this issue is cross-replicate analysis. Cross-replicate analysis does not work well with our approach for detecting ASE, because it would require running edgeR with only one sample at a time and therefore no biological replicates. Therefore, we tested if compensatory effects are likely to be common in our dataset using a subtraction approach to estimate *cis* and *trans* effects. In this method, *cis* effects are the logFC between allelic expression levels and *trans* effects are the difference between parental logFC and *cis* effects. We calculated logFC between parental FPKM averaged across all samples and logFC between alleles within individual F1 samples. When we tested for correlations between *cis* and *trans* effects calculated using the same sample, we saw strong negative correlations in both cell types (early: r = -0.41 to -0.44, P << 0.0001; late: r = -0.38 to -0.39, P << 0.0001). When we tested for correlations using *trans* effects calculated using a different sample than was used to calculate *cis* effects (i.e., cross-replicate analysis) we still saw negative correlations although they were weaker (early: r = -0.13 to -0,16, P << 0.0001; late: r = -0.12 to -0.15 P << 0.0001). This suggests that not all compensatory effects we observed were due to correlated error, and that there were more compensatory than reinforcing changes in our dataset. There is also a slight trend towards more negative correlations early, both with and without cross-replicate analysis, consistent with our results showing more compensatory changes early than late.

Reanalysis of whole-testis data from Mack, et al. (2016): Our pipeline assigned many more genes to the *trans* category and many fewer genes to the *cis* X *trans*

category than reported previously (Mack, et al. 2016; supplementary table S7). To investigate the reason for this potential inconsistency, we also reanalyzed these whole testes data using a binomial test following Mack, et al. (2016), which should be more sensitive (i.e., less conservative) than a negative binomial approach for detecting expression differences. The binomial test gave results more similar to those reported by Mack, et al. (2016), with proportionally fewer genes assigned to the *trans* category and a higher proportion of genes in the *cis* X *trans* category.

We also explored gene-level differences in category assignment between the two approaches and found large groups of genes that were assigned to one category using the binomial approach that were then consistently assigned to a different category using the negative binomial approach (supplementary table S14). For example, many genes assigned to *cis*+*trans* (same direction) using the binomial test were categorized as *trans* using the negative binomial test. These relatively subtle differences make sense because genes with significant expression divergence between parents and less extreme expression divergence between alleles in the F1s will be assigned to one of these two categories. The key distinction is that genes in the *trans* category do not show a significant difference in ASE, and therefore, our more conservative method for considering alleles to be differentially expressed in F1s will likely assign more genes to the *trans* category (supplementary fig. S10). Other common changes in category assignment between the binomial and negative binomial approaches are consistent with a more conservative method for calling genes DE or ASE (supplementary table S14, supplementary fig. S10).

However, we note that our more conservative analytical method likely explains most, but not all, of the quantitative differences between our results and those from (Mack, et al. 2016). Other inconsistencies likely result from differences in how F1 sequencing reads were bioinformatically assigned to parents. Although the general conceptual frameworks were similar, Mack, et al. (2016) used a custom script to assign reads, while we used the lapels and suspenders pipeline from modtools (Holt, et al. 2013; Huang, et al. 2014). The number of reads assigned to each parent were similar across both studies for most samples, but there were some notable differences and individual sample outliers that may have contributed to differences in regulatory category assignment for some genes (supplementary table S15).

Supplementary Figures

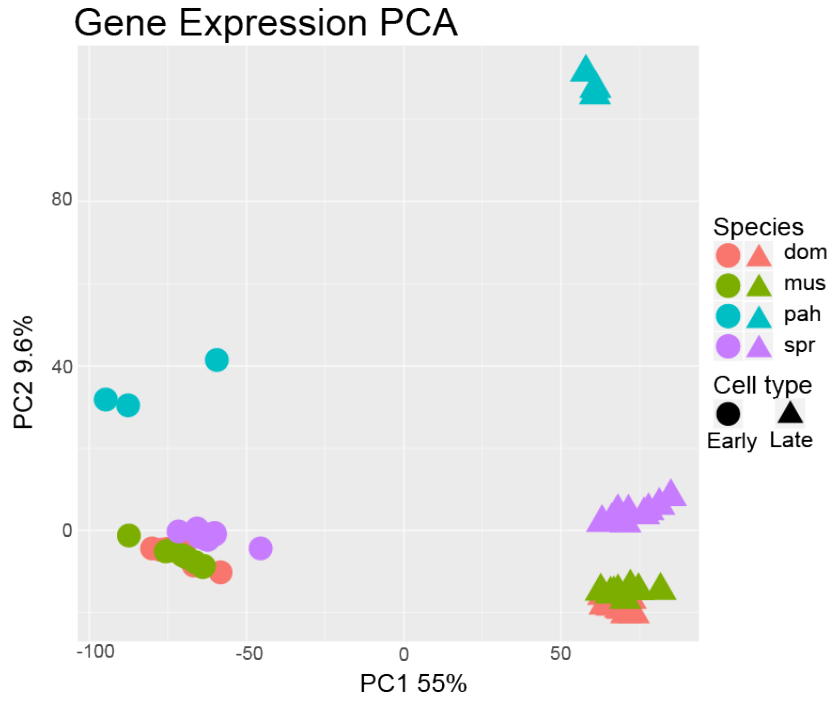

Fig. S1. Principal component analysis based on expression levels of genes expressed in either cell type. Circles represent the early cell type and triangles represent the late cell type. Colors correspond to different lineages. Cell type explains most of the variance (PC1, 55%) followed by lineage (PC2, 9.6%).

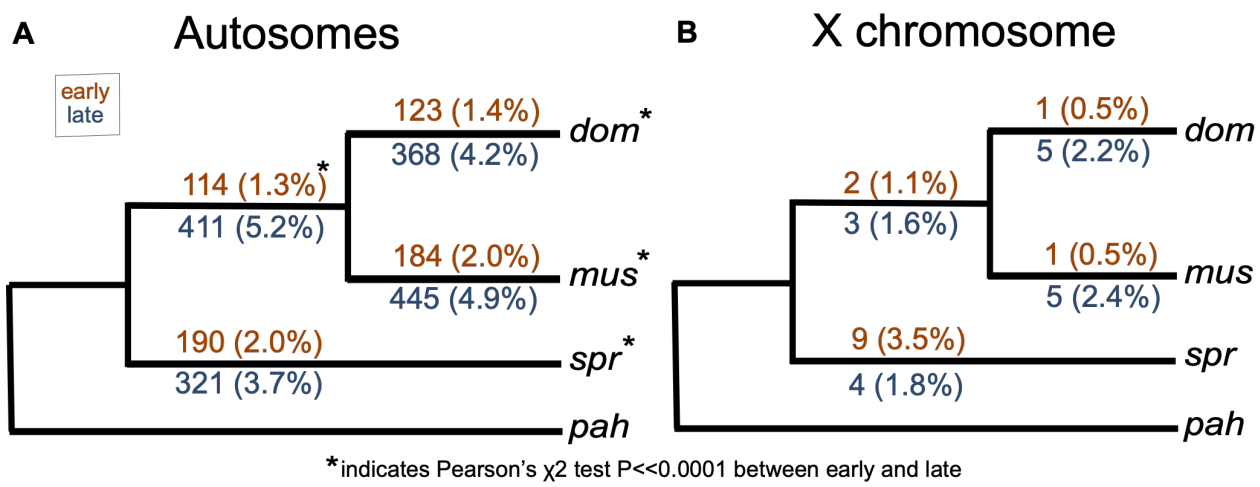

Fig. S2. Number of genes that are lineage-specific on each internal branch of the mouse phylogeny used in this study based on a logFC approach. Numbers in parentheses are the percent of active genes that are lineage-specific. Results are presented separately for the autosomes (*A*) and X chromosome (*B*). Orange values above each branch represent the early cell type and blue values below represent the late cell type. Asterisks indicate a significant difference between early and late on that branch based on a Pearson's χ^2 test.

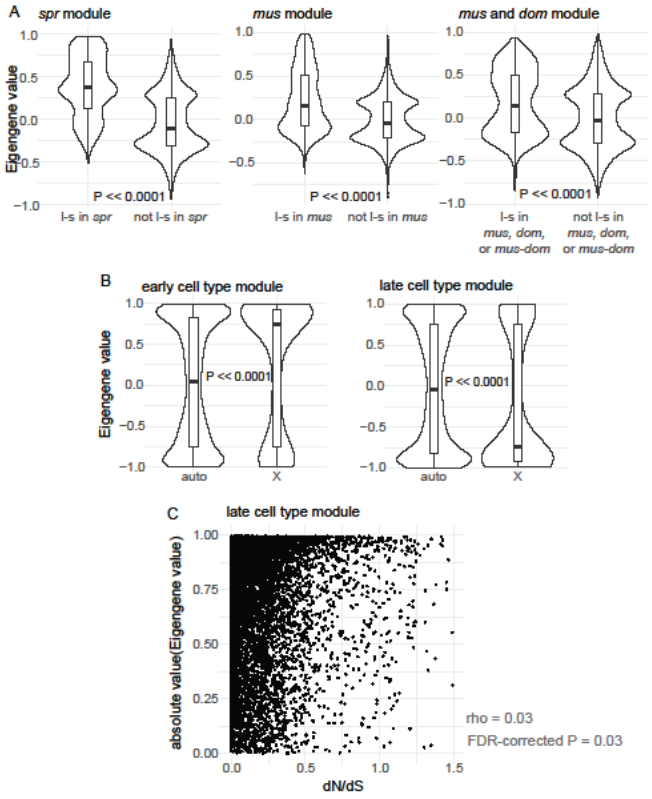

Fig. S3. Co-expression network analysis. (*A*) Eigengene value, a measure of module connectivity, plotted for lineage-specific genes. Modules are named for the lineages they were significantly associated with, based on linear models and post-hoc Tukey tests (FDR-corrected P < 0.05). (*B*) Eigengene value plotted for autosomal and X chromosome genes in the modules significantly associated with cell type based on linear models and post-hoc Tukey tests (FDR-corrected P < 0.05). (*C*) The absolute value of eigengene value for the late cell type module plotted against dN/dS. Rho and P-value are based on a Spearman's rank correlation test. l-s = lineage-specific

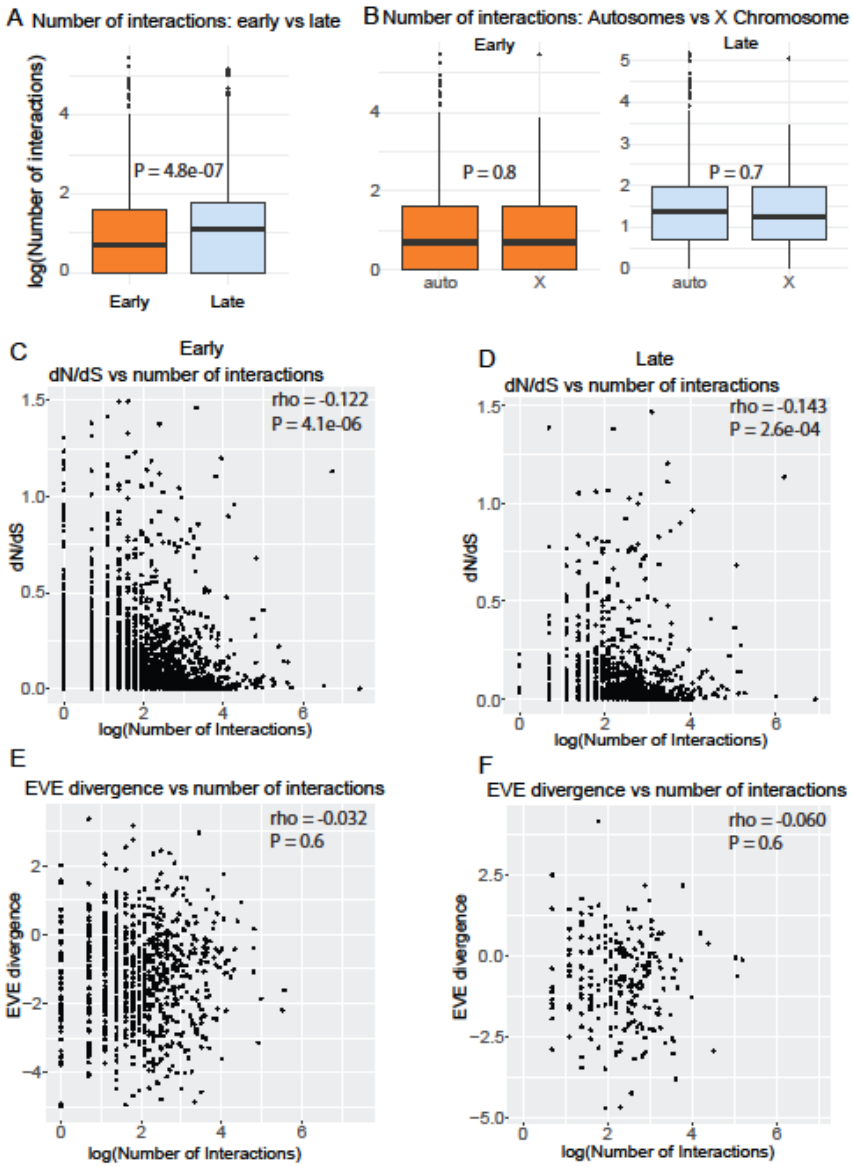

Fig. S4. Mouse protein-protein interactions in spermatogenesis cell types. A and B show comparisons in the number of interactions between cell types (A) or chromosome types (B). P-values are based on FDR-corrected Wilcoxon rank sum tests. C-F show correlations between the number of interactions and dN/dS (C and D) or EVE expression divergence (E and F). P-values are based on FDR-corrected Spearman's rho tests for correlation.

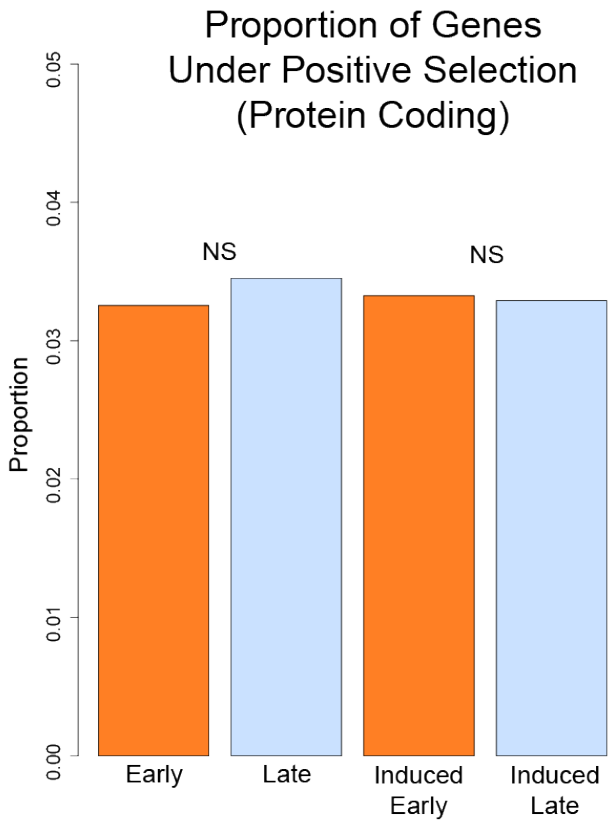

Fig. S5. Proportion of genes expressed or induced in each cell type under positive selection at the protein-coding level. NS = not significantly different (Pearson's χ^2 P > 0.05 after FDR correction)

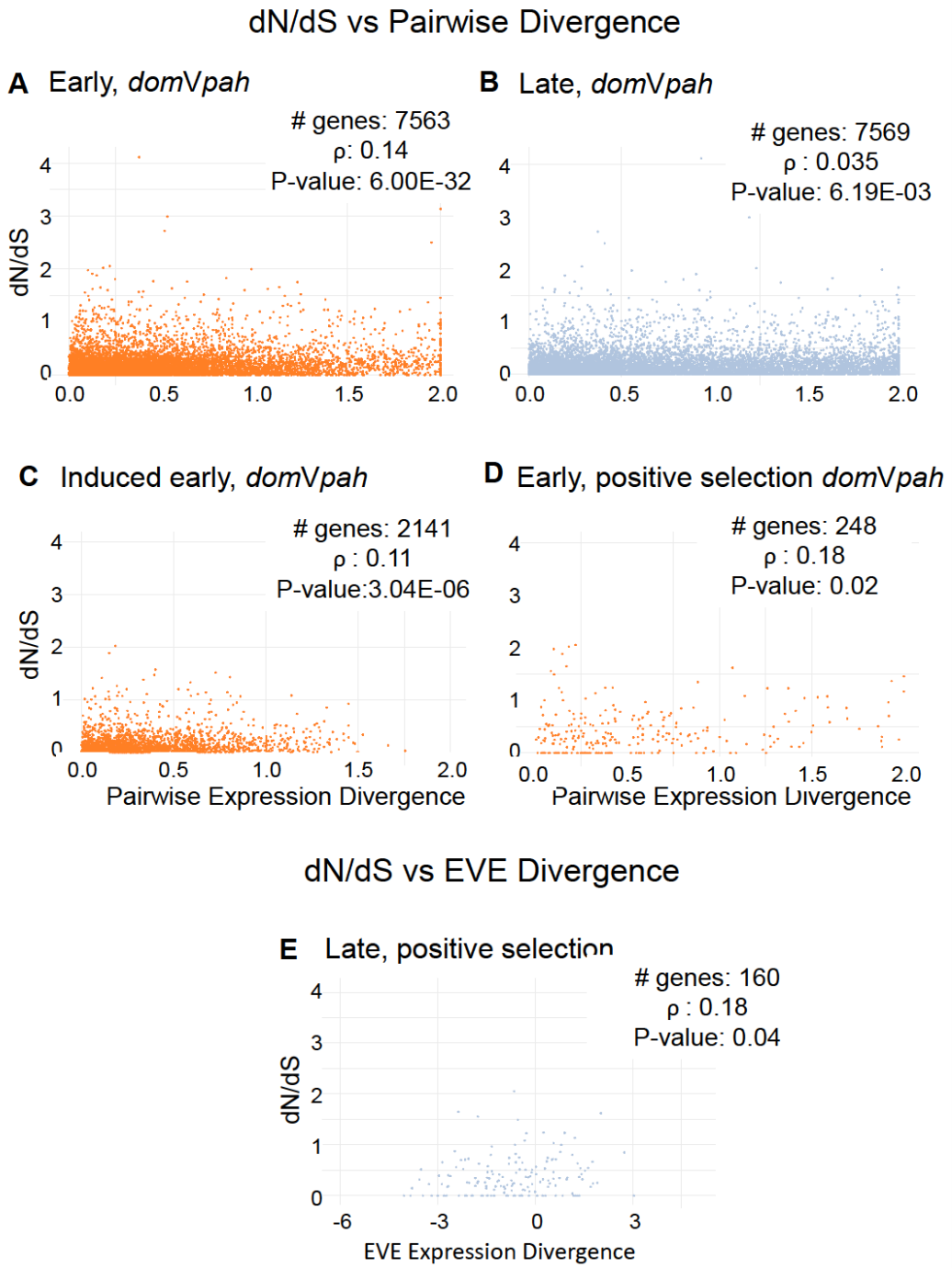

Fig. S6. Some examples showing weak positive correlations between protein-coding and expression level divergence. We chose to show the *dom* vs *pah* comparisons for the pairwise divergence plots, but other pairwise comparisons show similar patterns. See supplementary table S6 for Spearman's ρ and p-values for all comparisons. (*A-D*) dN/dS vs pairwise expression divergence for: (*A*) all genes expressed early, (*B*) all genes expressed late, (*C*) genes induced early, (*D*) genes expressed early and under positive selection for protein-coding; (*E*) dN/dS vs EVE expression divergence for all genes expressed late that are under positive selection for protein-coding.

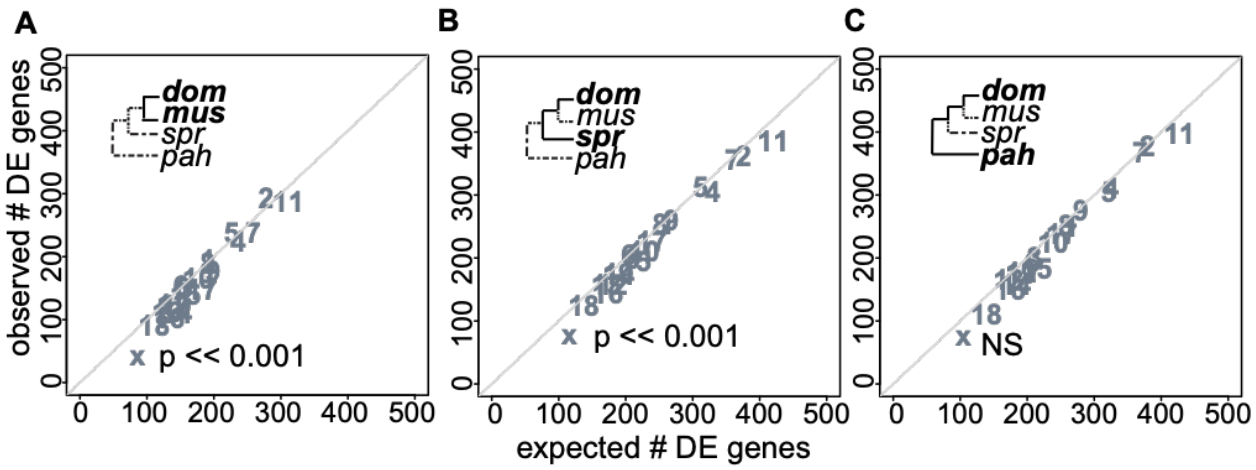

Fig. S7. Observed versus expected number of genes differentially expressed (DE) in late spermatogenesis for three pairwise comparisons at different levels of evolutionary divergence: (*A*) *dom* versus *mus*, (*B*) *spr* versus *dom,* and (*C*) *pah* versus *dom*. Each point represents a different chromosome. The diagonal line is the one-to-one line at which the observed number of DE genes equals the expected number. P-values are shown for the X chromosome only. They are based on a hypergeometric test for enrichment and corrected for multiple tests using a false discovery rate correction. A significant p-value indicates that the observed number of DE genes is different from the expected number.

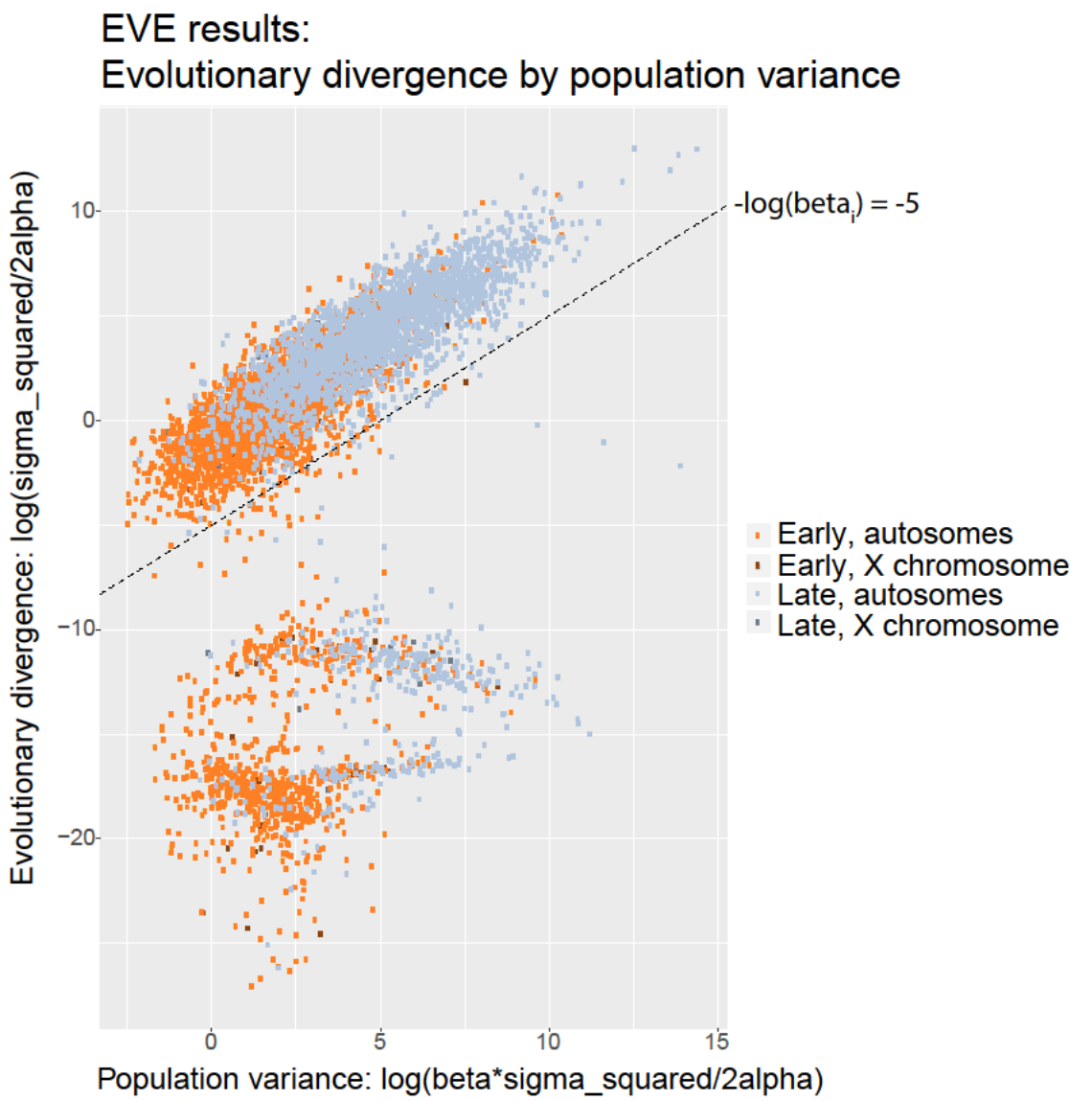

Fig. S8. EVE model evolutionary (between lineage) variance plotted against population (within lineage) variance in expression level. The EVE model assumes a linear relationship between these two values. Because genes with a divergence value [-log(*beta*$_i$)] below -5 violated this assumption, we excluded them from our analyses of expression divergence (below dashed line). This figure is based on fig. 1 from Rohlfs and Nielsen (2015). Each point represents a single gene, and points are colored by cell type and chromosome type (lighter for autosome or darker for X chromosome).

Supplementary Tables

Table S1: Number and percent of expressed genes that were lineage-specific in different *Mus* lineages under different methods of determining lineage-specific genes. Results from Zigzag, the Bayesian approach, are presented in Figure 2 of the main text. Results based on a logFC cutoff of 2 are presented in supplementary fig. S2. Autos = autosomes, X chr = X chromosome

		dom		***mus***		***dom-mus***		***spr***	
		early	late	early	late	early	late	early	late
Autos	Zigzag	49 (0.6%)	116 (1.4%)	248 (2.9%)	648 (6.7%)	104 (1.3%)	279 (3.5%)	305 (3.6%)	516 (5.7%)
	logFC > 1.5	142 (1.6%)	441 (4.9%)	195 (2.1%)	507 (5.5%)	129 (1.5%)	485 (6.0%)	214 (2.2%)	399 (4.5%)
	logFC > 2	129 (1.4%)	399 (4.4%)	189 (2.0%)	456 (5.0%)	116 (1.4%)	441 (5.5%)	195 (2.0%)	333 (3.8%)
	logFC > 3	123 (1.4%)	373 (4.1%)	186 (2.0%)	420 (4.6%)	108 (1.3%)	414 (5.2%)	190 (2.0%)	282 (3.2%)
	logFC > 5	123 (1.4%)	365 (4.0%)	186 (2.0%)	414 (4.5%)	106 (1.2%)	405 (5.0%)	188 (2.0%)	276 (3.1%)
	logFC > 10	123 (1.4%)	365 (4.0%)	186 (2.0%)	414 (4.5%)	105 (1.2%)	404 (5.0%)	188 (2.0%)	275 (3.1%)
X chr	Zigzag	5 (2.5%)	0 (0%)	7 (3.4%)	11 (5.6%)	6 (3.1%)	1 (0.6%)	11 (5.6%)	3 (1.6%)
	logFC > 1.5	1 (0.5%)	5 (2.2%)	1 (0.5%)	5 (2.4%)	2 (1.1%)	3 (1.6%)	9 (3.5%)	6 (2.7%)
	logFC > 2	1 (0.5%)	5 (2.2%)	1 (0.5%)	5 (2.4%)	2 (1.1%)	3 (1.6%)	9 (3.5%)	4 (1.8%)
	logFC > 3	1 (0.5%)	4 (1.8%)	1 (0.5%)	5 (2.4%)	2 (1.1%)	3 (1.6%)	9 (3.5%)	4 (1.8%)
	logFC > 5	1 (0.5%)	4 (1.8%)	1 (0.5%)	5 (2.4%)	2 (1.1%)	3 (1.6%)	9 (3.5%)	3 (1.4%)
	logFC > 10	1 (0.5%)	4 (1.8%)	1 (0.5%)	5 (2.4%)	2 (1.1%)	3 (1.6%)	9 (3.5%)	3 (1.4%)

Table S2. Number of genes and median dN/dS values for genes expressed, induced, testis-specific, or testis-specific and induced at different spermatogenesis stages and different chromosomes.

	Autosomes early		**X early**		**Autosomes late**		**X late**	
	n	dN/dS	n	dN/dS	n	dN/dS	n	dN/dS
Expressed	5729	0.131	167	0.182	5462	0.136	124	0.375
Induced	2046	0.105	54	0.251	1711	0.201	61	0.411
Testis-specific (TS)	350	0.282	16	0.587	424	0.297	24	0.579
TS and induced	32	0.259	6	0.745	329	0.306	19	0.543

Table S3. P-values for Wilcoxon Rank Sum tests comparing median dN/dS values between different groups of genes. Dark shaded boxes provide evidence for greater divergence in late spermatogenesis, and light shaded boxes provide evidence for faster X evolution.

	Early vs Late, Autosomes	**Early vs Late, X Chromosome**	**X vs Autosomes, Early**	**X vs Autosomes, Late**
Expressed	0.16399	0.00034	0.0061	7.10E-14
Induced	<2.00E-16	0.0488	0.00015	1.40E-07
Testis-specific (TS)	1	1	0.0015	5.10E-06
TS and Induced	0.2103	0.1009	0.0013	0.0013

Table S4: Protein coding divergence (dN/dS) and phylogeny-wide expression divergence (EVE) for different cell types and chromosome types using different methods of handling multiply mapped reads and different threshold cutoffs for expressed and induced genes. The first three rows compare different methods of handling multiply mapped reads. The middle four rows compare different threshold cutoffs for considering genes expressed. The last three rows compare different cutoffs for considering genes induced. n (prot) = number of genes in protein coding divergence analysis; n (exp) = number of genes in expression divergence analysis

	Autosomes early				X early				Autosomes late				X late			
	n (prot)	median dN/dS	n (exp)	median EVE divergence	n (prot)	median dN/dS	n (exp)	median EVE divergence	n (prot)	median dN/dS	n (exp)	median EVE divergence	n (prot)	median dN/dS	n (exp)	median EVE divergence
allow multiple mapping	5733	0.13	7211	-1.088	167	0.18	170	-1.835	5469	0.14	7638	-0.657	125	0.37	160	-0.803
allow multiple mapping; count fractionally	5732	0.13	7207	-1.064	166	0.18	170	-1.766	5469	0.14	7637	-0.656	126	0.37	158	-0.863
no multiple mapping	5729	0.13	7222	-1.045	164	0.18	174	-1.777	5462	0.14	7631	-0.646	126	0.37	160	-0.830
FPKM > 1	5733	0.13	7211	-1.088	167	0.18	170	-1.835	5462	0.14	7638	-0.657	125	0.37	160	-0.803
FPKM > 2	5127	0.13	6385	-1.061	121	0.19	120	-1.797	4886	0.14	6831	-0.649	109	0.39	139	-0.731

FPKM > 5	4091	0.13	5053	-1.009	68	0.26	64	-1.994	3912	0.14	5463	-0.633	84	0.44	103	-0.682
FPKM > 10	3166	0.12	3884	-0.934	40	0.32	30	-2.164	3071	0.15	4255	-0.632	62	0.47	70	-0.782
median expression > 2X	2046	0.11	2466	-1.094	54	0.25	44	-2.043	1711	0.20	2323	-0.704	61	0.41	69	-0.818
median expression > 5X	1181	0.11	1363	-1.125	44	0.21	33	-2.024	1076	0.24	1424	-0.770	51	0.47	53	-0.818
median expression > 10X	687	0.10	774	-1.127	35	0.18	30	-2.073	766	0.26	999	-0.810	46	0.50	46	-1.020

7

Table S5. Summary of gene expression divergence results using different methods to quantify expression divergence. Each row is a different method or set of genes, and each column is a different comparison between either cell types or chromosome types. "None" means no significant difference. For pairwise comparisons, comparisons that had a result different from the general trend are in parentheses.

	Early vs Late, Autosomes	**Early vs Late, X Chromosome**	**X vs Autosomes, Early**	**X vs Autosomes, Late**
EVE, all genes	faster late	faster late	slower-X	none
EVE, induced genes	faster late	faster late	slower-X	none
EVE, testis-specific genes	faster late	none	none	none
pairwise divergence, all genes	faster late	faster late (slower late domVspr)	faster-X (none domVpah and slower-X musVspr)	none (faster-X domVmus and musVpah, slower-X domVspr)
pairwise divergence, induced genes	faster late	none (slower late domVspr)	faster-X (none musVspr)	none
pairwise divergence, testis-specific genes	none	none	none	none (faster-X domVpah, musVpah, sprVpah)
logFC, all genes	faster late (none musVdom)	faster late (slower late domVspr)	faster-X (none musVspr)	faster-X (none domVspr and slower-X musVspr)
logFC, induced genes	faster late	faster late (slower late domVmus, domVspr, and sprVpah)	faster-X (none musVspr)	faster-X (none domVmus and musVspr)
logFC, testis-specific genes	slower late (faster late domVmus, none domVpah)	none (slower late domVmus)	none (faster-X domVmus)	none
proportion DE genes, all genes	NA	NA	none (slower-X musVspr and domVpah)	slower-X (none sprVpah)

proportion DE genes, induced genes	NA	NA	none (slower-X musVspr)	none (slower-X domVmus)
proportion DE genes, testis specific	NA	NA	none	none

Table S6. Relationship between protein-coding and expression level divergence. Rows in bold are significant based on Spearman's rank correlation. P-values are adjusted using an FDR correction for multiple tests. PW = pairwise expression divergence; ED = EVE phylogeny-wide expression divergence

Divergence comparison	Species comparison	Cell type	Induced?	Positive selection (PAML)?	No. of genes	ρ	P-value
dN/dS vs PW	**domVmus**	**early**	**no**	**no**	**7562**	**0.13**	**5.19E-27**
dN/dS vs PW	**domVspr**	**early**	**no**	**no**	**7569**	**0.16**	**3.67E-44**
dN/dS vs PW	**musVspr**	**early**	**no**	**no**	**7569**	**0.17**	**6.23E-50**
dN/dS vs PW	**domVpah**	**early**	**no**	**no**	**7563**	**0.14**	**6.00E-32**
dN/dS vs PW	**musVpah**	**early**	**no**	**no**	**7561**	**0.13**	**1.97E-29**
dN/dS vs PW	**sprVpah**	**early**	**no**	**no**	**7567**	**0.13**	**7.96E-28**
dN/dS vs PW	**domVmus**	**late**	**no**	**no**	**7570**	**0.03**	**2.17E-02**
dN/dS vs PW	**domVspr**	**late**	**no**	**no**	**7571**	**0.04**	**6.56E-04**
dN/dS vs PW	**musVspr**	**late**	**no**	**no**	**7569**	**0.05**	**2.27E-04**
dN/dS vs PW	**domVpah**	**late**	**no**	**no**	**7569**	**0.04**	**6.19E-03**
dN/dS vs PW	**musVpah**	**late**	**no**	**no**	**7562**	**0.03**	**7.23E-03**
dN/dS vs PW	**sprVpah**	**late**	**no**	**no**	**7562**	**0.05**	**2.95E-04**
dN/dS vs PW	**domVmus**	**early**	**yes**	**no**	**2141**	**0.07**	**6.19E-03**
dN/dS vs PW	**domVspr**	**early**	**yes**	**no**	**2141**	**0.09**	**2.49E-04**
dN/dS vs PW	**musVspr**	**early**	**yes**	**no**	**2141**	**0.09**	**2.49E-04**
dN/dS vs PW	**domVpah**	**early**	**yes**	**no**	**2141**	**0.11**	**3.04E-06**
dN/dS vs PW	**musVpah**	**early**	**yes**	**no**	**2141**	**0.08**	**6.97E-04**
dN/dS vs PW	**sprVpah**	**early**	**yes**	**no**	**2141**	**0.09**	**7.75E-05**
dN/dS vs PW	domVmus	late	yes	no	1760	0.02	0.57
dN/dS vs PW	domVspr	late	yes	no	1760	0.01	0.70
dN/dS vs PW	musVspr	late	yes	no	1760	0.01	0.87
dN/dS vs PW	domVpah	late	yes	no	1760	0.00	0.92
dN/dS vs PW	musVpah	late	yes	no	1760	-0.01	0.70
dN/dS vs PW	sprVpah	late	yes	no	1760	-0.04	0.23
dN/dS vs PW	domVmus	early	no	yes	248	0.09	0.31
dN/dS vs PW	domVspr	early	no	yes	250	0.12	0.10

dN/dS vs PW	**musVspr**	**early**	**no**	**yes**	**249**	**0.20**	**0.00**
dN/dS vs PW	**domVpah**	**early**	**no**	**yes**	**248**	**0.18**	**0.02**
dN/dS vs PW	**musVpah**	**early**	**no**	**yes**	**247**	**0.19**	**0.00**
dN/dS vs PW	**sprVpah**	**early**	**no**	**yes**	**248**	**0.21**	**0.00**
dN/dS vs PW	domVmus	late	no	yes	250	0.08	0.32
dN/dS vs PW	**domVspr**	**late**	**no**	**yes**	**250**	**0.17**	**0.02**
dN/dS vs PW	**musVspr**	**late**	**no**	**yes**	**249**	**0.18**	**0.00**
dN/dS vs PW	domVpah	late	no	yes	250	0.06	0.45
dN/dS vs PW	musVpah	late	no	yes	249	0.06	0.48
dN/dS vs PW	sprVpah	late	no	yes	250	0.09	0.25
dN/dS vs PW	domVmus	early	yes	yes	67	0.19	0.24
dN/dS vs PW	domVspr	early	yes	yes	67	0.00	1.00
dN/dS vs PW	musVspr	early	yes	yes	67	0.16	0.31
dN/dS vs PW	domVpah	early	yes	yes	67	0.08	0.60
dN/dS vs PW	musVpah	early	yes	yes	67	-0.02	0.90
dN/dS vs PW	sprVpah	early	yes	yes	67	0.09	0.60
dN/dS vs PW	domVmus	late	yes	yes	57	-0.15	0.39
dN/dS vs PW	domVspr	late	yes	yes	57	-0.21	0.23
dN/dS vs PW	musVspr	late	yes	yes	57	-0.04	0.84
dN/dS vs PW	domVpah	late	yes	yes	57	-0.10	0.58
dN/dS vs PW	musVpah	late	yes	yes	57	-0.10	0.57
dN/dS vs PW	sprVpah	late	yes	yes	57	-0.16	0.36
dN/dS vs ED	NA	early	no	no	4473	0.03	0.10
dN/dS vs ED	NA	late	no	no	4755	0.02	0.43
dN/dS vs ED	NA	early	yes	no	1544	0.02	0.56
dN/dS vs ED	NA	late	yes	no	1490	-0.01	0.89
dN/dS vs ED	NA	early	no	yes	144	-0.11	0.31
dN/dS vs ED	**NA**	**late**	**no**	**yes**	**160**	**0.18**	**0.04**
dN/dS vs ED	NA	early	yes	yes	55	-0.12	0.51
dN/dS vs ED	NA	late	yes	yes	48	0	1.00

19

Table S7. Proportion of autosomal genes in each regulatory category. P-values are based on a Pearson's chi-squared test for differences between the early and late cell types after FDR correction for multiple tests. The first column shows the results presented in Mack, et al. (2016). For the first two columns, genes in the "conserved" category are grouped into the "other" category for direct comparison with results from Mack, et al. (2016). *c*X*t* =*cis*X*trans;* comp = compensatory; c+*t,* opp = *cis* + *trans,* opposite; *c*+*t,* same = *cis* + *trans,* same

	fertile F1 hybrid, intra-subspecific F1 parents (binomial test)*		**fertile F1 hybrid, intra-subspecific F1 parents**				**fertile F1 hybrid, pure strain parents**			**sterile F1 hybrid, intra-subspecific F1 parents**				**sterile F1 hybrid, pure strain parents**			***dom* only**			***mus* only**		
	Mack et al. 2016	**re-analysis (Modtools)**	**whole testes**	**early**	**late**	**p-val**	**early**	**late**	**p-val**	**whole testes**	**early**	**late**	**p-val**	**early**	**late**	**p-val**	**early**	**late**	**p-val**	**early**	**late**	**p-val**
cis	24%	21%	22%	17%	30%	9.57E-47	11%	33%	4.86E-157	14%	18%	30%	6.59E-34	12%	31%	3.17E-145	14%	22%	5.42E-07	8%	12%	1.33E-04
trans	9%	17%	36%	53%	32%	2.46E-119	59%	28%	0.00E+00	34%	51%	33%	1.39E-84	57%	29%	3.31E-315	46%	29%	1.01E-27	59%	28%	1.14E-130
c*X*t	7%	10%	0%	0%	1%	1.08E-04	1%	2%	4.34E-04	0%	0%	1%	6.60E-03	1%	2%	1.33E-02	0%	3%	5.14E-05	3%	4%	7.63E-01
comp	13%	8%	16%	8%	12%	2.65E-01	7%	10%	5.62E-05	14%	10%	12%	1.94E-03	9%	10%	8.45E-04	12%	9%	9.94E-03	12%	18%	3.97E-07
***c+t*, opp**	16%	9%	12%	5%	6%	3.02E-02	2%	3%	2.88E-05	15%	4%	5%	3.17E-02	2%	4%	1.13E-09	3%	6%	5.55E-04	3%	9%	3.54E-12
***c+t*, same**	8%	17%	11%	3%	8%	1.20E-16	5%	9%	2.50E-14	13%	4%	7%	1.26E-09	7%	10%	4.72E-10	10%	11%	8.86E-01	7%	16%	2.23E-19
other	23%	18%	4%	15%	13%	3.28E-03	14%	15%	3.88E-01	11%	13%	13%	5.78E-01	13%	15%	2.12E-03	14%	21%	8.07E-07	8%	15%	3.58E-10
Total # genes	9851	9478	1430	2541	3676	NA	3291	4067	NA	1129	2416	3258	NA	3820	3859	NA	910	1768	NA	1215	1643	NA

*includes conserved genes in "other" category

Table S8. Number of genes that changed regulatory category between the early and late cell types. Rows indicate regulatory categories in the late cell type and columns indicate regulatory categories in the early cell type. Grey boxes along the diagonal indicate genes that do not change regulatory category between cell types.

		Early								
		not expressed	*cis*	*trans*	*cis* X *trans*	compensatory	*cis*+*trans*, opp	*cis*+*trans*, same	other	conserved
Late	not expressed	0	169	505	2	68	30	20	156	2483
	cis	268	132	85	0	22	31	22	49	480
	trans	188	7	275	1	8	2	5	30	642
	cis X *trans*	7	2	12	0	1	0	2	1	10
	compensatory	78	8	36	1	33	5	3	6	277
	cis+*trans*, opp	44	30	8	0	19	33	1	3	64
	cis+*trans*, same	52	19	52	0	2	0	11	16	122
	other	116	7	48	0	3	0	4	17	276
	conserved	859	45	325	1	43	14	12	99	2881

Table S9. Proportion of autosomal genes in each regulatory category that showed high pairwise expression divergence. High pairwise divergence is defined as genes in the top 25% of divergence values for a given pairwise comparison. Each row represents a different pairwise comparison and cell type. Highlighted boxes represent the proportion of genes in the *cis* + *trans* same (reinforcing) category that also were highly divergent between *dom* and *mus*. Of all genes in the reinforcing category, a higher proportion overlap with *dom* vs *mus* highly divergent genes than with genes highly divergent in other pairwise comparisons.

Early	***cis***	***trans***	***cis* X *trans***	**compensatory**	***cis* + *trans* opposite**	***cis* + *trans* same**	**other**
*dom*V*mus*	0.21	0.137	0	0.151	0.261	0.263	0.183
*dom*V*spr*	0.126	0.092	0	0.131	0.113	0.125	0.151
*mus*V*spr*	0.103	0.107	0	0.146	0.122	0.15	0.146
*dom*V*pah*	0.093	0.108	0.2	0.126	0.087	0.113	0.143
*mus*V*pah*	0.105	0.106	0.2	0.151	0.13	0.163	0.149
*spr*V*pah*	0.081	0.103	0.2	0.131	0.113	0.1	0.141
Late	***cis***	***trans***	***cis* X *trans***	**compensatory**	***cis* + *trans* opposite**	***cis* + *trans* same**	**other**
*dom*V*mus*	0.198	0.055	0.029	0.031	0.129	0.215	0.07
*dom*V*spr*	0.106	0.08	0.143	0.092	0.163	0.142	0.11
*mus*V*spr*	0.122	0.073	0.171	0.067	0.158	0.102	0.085
*dom*V*pah*	0.128	0.078	0.229	0.098	0.144	0.109	0.104
*mus*V*pah*	0.108	0.083	0.229	0.101	0.124	0.117	0.089
*spr*V*pah*	0.107	0.091	0.114	0.116	0.099	0.095	0.1

Table S10. Genes with evidence for rapid evolution during spermatogenesis (protein-coding, phylogeny-wide expression, or pairwise expression) that may also have testis-biased expression. Comparisons with high pairwise divergence have a divergence value > 1. Chr = chromosome; LZ = leptotene/zygotene spermatocytes ("early"); RS = round spermatids ("late"); EVE = -log(*beta$_i$*) value from the EVE model

Gene ID	Gene Name	Chr	dN/dS	Positive Selection (PAML)?	Induced ?	High Pairwise Expression Divergence (LZ)	High Pairwise Expression Divergence (RS)	EVE (LZ)	EVE (RS)	Expression Pattern
ENSMUSG00000022280	Rnf19a	15	0.14	no	yes, in RS	none	none	-0.266	4.159	primarily testis
ENSMUSG00000022602	Arc	15	0	no	no	none	none	0.023	4.251	primarily testis and brain
ENSMUSG00000056209	Npm3	19	0.12	no	no	none	none	3.388	2.687	primarily testis
ENSMUSG00000037101	Ttc29	8	2.05	yes	yes, in RS	none	none	-1.134	-0.643	primarily testis
ENSMUSG00000049761	Pmis2	7	1.55	yes	yes, in RS	none	none	-1.773	-1.76	testis-specific
ENSMUSG00000027317	Ppp1r14d	2	2.71	no	NA	domVmus, musVspr	domVmus, domVspr, musVspr, musVpah, sprVpah	NA	NA	has a testis-specific isoform

Table S11. RNAseq metadata for each sample.

Lineage	Strain	Sample Name	Cell Type	Cell Sort Date	Mouse Age at Sort Date	RNA concentration after cell sort RNA extraction (ng/µL)	RIN	# Raw Reads	# Mapped Reads	SRA Accession
dom	BIK/g	BIK_4665.1M_LZ	LZ	10/17/2013	95	8.5	9.6	28574278	15792758	SAMN19597717
dom	BIK/g	BIK_4665.1M_RS	RS	10/17/2013	95	4.2	8.5	38477703	29776396	SAMN19597718
dom	BIK/g	BIK_4665.2M_LZ	LZ	10/18/2013	96	12.4	9.7	11156170	9679873	SAMN19597719
dom	BIK/g	BIK_4665.2M_RS	RS	10/18/2013	96	3.4	8.1	38039743	30916495	SAMN19597720
dom	DGA	DGA_5406.1M_LZ	LZ	10/3/2013	94	8.2	9.8	24007340	14415796	SAMN19597721
dom	DGA	DGA_5406.1M_RS	RS	10/3/2013	94	4.7	8.7	10757881	9926211	SAMN19597722
dom	DGA	DGA_5406.2M_LZ	LZ	10/10/2013	101	7.4	9.7	16740461	14956442	SAMN19597723
dom	DGA	DGA_5406.2M_RS	RS	10/10/2013	101	7.7	8.6	10080556	9278016	SAMN19597724
dom	DGA	DGA_5406.3M_LZ	LZ	10/11/2013	102	10.6	9.6	26873371	22540332	SAMN19597725
dom	DGA	DGA_5406.3M_RS	RS	10/11/2013	102	9.7	8.4	50306765	45439841	SAMN19597726
dom	LEWES/EiJ	LL.LL125.1M.LZ	LZ	6/4/2015	80	14.2	9.7	41523441	32816986	SAMN19597727
dom	LEWES/EiJ	LL.LL125.1M.RS	RS	6/4/2015	80	17.6	7.9	32174124	28997946	SAMN19597728
dom	LEWES/EiJ	LL.LL125.2M.LZ	LZ	6/2/2015	78	6.1	8.8	38838198	34572862	SAMN19597729
dom	LEWES/EiJ	LL.LL125.2M.RS	RS	6/2/2015	78	5.6	8.5	16433187	15281121	SAMN19597730
dom	LEWES/EiJ	LL.LL125.3M.LZ	LZ	6/1/2015	77	7.4	9.5	25464043	23141321	SAMN19597731
dom	LEWES/EiJ	LL.LL125.3M.RS	RS	6/1/2015	77	5.6	8.6	32759463	29987663	SAMN19597732
dom	WSB/EiJ	WW.WW87.2M_LZ	LZ	5/13/2014	66	12.2	9.6	1783976	1570753	SAMN19597733
dom	WSB/EiJ	WW.WW87.2M_RS	RS	5/13/2014	66	4	8.6	15238105	14266130	SAMN19597734
dom	WSB/EiJ	WW.WW87.3M_LZ	LZ	5/14/2014	67	10.6	9.6	18188548	16621233	SAMN19597735

dom	WSB/EiJ	WW.WW87.3M_RS	RS	5/14/2014	67	7.7	8.6	11897490	11086144	SAMN19597736
dom	WSB/EiJ	WW.WW87.4M_LZ	LZ	5/30/2014	83	9.8	9.7	9943638	9062201	SAMN19597737
dom	WSB/EiJ	WW.WW87.4M_RS	RS	5/30/2014	83	3.1	8.7	13971859	12895786	SAMN19597738
dom	WSB/EiJ	WW.WW89.8M_LZ	LZ	4/29/2014	61	6.4	9.4	11641744	10377128	SAMN19597739
dom	WSB/EiJ	WW.WW89.8M_RS	RS	4/29/2014	61	4.3	8.3	18384101	16980022	SAMN19597740
mus	CZECHII/EiJ	CC.CC153.4M.LZ	LZ	5/20/2015	87	6.5	9.2	36237077	24720082	SAMN19597741
mus	CZECHII/EiJ	CC.CC153.4M.RS	RS	5/20/2015	87	3.3	8.2	8995430	8273438	SAMN19597742
mus	CZECHII/EiJ	CC.CC153.5M.LZ	LZ	5/22/2015	89	8	8.8	24808708	18264983	SAMN19597743
mus	CZECHII/EiJ	CC.CC153.5M.RS	RS	5/22/2015	89	7.5	7.3	39256981	32049212	SAMN19597744
mus	MBS	MBS_4527.1M_LZ	LZ	9/20/2013	63	6.8	9.5	22446803	20502488	SAMN19597745
mus	MBS	MBS_4527.1M_RS	RS	9/20/2013	63	10.9	8.2	12665610	11785311	SAMN19597746
mus	MBS	MBS_4527.2M_LZ	LZ	9/26/2013	69	25.9	8.8	18148591	16500321	SAMN19597747
mus	MBS	MBS_4527.2M_RS	RS	9/26/2013	69	10.5	8.2	24450942	22406598	SAMN19597748
mus	MBS	MBS_4527.3M_LZ	LZ	10/2/2013	75	18.3	9.5	26762674	24061812	SAMN19597749
mus	MBS	MBS_4527.3M_RS	RS	10/2/2013	75	6.9	8.3	15396056	14315196	SAMN19597750
mus	PWK/PhJ	PP.PP.98.4M.LZ	LZ	6/18/2015	73	35	9	11808242	10876393	SAMN19597751
mus	PWK/PhJ	PP.PP.98.4M.RS	RS	6/18/2015	73	13.8	8.6	72582309	67215275	SAMN19597752
mus	PWK/PhJ	PP.PP98.3M.LZ	LZ	6/17/2015	72	35.3	8.5	7372674	6519250	SAMN19597753
mus	PWK/PhJ	PP.PP98.3M.RS	RS	6/17/2015	72	13.8	8.1	19931551	18654440	SAMN19597754
mus	PWK/PhJ	PP.PP98.5M.LZ	LZ	6/24/2015	79	15.5	9.5	11848083	10572199	SAMN19597755
mus	PWK/PhJ	PP.PP98.5M.RS	RS	6/24/2015	79	6.8	8.7	16044407	15062711	SAMN19597756
spr	SEG	SEG_4130_LZ	LZ	3/5/2013	66	6	9.4	25968746	23559901	SAMN19597757
spr	SEG	SEG_4130_RS	RS	3/5/2013	66	5.2	8.3	25844180	23258335	SAMN19597758

spr	SEG	SEG_4156_LZ	LZ	9/19/2013	130	12.4	8.6	20360141	17686631	SAMN19597759
spr	SEG	SEG_4156_RS	RS	9/19/2013	130	4	7.6	22932882	20960724	SAMN19597760
spr	SEG	SEG_4176_LZ	LZ	2/26/2013	100	4.2	8.9	22777686	20079778	SAMN19597761
spr	SEG	SEG_4176_RS	RS	2/26/2013	100	6.8	7.2	14405903	13311983	SAMN19597762
spr	SEG	SEG_4197_LZ	LZ	3/11/2013	72	14	9	23944237	20931833	SAMN19597763
spr	SEG	SEG_4197_RS	RS	3/11/2013	72	5.7	8.4	10443918	9663280	SAMN19597764
spr	SEG	SEG_4700.1M_LZ	LZ	9/9/2013	148	16.7	8.8	13571853	12253820	SAMN19597765
spr	SEG	SEG_4700.1M_RS	RS	9/9/2013	148	6.1	7.8	24534179	22456069	SAMN19597766
spr	SFM	SFM_4513_LZ	LZ	3/7/2013	74	17.3	8.6	13616782	12154832	SAMN19597767
spr	SFM	SFM_4513_RS	RS	3/7/2013	74	7.4	8.4	6887627	6355085	SAMN19597768
spr	SFM	SFM_4514_LZ	LZ	3/12/2013	79	7.6	9	16693417	14970665	SAMN19597769
spr	SFM	SFM_4514_RS	RS	3/12/2013	79	11.3	7.8	13874143	12800911	SAMN19597770
spr	STF	STF_4495.1M_LZ	LZ	9/11/2013	150	13.7	9	7705953	7049222	SAMN19597771
spr	STF	STF_4495.1M_RS	RS	9/11/2013	150	4.3	7.9	22281038	20273895	SAMN19597772
spr	STF	STF_4515_LZ	LZ	3/20/2013	84	4.2	9.1	14293633	13197893	SAMN19597773
spr	STF	STF_4515_RS	RS	3/20/2013	84	1.8	9.3	8019452	7356732	SAMN19597774
spr	STF	STF_4516_LZ	LZ	3/4/2013	68	4.6	9.8	17387559	15715819	SAMN19597775
spr	STF	STF_4516_RS	RS	3/4/2013	68	6	8.8	15006751	13932584	SAMN19597776
spr	STF	STF_4517_LZ	LZ	3/8/2013	72	20.2	9.2	32586439	29128942	SAMN19597777
spr	STF	STF_4517_RS	RS	3/8/2013	72	3.8	7	27085325	25022314	SAMN19597778
pah	PAHARI/EiJ	PAH.New.1M_LZ	LZ	7/10/2014	75	9.5	8.2	8355637	7272288	SAMN19597779
pah	PAHARI/EiJ	PAH.New.1M_RS	RS	7/10/2014	75	4.6	6.9	30810159	28255790	SAMN19597780

pah	PAHARI/EiJ	PAH.New.2M_LZ	LZ	7/17/2014	82	11.7	8.6	23718990	20946764	SAMN19597781
pah	PAHARI/EiJ	PAH.New.2M_RS	RS	7/17/2014	82	7.3	7.5	10190552	9329587	SAMN19597782
pah	PAHARI/EiJ	PAH.New.4M_LZ	LZ	7/16/2014	81	6.4	8.3	96163007	84308080	SAMN19597783
pah	PAHARI/EiJ	PAH.New.4M_RS	RS	7/16/2014	81	4	7.5	10813786	9845873	SAMN19597784

Table S12. List of genes included in our analyses and whether they were considered expressed, induced, or active in each cell type. Available as a separate supplemental Excel file.

Table S13. Percent of genes assigned to each regulatory category using different approaches in edgeR. Proportions are presented for the LEWES♀ X PWK♂ F1 hybrids with intrasubspecific F1s as the parents. Columns labeled "with calcNormFactors()" indicate that normalization for relative expression levels of genes within samples was performed in edgeR. Columns labeled as "with blocking" mean that ASE tests were performed in edgeR with blocking by individual.

	with calcNormFactors()				without calcNormFactors()			
	without blocking		with blocking		without blocking		with blocking	
	early	late	early	late	early	late	early	late
cis	17%	30%	19%	31%	17%	30%	19%	31%
trans	52%	31%	31%	22%	51%	31%	31%	22%
cis X *trans*	0%	1%	2%	2%	0%	1%	2%	2%
compensatory	8%	12%	29%	19%	8%	12%	29%	19%
cis+*trans*, opp	5%	6%	4%	5%	5%	6%	3%	5%
cis+*trans*, same	3%	8%	10%	12%	3%	8%	10%	12%
other	15%	13%	6%	9%	16%	13%	6%	9%
total # genes	2582	3796	3083	4089	2556	3779	3100	4096

Table S14. Number of genes assigned to each regulatory category using both the binomial test and negative binomial test approaches for the fertile F1 mice. Gray boxes indicate the number of genes assigned to the same category using either approach.

		Negative Binomial Test for DE/ASE								
		NA	*cis*	*trans*	*cis* X *trans*	compensatory	*cis* + *trans* (opp)	*cis* + *trans* (same)	other	conserved
Binomial Test for DE/ASE	NA	0	116	53	0	38	48	26	19	655
	cis	203	200	0	0	0	0	0	43	1562
	trans	165	0	106	0	0	0	0	0	1298
	cis X *trans*	59	0	38	1	8	0	0	0	860
	compensatory	67	0	0	0	30	0	0	0	690
	cis + *trans* (opp)	51	0	0	0	149	118	0	0	527
	cis + *trans* (same)	94	0	310	0	0	0	127	0	1029
	other	250	0	0	0	0	0	0	0	1493

Table S15. Read counts from the lapels-suspenders pipeline output for whole testes data.

Sample	Mack et al. 2016 Sample ID	Total # Reads	Total # Assigned to Parent	LEWES	PWK
LLWW_SRR2060837	148	64851628	23459524	20941397	2518127
LLWW_SRR2060842	149	96823643	37201091	33111153	4089938
LLWW_SRR2060843	150	85112678	32593290	29021132	3572158
PPCC_SRR2060844	151	47030803	16925427	1776081	15149346
PPCC_SRR2060846	152	88496639	33850882	3394900	30455982
PPCC_SRR2060939	170	68604125	26294240	2627391	23666849
PPLL_SRR2060951	52	35628710	13361236	6461350	6899886
PPLL_SRR2060955	278	92806864	36956675	23978091	12978584
PPLL_SRR2060954	131	23146016	8661071	4279501	4381570
LLPP_SRR2060950	93	40057612	14199347	7360932	6838415
LLPP_SRR2060952	290	32200485	12058341	6153120	5905221
LLPP_SRR2060953	272	112352140	43526628	22241252	21285376

References

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31:166-169.
Benowitz KM, Coleman JM, Allan CW, Matzkin LM. 2020. Contributions of *cis*- and *trans*-Regulatory Evolution to Transcriptomic Divergence across Populations in the *Drosophila mojavensis* Larval Brain. *Genome Biol Evol*. 12:1407-1418.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.
Combs PA, Fraser HB. 2018. Spatially varying cis-regulatory divergence in *Drosophila* embryos elucidates cis-regulatory logic. *PLoS Genet*. 14:e1007631.
Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res*. 24:797-808.
Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol Ecol*. 17:5349-5363.
Holt J, Huang S, McMillan L, Wang W. 2013. Read Annotation Pipeline for High-Throughput Sequencing Data. Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics; Washington DC, USA: 2506645: ACM. p. 605-612.
Huang S, Holt J, Kao C-Y, McMillan L, Wang W. 2014. A novel multi-alignment pipeline for high-throughput sequencing data. *Database*. 2014.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36.
Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res*.
Rohlfs RV, Nielsen R. 2015. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Syst Biol*. 64:695-708.

The contribution of sex chromosome conflict to disrupted spermatogenesis in hybrid house mice

Emily E. K. Kopania*,[1], Eleanor M. Watson‡, Claudia C. Rathje§, Benjamin M. Skinner‡, Peter J. I. Ellis§, Erica L. Larson†,[1,2], Jeffrey M. Good*,[1,2]

*Division of Biological Sciences, University of Montana, Missoula, MT 59812
‡School of Life Sciences, University of Essex, Colchester, CO4 3SQ, U.K.
§School of Biosciences, University of Kent, Canterbury, CT2 7NJ, U.K
†Department of Biological Sciences, University of Denver, Denver, CO 80208

[1]Corresponding authors: emily.kopania@umconnect.umt.edu, erica.larson@du.edu, jeffrey.good@umontana.edu
[2]Co-senior authors

ORCID ID: 0000-0002-2710-2491 (EEKK)
ORCID ID: 0000-0002-1290-4360 (EMW)
ORCID ID: 0000-0002-7152-1167 (BMS)
ORCID ID: 0000-0001-9709-7934 (PJIE)
ORCID ID: 0000-0003-3006-645X (ELL)
ORCID ID: 0000-0003-0707-5374 (JMG)

Data Availability: The sequence data reported in this paper are available through the National Center for Biotechnology Information Sequence Read Archive under accession numbers PRJNA816542 and PRJNA816886**.** The phenotype data are available in the Supplemental Material, Table S2.

Running Title: Sex Chromosomes and Hybrid Sterility

Keywords: speciation, intragenomic conflict, ampliconic genes, hybrid male sterility, testis expression, FACS, sex chromosomes

Corresponding Authors:
Emily E. K. Kopania
Division of Biological Sciences
University of Montana
Missoula, MT 59812
Email: emily.kopania@umconnect.umt.edu

Erica L. Larson
Department of Biological Sciences
University of Denver
Denver, CO 80208
Email: erica.larson@du.edu

Jeffrey M. Good
Division of Biological Sciences
University of Montana
Missoula, MT 59812
Phone: (406) 243 - 5122
Email: jeffrey.good@umontana.edu

Abstract

Incompatibilities on the sex chromosomes play an important role in the evolution of hybrid male sterility, but the evolutionary forces underlying this phenomenon are largely unknown. Different lineages or subspecies of house mice (*Mus musculus*) have provided well-studied models for understanding the genetic basis of hybrid male sterility. X chromosome-autosome interactions cause strong F1 incompatibilities in *Mus musculus* hybrids, but variation in sterility phenotypes expressed across different genetic architectures also suggests a complex genetic basis. In parallel, X-Y chromosome conflict has emerged as a major driver of gene family evolution in house mice, resulting in rapid copy number evolution of ampliconic genes with dosage-dependent expression that is essential to spermatogenesis. Here we evaluated the contribution of X-Y lineage mismatch to disruption of sperm head development and genome-wide patterns of stage-specific gene expression in hybrid house mice. We performed backcrosses between two house mouse susbspecies to generate reciprocal Y-introgression strains and then used these consomic models to test the effects of X-Y mismatch in F1 and late-generation (introgressed) hybrids. We found evidence that X-Y mismatch contributed to some F1 male sterility phenotypes. However, these effects were subtle and transcriptome analyses of sorted postmeiotic cells (round spermatids) revealed widespread overexpression of the *M. musculus* X chromosome in sterile F1 hybrids independent of Y chromosome subspecies origin. Thus, widespread overexpression of the X chromosome commonly observed in sterile F1 mouse hybrids is likely a downstream consequence of disrupted X-inactivation during meiosis and is not caused by copy number divergence between coevolving X- and Y-linked ampliconic genes. Y-chromosome introgression did result in subfertility phenotypes and disrupted expression of several autosomal genes in mice with a non-hybrid autosomal and X-linked background. These results suggest that Y-linked incompatibilities contribute to reproductive barriers between these lineages, but likely not as a direct consequence of X-Y conflict. Collectively, these findings suggest that rapid X- and Y-linked gene family evolution driven genomic conflict has not resulted in strong male reproductive barriers in house mice.

Introduction

Sex chromosomes are often involved in the evolution of hybrid male sterility between animal species (Coyne and Orr 1989; Turelli and Orr 2000; Presgraves and Meiklejohn 2021). Referred to as the large X-effect in X-Y systems (Coyne and Orr 1989), it remains unclear to what extent this general pattern reflects common evolutionary processes or functional mechanisms unique to sex chromosomes (Meiklejohn and Tao 2010). For example, intrinsic reproductive barriers between nascent species are generally assumed to arise as an indirect consequence of rapid evolution within populations (Dobzhansky 1937; Muller 1942; Coyne and Orr 2004; Coughlan and Matute 2020). The outsized contribution of sex chromosomes to male sterility could be an inevitable consequence of rapid evolution due to recurrent genomic conflict, because selfish genetic elements are more likely to arise on sex chromosomes (i.e., meiotic drive *sensu lato;* Frank 1991; Hurst and Pomiankowski 1991; Meiklejohn and Tao 2010; Lindholm *et al.* 2016). Hemizygosity of the X chromosome is also expected to promote rapid evolution across a broad range of conditions independent of genomic conflict (i.e., the faster-X effect; Charlesworth *et al.* 1987; Vicoso and Charlesworth 2009). However, progress on understanding how these diverse evolutionary processes contribute to the large X-effect has been hampered by a lack of data on the genetic underpinnings of hybrid male sterility.

From a mechanistic perspective, the X and Y chromosomes are also subject to unique regulatory processes during mammalian spermatogenesis that are critical for normal male fertility and shape patterns of molecular evolution (Larson, et al. 2018a). Both the X and Y chromosomes are packaged into condensed chromatin early in meiosis, resulting in transcriptional silencing of most sex-linked genes known as meiotic sex chromosome inactivation (MSCI; McKee and Handel 1993). Repressive chromatin persists through the postmeiotic stages (Namekawa, et al. 2006), although many essential X- and Y-linked genes are highly expressed in haploid round spermatids prior to spermiogenesis (Mueller, et al. 2008; Sin and Namekawa 2013). Failure to broadly repress X-linked expression during these critical meiotic and postmeiotic stages can trigger spermatogenic arrest, reduced sperm production, and abnormal sperm

morphology (Burgoyne *et al.* 2009; Turner 2015). Interestingly, sex chromosome repression during both stages appears prone to disruption in hybrid mammals (Mihola *et al.* 2009; Good *et al.* 2010; Campbell *et al.* 2013; Davis *et al.* 2015; Larson *et al.* 2017), which may reflect common regulatory pathways underlying the evolution of hybrid male sterility (Bhattacharyya *et al.* 2013; Larson *et al.* 2021). Understanding how these intermediate developmental sterility phenotypes relate to genomic conflict and the broader evolutionary dynamics of the sex chromosomes awaits more data.

House mice (*Mus musculus*) have emerged as predominant models for understanding both the basic molecular control of spermatogenesis and the evolution of hybrid male sterility in mammals (Phifer-Rixey and Nachman 2015). Closely related subspecies of mice, *Mus musculus musculus* and *M. m. domesticus* (hereafter, "*musculus*" and "*domesticus*"), readily hybridize in both the lab and along a natural hybrid zone in Europe (Janoušek *et al.* 2012). Hybrid male sterility is the strongest and likely primary reproductive barrier isolating these incipient species in nature (Vyskočilová, et al. 2005; Turner, et al. 2012) and in the lab (Good *et al.* 2008b; Vyskočilová *et al.* 2009; but see Suzuki and Nachman 2015), following Haldane's rule (i.e., hybrid breakdown primarily occurs in the heterogametic sex; Haldane 1922). Male sterility is polymorphic with laboratory crosses yielding sterile, subfertile, or fertile male hybrids depending on genotype and cross direction (Good *et al.* 2008b; Balcova *et al.* 2016; Larson *et al.* 2018b; Widmayer *et al.* 2020); *musculus*♀ × *domesticus*♂ crosses usually result in sterile F1 males, while the reciprocal cross tends to be more fertile (Good *et al.* 2008b). This asymmetry is caused by epistatic incompatibilities that are exposed on the *musculus* X chromosome in hybrid males (Storchová *et al.* 2004; Good *et al.* 2008a; Turner and Harr 2014). House mice also remain the only mammalian system where the evolution of a specific gene, *Prdm9*, has been directly linked to the evolution of intrinsic reproductive barriers (Mihola *et al.* 2009; Bhattacharyya *et al.* 2013; Mukaj *et al.* 2020). *Prdm9* is an autosomal gene encoding a DNA-binding protein that directs double stranded breaks where meiotic recombination occurs (Grey *et al.* 2011). PRDM9 binding sites evolve rapidly (Oliver *et al.* 2009; Baker *et al.* 2015), leading to asymmetric binding in hybrid mice that triggers autosomal asynapsis and disruption of MSCI during early pachytene of Meiosis I (Mihola *et al.* 2009; Davies *et al.* 2016).

Prdm9-related sterility depends on *Prdm9* heterozygosity and epistatic interactions with other unlinked factors, including a major incompatibility locus, *Hstx2*, located near the middle the *musculus* X chromosome (Forejt *et al.* 2021). This same X-linked region also influences hybrid male sterility in backcrossed consomic models (i.e., presumably independent of *Prdm9*; Storchová *et al.* 2004; Good *et al.* 2008a), and recombination rate variation between *M. m. musculus* and another subspecies, *M. m. castaneus* (Dumont and Payseur 2011).

This broad foundation on the genetics of hybrid male sterility provides an opportunity to further unravel the various evolutionary and mechanistic processes that contribute to the large X-effect in mice. *Prdm9*-related sterility plays a central role in the evolution of hybrid male sterility and the disruption of MSCI in F1 mouse hybrids (Forejt *et al.* 2021; Larson *et al.* 2021). However, X- and Y-linked hybrid sterility arises across a broader range of genetic architectures and phenotypes than cannot be easily ascribed to *Prdm9*-related interactions (Campbell *et al.* 2012; Campbell and Nachman 2014; Larson *et al.* 2018b; Larson *et al.* 2021). The mouse X and Y chromosomes also contain clusters of several high copy ampliconic genes (Mueller *et al.* 2008; Soh *et al.* 2014; Case *et al.* 2015; Morgan and Pardo-Manuel De Villena 2017; Larson *et al.* 2021) that appear to have evolved in response to intense intragenomic conflict (Cocquet *et al.* 2009; Ellis *et al.* 2011; Cocquet *et al.* 2012). These X- and Y-linked gene clusters are primarily expressed in postmeiotic cells with repressed sex chromatin (Namekawa *et al.* 2006; Sin *et al.* 2012) and thus increases in copy number may help counteract repressive chromatin (Ellis *et al.* 2011; Mueller *et al.* 2013; Sin and Namekawa 2013). Conflict arises because the maintenance of repressive postmeiotic sex chromatin appears to be controlled by dosage dependent interactions between X-linked (*Slx* and *Slxl1*) and Y-linked (*Sly*) gene families (Cocquet *et al.* 2012; Kruger *et al.* 2019). Experimental knockdowns of *Slx* and *Slxl1* showed increased sex chromosome repression, abnormal sperm head morphology, and an excess of male offspring. In contrast, knockdowns of *Sly* showed sex chromosome overexpression, abnormal sperm head morphology, and an excess of female offspring (Cocquet *et al.* 2009; Cocquet *et al.* 2012) due to reduced motility of Y-bearing sperm (Rathje *et al.* 2019). CRISPR-based deletions have further shown that sex-ratio distortion is primarily mediated by

Slxl1 versus *Sly* competition for the spindlin proteins (SPIN1, STY1/2; Kruger et al. 2019).

Copy numbers of *Slx*, *Slxl1,* and *Sly* genes have co-evolved in different mouse lineages (Ellis *et al.* 2011; Good 2012; Morgan and Pardo-Manuel De Villena 2017), such that hybrids could have copy number mismatch sufficient to generate dosage-based sterility phenotypes seen in genetic manipulation studies (Ellis *et al.* 2011). In support of this model, hybrid interactions between the *musculus* X and the *domesticus* Y have been shown to cause abnormal sperm head morphology (Campbell *et al.* 2012; Campbell and Nachman 2014), and male sterility is associated with extensive overexpression of the sex chromosomes in postmeiotic round spermatids in *musculus*♀ × *domesticus*♂ mice (Larson *et al.* 2017). These hybrids have proportionally higher numbers of *Slx* and *Slxl1* relative to *Sly* copies compared to non-hybrids, qualitatively consistent with the overexpression phenotypes observed in *Sly* knockdown and *Slx/Slxl1* duplication mice (Cocquet *et al.* 2012; Kruger *et al.* 2019). However, postmeiotic sex chromatin repression is thought to partially depend on repressive histone marks established during meiosis (Turner *et al.* 2006), and the same direction of the hybrid cross also shows disrupted MSCI in meiotic spermatocytes (Campbell *et al.* 2013; Larson *et al.* 2017). Thus, it remains unclear if the disruption of repressive postmeiotic chromatin is a consequence of X-Y mismatch or primarily a downstream epigenetic effect of deleterious interactions between the *musculus* X chromosome and *Prdm9* during meiosis (Larson *et al.* 2021).

Here, we advance understanding of the basis of hybrid male sterility in this system using a reciprocal backcrossing scheme to generate mice with the Y chromosome of one *Mus musculus* subspecies on the genomic background of another (Figure 1A). We used these Y-consomic genetic models to perform two reciprocal cross experiments while controlling for the effects of inbreeding. First, we tested for the potential rescue of sterility phenotypes in hybrid males with F1 autosomal genotypes but with matching X and Y chromosomes from the same subspecies (Experiment 1; Figure 1B). This experiment allowed us to tease apart X-Y interactions (i.e., *Slx* and *Slxl1* versus *Sly*) from X-autosomal interactions (i.e., *Prdm9*-related sterility). Second, we tested the effects of X-Y mismatch on different subspecific backgrounds (Experiment 2;

Figure 1B). This experiment allowed us to test for incompatibilities exposed on introgressed Y chromosomes that occur independently of other hybrid interactions. We used genome sequencing to quantify X- and Y-linked gene copy numbers, collected male reproductive phenotypes (testis weight and high-resolution sperm head morphology), and used fluorescence-activated cell sorting (FACS) to isolate cell populations enriched for either early meiotic leptotene-zygotene spermatocytes or postmeiotic round spermatids. We used these experiments to address three main questions: (i) Does X-Y mismatch cause abnormal male reproductive traits? (ii) Do differences in copy number predict differences in ampliconic gene family expression levels during late spermatogenesis? (iii) Is X-Y mismatch associated with disrupted gene expression during late spermatogenesis, particularly on the sex chromosomes?

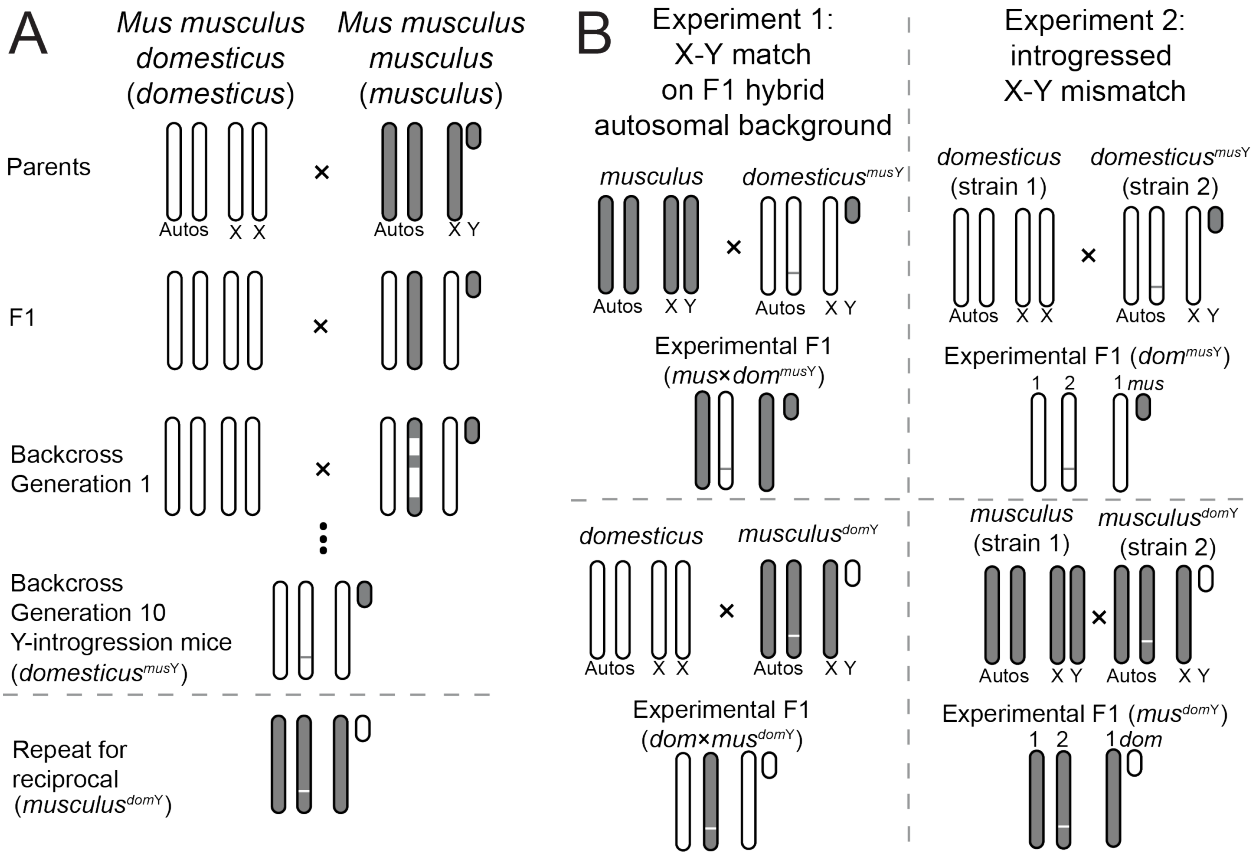

Figure 1: Experimental design. (A) Backcrosses used to generate Y-introgression mouse strains. We performed 10 generations of backcrosses in reciprocal directions to

generate mice with a *Mus musculus domesticus* (*domesticus*) genetic background and *Mus musculus musculus (musculus)* Y chromosome (*domesticus*^{musY}) and mice with a *musculus* genetic background and *domesticus* Y chromosome (*musculus*^{domY}). (B) Crosses were performed with Y-introgression mice to produce two types of experimental F1 mice. In Experiment 1, we crossed Y-introgression males to females from the other subspecies to generate F1 mice with hybrid autosomes but matched sex chromosomes. In Experiment 2, we crossed Y-introgression males to females from a different strain but the same subspecies to generate F1 mice with X-Y mismatch and non-hybrid autosomes. Autos = autosomes, X = X chromosome, Y = Y chromosome.

Materials and Methods

Mouse resources and experimental design

We used publicly available whole genome sequence data to estimate copy number in wild house mice (PRJEB9450 for *domesticus,* Pezer *et al.* 2015; PRJEB11742 for *musculus,* Harr *et al.* 2016) and wild-derived inbred laboratory mouse strains representing *musculus* (PWK/PhJ and CZECHII/EiJ) and *domesticus* (LEWES/EiJ and WSB/EiJ; PRJNA732719; Larson *et al.* 2021). We generated reciprocal consomic introgression strains with the Y chromosome from one subspecies on the genetic background of the other by backcrossing *musculus* (PWK) and *domesticus* (LEWES) for 10 generations, which we refer to as *musculus*^{domY} and *domesticus*^{musY} (Figure 1A). We then used these Y-introgression strains to perform two experiments and test the effects of X-Y mismatch on hybrid sterility independent of X-autosomal incompatibilities (Figure 1B).

Experiment 1: To test the effects of X-autosomal F1 incompatibilities without the effect of sex chromosome mismatch, we crossed Y-introgression males to females with the same autosomal and X chromosome type as the male Y chromosome (LEWES or PWK). This generated mice with an F1 hybrid autosomal background and X-autosomal mismatch but X and Y chromosomes from the same subspecies. Throughout the text, we refer to these mice as *mus*×*dom*^{musY} and *dom*×*mus*^{domY}. We compared these mice to standard F1

hybrid mice with the same X chromosome and autosomal background but no Y chromosome introgression (PWK♀ × LEWES♂, hereafter "*mus*×*dom*" and LEWES♀ × PWK♂, hereafter "*dom*×*mus*").

Experiment 2: To test the effects of X-Y mismatch while controlling for inbreeding effects, we crossed Y-introgression males to females from the same subspecies but a different strain from the genomic background of the Y-introgression strain (CZECHII or WSB). This generated mice with a non-hybrid (intrasubspecific) F1 autosomal background and mismatched sex chromosomes (i.e., no X-autosomal mismatch), which we will refer to as mus^{domY} and dom^{musY}. We compared these to intrasubspecific F1 mice with the same autosomal background as these F1 Y-introgression mice, but without sex chromosome mismatch (CZECHII♀ × PWK♂, hereafter "*mus*" and WSB♀ × LEWES♂, hereafter "*dom*"). Note that these Experiment 2 mice had X chromosomes from different laboratory strains than the Experiment 1 mice of the same subspecies as a necessary consequence of breeding mice with a heterozygous F1 background.

All mice from wild-derived inbred strains, Y-introgression strains, and experimental crosses were maintained in breeding colonies at the University of Montana (UM) Department of Laboratory Animal Resources (IACUC protocols 002-13, 050-15, and 062-18), which were initially purchased from The Jackson Laboratory, Bar Harbor, ME in 2010. Replacement stock of LEWES/EiJ mice were ordered in 2013, and these mice were used for the backcrosses to generate the dom^{musY} Y-introgression strains, as dames in the *dom* intrasubspecific F1s, and as sires in the *dom*×*mus* and *dom*×mus^{domY} crosses.

Whole genome sequencing and copy number estimation

We sequenced whole genomes from Y-introgression mice to estimate ampliconic gene family copy numbers. We extracted DNA from mouse liver using a Qiagen DNeasy kit and sent samples to Novogene (Novogene Corporation Inc., Sacramento, California) for library preparation and sequencing using Illumina HiSeq paired-end 150bp. Libraries

were prepared and sequenced twice to increase unique coverage. We trimmed raw reads with Trimmomatic version 0.39 (Bolger *et al.* 2014). We mapped reads to the mouse reference genome build GRCm38 using bwa mem version 0.7.17 (Li and Durbin 2009) and used picard version 2.18.29 to fix mates and mark duplicates (Picard Toolkit). Data from the two sequencing runs were then merged for each sample.

To identify paralogs of ampliconic gene families, we extracted known X (*Slx*, *Slxl1*, *Sstx*), Y (*Sly, Ssty1, Ssty2*), and autosomal (*Speer,* and *α-takusan*) ampliconic gene sequences from the mouse reference GRCm38 using Ensembl annotation version 102 (Yates *et al.* 2019). We used the predicted gene *Gm5926* for *Sstx* because *Sstx* was not annotated in this version of Ensembl. For the autosomal gene families, we used the longest annotated genes in the gene family (*α7-takusan* and *Speer4f2*). We performed Ensembl BLAT searches with these sequences against the GRCm38 mouse reference, allowing up to 1000 hits. We then extracted all BLAT hits with greater than or equal to 97% sequence identity and an e-value of 0.0 and considered these filtered BLAT hits to be gene family paralogs for downstream copy number estimation.

We estimated copy numbers using a relative coverage approach similar to (Morgan and Pardo-Manuel De Villena 2017) and AmpliCoNE (Vegesna *et al.* 2020). For the relative coverage approach, we used Mosdepth v0.3.2 (Pedersen and Quinlan 2017) to estimate coverage across paralogous regions and divided this sum by half the genome-wide average coverage to account for hemizygosity of the sex chromosomes in males.

AmpliCoNE also estimates copy number based on relative coverage, while also controlling for GC content and only using informative regions based on repeat masking and mappability. AmpliCoNE was developed for estimating copy number on the assembly and annotation of the human Y, so we made some modifications to allow AmpliCoNE to work with the mouse sex chromosomes (Larson *et al.* 2021; https://github.com/ekopania/modified-AmpliCoNE). Specifically, we replaced AmpliCoNE's method for identifying informative sites with an approach more suitable

for the mouse assembly. For each ampliconic gene family, we extracted all k-mers of length 101bp from the sequence of one gene representing the ampliconic family and mapped these back to the mouse reference genome using Bowtie2 and allowing up to 500 multiple mapping hits. For each gene, we identified the most frequent number of times (*m*) k-mers mapped to the mouse genome and kept only k-mers that mapped *m* times. We identified all locations where these k-mers mapped with 2 or fewer mismatches. We considered the start locations of these k-mer mapping hits to be "informative sites."

Identifying autosomal introgression in the Y-introgression strains

A small amount of autosomal material (~0.1%) is expected to have introgressed along with the Y chromosome in our backcross experiments. To test this theoretical expectation and identify regions of introgression, we mapped whole genome sequence data from Y-introgression strains to both parental genomes using bwa mem v0.7.17-r1188 (Li and Durbin 2009) and called variants with GATK HaplotypeCaller v4.2.2.0. We then counted the number of variants in 100kb windows across the autosomes and identified regions where the number of variants when mapped to the maternal parent (autosomal background) genome exceeded the number of variants when mapped to the paternal parent (Y-introgression) genome. We repeated this analysis using whole genome sequence data from PWK and LEWES samples in our mouse colony. We excluded regions that had more variants when mapped to the opposite strain than when mapped to the same strain, as these are likely regions where genotype calls are unreliable due to assembly issues. After excluding these regions, 100kb windows with at least two more variants when mapped to the maternal parent compared to the paternal parent were considered introgressed in Y-introgression strains, reflecting the 95th percentile of differences in the number of variants within a window.

Reproductive phenotypes

We phenotyped unmated male mice that were weaned at 21 days post-partum (dpp) into same-sex sibling groups and housed individually starting at 45 dpp to minimize effects of social dominance. We weighed paired testes and paired seminal vesicles and

calculated their mass relative to body weight. We compared offspring sex ratios from Y-introgression mice by recording the number of offspring of each sex at weaning. We then tested for a significant difference from an even sex ratio using a Pearson's chi-squared test in R, and did a power analysis for this chi-squared test using the *pwr.chisq.test* function in the pwr package in R.

To quantify sperm morphology, we extracted sperm from each cross type from cauda epididymides diced in 1mL Dulbecco's PBS (Sigma) and incubated at 37℃ for 10 minutes. Sperm were fixed in 2% PFA, then dropped onto a slide with DAPI solution to stain the sperm nuclei. We imaged greater than 400 nuclei per genotype and analyzed the images using the Nuclear Morphology Analysis software (Skinner *et al.* 2019). We used two microscopes but performed clustering analysis on combined nuclei imaged from both microscopes to ensure that nuclei imaged on one scope were not clustering separately from those taken on the other microscope (Supplemental Material, Figure S1). The Nuclear Morphology Analysis software uses a Canny edge detection algorithm to detect objects (nuclei) within images, orients and aligns the nuclei, and uses a modification of the Zahn-Roskies transformation of the nucleus outlines to automatically detect landmarks. The software estimates area, perimeter, bounding height, bounding width, regularity, difference from median, and a consensus shape of the nuclei for each genotype. We tested for significant differences among cross types for each of these parameters using a Wilcoxon rank sum test in R. Using this automated morphology analysis software, we were able to analyze 5652 nuclei and detect subtle but significant differences that may not be measurable by eye or qualitative analysis.

Testis sorting and RNA sequencing

We collected testes from mice immediately following euthanization and isolated cells at different stages of spermatogenesis using Fluorescence Activated Cell Sorting (FACS; Getun *et al.* 2011). The full FACS protocol is available on GitHub (https://github.com/goodest-goodlab/good-protocols/tree/main/protocols/FACS). Briefly, we decapsulated testes and washed them twice with 1mg/mL collagenase (Worthington Biochemical), 0.004mg/mL DNase I (Qiagen), and GBSS (Sigma), followed by disassociation with 1mg/mL trypsin (Worthington Biochemical) and 0.004mg/mL DNase

I. We then inactivated trypsin with 0.16mg/mL fetal calf serum (Sigma). For each wash and disassociation step, we incubated and agitated samples at 33°C for 15 minutes on a SciGene Model 700 Microarray Oven at approximately 10rpm. We stained cells with 0.36mg/mL Hoechst 33324 (Invitrogen) and 0.002mg/mL propidium iodide and filtered with a 40μm cell filter. For Experiment 1, we sorted using a FACSAria Fusion flow cytometer, and for Experiment 2 we sorted cells using a FACSAria IIu cell sorter (BD Biosciences), both at the UM Center for Environmental Health Sciences Fluorescence Cytometry Core. We periodically added 0.004mg/mL DNase I as needed during sorting to prevent DNA clumps from clogging the sorter. We sorted cells into 15μL beta-mercaptoethanol (Sigma) per 1mL of RLT lysis buffer (Qiagen) and kept samples on ice whenever they were not in the incubator or the cell sorter. We focused on two cell populations: early meiotic spermatocytes (leptotene/zygotene) and postmeiotic round spermatids. We extracted RNA using the Qiagen RNeasy Blood and Tissue Kit and checked RNA integrity with a TapeStation 2200 (Agilent). Only two samples had RNA integrity numbers (RIN) less than 8 (RIN = 7 and 7.1; Supplemental Material, Table S1). We prepared RNAseq libraries using the KAPA mRNA hyperprep kit and sequenced samples with Novogene (Illumina NovaSeq6000 PE 150). Samples were prepared and sequenced together, but Experiments 1 and 2 were done on different FACS sort machines, so to minimize experimental batch effects we analyzed these two experiments separately unless otherwise noted.

Gene expression analyses

We performed gene expression analyses on FACS expression data representing two cell populations: early meiosis (leptotene-zygotene, hereafter “early”) and postmeiosis (round spermatids, hereafter “late”). For the early cell type, a few samples did not group with others of the same cross type in multidimensional scaling (MDS) plots (Supplemental Material, Figure S2). These samples were likely contaminated with other cell types based on their relative expression levels of cell-type marker genes from *Mus musculus* testes single-cell RNAseq experiments (Supplemental Material, Figure S3; Green *et al.* 2018; Hunnicutt *et al.* 2021), and were therefore removed from expression analyses. Because sex chromosome ampliconic genes are primarily expressed in late

spermatogenesis (Mueller *et al.* 2013; Larson *et al.* 2018a), and disrupted sex chromosome expression in hybrid males primarily occurs after the early cell type stage (Larson *et al.* 2017), we focus on data from the late cell type in the main text and report results from the early cell type in the Supplemental Material.

We performed gene expression analyses using mice from both Experiments 1 and 2, and reanalyzed expression data from (Larson *et al.* 2017), which generated spermatogenesis cell-type enriched gene expression data from the same F1 hybrid crosses (PWK♀ × LEWES♂ and LEWES♀ × PWK♂) and intrasubspecific F1 crosses (CZECHII♀ × PWK♂ and WSB♀ × LEWES♂) used in this study.

We trimmed RNAseq reads using trimmomatic v0.39 (Bolger *et al.* 2014). One sample (PP.LL30.7MLZ) had about an order of magnitude more reads than any other sample (> 900 million raw reads), so we downsampled to the mean number of reads after trimming using fastq-sample version 0.8.3 and verified that reads were properly paired after downsampling using fastq_pair (Edwards and Edwards 2019). We quantified reads using a kmer-based quasi-mapping approach implemented in salmon v1.4.0 (Patro *et al.* 2017) and a salmon index based on the mouse reference transcriptome version GRCm38. We then converted from transcript-level counts to gene-level counts using the R packages tximport 1.14.2 and EnsDb.Mmusculus.v79. We used EdgeR version 3.32.1 to normalize expression data. First, we filtered out genes with low expression by only including genes that had an FPKM > 1 in at least 4 samples. Then, we normalized expression data following the recommendations in the tximport documentation.

We quantified expression levels of ampliconic gene families by calculating transcripts per million (TPM) for each gene separately then summing TPM values for all paralogs of a gene family (≥97% sequence identity). We used linear mixed-effect models to test if gene family expression level was significantly associated with copy number for *Slx, Slxl1, Sly, Ssty1, Ssty2,* and α*-takusan.* We compared disrupted expression levels on the autosomes, X chromosome, and Y chromosome by subtracting normalized FPKM values in control mice from normalized FPKM values in X-Y mismatch mice and control mice for every gene (Good *et al.* 2010). We then used a Mann-Whitney U test to compare the distribution of normalized FPKM differences

among the chromosome types. To identify differentially expressed (DE) genes between cross types, we used the likelihood ratio test approach with false-discovery rate (FDR) correction in EdgeR and visualized overlaps in DE genes among cross types using the R package UpSetR (Conway *et al.* 2017). We removed DE genes in autosomal regions we identified as putatively introgressed, because these genes may be DE due to introgressed autosomal variants rather than incompatibilities resulting from mismatching sex chromosomes. We further investigated genome-wide expression differences among cross types using weighted correlation network analyses (WGCNA; Langfelder and Horvath 2008). We identified correlated expression modules significantly associated with different cross types using a linear model and Tukey's honest significant difference (HSD) test. We used R version 4.0.3 for all statistical tests and to implement all R packages (R Core Team).

Data availability

Whole genome sequence data from Y-introgression strains and RNAseq data from testes cell sort populations are publicly available on the Sequence Read Archive, accession numbers PRJNA816542 (whole genome) and PRJNA816886 (RNAseq). Raw phenotype data are available in the Supplemental Material, Table S2.
Scripts used to modify the AmpliCoNE program for copy number estimation are publicly available at: https://github.com/ekopania/modified-AmpliCoNE. Scripts used for gene expression analyses are available at: https://github.com/ekopania/xy_mismatch_expression_analyses.

Results

Copy Number Imbalance in Y-introgression Mice

We first estimated ampliconic gene family copy numbers in wild mice, wild-derived inbred strains, and Y-introgression mice using whole genome sequencing. The samples that we sequenced had genome-wide average coverages of 10-15×, and samples with publicly available data all had coverage >5×. We found that *musculus* tended to have higher *Slx* and *Sly* copy numbers than *domesticus* (median *Slx* copy number in *musculus*: 62, in *domesticus*: 17, FDR-corrected Wilcoxon rank sum $P < 0.01$; median

Sly copy number in *musculus*: 226, in *domesticus*: 109, FDR-corrected Wilcoxon rank sum P < 0.01), qualitatively consistent with previous studies (Ellis *et al.* 2011; Case *et al.* 2015; Morgan and Pardo-Manuel De Villena 2017; Figure 2A). *Slxl1* copy numbers also tended to be higher in *musculus,* but there was high copy number variation for this gene family in *domesticus* with some samples reaching copy numbers as high as those found in *musculus* (median *Slxl1* copy number in *musculus*: 37, in *domesticus*: 31, FDR-corrected Wilcoxon rank sum P < 0.01; Figure 2B). *Slx, Slxl1,* and *Sly* copy numbers for wild-derived inbred strains were representative of those found in wild mice (Figures 2A and 2B; Supplemental Material, Table S3), consistent with previous results (Larson *et al.* 2021). Our Y-introgression mice retained copy numbers similar to those of pure strains with the same X and Y chromosome genotypes, so they had *Slx*-*Sly* and *Slxl1*-*Sly* dosage imbalance similar to that expected in natural hybrids (Figures 2A and 2B; Supplemental Material, Table S3).

Additional ampliconic gene families showed copy number differences between *musculus* and *domesticus* that were also represented in our Y-introgression mice. *Sstx* had similar copy numbers in *musculus* and *domesticus,* but its two Y-linked homologs showed differences between subspecies, with *Ssty1* having more copies in *domesticus* and *Ssty2* having more copies in *musculus* (median *Sstx* copy number in *musculus*: 48, in *domesticus*: 39, FDR-corrected Wilcoxon rank sum P = 0.57; median *Ssty1* copy number in *musculus*: 74, in *domesticus*: 139, FDR-corrected Wilcoxon rank sum P < 0.01; median *Ssty2* copy number in *musculus*: 145, in *domesticus*: 92, FDR-corrected Wilcoxon rank sum P < 0.01; Figure 2C, 2D).

We also estimated copy number for α*-takusan* and *Speer*, two autosomal ampliconic gene families thought to be regulated by sex chromosome ampliconic genes (Moretti *et al.* 2020). In both males and females, α*-takusan* showed a high correlation in copy number with *Slx* (r = 0.95; Pearson's correlation P < 0.001), suggesting that it was co-amplified with the *Slx* gene family (Figure 2E). Note that correlation tests were performed without phylogenetic correction, because we wanted to test if gene families were co-amplified regardless of whether this was a result of shared evolutionary history. *Speer* copy number was more difficult to estimate using our approaches due to lower sequence similarity among *Speer* paralogs compared to other ampliconic gene families,

but our estimates suggested that *Speer* may also have higher copy number in *musculus* relative to *domesticus* (Supplemental Material, Table S3). To verify our computational copy number estimates, we also performed digital droplet PCR (ddPCR) on a subset of *dom* samples using the *Slxl1* primers from (Kruger *et al.* 2019). We found 15 *Slxl1* copies with ddPCR, consistent with findings in (Kruger *et al.* 2019). While our computational estimates are higher than this, we found similar results if we imposed a stricter cutoff for considering genes paralogs (98-99% sequence identity), likely reflecting a high specificity of the primers we used. We also found similar results using a different computational approach based on relative coverage (Supplemental Material, Table S3; Larson *et al.* 2021).

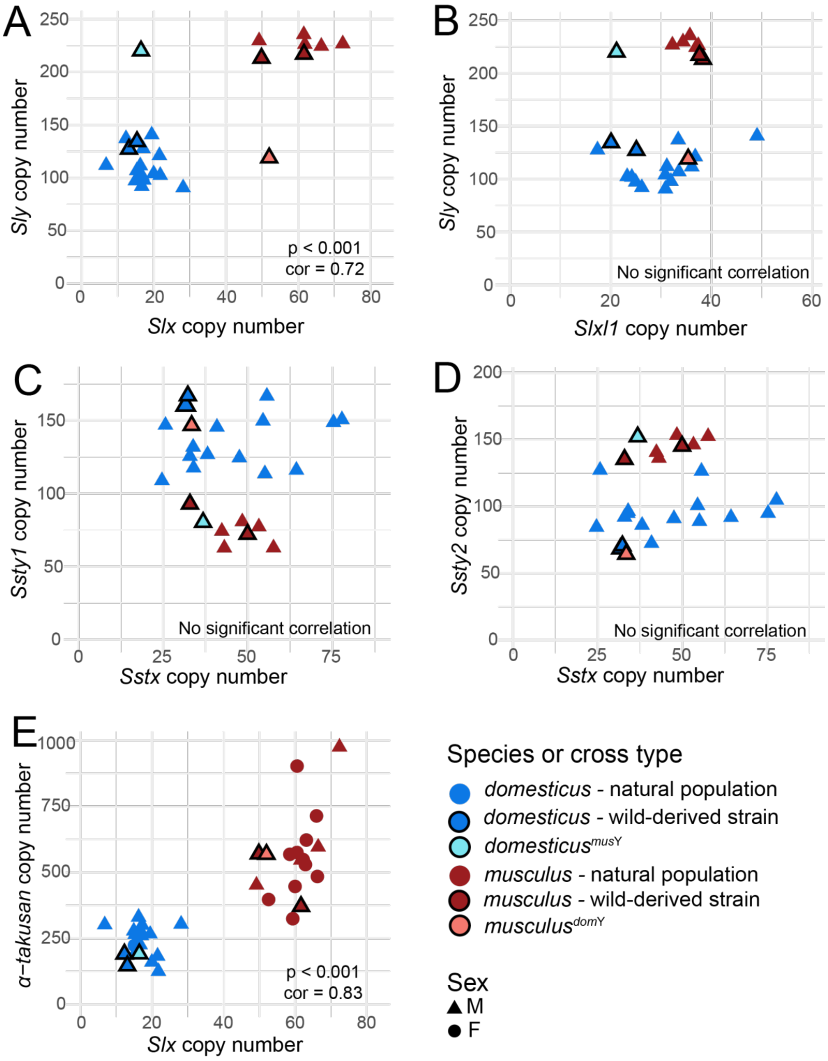

Figure 2: Copy number estimates for ampliconic gene families in wild mice, wild-derived inbred strains, and Y-introgression strains. Copy number was estimated using a 97% identity cutoff for paralogs. (A-D) show copy numbers in male mice, with Y chromosome genes on the y-axis and their X chromosome homologs on the x-axis. (E) includes both males and females and shows haploid copy number for the autosomal gene family α-*takusan* on the y-axis and haploid copy number for the X-linked family *Slx* on the x-axis. Points outlined in black represent wild-derived inbred strains or Y-introgression strains used in experimental and control crosses in this study. Correlations and p-values are based on a Pearson's correlation test. P-values were FDR-corrected for multiple tests.

Residual Autosomal Introgression in Y-introgression strains

We divided the reference genome autosomal regions into 24,639 100kb windows and mapped samples representing the LEWES and PWK inbred strains to both reference genomes. We found evidence for introgression in 105 windows in *domesticus*^{musY}, and 33 windows in *musculus*^{domY}, representing 0.43% and 0.13% of the autosomal windows that passed filtering, respectively (Supplemental Material, Table S4). Thus, the *domesticus*^{musY} strain had approximately four times more introgression than the theoretical expectation of 0.1% based on the number of backcross generations. Note that this theoretical expectation is likely conservative, because the Y-introgression strains were sequenced after about six to nine generations of additional inbreeding. The relatively large difference in percentages of introgression between the strains was primarily due to an ~7.6 Mbp introgressed region on chromosome 2 in *domesticus*^{musY} (Supplemental Material, Figure S4). Four introgressed windows were the same between the two reciprocal Y-introgression strains, which is more than expected by chance (hypergeometric test $P << 0.001$; Supplemental Material, Table S4). These included one region on chromosome 12 and three nearby regions on chromosome 13, one of which contains the gene *Nlrp4f*, which is involved in female fertility (Smith *et al.* 2019). Of the putative introgressed regions, 29 windows in *domesticus*^{musY} and 28 windows in *musculus*^{domY} were not adjacent to any other window with evidence for introgression, so they likely do not represent long tracks of introgression. Additionally, the median

difference in number of SNPs across autosomal introgressed windows was eight, after excluding the large introgressed region on chromosome 2, so there was unlikely to be introgression across the full 100kb window for most windows. In contrast, the large introgressed region on chromosome 2 had an average difference of 958 SNPs. Thus, the introgressed region on chromosome 2 in the *domesticus*$^{mus Y}$ strain likely represents the only large track of autosomal introgression, with some evidence for additional, smaller amounts of introgression throughout the autosomes in both reciprocal Y-introgression strains.

Some of the putatively introgressed regions we identified may be prone to introgression more generally. The large area on chromosome 2 overlapped with a region with evidence for introgression from *musculus* into the *domesticus* wild-derived inbred strains STRA and STRB (Mukaj *et al.* 2020). We used the Mouse Phylogeny Viewer (Yang *et al.* 2011) to identify an additional nine mouse inbred strains with introgression from *musculus* into a *domesticus* background in this region (Supplemental Material, Figure S4C). In one area of the mouse hybrid zone, a SNP contained within this introgressed region showed evidence for excess of the *musculus* allele in mice with primarily *domesticus* backgrounds, suggesting that introgression of this region from *musculus* into *domesticus* may have occurred in wild populations (Teeter *et al.* 2010). This region is also adjacent to *R2d2*, a copy number variant in mice that shows transmission ratio distortion in females heterozygous for the high copy number *R2d2* drive allele (Didion *et al.* 2016). We also identified 5 different 100kb windows near each other on chromosome 14 with evidence for introgression in *musculus*$^{dom Y}$ mice that overlap with a region in the *musculus* wild-derived strain PWD with evidence for introgression from *domesticus* (41.3-41.4Mb, 41.8-41.9Mb, 42.2-42.3Mb, 42.3-43.4Mb, and 44.2-44.3Mb; Mukaj *et al.* 2020).

X-Y Mismatch Contributed to Male Sterility Phenotypes

We next asked if X-Y mismatch was associated with male sterility phenotypes (Table 1). For Experiment 1, where we compared hybrid mice both with and without sex chromosome mismatch, hybrids with a *musculus*♀ × *domesticus*♂ background had lower relative testes mass than hybrids with the reciprocal *domesticus*♀ × *musculus*♂

background regardless of whether they had X-Y mismatch or not (Figure 3A). These results were consistent with previous studies showing more severe hybrid sterility in the *musculus*♀ × *domesticus*♂ direction of this cross (Good *et al.* 2008b; Good *et al.* 2010; Campbell *et al.* 2012; Larson *et al.* 2017). For Experiment 1 mice, *dom*×*mus*^{domY} mice had higher relative testis mass than *dom*×*mus* mice, suggesting that X-Y match partially rescued relative testes mass in some mice with a hybrid autosomal background (Figure 3A). In the reciprocal direction, however, X-Y match had no significant effect on relative testes mass (Figure 3A). For Experiment 2, we found that mice with X-Y mismatch had reduced relative testis mass compared to control mice with the same non-hybrid X and autosomal background (Figure 3A). In summary, we found little effect of X-Y mismatch on testis mass in the most sterile F1 cross (*musculus*♀ × *domesticus*♂), where sterility is therefore likely due to X-autosomal or autosomal-autosomal incompatibilities (Campbell and Nachman 2014). However, in the reciprocal and more fertile F1 direction X-Y mismatch seemed to have an important effect on testis mass. Furthermore, in the absence of any autosomal or X-autosomal incompatibilities, X-Y mismatch resulted in slightly but significantly decreased relative testis mass.

We saw severe sperm head abnormalities in our Experiment 1 crosses with a *musculus*♀ × *domesticus*♂ background (*mus*×*dom* and *mus*×*dom*^{musY}). Sperm from both these cross types had significantly lower bounding height and bounding width compared to all other cross types (FDR-corrected Wilcoxon rank sum $P << 0.0001$; Table 1), largely due to their shortened hook and consistent with hybrid sterility in this direction of the cross (Figure 3B). This was also consistent with previous manual (categorical) observations of abnormal sperm head morphology in this cross type in other studies (Good *et al.* 2008a; Campbell and Nachman 2014; Larson *et al.* 2017; Larson *et al.* 2018b). The reciprocal *dom*×*mus* F1 hybrids had sperm with much higher bounding height and bounding width compared to sperm from all other cross types, including the reference subspecies (FDR-corrected Wilcoxon rank sum $P < 0.01$; Table 1; Figure 3B; Supplemental Material, Figure S5). This direction of the cross is generally considered more fertile but sometimes shows reduced fertility compared to non-hybrid mice (Larson *et al.* 2018b). It is possible that the larger overall size of these sperm may reflect abnormal nuclear packaging and could contribute to reduced fertility in *domesticus*♀ ×

musculus♂ F1 mice. When comparing X-Y match mice to F1 hybrids with abnormally small sperm heads, *mus*×*dom*^{musY} mice had significantly higher bounding width and bounding height than *mus*×*dom* mice (FDR-corrected Wilcoxon rank sum $P < 0.01$; Table 1; Figure 3B; Supplemental Material, Figure S5). These results suggest that X-Y match rescued some of the aberrant sperm head morphology associated with hybrid sterility in *musculus*♀ × *domesticus*♂ F1s, but the effects of X-Y match rescue were subtle, consistent with previous observations (Campbell and Nachman 2014). In the reciprocal cross direction, *dom*×*mus*^{domY} had lower bounding width and bounding height than the abnormally large *dom*×*mus* sperm heads (FDR-corrected Wilcoxon rank sum $P << 0.0001$; Table 1; Figure 3B; Supplemental Material, Figure S5), so X-Y match rescued some of the oversized sperm head morphology we observed in *dom*×*mus*.

In Experiment 2, we observed subtle effects of X-Y mismatch consistent with our Experiment 1 observations. Sperm from *mus*^{domY} mice had slightly lower bounding height and bounding width compared to sperm from *mus* (FDR-corrected Wilcoxon rank sum $P < 0.01$; Table 1; Figure 3B; Supplemental Material, Figure S5), consistent with lower bounding height and bounding width in sperm from *mus*×*dom* mice that also had a *mus* X chromosome and *dom* Y chromosome. However, *mus*^{domY} sperm were more similar in size to *mus* sperm than *mus*×*dom* sperm and qualitatively had a hook morphology more similar to that of fertile *mus* than sterile *mus*×*dom* mice, so the contribution of X-Y mismatch to sperm head morphology is small compared to the effect of X-autosomal interactions. In the reciprocal direction, *dom*^{musY} mice had sperm with higher bounding height and bounding width compared to sperm from *dom* mice (FDR-corrected Wilcoxon rank sum $P << 0.0001$; Table 1; Figure 3B; Supplemental Material, Figure S5), consistent with the higher bounding height and bounding width in *dom*×*mus* hybrids. Sperm from *dom*^{musY} mice also had smaller areas (FDR-corrected Wilcoxon rank sum $P << 0.0001$; Table 1; Supplemental Material, Figure S5), so the larger bounding height and bounding width are primarily the result of a slightly elongated hook rather than an overall increase in the sperm head size. Other sperm head morphology parameters, including area, perimeter, and differences from median, showed similar subtle differences or no differences among cross types (Table 1; Supplemental Material, Figures S1 and S5).

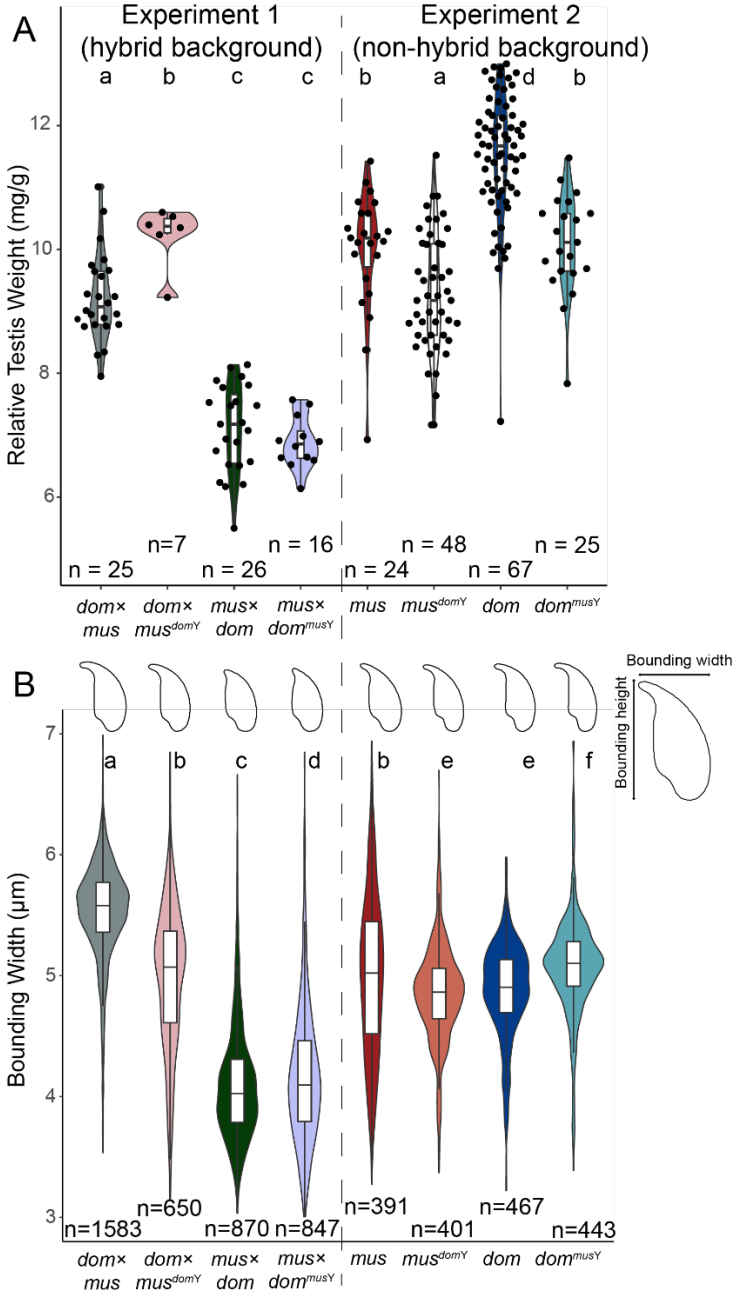

Figure 3: (A) Relative testes mass (mg/g) and (B) sperm nucleus bounding width (µm) by cross type. Letters above each violin plot indicate significant differences (FDR-corrected $P < 0.05$) based on a Welch's t-test (relative testes mass) or Wilcoxon rank-sum test (bounding width). Sample size for each cross type is indicated below each violin plot. Bounding width sample sizes indicate the number of sperm nuclei observed. Representative sperm nuclei morphologies for each cross type are depicted above each violin plot in (B).

Genetic manipulation studies have shown offspring sex ratio skews under *Slxl1-Sly* dosage imbalance, contributing to evidence for *Slxl1-Sly* intragenomic conflict. Male mice with an excess of *Sly* relative to *Slxl1* produce more male offspring, while mice with an excess of *Slxl1* produce more female offspring (Cocquet *et al.* 2012; Kruger *et al.* 2019) due to reduced motility of Y-bearing sperm (Rathje *et al.* 2019). We asked if more subtle imbalances in relative copy numbers expected in natural hybrid mice also result in sex ratio skews and did not see a significant difference from a 50:50 sex ratio for offspring of X-Y mismatch mice (Supplemental Material, Table S5). A more extreme dosage imbalance than that seen in our X-Y mismatch experimental mice (and in natural hybrids) is probably required to produce a large sex ratio skew. However, it is important to note that we had very little power to detect differences in sex ratio, with type II error probabilities over 0.8 (Supplemental Material, Table S5).

Table 1: Reproductive phenotypes for experimental X-Y mismatch mice and controls. Median values are presented +/- 1 standard error. Sample sizes are in parentheses. For sperm morphology parameters (bounding height, bounding width, area, perimeter, difference from median [a measure of the variability of nuclear shapes within the sample]), sample sizes indicate the number of sperm heads observed, and variance is depicted in violin plots (Figure 3; Supplemental Material, Figure S5). Gray boxes indicate significant differences (FDR-corrected Wilcoxon rank sum test $P < 0.05$) between X-Y mismatch cross types and control cross types with the same autosomal background. (‡) Indicates phenotypes with significant differences (FDR-corrected pairwise Wilcoxon rank sum test $P < 0.05$) between *mus*×*dom* F1 hybrids and both parental subspecies (*mus* and *dom*). (*) Indicates phenotypes with significant differences (FDR-correct pairwise Wilcoxon rank sum test $P < 0.05$) between *dom*×*mus* F1 hybrids and both parental subspecies (*mus* and *dom*). Testes and seminal vesicle weights are both paired. SV = seminal vesicle

	Experiment 1 (hybrid background)				**Experiment 2 (non-hybrid background)**			
Phenotype	***dom*×*mus***	***dom*× *mus*^{domY}**	***mus*×*dom***	***mus*× *dom*^{musY}**	***mus***	***mus*^{domY}**	***dom***	***dom*^{musY}**

Body mass (g)	20 +/- 0.3 (24)	19.6 +/- 0.3 (7)	17.9 +/- 0.4 (24)	18 +/- 0.4 (12)	19 +/- 0.4 (23)	18.4 +/- 0.3 (47)	18 +/- 0.2 (67)	19 +/- 0.5 (21)
Testes mass (mg)‡	186.4 +/- 3 (24)	200.7 +/- 2 (6)	123.9 +/- 2 (23)	125.6 +/- 3 (12)	193.2 +/- 5 (23)	172.7 +/- 2 (47)	209.1 +/- 3 (67)	189.3 +/- 6 (21)
Relative testes mass (mg/g)*‡	9.1 +/- 0.1 (24)	10.4 +/- 0.2 (6)	7.2 +/- 0.1 (23)	6.9 +/- 0.1 (12)	10.2 +/- 0.2 (23)	9.2 +/- 0.1 (47)	11.7 +/- 0.1 (67)	10.1 +/- 0.2 (21)
Relative SV mass (mg/g)	6.6 +/- 0.2 (23)	7.3 +/- 0.6 (6)	5.2 +/- 0.3 (24)	5.3 +/- 0.3 (12)	6 +/- 0.3 (23)	6.7 +/- 0.2 (47)	5.2 +/- 0.2 (65)	5.9 +/- 0.3 (21)
Bounding height‡*	8.39 (1583)	8.14 (650)	7.46 (870)	7.52 (847)	8.21 (391)	8.02 (401)	8.11 (467)	8.23 (443)
Bounding width‡*	5.58 (1583)	5.07 (650)	4.02 (870)	4.09 (847)	5.02 (391)	4.87 (401)	4.9 (467)	5.11 (443)
Area‡*	24.5 (1583)	21.6 (650)	20.1 (870)	20.1 (847)	22.1 (391)	20 (401)	21.3 (467)	20.4 (443)
Perimeter‡*	23.8 (1583)	22.7 (650)	19.8 (870)	20.2 (847)	22.7 (391)	21.9 (401)	22.3 (467)	23 (443)
Difference from median‡*	6.22 (1583)	8.67 (650)	8.22 (870)	10.8 (847)	8.88 (391)	5.77 (401)	5.86 (467)	6.72 (443)

Slx- and *Slxl1-Sly* Dosage Imbalance Did Not Lead to Ampliconic Gene Family Overexpression

Copy number imbalance of *Slx* and *Slxl1* relative to *Sly* is thought to disrupt expression of these gene families in late spermatogenesis, with particularly strong evidence for *Slx* and *Slxl1* overexpression when *Sly* is knocked down (Cocquet *et al.* 2009; Cocquet *et al.* 2012) and *Slxl1* overexpression when *Slx* and *Slxl1* are duplicated (Kruger *et al.* 2019). *Slx*, *Slxl1,* and *Sly* appear to be involved in the regulation of sex chromatin which impacts the regulation of many genes during late spermatogenesis (Kruger *et al.* 2019). Therefore, we predicted that their misregulation may disrupt the expression of additional genes, including additional Y-linked ampliconic gene families *Ssty1/2* and the autosomal ampliconic gene family α*-takusan* (Larson *et al.* 2017; Moretti *et al.* 2020). To test if *Slx, Slxl1, Sly, Ssty1, Ssty2,* and α*-takusan* expression was disrupted under less extreme copy number differences in hybrid mice, we compared ampliconic gene family expression levels in round spermatids among cross types. We did not directly quantify

copy number for the mice that were FACS sorted, so we used our previous copy number estimates from pure strains sharing the same sex chromosomes as our experimental mice (Larson *et al.* 2021). For all six gene families, expression level was significantly associated with copy number based on a linear mixed-effects model with experiment as a random effect to control for batch effects (FDR-corrected $P < 0.05$; Figure 4). However, for *Slxl1,* this association was negative, suggesting that copy number was not the primary determinant of *Slxl1* expression. This is interesting given that we found high overlap in the range of *Slxl1* copy numbers in naturally occurring *musculus* and *domesticus* (Figure 2B), and the previous demonstration that *Slxl1* plays a more direct role in sex ratio bias than *Slx* (Kruger *et al.* 2019). We then tested if X-Y mismatch had a significant effect on expression level using a linear mixed-effects model with both copy number and presence of X-Y mismatch as fixed effects and experiment as a random effect. We used an ANOVA to compare this model to a null model with copy number as the only fixed effect and experiment as a random effect. For all six genes, X-Y mismatch was not significantly associated with ampliconic gene expression levels (FDR-corrected ANOVA $P > 0.05$). When we specified the direction of X-Y mismatch (i.e., *musculus* X and *domesticus* Y, the direction with an excess of *Slx* relative to *Sly*), only *Ssty2* expression was significantly associated with X-Y mismatch in this direction (FDR-corrected ANOVA $P > 0.05$).

We also tested if X-autosomal background was significantly associated with expression levels using the same mixed-effects model approach. For *Slx, Slxl1, Sly*, *Ssty1,* and *Ssty2,* the sterile hybrid background (*musculus*♀ × *domesticus*♂) was significantly associated with expression levels after FDR-correction (*Slx* ANOVA $P << 0.0001$; *Slxl1* $P < 0.001$; *Sly* $P = 0.01$; *Ssty1* $P < 0.001$; *Ssty2* $P = 0.001$). We observed overexpression of *Slx, Slxl1, Sly, Ssty1,* and *Ssty2* relative to their copy numbers for mice with *musculus*♀ × *domesticus*♂ backgrounds (*mus*×*dom* and *mus*×*dom*$^{mus\mathrm{Y}}$; Figure 4A-E), consistent with previous studies showing that these hybrid mice exhibit widespread overexpression on the sex chromosomes (Good *et al.* 2010; Campbell *et al.* 2013; Larson *et al.* 2017). Both *mus*×*dom* and *mus*×*dom*$^{mus\mathrm{Y}}$ mice in our study overexpressed *Slx, Slxl1,* and *Sly* (Figure 4A, 4B, and 4C), suggesting that matching X and Y chromosomes from *musculus* did not rescue *Slx, Slxl1,* or *Sly* upregulation, and

that the overexpression we observed likely results from X-autosomal incompatibilities that disrupt MSCI rather than *Slx*- or *Slxl1-Sly* dosage imbalance. Additionally, *mus*$^{dom\mathrm{Y}}$ mice from our Experiment 2 also had a *musculus* X and *domesticus* Y, the same X and Y chromosome combination found in sterile hybrids that results in an excess of *Slx* and *Slxl1* copies relative to *Sly* copies. If *Slx*- or *Slxl1-Sly* dosage imbalance contributed to *Slx, Slxl1,* and *Sly* overexpression, we would expect *mus*$^{dom\mathrm{Y}}$ mice to have higher expression than *mus* controls. We observed the opposite effect, with *mus*$^{dom\mathrm{Y}}$ mice showing slightly lower *Slx, Slxl1,* and *Sly* expression levels (Figure 4A, 4B, and 4C). This result provides further evidence that postmeiotic *Slx, Slxl1*, and *Sly* overexpression in sterile F1 hybrids is unlikely to be primarily due to *Slx*- or *Slxl1-Sly* dosage imbalance, and that X-Y mismatch in the absence of autosomal mismatch is not sufficient to cause overexpression of *Slx, Slxl1,* and *Sly*.

Given that *Slx, Slxl1,* and *Sly* are thought to regulate the α*-takusan* ampliconic family, we predicted that α*-takusan* expression levels would also be associated with a *musculus*♀ × *domesticus*♂ background. Surprisingly, this association was not significant (ANOVA P = 0.40). Instead, we observed that α*-takusan* was overexpressed in all cross types with an F1 autosomal background regardless of cross direction (Figure 4F), and that expression was significantly associated with an F1 autosomal background (ANOVA P < 0.01). This suggests that α*-takusan* regulation likely involves autosomal loci in addition to SLX, SLXL1, SLY, SSTY1, and SSTY2 (Moretti *et al.* 2020).

Sex-linked ampliconic genes are primarily expressed during postmeiotic spermatogenesis, in mice and more generally across mammals (Cocquet *et al.* 2012; Mueller *et al.* 2013; Sin and Namekawa 2013). Our non-hybrid expression data supported this, with little to no expression of *Slx, Slxl1, Sly,* or *Ssty1/2* in early meiotic cells in our *mus* and *dom* samples. However, we did detect some meiotic expression of *Slx, Slxl1, Sly,* and *Ssty2* in mice with hybrid autosomal backgrounds, and expression levels of these gene families in early meiosis was significantly associated with F1 autosomal background (ANOVA P < 0.05, Supplemental Material, Figure S6). X chromosome expression has been shown to be disrupted throughout spermatogenesis in F1 hybrids, although the effect was smaller during earlier spermatogenic stages (Larson *et al.* 2017). Our results suggest that disruption of early spermatogenesis

regulatory networks may result in spurious expression of sex-linked ampliconic genes during early meiotic stages when they are normally silenced.

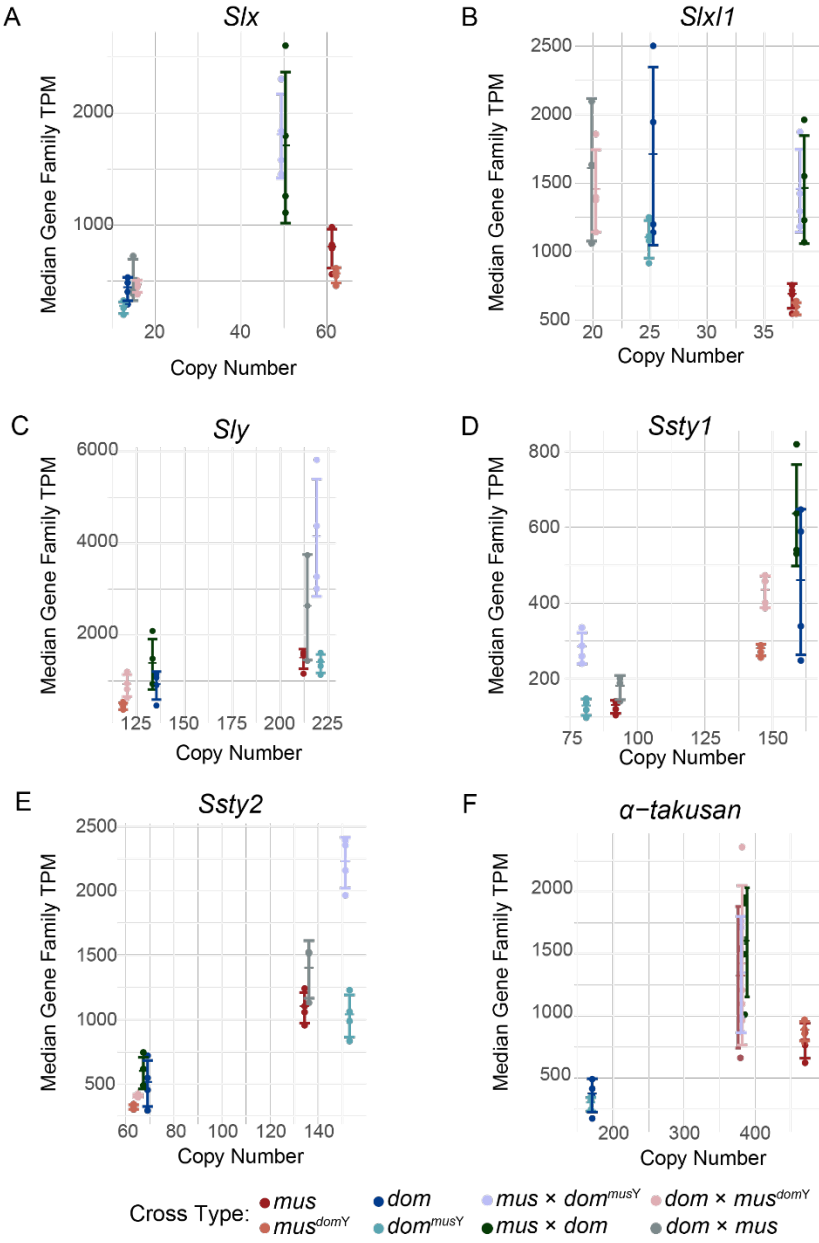

Figure 4: Normalized expression levels of *Slx* (A), *Slxl1* (B), *Sly* (C), *Ssty1* (D), *Ssty2* (E), and *α-takusan* (F) ampliconic gene families in different cross types plotted against their copy numbers. Copy number estimates are based on estimates from wild-derived strains used in experimental and control crosses (see Figure 2). Cross types with the same sex chromosome and therefore same copy number estimate are jittered slightly along the x-axis for clarity. Expression level was calculated by summing transcripts-per

million (TPM) for each paralog of the gene family with at least 97% sequence identity to the ampliconic gene. Points represent values for individual samples, and lines indicate median and standard deviation for each cross type.

X-Y Mismatch was not associated with Sex Chromosome Overexpression in Sterile F1 Hybrids

Next we sought to differentiate if widespread postmeiotic overexpression in sterile hybrids was a direct result of sex chromosome mismatch, a continuation of disrupted meiotic sex chromosome inactivation (MSCI), or a combination of both (Larson *et al.* 2017; Larson *et al.* 2021). We first reanalyzed data from (Larson *et al.* 2017) and repeated their result showing sex chromosome upregulation in late spermatogenesis in sterile F1 hybrids (*mus*×*dom*, Figure 5A and 5D). We then tested if upregulation was due to X-Y mismatch by comparing relative expression levels in F1 hybrids to those in our Experiment 1 mice, which had sex chromosomes from the same subspecies. If X-Y mismatch contributed to sex chromosome upregulation in sterile hybrids, we would expect to see some rescue from disrupted postmeiotic expression in these Experiment 1 mice, with *mus*×*dom*^{musY} mice having lower expression on the X chromosome relative to *mus*×*dom* F1s. Contrary to this prediction, the X chromosome showed similar expression levels when comparing expression in these two cross types. Therefore, restoring matching sex chromosomes did not rescue expression levels on the *musculus* X chromosome from overexpression in hybrids (Figure 5B). We further tested the effects of sex chromosome mismatch using our Experiment 2 mice, which had introgressed Y chromosomes on a non-hybrid autosomal background. If mismatch between a *musculus* X chromosome and *domesticus* Y chromosome was sufficient to induce postmeiotic sex chromosome overexpression, then we would expect to see higher X chromosome expression in *mus*^{domY} mice. Instead, we observed slight under expression on the X chromosome compared to the autosomes in *mus*^{domY} mice, confirming that sex chromosome mismatch does not cause X chromosome overexpression in late spermatogenesis (Figure 5C).

We also found evidence that sex chromosome mismatch does not contribute to Y chromosome overexpression in late spermatogenesis in sterile *musculus*♀ ×

domesticus♂ hybrids. The Y chromosome was upregulated in *mus*×*dom* sterile hybrids relative to *dom*×*mus*^{domY} mice. This could be due to rescue of *domesticus* Y chromosome expression when paired with the *domesticus* X, but it could also be due to overall lower sex chromosome expression in mice with a *domesticus*♀ × *musculus*♂ background (Figure 5E). In Experiment 2, we saw that *mus*^{domY} mice had lower expression on the Y chromosome compared to *dom* controls, in contrast to the Y chromosome overexpression observed in *mus*×*dom* hybrids (Figure 5F). Thus, X-Y mismatch does appear to influence Y chromosome expression, but in the opposite direction of that observed in sterile hybrids.

In the reciprocal cross (*domesticus*♀ × *musculus*♂ F1 hybrids), we found some evidence that X-Y mismatch may contribute to disrupted expression of X-linked genes. Here Y chromosome expression was not different from that on the autosomes (Figure 5G), but the X chromosome tended to be downregulated (Figure 5J; Larson *et al.* 2017). There was no evidence that X-Y match restored normal X chromosome expression levels in *dom*×*mus*^{domY} (Experiment 1), with this cross type showing similar or even slightly lower expression levels on the X chromosome relative to *dom*×*mus* hybrids (Figure 5K). However, in Experiment 2 we observed lower expression on the X chromosome in *dom*^{musY} mice relative to *dom* controls (Figure 5L). Therefore, a *domesticus* X paired with a *musculus* Y can result in suppression of X-linked gene expression even in the absence of autosomal incompatibilities.

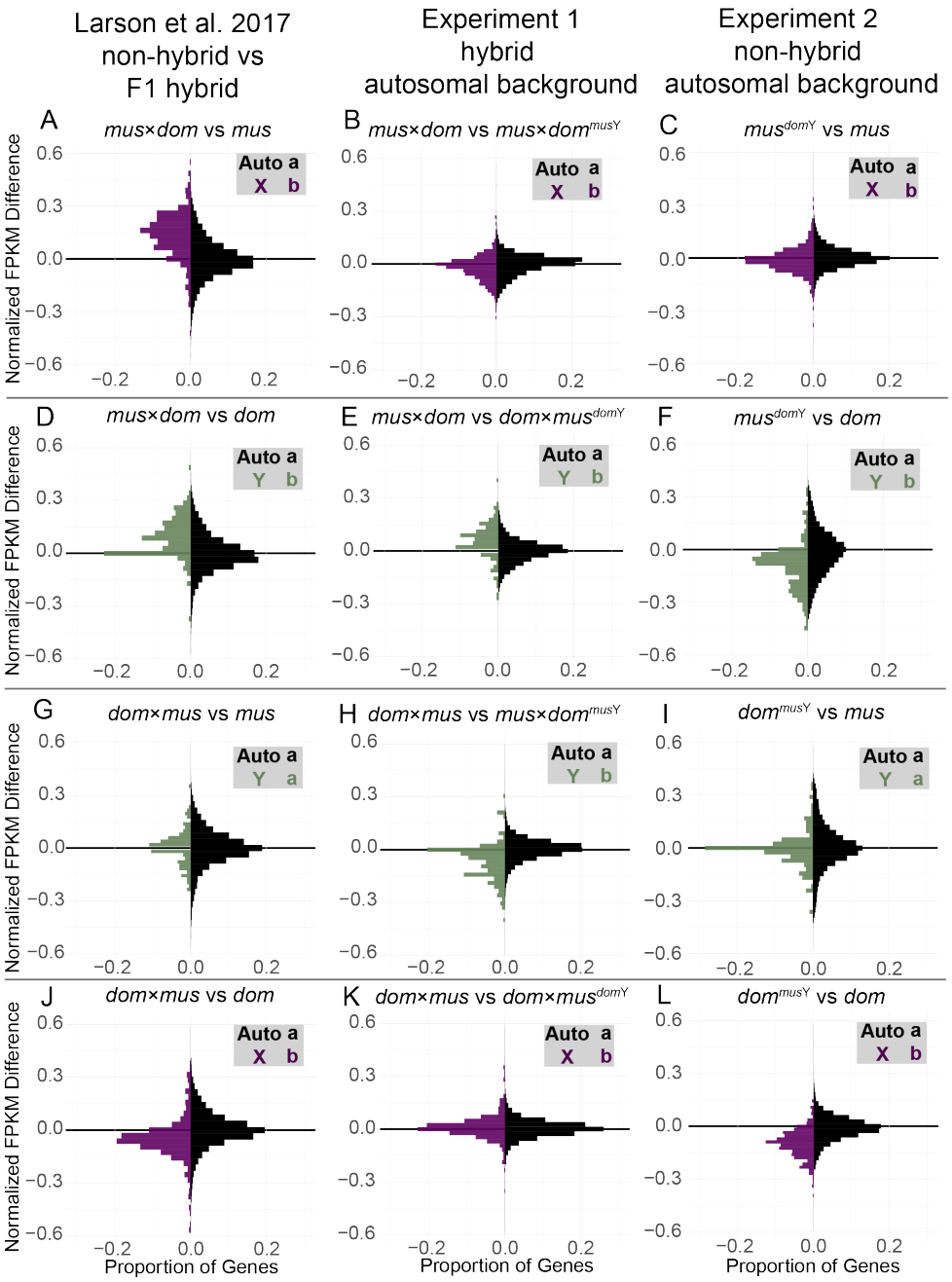

Figure 5: Histograms of relative expression levels between experimental cross types and control mice. (A-C) Contrasts that all have a *musculus* X chromosome, (D-F) contrasts with a *domesticus* Y chromosome (G-I) contrasts with a *musculus* Y chromosome, and (J-L) contrasts with a *domesticus* X chromosome. (A-F) represent sex chromosome mismatch present in sterile hybrids (*musculus* X and *domesticus* Y), while (G-L) represent sex chromosome mismatch present in more fertile hybrids (*domesticus* X and *musculus* Y). The first column (A, D, G, and J) shows data

reanalyzed from (Larson *et al.* 2017). The second column (B, E, H, K) tests if gene expression levels are rescued when the sex chromosomes are matched but on a hybrid autosomal background (Experiment 1). The third column (C, F, I, L) tests for disrupted expression due to sex chromosome mismatch alone, on a non-hybrid autosomal background (Experiment 2). The y-axis shows the difference in normalized expression levels between the two cross types being compared. The x-axis shows the proportion of genes in each expression difference bin. Black bars represent the autosomes, purple bars represent the X chromosome, and green bars represent the Y chromosome. Letters indicate significant differences in median expression differences among the chromosome types based on a Mann-Whitney U test (FDR-corrected $P < 0.05$).

X-Y Mismatch Disrupted the Expression of Several Genes during Late Spermatogenesis

We also tested for effects of X-Y mismatch on individual genes by identifying differentially expressed (DE) genes in X-Y mismatch mice compared to controls. In our reanalysis, we identified many more overexpressed genes in sterile *mus×dom* hybrids compared to *mus* and many more underexpressed genes in the reciprocal *dom×mus* hybrids compared to *dom* on the X chromosome (Table 2), consistent with previous results (Larson *et al.* 2017) and with our observations of overall expression differences (Figure 5). We then asked if any of these X-linked DE genes were associated with X-Y mismatch. If so, then we would expect our Experiment 1 *mus×dom*^{musY} to rescue some of the disrupted X-linked expression, and thus manifest as DE genes in comparisons between *mus×dom* and *mus×dom*^{musY}. These genes should also overlap with genes DE between *mus×dom* and *mus*. However, there were only two X-linked DE genes in the *mus×dom* versus *mus×dom*^{musY} comparison (Table 2), and only one was also DE in the *mus×dom* versus *mus* comparison (Figure 6). This gene is a predicted protein coding gene, *Gm10058*, that shares 97% sequence identity with *Slx* and is therefore likely a paralog of this gene family. The other DE gene was *Btbd35f17*, a gene with a protein-protein binding domain that is specifically expressed in male reproductive tissues (Smith *et al.* 2019). In Experiment 2, we only observed one X-linked DE gene in *mus*^{domY} compared to *mus*, and this gene was not DE in any other comparisons. Taken together,

both Experiment 1 and Experiment 2 results suggest that almost all DE genes on the X chromosome in sterile *musculus*♀ × *domesticus*♂ hybrids are disrupted due to X-autosomal or autosomal-autosomal incompatibilities, rather than Y-linked incompatibilities.

On the X chromosome, very few DE genes were shared across multiple comparisons. However, 57 DE genes were shared between the *mus*×*dom* versus *mus* and *dom*×*mus* versus *dom* comparisons. When we looked at DE genes separated by direction of expression difference, only eight were shared between these two comparisons (Supplemental Material, Figure S7), so most of the overlap represented genes overexpressed in *mus*×*dom* but underexpressed in *dom*×*mus*. This could indicate that similar regulatory networks are disrupted in reciprocal F1 hybrids, but in ways that disrupt gene expression levels in opposite directions.

In contrast to the X chromosome, more Y-linked DE genes were shared across comparisons (Figure 6). Sterile *mus*×*dom* hybrids had 17 Y-linked DE genes that showed a clear bias towards overexpression (Table 2). Of these 17 DE genes, 5 were shared with the Experiment 1 comparison *mus*×*dom* versus *dom*×*mus*^{domY}, so having *domesticus* X and Y chromosomes partially rescued expression levels on the Y chromosome in *dom*×*mus*^{domY} mice. However, none of the 17 Y-linked genes DE in sterile hybrids were also DE in the Experiment 1 comparison (*mus*^{domY} versus *dom*), so it is unlikely that X-Y mismatch alone disrupts expression of these genes. Instead, there may be a complex interaction between X-Y mismatch and a hybrid autosomal background that disrupts Y chromosome expression. Consistent with this, we found the most Y-linked DE genes in comparisons between cross types with reciprocal hybrid autosomal backgrounds but the same Y chromosome (Table 2). Of these, 78 Y-linked DE genes were shared between these two comparisons (Figure 6), suggesting that reciprocal hybrid autosomal backgrounds may have resulted in disrupted expression for many of the same Y-linked genes, regardless of the subspecies origin of the Y chromosome.

We also found several autosomal genes that were DE between cross types with the same autosomal background but different sex chromosome combinations (Table 2). We excluded autosomal genes that overlapped with putatively introgressed regions, so

the DE that we detected was unlikely due to *cis*-regulatory effects of variants from the opposite subspecies that introgressed along with the Y chromosome. In Experiment 1, 104 autosomal genes were DE when comparing *mus*×*dom* to *dom*×*mus*^{domY} and 494 autosomal genes were DE when comparing *dom*×*mus* to *mus*×*dom*^{musY} (Table 2). These comparisons involved reciprocal crosses with the same autosomal and Y chromosome genotypes, and so DE presumably resulted from X-autosomal incompatibilities. Although overexpression on the X chromosome tends to be the most notable expression pattern associated with X-autosomal incompatibilities, previous studies have shown disrupted postmeiotic autosomal expression in sterile hybrids as well (Larson *et al.* 2017). We detected only six (non-overlapping) DE genes in each comparison with different Y chromosomes but the same autosomal and X chromosome genotypes (*mus*×*dom* versus *mus*×*dom*^{musY} and *dom*×*mus* versus *dom*×*mus*^{domY}; Table 2).

In In Experiment 2, we identified some autosomal DE genes in comparisons that had different Y chromosomes but the same autosomal and X backgrounds, suggesting that interactions involving the Y chromosome disrupted some autosomal expression (Table 2). These autosomal DE genes tended to be underexpressed in the cross type with X-Y mismatch regardless of the direction of the cross (Table 2) and must result from direct interactions with the Y chromosome or indirect interactions with X-Y mediated expression changes. Only one autosomal gene, *Babam2*, was DE in both reciprocal comparisons. It is a member of the BRCA1-A complex, which is involved in DNA double-strand break repair (The Uniprot Consortium 2020).

Table 2: Number of differentially expressed genes in round spermatids for different cross type comparisons. “Higher” indicates higher expression (i.e., overexpressed) in the cross type with X-Y mismatch (F1 hybrids in Larson et al. 2017 and Experiment 1, Y-introgression F1 crosses in Experiment 2). “Lower” indicates lower expression (i.e., underexpressed) in the cross type with X-Y mismatch. Gray boxes indicate chromosomes that are from the same subspecies in the two cross types being compared. Reciprocal F1s were considered as having the same autosomal

backgrounds. Autosomal DE genes overlapping with putatively introgressed regions were excluded.

		Autosomes		X Chromosome		Y Chromosome	
		Higher	**Lower**	**Higher**	**Lower**	**Higher**	**Lower**
Larson et al. 2017	*mus*×*dom* vs *mus*	1518	1476	252	13	109	66
	mus×*dom* vs *dom*	1357	1241	190	55	15	2
	dom×*mus* vs *mus*	1360	1009	62	73	6	8
	dom×*mus* vs *dom*	1237	878	27	73	69	86
Exp. 1	*mus*×*dom* vs *mus*×*dom*^{musY}	3	3	2	0	74	70
	mus×*dom* vs *dom*×*mus*^{domY}	21	83	38	96	68	84
	dom×*mus* vs *mus*×*dom*^{musY}	372	122	44	101	76	85
	dom×*mus* vs *dom*×*mus*^{domY}	2	4	1	0	71	66
Exp. 2	*mus*^{domY} vs *mus*	13	34	1	0	52	63
	mus^{domY} vs *dom*	1820	2269	28	179	3	69
	dom^{musY} vs *mus*	1634	1679	70	55	10	7
	dom^{musY} vs *dom*	13	63	0	10	14	70

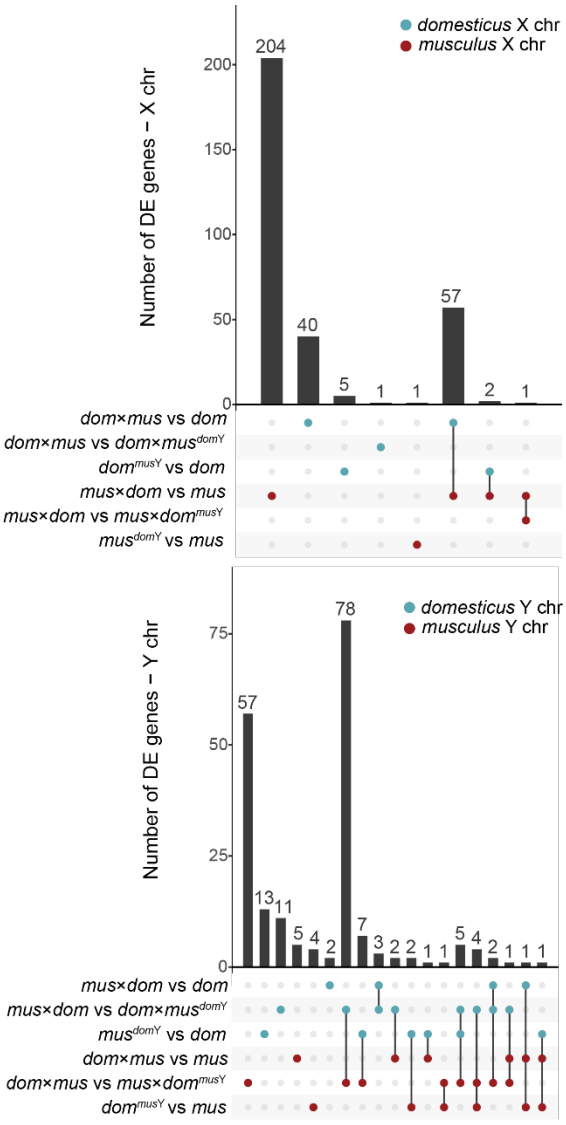

Figure 6: Upset plots showing the number of differentially expressed (DE) genes in each cross type comparison, and genes that are DE across multiple comparisons. (A) DE genes on the X chromosome. (B) DE genes on the Y chromosome. Bars corresponding to multiple dots connected by lines indicate genes that are DE across multiple comparisons. Bars corresponding to single dots indicate genes that are DE in only one comparison. Blue dots indicate comparisons on the *domesticus* X chromosome (A) or *domesticus* Y chromosome (B), and red dots indicate comparisons on the *musculus* X chromosome (A) or *musculus* Y chromosome (B). Genes that were DE in opposite directions across multiple comparisons of the same sex chromosome were excluded.

Finally, we tested if DE genes tended to be in the same co-expression networks using weighted correlation network analysis (WGCNA). We found one module in Experiment 1 associated with the *mus*×*dom* autosomal background, one module in Experiment 2 associated with the *musculus* background, and one module in Experiment 2 associated with the *domesticus* background (Figure 7A, B, D). These modules were significantly enriched for genes DE between cross types with different autosomal backgrounds (Table 3). There were also multiple modules enriched for DE genes despite not having a significant association with cross type (Table 3). For example, Module 5 was significantly enriched for DE genes in all pairwise comparisons in Experiment 1. Although we did not detect a significant cross type association for this module, there was a trend towards an autosomal background by sex chromosome effect for this module, with *mus*×*dom* background cross types tending to have lower module membership in general, but with *mus*×*dom*^{musY} mice tending to have higher module membership than *mus*×*dom* mice (Figure 7E). Another Experiment 1 module showed a similar pattern (Module 3, Figure 7C) and was enriched for genes DE between *dom*×*mus* and *mus*×*dom*^{musY} (Table 3). In Experiment 2, Module 5 was enriched for genes DE between *mus*^{domY} and either subspecies (*mus* or *dom*; Table 3), and X-Y mismatch mice tended to have lower associations with this module (Figure 7). We likely did not have enough power to detect significant module associations with complex autosome by sex chromosome interactions given our sample size, especially because these effects on gene expression tended to be subtle and affect relatively few genes (Figure 5, Table 2). Despite low power, the fact that certain modules were enriched for DE genes suggests that groups of genes were disrupted in similar ways in X-Y mismatch mice, and that particular gene networks may be disrupted under X-Y mismatch. Additionally, we found a significant positive correlation in module eigengene values between Experiment 1 and Experiment 2 (Module 5 in both experiments, r = 0.64; FDR-corrected Pearson's correlation $P < 0.001$; Supplemental Figure S8) and a significant overlap in genes (279 genes, FDR-corrected Fisher's Exact Test $P < 0.001$), suggesting that these two modules represent genes with similar expression patterns between the two experiments. Interestingly, these modules trended towards a negative association with cross types that had a *musculus* X chromosome and *domesticus* Y

chromosome (Figure 7E, 7F), and may represent genes with similar expression patterns under X-Y mismatch regardless of autosomal background.

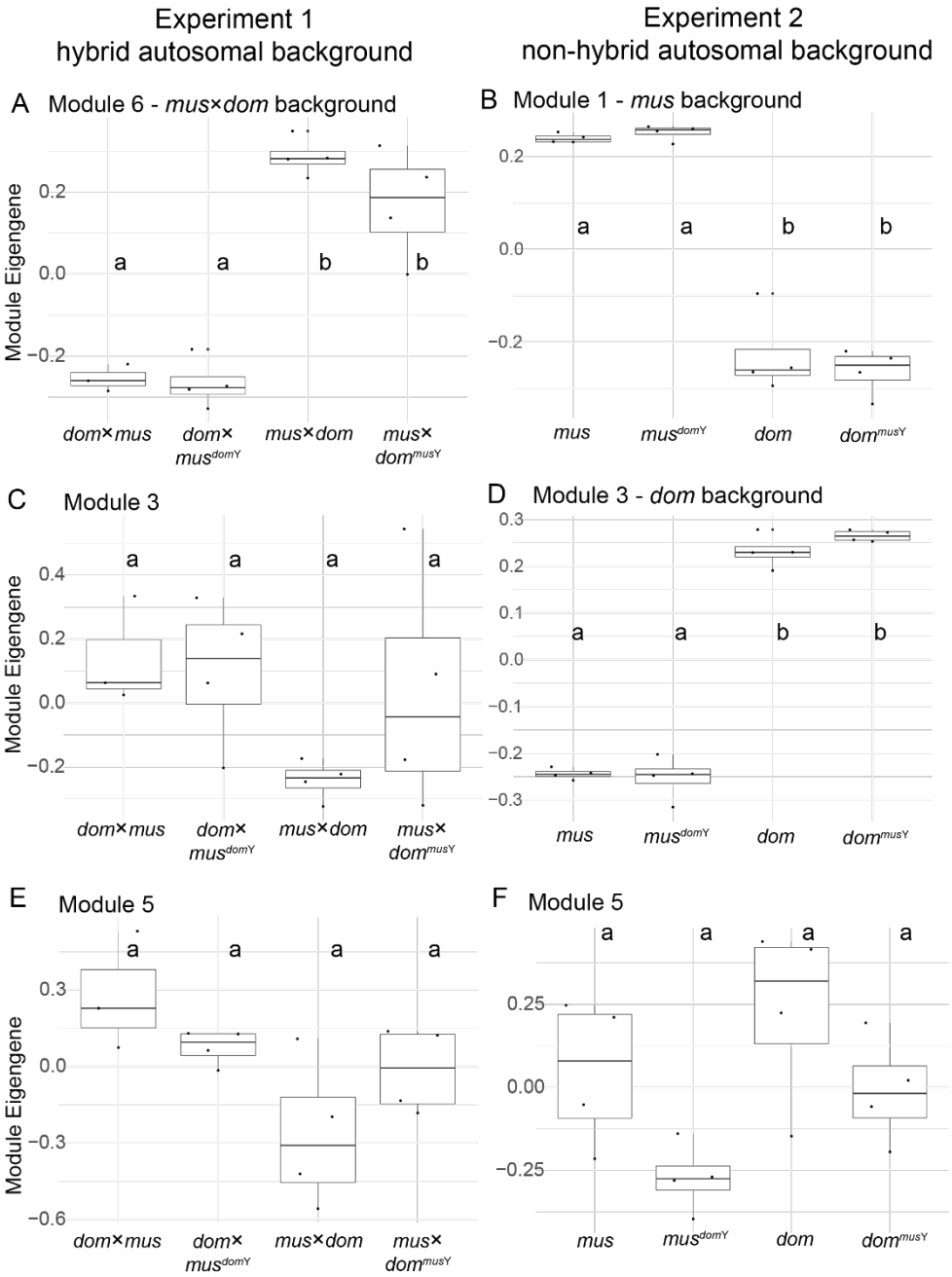

Figure 7: Example WGCNA module eigengene values plotted by cross type. Note that WGCNA was performed separately for each experiment, so there is not necessarily a relationship between Experiment 1 and Experiment 2 modules with the same number. Modules that were significantly associated with cross types are also labeled based on these associations (A, B, and D). Other modules shown were not significantly

associated with a cross type but trended towards an association with X-autosomal background by Y chromosome type interaction and were enriched for DE genes in at least one comparison (C, E, and F; Table 3). Letters indicate significant differences in module association based on linear models with post-hoc Tukey tests ($P < 0.05$).

Table 3: Number of differentially expressed genes in each WGCNA module. Rows indicate WGCNA modules and columns indicate comparisons between cross types used to identify differentially expressed (DE) genes. Module associations with cross types are based on linear models with post-hoc Tukey tests. Shaded boxes indicate a significant enrichment for DE genes based on a hypergeometric test with FDR-correction ($P < 0.05$). Note that there is not necessarily a relationship between Experiment 1 and Experiment 2 modules with the same module number.

			Number of DE genes in module			
	Module	**Significant cross type associations**	***mus*×*dom* vs *mus*×*dom*^{musY}**	***mus*×*dom* vs *dom*×*mus*^{domY}**	**dom×*mus* vs *mus*×*dom*^{musY}**	***dom*×*mus* vs *dom*×*mus*^{domY}**
Exp. 1	1	none	0	13	9	0
	2	none	1	19	35	2
	3	none	1	10	170	1
	4	none	4	3	11	2
	5	none	7	21	155	5
	6	*mus*×*dom* background	2	87	102	1
	Module	**Significant cross type associations**	***mus*^{domY} vs *mus***	***mus*^{domY} vs *dom***	***dom*^{musY} vs *mus***	***dom*^{musY} vs *dom***
Exp. 2	1	*mus* background	0	1039	972	40
	2	none	0	168	133	4
	3	*dom* background	9	913	970	4
	4	none	4	91	358	2
	5	none	23	532	28	9
	6	none	3	329	17	5
	7	none	1	220	55	1
	8	none	1	22	77	8
	9	none	0	106	3	0
	10	none	1	1	104	1
	11	none	0	111	24	0
	12	none	1	3	6	0
	13	none	0	1	0	0
	14	none	1	1	12	1
	15	none	0	0	0	0
	16	none	0	0	5	0
	17	none	0	18	0	1
	18	none	0	0	0	0

Discussion

The large X-effect and Haldane's rule are prevalent patterns observed in intrinsic hybrid incompatibilities across divergent taxa and suggest that sex chromosomes play a predominant role in speciation, but the evolutionary forces underlying rapid sex chromosome divergence that leads to hybrid incompatibilities remain unclear (Presgraves and Meiklejohn 2021). One compelling hypothesis is intragenomic conflict between sex chromosomes (Frank 1991; Hurst and Pomiankowski 1991; Lindholm *et al.* 2016). Some empirical studies have identified loci involved in both intragenomic conflict and hybrid incompatibilities (Tao *et al.* 2001; Phadnis and Orr 2009; Wilkinson *et al.* 2014; Zanders *et al.* 2014; Case *et al.* 2015; Zhang *et al.* 2015; Larson *et al.* 2017), but it remains unknown how prevalent these systems are in natural populations and if intragenomic conflict is the primary causative force behind the evolution of these incompatibilities. While X-autosomal incompatibilities are known to play a central role in house mouse hybrid sterility, previous work has shown that house mouse speciation likely has a more complex genetic basis (Vyskočilová *et al.* 2005; Good *et al.* 2008b; Turner *et al.* 2012; Turner and Harr 2014; Larson *et al.* 2018b) and may involve sex chromosome intragenomic conflict (Ellis *et al.* 2011; Campbell *et al.* 2012; Larson *et al.* 2017). In this study, we showed that intragenomic conflict between the sex chromosomes may contribute to some hybrid incompatibilities in house mice, but not in a simple dosage-dependent manner, and with subtle effects relative to other components of F1 hybrid incompatibilities. Below, we discuss the implications of our findings for the genetic basis of house mouse male hybrid sterility and the potential role of intragenomic conflict in mouse speciation.

Insights into the Genetic Basis of Mouse Male Hybrid Sterility

We performed Y introgression experiments to test the effects of X-Y mismatch on house mouse male hybrid sterility, and our results did not support the model of *Slx*- and *Slxl1*-*Sly* dosage imbalance leading to X chromosome overexpression in mouse F1 hybrids. In Experiment 1, we showed that X-Y match on an F1 background did not restore postmeiotic X chromosome repression (Figure 5). In Experiment 2, we directly tested

the effects of X-Y mismatch in the absence of X-autosomal mismatch on postmeiotic spermatogenesis gene expression. We found some evidence for disrupted expression in X-Y mismatch mice (Figure 5, Table 2), but the effects were relatively subtle and often in the opposite direction than expected based on genetic manipulation studies (Cocquet *et al.* 2012; Kruger *et al.* 2019) or disrupted expression in sterile F1 mice (Larson *et al.* 2017; Figures 4, 5, and 6).

Our results contrast those of genetic manipulation studies, which performed nearly complete knockdowns or duplications and therefore do not represent more subtle copy number differences expected to occur in natural hybrids (Cocquet *et al.* 2009; Cocquet *et al.* 2012; Kruger *et al.* 2019). Another important difference from genetic manipulation studies is that we used wild-derived inbred strains instead of the C57BL/6J classic laboratory mouse, which has a mostly *domesticus* background but some *musculus* introgression throughout, including the Y chromosome (Nagamine *et al.* 1992). Because C57BL/6J is mostly *domesticus* with a *musculus* Y chromosome, it is similar to our *dom*$^{mus\mathrm{Y}}$ mice and therefore may have some of the subtle disruptions to gene expression and sperm morphology that we observed compared to pure *domesticus* mice. We also introgressed the entire Y chromosome, so there should not have been dosage imbalances among ampliconic genes on the same sex chromosome. However, our Y-introgression mice also had imbalance between all Y-linked ampliconic genes and interacting genes on the X chromosome and autosomes, so it is unclear if introgressing the entire Y chromosome should cause larger or smaller effects on spermatogenesis expression. SLX, SLXL1, and SLY proteins interact with other sex-linked and autosomal ampliconic genes, including *Ssty1/2,* α*-takusan,* and *Speer,* so additional gene families may be involved in intragenomic conflict with *Slx, Slxl1,* and *Sly* (Kruger *et al.* 2019; Moretti *et al.* 2020). Our autosomal gene family expression results further complicate understanding of ampliconic gene conflict, because we found that the α*-takusan* gene family is overexpressed in F1 hybrids regardless of cross direction or sex chromosome type (Figure 4F). Sex chromosome mismatch, however, did not disrupt α*-takusan* expression when the autosomal background was non-hybrid. This was somewhat puzzling because protein products of sex-linked ampliconic genes are thought to regulate α*-takusan* expression in late spermatogenesis.

Our results suggest that differences in *Slx*- or *Slxl1*-*Sly* dosage may not result in hybrid incompatibilities; we did not observe sex chromosome overexpression with an excess of *Slx* and *Slxl1* copies or underexpression with an excess of *Sly* copies that this model predicts. Therefore, the primary mechanisms underlying postmeiotic X chromosome overexpression in sterile F1 hybrids likely do not involve X-Y interactions. Instead, disrupted postmeiotic repression is likely a continuation of *Prdm9*-mediated MSCI disruption, or perhaps another mechanism that involves X chromosome and autosome incompatibilities. Although our results showed that *Slx*- or *Slxl1*-*Sly* copy number imbalance is unlikely to explain disrupted postmeiotic repression in F1 hybrids, sex chromosome mismatch is still likely to play a role, albeit more subtle, in house mouse hybrid sterility. We showed that X-Y mismatch can lead to disrupted expression of ampliconic genes and other genes throughout the genome (Figure 4, Figure 6, Table 2), and some of these genes are thought to be essential for spermatogenesis. For example, *Taf7l* knockouts have abnormal sperm morphology (Cheng *et al.* 2007), *Prdx4* knockouts have reduced sperm counts (Iuchi *et al.* 2009), and both these genes were DE in dom^{musY} mice. We also showed that sex chromosome mismatch is associated with subfertility phenotypes (Table 1), consistent with previous studies (Campbell *et al.* 2012; Campbell and Nachman 2014). We focused on interactions between the sex chromosomes because the ampliconic gene conflict model established a clear prediction for X-Y incompatibilities, but we could not distinguish X-Y incompatibilities from Y-autosomal incompatibilities in our experimental crosses and we note that some of our observations could result from Y-autosomal interactions.

These results are likely important in the context of mouse speciation in nature. Mice sampled from the European hybrid zone are often advanced generation hybrids with complex patterns of ancestry from both *musculus* and *domesticus*, and true F1 genotypes are exceptionally rare (Teeter *et al.* 2010; Turner *et al.* 2012). Therefore, understanding mechanisms of hybrid incompatibility in addition to F1 X-autosomal incompatibilities is essential for understanding the complex genetic basis of mouse speciation occurring in nature. Experiment 2 demonstrated that disrupted gene expression phenotypes can occur in the absence of an F1 autosomal background. Previous studies have shown that advanced intercrosses of hybrid mice show different

sterility phenotypes than F1s (Campbell *et al.* 2012), and *Prdm9*-mediated hybrid sterility requires an F1 autosomal background, leading others to speculate that genetic incompatibilities underlying hybrid sterility may be different in later hybrid generations (Campbell and Nachman 2014; Mukaj *et al.* 2020). Our results show that Y chromosome introgression can contribute to reduced fertility (consistent with Campbell *et al.* 2012) and disrupted spermatogenesis gene expression in later generation hybrids with non-F1 autosomal backgrounds.

What is the Contribution of Sex Chromosome Conflict to Mouse Speciation?

The exact mechanisms underlying reduced fertility associated with Y chromosome mismatch is unknown, and it is still unclear what role, if any, sex chromosome intragenomic conflict may play (Ellis *et al.* 2011; Campbell *et al.* 2012; Larson *et al.* 2017). Intragenomic conflict among sex chromosome ampliconic genes has been proposed as a mechanism through which hybrid incompatibilities evolved in several mammalian species (Davis *et al.* 2015; Dutheil *et al.* 2015; Larson *et al.* 2018a; Kruger *et al.* 2019). Ampliconic genes are a common feature of mammalian sex chromosomes, and they tend to be expressed specifically during spermatogenesis (Li *et al.* 2013; Soh *et al.* 2014; Skinner *et al.* 2016; Lucotte *et al.* 2017; Bellott *et al.* 2017; Hughes *et al.* 2020; reviewed in Larson *et al.* 2018a). In cats, loci associated with hybrid sterility tend to be in or near high copy number genes (Davis *et al.* 2015). In great apes, sex chromosome amplicon copy number can evolve rapidly (Lucotte *et al.* 2017; Cechova *et al.* 2020), and ampliconic regions on the X chromosome are thought to have experienced selective sweeps as a result of strong selection pressures imposed by intragenomic conflict with the Y chromosome (Nam *et al.* 2015). These regions also overlap sections of the modern human X chromosome that lack Neandertal introgression, and therefore may represent regions involved in genetic incompatibilities between modern humans and Neandertals (Dutheil *et al.* 2015).

Theoretical work introducing the idea that sex chromosome intragenomic conflict could contribute to hybrid incompatibilities did so directly as an explanation for Haldane's rule and the large X-effect (Frank 1991; Hurst and Pomiankowski 1991). However, conflict between the sex chromosomes cannot explain some observations,

such as the applicability of Haldane's rule and the large X-effect to hybrid inviability and the important role of the X chromosome in many incompatibilities that occur in homogametic hybrids (Coyne 1992). In this study, we showed that X-Y conflict may have a small effect on male hybrid sterility, but X-autosomal incompatibilities that do not appear to be involved in intragenomic conflict probably play the most important role in the observations consistent with Haldane's rule and the large X-effect in house mice.

It remains unknown if ampliconic genes are involved in intragenomic conflict and if they frequently underlie hybrid sterility broadly across mammals. If so, intragenomic conflict may be much more important in the evolution of hybrid incompatibility loci than once thought (Johnson and Wu 1992; Coyne and Orr 2004). Some recent empirical studies support this hypothesis in both flies and mammals (Presgraves and Meiklejohn 2021), however, our study did not provide direct support for this hypothesis. X-Y mismatch likely contributes to hybrid male sterility and disrupted expression, but in more complex ways than the *Slx*, *Slxl1,* and *Sly* dosage-based conflict model, and with relatively small effects on hybrid sterility. Further work is required to identify loci involved in these X-Y or Y-autosomal incompatibilities, but it is plausible that intragenomic conflict among ampliconic genes still plays a role given that these genes are the primary sex chromosome genes expressed in the postmeiotic stages during which spermatogenesis expression is highly disrupted (Sin and Namekawa 2013; Larson *et al.* 2017).

Acknowledgements

We would like to thank Pamela K. Shaw and the UM Fluorescence Cytometry Core supported by an Institutional Development Award from the NIGMS (P30GM103338 and S10-OD025019), the UM Genomics Core supported by the M.J. Murdock Charitable Trust, and the UM Lab Animal Resources staff. This work was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (R01-HD073439, R01-HD094787 to JMG). E.E.K.K. was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1313190) and a Rosemary Grant Award for Graduate Student Research from the Society for the Study of Evolution. E.L.L. was supported by the National

Science Foundation (DEB-2012041). C.C.R. was supported by the BBSRC (BB/N000463/1 to PE). P.E. also acknowledges funding from the Leverhulme Trust (RPG-2019-414 194). E.M.W. was supported by a faculty studentship from the University of Essex. B.M.S was supported by UKRI (University of Essex). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the National Institutes of Health, or the Society for the Study of Evolution.

Author Contributions

J.M.G. and E.L.L. conceived and funded the project. E.E.K.K., E.L.L., and J.M.G. designed the experiments. E.L.L. and E.K.K.K. did the mouse husbandry and breeding. E.E.K.K. performed the mouse dissections, cell sorts, and sequencing library preparation. E.M.W, B.M.S, and P.J.I.E. performed and analyzed the sperm morphology assays. E.E.K.K. analyzed the data. E.E.K.K., E.L.L., and J.M.G. wrote the manuscript with input from all authors.

Conflicts of Interest

The authors declare no conflicts of interest.

References

Baker, C. L., S. Kajita, M. Walker, R. L. Saxl, N. Raghupathy *et al.*, 2015 PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. PLOS Genetics 11**:** e1004916.

Balcova, M., B. Faltusova, V. Gergelits, T. Bhattacharyya, O. Mihola *et al.*, 2016 Hybrid Sterility Locus on Chromosome X Controls Meiotic Recombination Rate in Mouse. PLOS Genetics 12**:** e1005906.

Bellott, D. W., H. Skaletsky, T.-J. Cho, L. Brown, D. Locke *et al.*, 2017 Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. Nat Genet 49**:** 387-394.

Bhattacharyya, T., S. Gregorova, O. Mihola, M. Anger, J. Sebestova *et al.*, 2013 Mechanistic basis of infertility of mouse intersubspecific hybrids. Proceedings of the National Academy of Sciences of the United States of America 110**:** E468-E477.

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

Burgoyne, P. S., S. K. Mahadevaiah and J. M. A. Turner, 2009 The consequences of asynapsis for mammalian meiosis. Nat Rev Genet 10**:** 207-216.
Campbell, P., J. M. Good, M. D. Dean, P. K. Tucker and M. W. Nachman, 2012 The Contribution of the Y Chromosome to Hybrid Male Sterility in House Mice. Genetics 191**:** 1271-1281.
Campbell, P., J. M. Good and M. W. Nachman, 2013 Meiotic Sex Chromosome Inactivation Is Disrupted in Sterile Hybrid Male House Mice. Genetics 193**:** 819-828.
Campbell, P., and M. W. Nachman, 2014 X–Y Interactions Underlie Sperm Head Abnormality in Hybrid Male House Mice. Genetics 196**:** 1231-1240.
Case, L. K., E. H. Wall, E. E. Osmanski, J. A. Dragon, N. Saligrama *et al.*, 2015 Copy number variation in Y chromosome multicopy genes is linked to a paternal parent-of-origin effect on CNS autoimmune disease in female offspring. Genome Biology 16**:** 28.
Cechova, M., R. Vegesna, M. Tomaszkiewicz, R. S. Harris, D. Chen *et al.*, 2020 Dynamic evolution of great ape Y chromosomes. Proceedings of the National Academy of Sciences**:** 202001749.
Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The Relative Rates of Evolution of Sex Chromosomes and Autosomes. American Naturalist 130**:** 113-146.
Cheng, Y., M. G. Buffone, M. Kouadio, M. Goodheart, D. C. Page *et al.*, 2007 Abnormal Sperm in Mice Lacking the *Taf7l* Gene. Molecular and Cellular Biology 27**:** 2582-2589.
Cocquet, J., P. J. Ellis, S. K. Mahadevaiah, N. A. Affara, D. Vaiman *et al.*, 2012 A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet 8**:** e1002900.
Cocquet, J., P. J. Ellis, Y. Yamauchi, S. K. Mahadevaiah, N. A. Affara *et al.*, 2009 The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. PLoS Biol 7**:** e1000244.
Conway, J. R., A. Lex and N. Gehlenborg, 2017 UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33**:** 2938-2940.
Coughlan, J. M., and D. R. Matute, 2020 The importance of intrinsic postzygotic barriers throughout the speciation process. Philosophical Transactions of the Royal Society B: Biological Sciences 375**:** 20190533.
Coyne, J. A., 1992 Genetics and Speciation. Nature 355**:** 511-515.
Coyne, J. A., and H. A. Orr, 1989 Patterns of speciation in *Drosophila*. Evolution 43**:** 362-381.
Coyne, J. A., and H. A. Orr, 2004 *Speciation*. W.H. Freeman.
Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto *et al.*, 2016 Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature 530**:** 171-176.
Davis, B. W., C. M. Seabury, W. A. Brashear, G. Li, M. Roelke-Parker *et al.*, 2015 Mechanisms Underlying Mammalian Hybrid Sterility in Two Feline Interspecies Models. Molecular Biology and Evolution 32**:** 2534-2546.
Didion, J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan *et al.*, 2016 R2d2 Drives Selfish Sweeps in the House Mouse. Molecular Biology and Evolution 33**:** 1381-1395.
Dobzhansky, T., 1937 Genetic Nature of Species Differences. The American Naturalist 71**:** 404-420.
Dumont, B. L., and B. A. Payseur, 2011 Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice. PLOS Genetics 7**:** e1002116.
Dutheil, J. Y., K. Munch, K. Nam, T. Mailund and M. H. Schierup, 2015 Strong Selective Sweeps on the X Chromosome in the Human-Chimpanzee Ancestor Explain Its Low Divergence. PLOS Genetics 11**:** e1005451.
Edwards, J. A., and R. A. Edwards, 2019 Fastq-pair: efficient synchronization of paired-end fastq files. bioRxiv**:** 552885.

Ellis, P. J., J. Bacon and N. A. Affara, 2011 Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? Hum Mol Genet 20**:** 3010-3021.
Forejt, J., P. Jansa and E. Parvanov, 2021 Hybrid sterility genes in mice (Mus musculus): a peculiar case of PRDM9 incompatibility. Trends in Genetics.
Frank, S. A., 1991 Divergence of Meiotic Drive-Suppression Systems as an Explanation for Sex- Biased Hybrid Sterility and Inviability. Evolution 45**:** 262-267.
Getun, I. V., B. Torres and P. R. J. Bois, 2011 Flow Cytometry Purification of Mouse Meiotic Cells. Journal of Visualized Experiments: JoVE**:** 2602.
Good, J. M., 2012 The Conflict within and the Escalating War between the Sex Chromosomes. PLoS Genet 8**:** e1002955.
Good, J. M., M. D. Dean and M. W. Nachman, 2008a A Complex Genetic Basis to X-Linked Hybrid Male Sterility Between Two Species of House Mice. Genetics 179**:** 2213-2228.
Good, J. M., T. Giger, M. D. Dean and M. W. Nachman, 2010 Widespread Over-Expression of the X Chromosome in Sterile F_1 Hybrid Mice. PLoS Genet 6**:** e1001148.
Good, J. M., M. A. Handel and M. W. Nachman, 2008b Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evolution 62**:** 50-65.
Green, C. D., Q. Ma, G. L. Manske, A. N. Shami, X. Zheng *et al.*, 2018 A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Developmental Cell.
Grey, C., P. Barthès, G. Chauveau-Le Friec, F. Langa, F. Baudat *et al.*, 2011 Mouse PRDM9 DNA-Binding Specificity Determines Sites of Histone H3 Lysine 4 Trimethylation for Initiation of Meiotic Recombination. PLOS Biology 9**:** e1001176.
Haldane, J. B. S., 1922 Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics 12**:** 101-109.
Harr, B., E. Karakoc, R. Neme, M. Teschke, C. Pfeifle *et al.*, 2016 Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. Sci Data 3**:** 160075.
Hughes, J. F., H. Skaletsky, T. Pyntikova, N. Koutseva, T. Raudsepp *et al.*, 2020 Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. Genome Research 30**:** 1716-1726.
Hunnicutt, K. E., J. M. Good and E. L. Larson, 2021 Unraveling patterns of disrupted gene expression across a complex tissue. Evolution n/a.
Hurst, L. D., and A. Pomiankowski, 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**:** 841-858.
Iuchi, Y., F. Okada, S. Tsunoda, N. Kibe, N. Shirasawa *et al.*, 2009 Peroxiredoxin 4 knockout results in elevated spermatogenic cell death via oxidative stress. Biochemical Journal 419**:** 149-158.
JanouŠEk, V., L. Wang, K. E. N. Luzynski, P. DufkovÁ, M. VyskoČIlovÁ Martina *et al.*, 2012 Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus. Molecular Ecology 21**:** 3032-3047.
Johnson, N. A., and C. I. Wu, 1992 An empirical test of the meiotic drive models of hybrid sterility: sex-ratio data from hybrids between Drosophila simulans and Drosophila sechellia. Genetics 130**:** 507-511.
Kruger, A. N., M. A. Brogley, J. L. Huizinga, J. M. Kidd, D. G. de Rooij *et al.*, 2019 A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. Current Biology.
Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9**:** 559.

Larson, E. L., S. Keeble, D. Vanderpool, M. D. Dean and J. M. Good, 2017 The Composite Regulatory Basis of the Large X-Effect in Mouse Speciation. Molecular Biology and Evolution 34**:** 282-295.
Larson, E. L., E. E. K. Kopania and J. M. Good, 2018a Spermatogenesis and the Evolution of Mammalian Sex Chromosomes. Trends Genet 34**:** 722-732.
Larson, E. L., E. E. K. Kopania, K. E. Hunnicutt, D. Vanderpool, S. Keeble *et al.*, 2021 Stage-specific disruption of X chromosome expression during spermatogenesis in sterile house mouse hybrids. G3 Genes|Genomes|Genetics.
Larson, E. L., D. Vanderpool, B. A. J. Sarver, C. Callahan, S. Keeble *et al.*, 2018b The Evolution of Polymorphic Hybrid Incompatibilities in House Mice. Genetics.
Li, G., B. W. Davis, T. Raudsepp, A. J. Pearks Wilkerson, V. C. Mason *et al.*, 2013 Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Research 23**:** 1486-1495.
Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25**:** 1754-1760.
Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier *et al.*, 2016 The Ecology and Evolutionary Dynamics of Meiotic Drive. Trends in Ecology & Evolution 31**:** 315-326.
Lucotte, E. A., L. Skov, M. Coll Macia, K. Munch and M. H. Schierup, 2017 Dynamic copy number evolution of X- and Y-linked ampliconic genes in human populations. bioRxiv.
McKee, B. D., and M. A. Handel, 1993 Sex chromosomes, recombination, and chromatin conformation. Chromosoma 102**:** 71-80.
Meiklejohn, C. D., and Y. Tao, 2010 Genetic conflict and sex chromosome evolution. Trends in Ecology & Evolution 25**:** 215-223.
Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti and J. Forejt, 2009 A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. Science 323**:** 373-375.
Moretti, C., M. Blanco, C. Ialy-Radio, M.-E. Serrentino, C. Gobé *et al.*, 2020 Battle of the Sex Chromosomes: Competition between X and Y Chromosome-Encoded Proteins for Partner Interaction and Chromatin Occupancy Drives Multicopy Gene Expression and Evolution in Muroid Rodents. Molecular Biology and Evolution.
Morgan, A. P., and F. Pardo-Manuel de Villena, 2017 Sequence and Structural Diversity of Mouse Y Chromosomes. Molecular Biology and Evolution 34**:** 3186-3204.
Mueller, J. L., S. K. Mahadevaiah, P. J. Park, P. E. Warburton, D. C. Page *et al.*, 2008 The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat Genet 40**:** 794-799.
Mueller, J. L., H. Skaletsky, L. G. Brown, S. Zaghlul, S. Rock *et al.*, 2013 Independent specialization of the human and mouse X chromosomes for the male germ line. Nat Genet 45**:** 1083-1087.
Mukaj, A., J. Piálek, V. Fotopulosova, A. P. Morgan, L. Odenthal-Hesse *et al.*, 2020 Prdm9 Intersubspecific Interactions in Hybrid Male Sterility of House Mouse. Molecular Biology and Evolution 37**:** 3423-3438.
Muller H. 1942. Isolating mechanisms, evolution, and temperature. Biol Symp. 6:71–125.Nagamine, C. M., Y. Nishioka, K. Moriwaki, P. Boursot, F. Bonhomme *et al.*, 1992 The musculus-type Y Chromosome of the laboratory mouse is of Asian origin. Mammalian Genome 3**:** 84-91.
Nam, K., K. Munch, A. Hobolth, J. Y. Dutheil, K. R. Veeramah *et al.*, 2015 Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc Natl Acad Sci U S A 112**:** 6413-6418.
Namekawa, S. H., P. J. Park, L.-F. Zhang, J. E. Shima, J. R. McCarrey *et al.*, 2006 Postmeiotic Sex Chromatin in the Male Germline of Mice. Current Biology 16**:** 660-667.

Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach *et al.*, 2009 Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. PLOS Genetics 5**:** e1000753.
Patro, R., G. Duggal, M. I. Love, R. A. Irizarry and C. Kingsford, 2017 Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature methods 14**:** 417-419.
Pedersen, B. S., and A. R. Quinlan, 2017 Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34**:** 867-868.
Pezer, Ž., B. Harr, M. Teschke, H. Babiker and D. Tautz, 2015 Divergence patterns of genic copy number variation in natural populations of the house mouse (Mus musculus domesticus) reveal three conserved genes with major population-specific expansions. Genome Research 25**:** 1114-1124.
Phadnis, N., and H. A. Orr, 2009 A Single Gene Causes Both Male Sterility and Segregation Distortion in Drosophila Hybrids. Science 323**:** 376-379.
Phifer-Rixey, M., and M. W. Nachman, 2015 Insights into mammalian biology from the wild house mouse *Mus musculus*. Elife 4.
“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute
Presgraves, D. C., and C. D. Meiklejohn, 2021 Hybrid Sterility, Genetic Conflict and Complex Speciation: Lessons From the Drosophila simulans Clade Species. Frontiers in Genetics 12.
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Rathje, C. C., E. E. P. Johnson, D. Drage, C. Patinioti, G. Silvestri *et al.*, 2019 Differential Sperm Motility Mediates the Sex Ratio Drive Shaping Mouse Sex Chromosome Evolution. Current Biology 29**:** 3692-3698.e3694.
Sin, H. S., Y. Ichijima, E. Koh, M. Namiki and S. H. Namekawa, 2012 Human postmeiotic sex chromatin and its impact on sex chromosome evolution. Genome Res 22**:** 827-836.
Sin, H. S., and S. H. Namekawa, 2013 The great escape: Active genes on inactive sex chromosomes and their evolutionary implications. Epigenetics 8**:** 887-892.
Skinner, B. M., C. C. Rathje, J. Bacon, E. E. P. Johnson, E. L. Larson *et al.*, 2019 A high-throughput method for unbiased quantitation and categorization of nuclear morphology. Biology of Reproduction 100**:** 1250-1260.
Skinner, B. M., C. A. Sargent, C. Churcher, T. Hunt, J. Herrero *et al.*, 2016 The pig X and Y Chromosomes: structure, sequence, and evolution. Genome Research 26**:** 130-139.
Smith, C. M., T. F. Hayamizu, J. H. Finger, S. M. Bello, I. J. McCright *et al.*, 2019 The mouse Gene Expression Database (GXD): 2019 update. Nucleic acids research 47**:** D774-D779.
Soh, Y. Q., J. Alfoldi, T. Pyntikova, L. G. Brown, T. Graves *et al.*, 2014 Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159**:** 800-813.
Storchová, R., S. Gregorová, D. Buckiová, V. Kyselová, P. Divina *et al.*, 2004 Genetic analysis of X-linked hybrid sterility in the house mouse. Mammalian Genome 15**:** 515-524.
Suzuki, T. A., and M. W. Nachman, 2015 Speciation and reduced hybrid female fertility in house mice. Evolution 69**:** 2468-2481.
Tao, Y., D. L. Hartl and C. C. Laurie, 2001 Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. Proceedings of the National Academy of Sciences 98**:** 13183-13188.
Teeter, K. C., L. M. Thibodeau, Z. Gompert, C. A. Buerkle, M. W. Nachman *et al.*, 2010 The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64**:** 472-485.

The UniProt Consortium, 2020 UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49**:** D480-D489.
Turelli, M., and H. A. Orr, 2000 Dominance, Epistasis and the Genetics of Postzygotic Isolation. Genetics 154**:** 1663-1679.
Turner, J. M. A., 2015 Meiotic Silencing in Mammals. Annual Review of Genetics 49**:** 395-412.
Turner, J. M. A., S. K. Mahadevaiah, P. J. I. Ellis, M. J. Mitchell and P. S. Burgoyne, 2006 Pachytene Asynapsis Drives Meiotic Sex Chromosome Inactivation and Leads to Substantial Postmeiotic Repression in Spermatids. Developmental Cell 10**:** 521-529.
Turner, L. M., and B. Harr, 2014 Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. eLife 3**:** e02504.
Turner, L. M., D. J. Schwahn and B. Harr, 2012 Reduced Male Fertility Is Common but Highly Variable in Form and Severity in a Natural House Mouse Hybrid Zone. Evolution 66**:** 443-458.
Vegesna, R., M. Tomaszkiewicz, O. A. Ryder, R. Campos-Sánchez, P. Medvedev *et al.*, 2020 Ampliconic genes on the great ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. Genome Biology and Evolution.
Vicoso, B., and B. Charlesworth, 2009 Effective population size and the faster-x effect: An extended model. Evolution 63**:** 2413-2426.
Vyskočilová, M., G. Pražanová and J. Piálek, 2009 Polymorphism in hybrid male sterility in wild-derived Mus musculus musculus strains on proximal chromosome 17. Mammalian Genome 20**:** 83.
Vyskočilová, M., Z. Trachtulec, J. Forejt and J. Piálek, 2005 Does geography matter in hybrid sterility in house mice? Biological Journal of the Linnean Society 84**:** 663-674.
Widmayer, S. J., M. A. Handel and D. L. Aylor, 2020 Age and Genetic Background Modify Hybrid Male Sterility in House Mice. Genetics 216**:** 585-597.
Wilkinson, G. S., S. J. Christianson, C. L. Brand, G. Ru and W. Shell, 2014 Haldane's Rule Is Linked to Extraordinary Sex Ratios and Sperm Length in Stalk-Eyed Flies. Genetics 198**:** 1167-1181.
Yang, H., J. R. Wang, J. P. Didion, R. J. Buus, T. A. Bell *et al.*, 2011 Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet 43**:** 648-655.
Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen *et al.*, 2019 Ensembl 2020. Nucleic Acids Research 48**:** D682-D688.
Zanders, S. E., M. T. Eickbush, J. S. Yu, J.-W. Kang, K. R. Fowler *et al.*, 2014 Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. eLife 3**:** e02630.
Zhang, L., T. Sun, F. Woldesellassie, H. Xiao and Y. Tao, 2015 Sex Ratio Meiotic Drive as a Plausible Evolutionary Mechanism for Hybrid Male Sterility. PLOS Genetics 11**:** e1005073.

Supplementary Figures

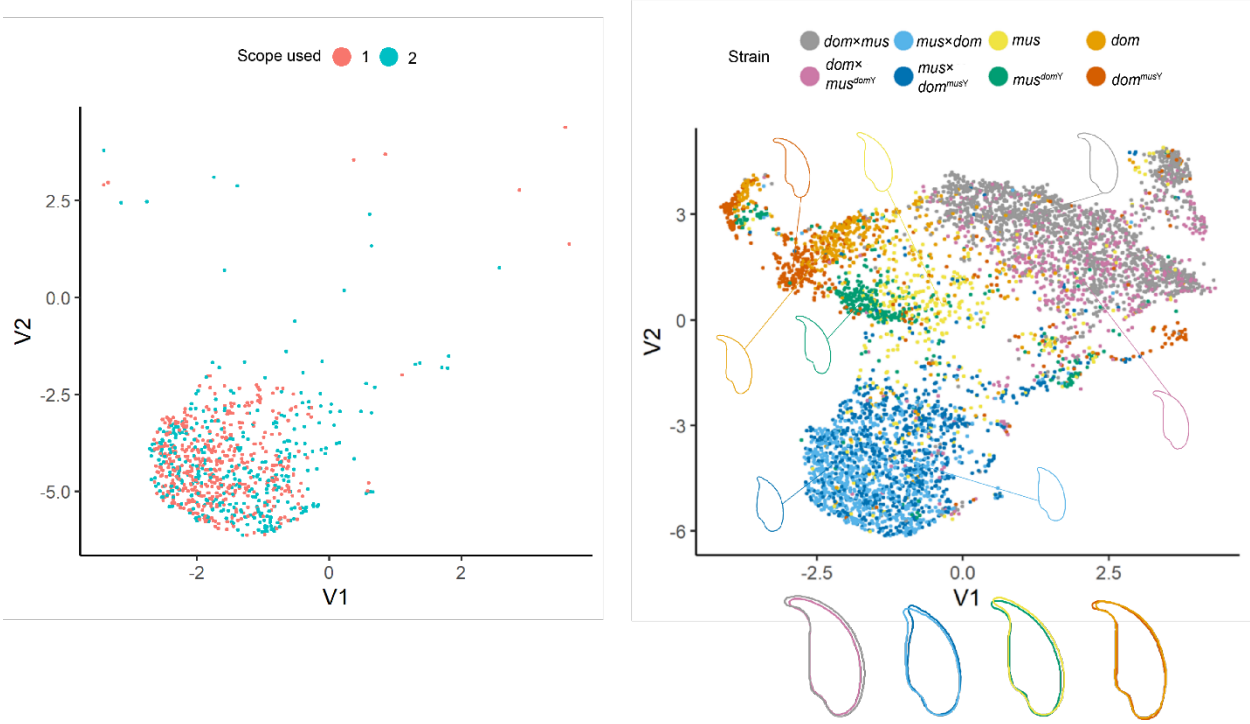

Figure S1: Dimensionality reduction plots (UMAP) of sperm nuclei morphology. (A) Nuclei from the *mus*×*dom* samples, which were imaged using two different microscopes, are evenly distributed within their cluster regardless of which microscope was used. This shows us that there was no experimental bias occurring based on which microscope was used, which allowed consistent data collection of the remaining samples from both scopes. (B) Clustering for all imaged nuclei colored by cross type.

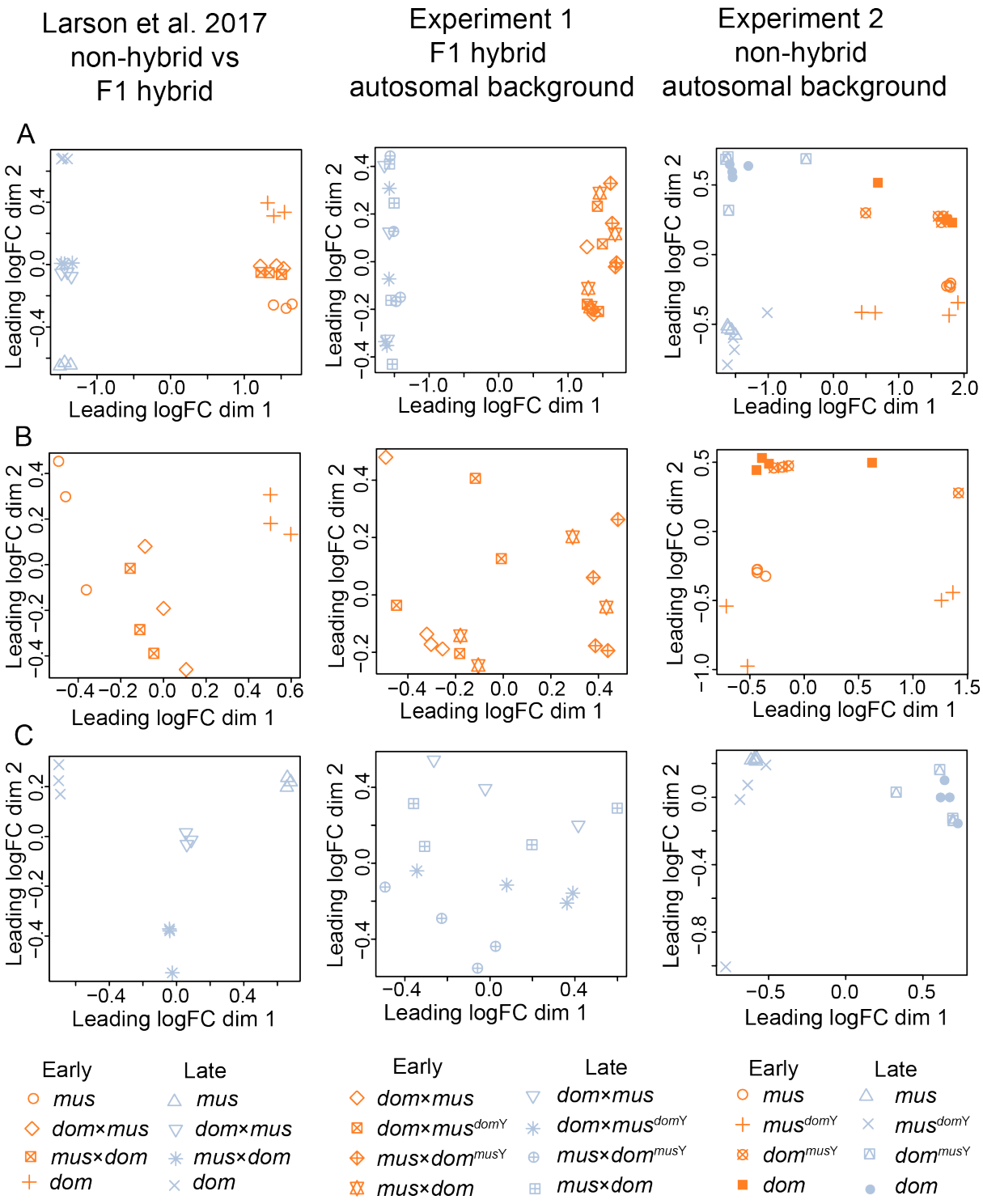

Figure S2: Multidimensional scaling (MDS) plots of distances between expression data from sorted cells. (A) Data from leptotene-zygotene (LZ, orange) and round spermatids (RS, blue) combined. (B) Data from LZ only. (C) Data from RS only. The first column shows data from our reanalysis of data from (Larson, et al. 2017). The second two columns show data collected from our Experiments 1 and 2. Each shape represents a different cross type, which are different across experiments. See legend at the bottom of the figure.

Figure S3: Expression levels of cell type marker genes in Experiment 2 leptotene-zygotene (LZ) samples. Each plot represents a different individual, labeled by sample ID. Each column contains samples from the same cross type. Marker genes on the x-axis are colored by the cell type they are preferentially expressed in, and are based on single-cell RNAseq data (Green, et al. 2018). The y-axis indicates gene expression level in FPKM. LZ are known to have similar expression profiles to spermatogonia, so the high expression levels of spermatogonia marker genes is expected (Larson, et al. 2016). Note that the absolute expression level of these marker genes in the cell types they represent is highly variable, and a previous study showed that the LZ marker gene has a median FPKM value of about 25, so the relatively low FPKM value for the LZ marker in these plots is also expected (Hunnicutt, et al. 2021). Red boxes indicate samples that appear to have contamination from diplotene and elongating spermatid cells based on their relative expression levels of marker genes for the cell types compared to other LZ samples. These are the same 4 samples that separate from other LZ samples on leading logFC dim in Figure S1A and B (3rd column), indicating that enough cell type contamination occurred in these samples to affect their overall expression profiles.

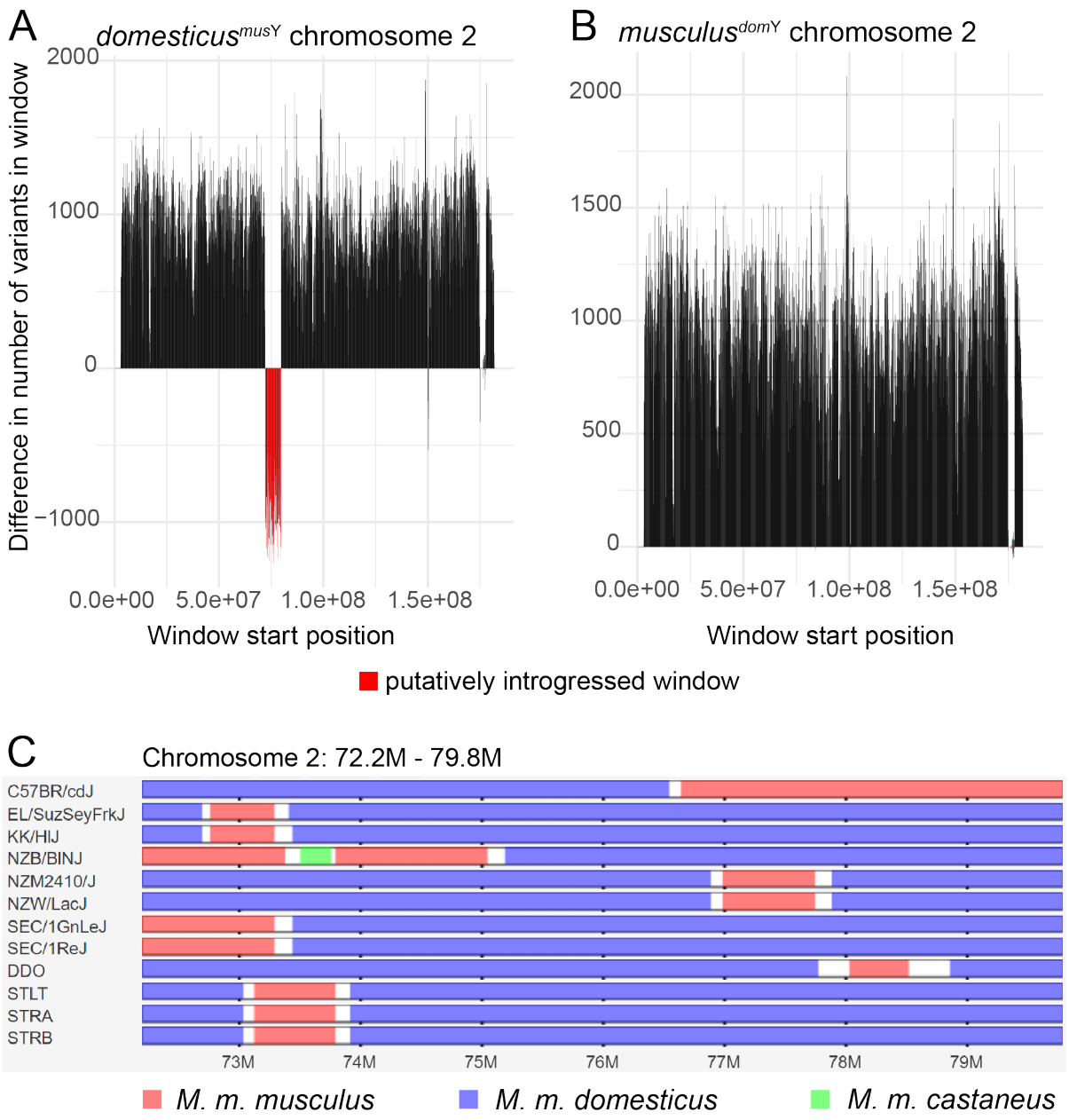

Figure S4: Evidence for introgression from *musculus* into *domesticus* on chromosome 2. (A) and (B) show the difference in the number of variants when Y introgression strains were mapped to their Y chromosome origin reference genome compared to their autosomal background reference genome in 100kb windows. Regions with evidence for introgression had more variants compared to the autosomal background reference than compared to the Y chromosome reference and are shown in red. Chromosome 2 has a large introgressed region in *domesticus*^{musY} (A) but not *musculus*^{domY} (B). (C) shows a screenshot from the Mouse Phylogeny Viewer (Yang, et al. 2011) depicting mouse inbred strains with evidence for introgression from *musculus* (red) into *domesticus* (blue) in the region of chromosome 2 where we found evidence for introgression.

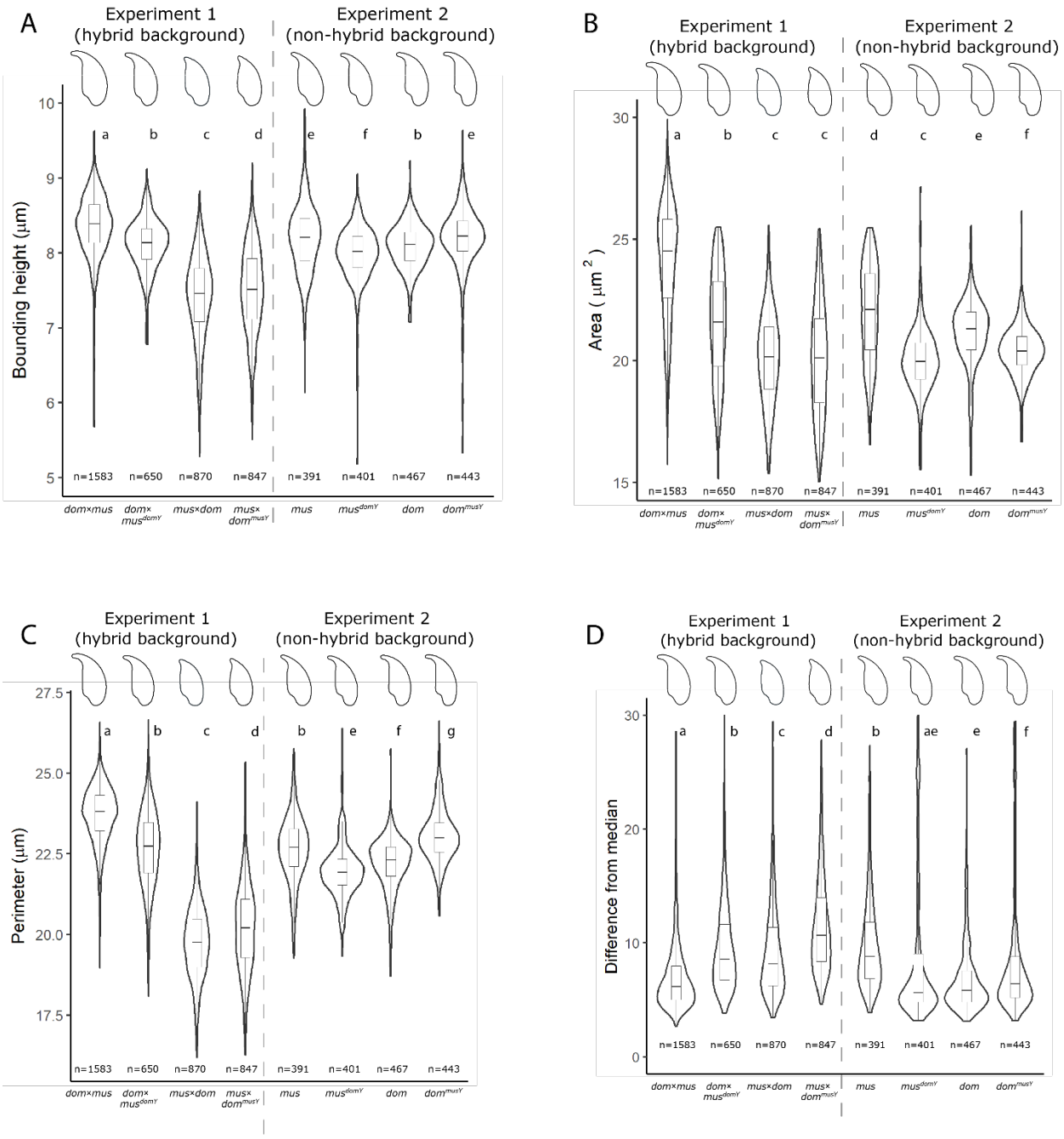

Figure S5: Violin plots showing sperm nuclear morphology parameters: (A) bounding height (μm), (B) area (μm^2), (C) perimeter (μm), and (D) difference from median. Difference from median is a measure of variance within cross types. Letters above each violin plot indicate significant differences among cross types based on an FDR-corrected pairwise Wilcoxon rank sum test. Numbers below each violin plot represent the number of sperm head nuclei observed for each cross type.

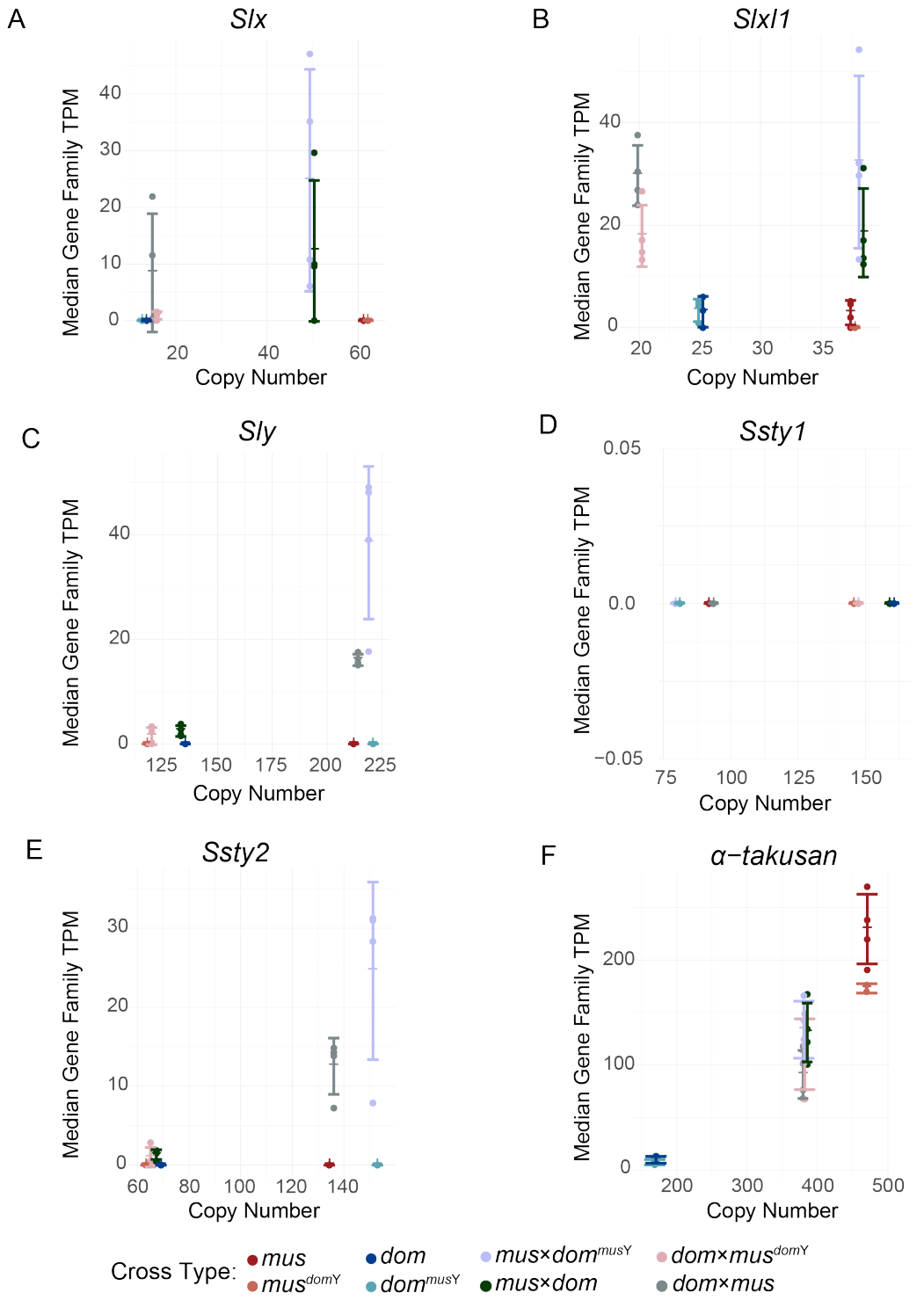

Figure S6: Normalized expression levels in leptotene-zygotene of *Slx* (A), *Slxl1* (B), *Sly* (C), *Ssty1* (D), *Ssty2* (E), and *α-takusan* (F) ampliconic gene families in different cross types plotted against their copy numbers. Expression level was calculated by summing transcripts-per million (TPM) for each paralog of the gene family with at least 97% sequence identity to the ampliconic gene. Points represent values for individual samples, and lines indicate median and standard deviation for each cross type.

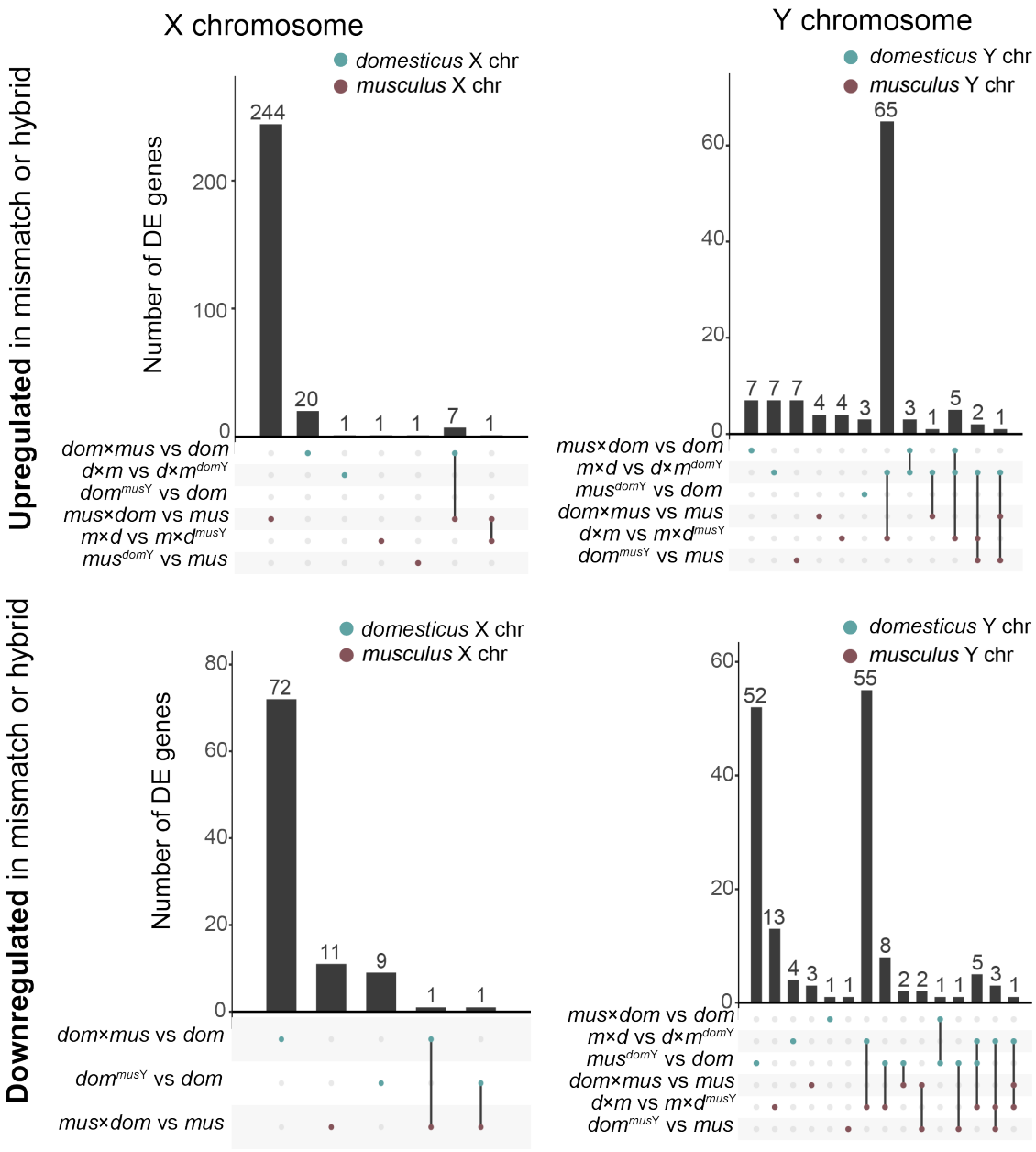

Figure S7: Upset plots showing the number of DE genes in each cross type comparison, and genes that are DE across multiple comparisons. (A) DE genes on the X chromosome overexpressed in F1 hybrids or XY mismatch mice relative to controls. (B) DE genes on the Y chromosome overexpressed in F1 hybrids or XY mismatch mice relative to controls. (C) DE genes on the X chromosome underexpressed in F1 hybrids or XY mismatch mice relative to controls. (D) DE genes on the Y chromosome underexpressed in F1 hybrids or XY mismatch mice relative to controls. Bars corresponding to multiple dots connected by lines indicate genes that are DE across multiple comparisons. Bars corresponding to single dots indicate genes that are DE in only one comparison. Blue dots indicate comparisons on the *domesticus* X chromosome (A and C) or *domesticus* Y chromosome (B and D), and red dots indicate comparisons on the *musculus* X chromosome (A and C) or *musculus* Y chromosome (B and D).

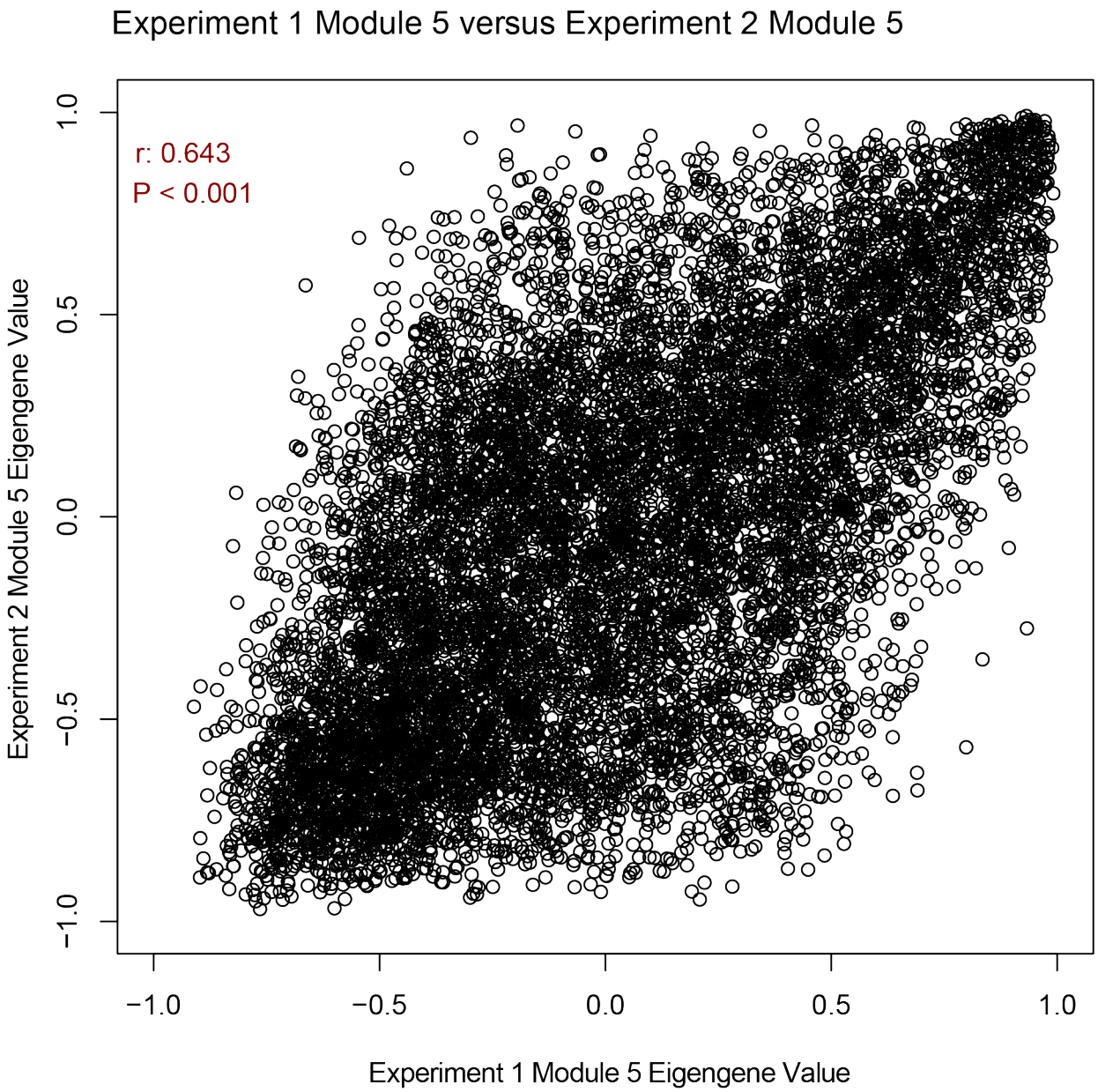

Figure S8: Plot showing the correlation between per-gene eigengene value in Experiment 1 Module 5 and Experiment 2 Module 5. Each point represents a gene, with its module membership (module eigengene value) in Experiment 1 Module 5 on the x-axis and its module membership in Experiment 2 Module 5 on the y-axis. Correlation coefficient and p-value are based on a Pearson's correlation test with FDR correction for multiple tests.

Supplementary Tables

Table S1: RNAseq metadata. Available as a separate attachment.

Table S2: Male reproductive trait raw phenotype data for each mouse sample. Available as a separate attachment.

Table S3: Copy number estimates in wild-derived inbred laboratory strains and Y-introgression strains. Results are presented using two different methods: a relative coverage approach using Mosdepth (Pedersen and Quinlan 2017) and a k-mer based coverage approach as implemented in AmpliCoNE (Vegesna, et al. 2019). For all genes except *Speer*, paralogs were based on a BLAT search and 97% sequence identity threshold cutoff. For *Speer*, we used a 90% sequence identity threshold cutoff, indicated by (*), because many annotated *Speer* genes had ~90-97% sequence identity with each other based on a BLAT search. Unlike other gene families, *Speer* copy number estimates were very different between the Mosdepth and AmpliCoNE approaches, likely because AmpliCoNE involves a mapping step that requires high sequence identity among paralogs (See Materials and Methods).

		Mosdepth								**AmpliCoNE**							
Cross type	**Strain**	***Slx***	***Slxl1***	***Sly***	***Sstx***	***Ssty1***	***Ssty2***	***α-takusan***	***Speer****	***Slx***	***Slxl1***	***Sly***	***Sstx***	***Ssty1***	***Ssty2***	***α-takusan***	***Speer****
musculus	PWK	48	34	192	36	136	123	729	238	50	38	213	33	93	135	570	3
musculus^{domY}	PWK.LY	50	38	148	40	200	78	729	240	52	35	119	33	147	64	569	3
domesticus^{musY}	LEWES.PY	17	23	211	45	139	135	259	119	16	21	220	37	80	152	191	3
domesticus	LEWES	16	22	152	43	201	82	254	127	15	20	134	32	160	68	195	3

Table S4: 100kb windows with evidence for introgression. Available as a separate attachment.

Table S5: Sex ratios produced by Y-introgression male mice with X-Y mismatch. Each column represents a different cross type involving Y introgression mice. Note that *domesticus*$^{musY,♀}$ and *musculus*$^{domY,♀}$ are females produced from reciprocal backcrosses for generating Y introgression males, but do not have an introgressed Y chromosome because they are females. P-values and chi-squared values are based on a Pearson's chi-squared test for a significant difference from a 50:50 sex ratio. We did not perform a correction for multiple tests because none of the p-values were significant. Power was calculated based on degrees of freedom = 1 and a significance level = 0.05. Effect sizes for power calculations were calculated by dividing the chi-squared value by the sample size and taking the square root.

	domesticus$^{musY,♀}$ × *domesticus*$^{musY,♂}$	*domesticus*♀ × *domesticus*$^{musY,♂}$	*musculus*$^{domY,♀}$ × *musculus*$^{domY,♂}$	*musculus*♀ × *musculus*$^{domY,♂}$
# male offspring	40	19	55	29
# female offspring	30	24	49	36
P-value	0.28	0.54	0.62	0.46
Chi-squared	1.157	0.372	0.240	0.554
Power (1 – Type II Error Probability)	0.19	0.09	0.08	0.12

References

Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB, et al. 2018. A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev Cell*.

Hunnicutt KE, Good JM, Larson EL. 2021. Unraveling patterns of disrupted gene expression across a complex tissue. *Evolution*. n/a.

Larson EL, Keeble S, Vanderpool D, Dean MD, Good JM. 2017. The Composite Regulatory Basis of the Large X-Effect in Mouse Speciation. *Mol Biol Evol*. 34:282-295.

Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BAJ, Smith AD, Dean MD, Good JM. 2016. Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. *Genetics*. 203:1841-1857.

Pedersen BS, Quinlan AR. 2017. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 34:867-868.

Vegesna R, Tomaszkiewicz M, Medvedev P, Makova KD. 2019. Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLoS Genet*. 15:e1008369.

Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, Bonhomme F, Yu AH, Nachman MW, Pialek J, et al. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet*. 43:648-655.

Rodents of Unusual Sperm: Molecular and Phenotypic Evolution of Male Reproduction in Murine Rodents

Authors: Emily E. K. Kopania, Gregg W. C. Thomas, Carl R. Hutter, Sebastian Mortimer, Colin Callahan, William G. Breed, Jacob A. Esselstyn, Kevin C. Rowe, Jeffrey M. Good

Abstract

Male reproductive traits can evolve extremely rapidly, and sperm competition is thought to underlie this rapid evolution. In parallel, genes expressed in reproductive tissues tend to diverge rapidly in protein coding sequence, which is often attributed to positive selection due to postmating sexual selection. However, few studies have connected rapid phenotypic evolution to rapid molecular evolution for reproductive traits and directly tested the role of positive selection in shaping rapid divergence. Furthermore, studies have shown that genes expressed in some male reproductive tissues or cell types are highly conserved, suggesting that evolutionary forces causing rapid divergence are more intense for particular reproductive functions or at different stages of sperm development. Investigating the causes of rapid phenotypic and molecular evolution for reproductive traits requires a well-resolved phylogeny with differences among species in reproductive phenotypes, as well as a nuanced understanding of genes involved in different tissues or cell types. Murine rodents provide an ideal system for studying reproductive evolution because they represent a rapid radiation comprising over 10% of mammal species with striking differences in reproductive phenotypes among species. Murines also include two model organisms, the house mouse and Norway rat, providing vast genomic resources and a good understanding of tissue- and developmental-specificity in this taxonomic group. In this study, we performed exome sequencing of over 200 murine species to infer a well-resolved phylogeny. For a subset of species with available phenotype data, we showed that relative testes mass was evolving independently of phylogeny. Most murine species have a hook on the sperm head, and our analyses showed that hook length and angle were correlated with relative testes mass after controlling for phylogeny, suggesting that these traits may be evolving

in response to sperm competition. We also showed that genes predominantly expressed in the seminal vesicles and during the postmeiotic stages of spermatogenesis tended to be the most rapidly evolving male reproductive genes across the murine phylogeny, and that the rapid evolution of postmeiotic spermatogenesis genes was due in part to positive selection. Collectively, these results demonstrate that sperm competition and positive selection likely play a central role in the rapid phenotypic and molecular evolution of male reproductive traits in murine rodents, but the intensity of these forces and their importance in shaping evolutionary patterns is highly variable across traits, tissues, and developmental stages.

Introduction

Sperm competition is thought to select for extreme reproductive traits in males (Pitnick *et al.* 2009, Simmons 2019), which may drive both rapid phenotypic divergence within populations and the evolution of reproductive barriers between nascent species. In many taxa, the rapid evolution of reproductive traits is also seen at the molecular level, with genes involved in reproduction tending to show rapid protein sequence evolution (Swanson *et al.* 2001; Clark and Swanson 2005; Ahmed-Braimah *et al.* 2017; Dean *et al.* 2017; Roycroft *et al.* 2021). However, it remains unclear how trait evolution and molecular evolution relate across the complex developmental process of spermatogenesis or among different male reproductive tissues (Ramm *et al.* 2008; Wong 2011; Good *et al.* 2013; Claw *et al.* 2018; but see Wong 2014).

The hypothesis that sexual selection is the leading cause of rapid reproductive phenotype divergence predicts that phenotypic evolution should be correlated with the intensity of sperm competition (Breed and Taylor 2000; Wong 2011; Simmons and Fitzpatrick 2012; Lüpold *et al.* 2016). In some systems, there is a direct correlation between the intensity of sperm competition and reproductive phenotypes, such as larger relative testes mass in primates (Harcourt *et al.* 1981). In rodents, the intensity of sperm competition is sometimes correlated with relative testes mass across conspecific populations (Firman and Simmons 2008), and some studies have used relative testes mass as a proxy for the intensity of sperm competition across different species (Gómez Montoto *et al.* 2011; Pahl *et al.* 2018), but other factors could underlie the evolution of

increased testes size (Ramm and Schärer 2014). It is also unclear if or how other male reproductive traits, such as sperm form and function, relate to sperm competition (Simmons and Fitzpatrick 2012). Directly measuring the level of sperm competition is challenging in most taxa, but we can infer traits likely evolving in response to mating system by identifying reproductive traits with evolutionary divergence that cannot be explained by phylogeny (Pahl *et al.* 2018). This approach has rarely been applied because it requires a well-resolved phylogeny that includes many species with variation in their reproductive traits.

At the molecular level, rapid evolution is also often attributed to sexual selection, but other evolutionary forces may shape the molecular evolution of reproductive traits, and there is heterogeneity in the molecular evolutionary rates of genes enriched in different reproductive tissues or cell types (Dean *et al.* 2009; Larson *et al.* 2016; Finseth and Harrison 2018; Kopania *et al.* 2022). Spermatogenesis provides a key example of these contrasting forces. Sperm competition is thought to act on both the rate of sperm production and sperm head morphology (Pitnick *et al.* 2009), so sexual selection is predicted to be more intense during early spermatogenesis stages in which rates of cell division can determine the overall rate of sperm production, and during late stages in which sperm elongate and form their mature shape (Larson *et al.* 2018). However, developmental constraints may also shape spermatogenesis evolution, because genes expressed during late spermatogenesis stages tend to be tissue specific and therefore less subject to pleiotropic constraint imposed by their roles in other tissues (Eddy 2002; Green *et al.* 2018; Murat *et al.* 2021). We previously showed that genes expressed during the late postmeiotic stages of spermatogenesis diverge rapidly relative to genes expressed during the early meiotic stages of spermatogenesis, thus supporting this developmental model (Kopania *et al.* 2022). Other studies have shown differences in evolutionary rates among accessory male reproductive tissues, with seminal vesicle genes tending to evolve more rapidly (Dean *et al.* 2009). Seminal vesicles produce many seminal fluid proteins (SFPs) that are ejaculated with sperm and interact with the female reproductive tract, so rapid seminal vesicle divergence may reflect relaxed pleiotropic constraint due to the highly specialized function of this tissue or sexual selection on SFPs (Dean *et al.* 2009). For both spermatogenesis cell types and male

reproductive tissues, the relative importance of relaxed constraint and positive selection to rapid evolution remain unclear. Previous studies have been underpowered to test for positive selection using rates of protein sequence evolution (dN/dS) in a phylogenetic framework, because these tests require many taxa to have enough power (Anisimova *et al.* 2001). Because male reproductive genes tend to evolve rapidly, it is difficult to identify orthologs and produce alignments for many of them if taxa are too distantly related (Dean *et al.* 2009). Thus, testing for positive selection on male reproductive genes in a phylogenetic framework requires a large phylogeny that includes many, relatively closely related taxa.

Murine rodents are an excellent system in which to study male reproductive biology and test for positive selection on different groups of genes involved in reproduction (Roycroft *et al.* 2021). Murines, or species in the subfamily Murinae, include over 700 species and make up >10% of extant mammal species (Rowe *et al.* 2016; Roycroft *et al.* 2021). They have diverged in approximately the last 12 million years, representing one of the most rapid radiations within mammals (Rowe *et al.* 2019; Roycroft *et al.* 2021). Murines show striking diversity in their reproductive biology, with variation in litter size, number of mammae, relative testes mass, and sperm morphology (Breed *et al.* 2019; Roycroft *et al.* 2021). This diversity in sperm head morphology is particularly remarkable (Figure 1; Breed 1997; Breed 2005; Peirce *et al.* 2018; Breed *et al.* 2019). Previous studies have shown that some sperm morphology traits are correlated with relative testes mass in murines, and therefore these traits are thought to be evolving in response to sperm competition (Gómez Montoto *et al.* 2011; Pahl *et al.* 2018). However, these studies were done in smaller subsets of murines, so it is unclear if the same traits are evolving in response to sperm competition across the whole Murinae subfamily. Two model organisms, the house mouse (*Mus musculus*) and Norway rat (*Rattus norvegicus*) are part of the Murinae subfamily, providing extensive genomic resources to combine molecular and phenotypic evolution studies (Gibbs *et al.* 2004; Keane *et al.* 2011). Furthermore, we have a detailed understanding of spermatogenesis and reproductive biology in these model organisms, allowing us to study murine sperm evolution in both a developmental and ecological context (Green *et al.* 2018; Firman 2020). Thus, murines are an ideal taxonomic group for studying the

evolution of male reproductive traits and testing for positive selection on genes involved in specific spermatogenesis stages or reproductive tissues.

We generated whole genome or exome capture datasets for 188 species of murine rodents and used these data to generate gene trees for 11,775 coding loci. Of these, we identified a subset of 96 taxa that had available reproductive phenotype data and generated a summary species tree from gene trees. This provided a powerful dataset for investigating the evolution of male reproductive traits across murines and testing for positive selection on reproductive genes. We used this dataset to address three main questions: (i) Does the intensity of sperm competition underlie variation in sperm head morphology in murines, or is this variation largely explained by evolutionary history? (ii) Are genes enriched in late spermatogenesis and seminal vesicles that evolve rapidly in *Mus* also rapidly evolving across murines? (iii) What are the relative roles of positive selection and relaxed purifying selection in shaping the rapid molecular evolution of male reproductive genes?

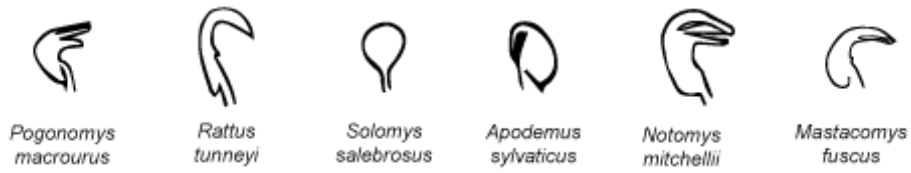

Figure 1: Examples of sperm head morphological diversity in murine rodents. Images are traced from micrographs. Typical murine sperm have an apical hook, with variation among species in the length and angle of this hook. Additionally, some species have evolved additional hooks, called ventral processes, or lost the hook altogether.

Materials and Methods

Samples and data collection

We sampled extensively from the field to obtain wild-caught individuals representing all of the major tribes within Murinae, with a particular emphasis on species in Southeast Asia, Australia, and Papua New Guinea. All samples collected for this study have been cataloged in museum collections, and we also obtained additional tissue samples from museum collections (museum accessions available on request). All specimens were

collected following the legal and ethical requirements for their country of origin, and tissues were preserved in ethanol. We extracted DNA using a Qiagen DNeasy kit with the following modifications to account for ethanol preservation. We rehydrated tissues with two 30-minute incubations of 1mL 1× STE buffer followed by three to five minutes of vortexing (Bi *et al.* 2013). To lyse samples, we added 20µL 1M DTT and 10µL 0.5M EDTA (Shapiro and Hofreiter 2012), in addition to the Buffer ATL and proteinase K provided in the Qiagen kit. We performed library preparation using a Kapa Biosystems HyperPrep Kit (Roche Diagnostics Corporation, Indianapolis, IN) and exome capture using SeqCap EZ Developer Probes (Roche Diagnostics Corporation, Indianapolis, IN). These probes were custom designed to target 203,188 exons based on the mm9 mouse reference genome. We used liftOver to transfer the coordinates for targeted regions to match the mm10 reference genome (Hinrichs *et al.* 2006).

Phenotypic data were compiled from several previous studies (Breed and Taylor 2000; McLennan *et al.* 2017; Pahl *et al.* 2018; Peirce *et al.* 2018; Breed *et al.* 2020). Relative testes mass was reported as percent of body mass (paired testes mass / body mass; Breed and Taylor 2000; McLennan *et al.* 2017; Peirce *et al.* 2018; Breed *et al.* 2020). In some cases, only one testis was weighed and its mass was doubled to approximate paired testes mass (Pahl *et al.* 2018; Breed *et al.* 2019). Sperm morphological traits were measured from scanning electron microscope images (McLennan *et al.* 2017; Pahl *et al.* 2018; Peirce *et al.* 2018; Breed *et al.* 2019). Sperm head length was measured from the base of the head to the base of the apical hook, and head width was measured perpendicular to head length at the widest part of the head (McLennan *et al.* 2017; Breed *et al.* 2019). Sperm head area was measured by tracing the outer surface of the sperm head including apical hooks and ventral processes and using an area tool (Pahl *et al.* 2018). Apical hook and ventral processes lengths were measured from the base to the tip of the hook along the concave surface of the hook (McLennan *et al.* 2017; Pahl *et al.* 2018; Breed *et al.* 2019). For sperm with multiple ventral processes, the length of the longest ventral process was used (McLennan *et al.* 2017). Apical hook and ventral processes angles were reported as the angle between the tangent line from the tip to the base of the hook and the line along

the sperm head longitudinal axis (Immler *et al.* 2007; McLennan *et al.* 2017; Pahl *et al.* 2018).

Assembly, alignment, and phylogenetic inference

We performed *de novo* exome assembly on samples from 210 species using SPAdes (Bankevich *et al.* 2012) and assessed assembly quality and corrected low-quality bases using Referee, which re-maps reads to assembled contigs (Thomas and Hahn 2019). We also re-mapped reads to the corrected assemblies to call genotypes and identify heterozygous sites in each sample. To annotate these assemblies, we selected transcripts based on the mouse and rat reference genomes (mm10 and rnor6) using the following criteria: (1) The transcript exists in both the mouse and rat references as a one-to-one ortholog with an orthology confidence of 1. (2) The transcript had a dS below 0.5 between mouse and rat. (3) If multiple transcripts from a gene passed the first two filters, we kept the one containing the highest number of probe targets from our exome capture probe set.

To generate alignments, we first identified homologous regions between mouse reference exons and our assembled contigs using BLAST. We then converted mouse reference exons to trimmed amino acid sequences to keep coding sequences between exons in-frame, and we used exonerate to identify exons in our assembled contigs that were homologous to mouse reference exons. We filtered out exons that had fewer than 175 samples with a matching BLAST hit and aligned the remaining exons using MAFFT (Katoh *et al.* 2002). We then back-translated to nucleotide sequences and filtered out sequences that were >20% gaps or 3-codon windows in which 2 or more codons had 2 or more gaps in over half of sequences for each alignment. Lastly, we removed samples with premature stop-codons. After filtering, our alignments had on average > 150 aligned sequences per protein, and the average non-gapped sequence length per protein was about 250 codons. Our final dataset, which we will refer to as the “full coding dataset”, included 188 species and 11,775 protein-coding genes. From these alignments, we inferred gene trees using IQtree v2.0.4 (Nguyen *et al.* 2014; Kalyaanamoorthy *et al.* 2017) with 1000 bootstrap iterations (Hoang *et al.* 2017). We then inferred a species tree from the gene trees using ASTRAL-MP v5.15.2 (Yin *et al.*

2019). We also inferred gene trees and an ASTRAL species tree for a subset of taxa for which we had reproductive phenotype data, giving us a dataset of 96 species and 12,993 protein-coding genes that we will refer to as the “reproductive phenotype dataset”.

Phylogenetic analyses of trait evolution

These analyses were performed using the reproductive phenotype dataset that contained 96 species. We did not have phenotype data for all species, so we pruned the tree for each trait to only include taxa with phenotype data. Sample sizes for each trait are reported in Table 1. We used the R package phytools v0.7-90 (Revell 2012) and R v4.0.3 for all phylogenetic analyses unless otherwise noted. We overlaid relative testes mass on the species tree using the function *dotTree*. We traced micrographs of sperm heads to generate the images in Figures 1 and 2 and counted the number of independent transitions to the multiple hooks and no hook phenotypes using the function *countSimmap.* For each phenotypic trait in Table 1, we estimated phylogenetic signal using the function *phylosig* to calculate Pagel’s ƛ (Pagel 1999) and perform a likelihood ratio test to test if ƛ was significantly different from zero and used false-discovery rate (FDR) correction for multiple tests. We tested for significant correlations between phenotypic traits after controlling for phylogeny using phylogenetic generalized least squares (pgls) implemented in the R package nlme v3.1-153 (Pinheiro J 2020) using the function *gls* with a Brownian motion model. We generated the Brownian motion model using the function *corBrownian* with value set to 1. For sperm hook presence/absence, we used a phylogenetic logistic regression because it is a binary trait (Ives and Garland 2009) implemented using the function *phyloglm* in the R package phylolm v2.6.2 (Tung Ho and Ané 2014).

Molecular evolution analyses

We used HyPhy 2.5 to calculate rates of protein sequence evolution (dN/dS; Kosakovsky Pond *et al.* 2019) using the full coding dataset consisting of 188 species. We previously showed that protein sequence divergence is higher for genes predominantly expressed during late spermatogenesis compared to early in *Mus,* the

genus that includes house mice (Kopania *et al.* 2022). To test if this pattern holds for a much broader sample across murines, we compared dN/dS for genes enriched for expression in different testes cell types. We estimated dN/dS for genes induced in early meiotic leptotene-zygotene and postmeiotic round spermatid cell types in *Mus* based on Kopania *et al.* (2022) and for testis-specific genes based on Chalmel *et al.* (2007). We also used supplemental table S3 from Green *et al.* (2018) to identify marker genes associated with each cell type cluster in their *Mus musculus* single-cell RNAseq dataset. We then used Figure 2B from Green *et al.* (2018) to merge genes from these clusters into five cell type categories: spermatogonia, pre-leptotene, spermatocytes, spermatids, and elongating spermatids. We identified somatic cell marker genes using supplemental table S4D from Green *et al.* (2018).

Previous work has shown that proteins enriched in seminal vesicles tend to evolve more rapidly than those in other male reproductive tissues in *Mus musculus domesticus* (Dean *et al.* 2009), so we also wanted to compare evolutionary rates across male reproductive tissues in our large sample of murines. Using supplementary tables 1 and 2 from Dean *et al.* (2009), we obtained lists of proteins associated with each tissue. We note that these datasets represent genes expressed in particular tissues and cell types in the *Mus* genus, so some genes in these datasets likely are not expressed in the same cell types across all of Murinae. However, these datasets provide a reasonable proxy for genes enriched in these cell types and tissues given that it is not feasible to generate expression data for all species in our dataset.

We sought to average dN/dS across branches in the murine species tree for sets of genes enriched in different male reproductive tissues or cell types. Because gene tree topologies do not always match the topology of the species tree, some branches in our species tree do not exist in the gene trees of some loci. Thus, there is high variation in the number of genes for which we can calculate a dN/dS value across branches, which can lead to biases in average dN/dS estimates. Some studies estimate dN/dS by using the species tree topology for every gene, but this can lead to overestimates of substitution rates, particularly for large datasets with a high frequency of ILS such as our dataset (Mendes and Hahn 2016). Others have addressed this issue by only estimating dN/dS for genes that have tree topologies matching the species tree

topology, but this requires filtering out a large amount of data and therefore is impractical for large datasets that are likely to have lots of gene tree discordance (Jarvis *et al.* 2014; Pease *et al.* 2016). To address these challenges, we developed a method to estimate a "concatenated" dN/dS across branches for each gene, which minimizes both errors in inferring substitutions and data loss. We calculated the number of synonymous (ES) and nonsynonymous (EN) sites and the number of synonymous (S) and nonsynonymous (N) changes across each gene. For each branch and gene set, we then summed ES, EN, S, and N across all genes that had that branch present and were in the relevant gene set. We then calculated concatenated dN/dS as:

$$\frac{\Sigma N/\Sigma EN}{\Sigma S/\Sigma ES} \quad (1)$$

In this formula, the numerator is the proportion of nonsynonymous sites that had a substitution across all genes in a set on a given branch; in other words, it is a "concatenated" form of dN. Similarly, the denominator is a "concatenated" dS. Unless otherwise stated, all dN/dS values reported in the text are concatenated dN/dS values. ES, EN, S, and N for each gene were calculated using the HyPhy program SLAC (Kosakovsky Pond and Frost 2005).

We also sought to test for positive directional selection on groups of genes enriched in male reproductive tissues and cell types. Running maximum likelihood tests for selection on the complete reproductive phenotype dataset was computationally demanding and would have resulted in unreasonable runtimes, so we pared the species tree from this dataset down to a subset of 41 taxa using Bonsai (https://github.com/gwct/bonsai). We pruned gene trees and alignments to only include taxa in this pared species tree, and we used these to identify sites under positive selection with the PAML site test (M1a vs M2a; Yang 2007). We then compared the proportions of genes with evidence for selection for genes enriched in reproductive cell types or tissues to the genome-wide average using a Pearson's chi-squared test with FDR-correction for multiple tests.

Results

Remarkable diversity of male reproductive traits across murine rodents

We used exome capture data to generate assemblies, alignments, gene trees, and a species tree for 188 Murinae species. Across samples, about 90% of reads on average mapped to our *de novo* assemblies. Based on Referee, our assemblies had on average fewer than 0.0005 errors per base and fewer than 2,500,000 low quality positions, indicating that our assemblies were of high quality. We generated a species tree for a subset of murine species that had available phenotype data (Figure 2). This dataset had 96 species, including two outgroup species. Our species tree was well-supported, with all but four nodes having an ASTRAL support value of 1. Of these four nodes, all had an ASTRAL support value > 0.92. However, concordance factors were relatively low, with a median gene concordance factor of 46.7 and a median site concordance factor of 52.5. These low concordance factors may reflect frequent incomplete lineage sorting or introgression, which are both known to cause high gene-tree/species-tree discordance in rapid species radiations (Martin *et al.* 2018; Hibbins *et al.* 2020).

We focused on two male reproductive traits: relative testis mass and sperm head morphology. Relative testes mass is often used as a proxy for the intensity of sperm competition, because higher levels of sperm competition are thought to select for increased rates of sperm production and therefore larger testes (Harcourt *et al.* 1981). However, cell type composition in the testes is highly variable across species, and it is unclear how well relative testes mass reflects the intensity of sperm competition across taxa (Ramm and Schärer 2014). Relative testes mass was highly variable across species in our dataset, ranging from 0.1% to 4.8% of body mass (Figure 2). Sperm morphology is thought to be important for sperm swimming speed and ability to fertilize the egg. Many murine rodents have sperm with an apical hook, or even multiple hooks called ventral processes, that vary considerably in length and angle (Figure 1, Figure 2). The exact functions of these hooks in murine rodents are not well understood, but they are thought to play a role in sperm competition (Immler *et al.* 2007; Firman and Simmons 2009). Previous studies showed that these sperm morphological characteristics have evolved multiple times in murines, but these studies used phylogenies based on one or a few genes, so evolutionary relationships were not as well resolved (Breed 1997; Breed 2005). Using our well-supported and fully resolved phylogeny, we confirmed that these traits have evolved multiple times in murines.

Across species in our dataset, we found evidence that the multiple hooks phenotype evolved independently at least three times, and there were also at least four independent reversions to the no hook phenotype (Figure 2).

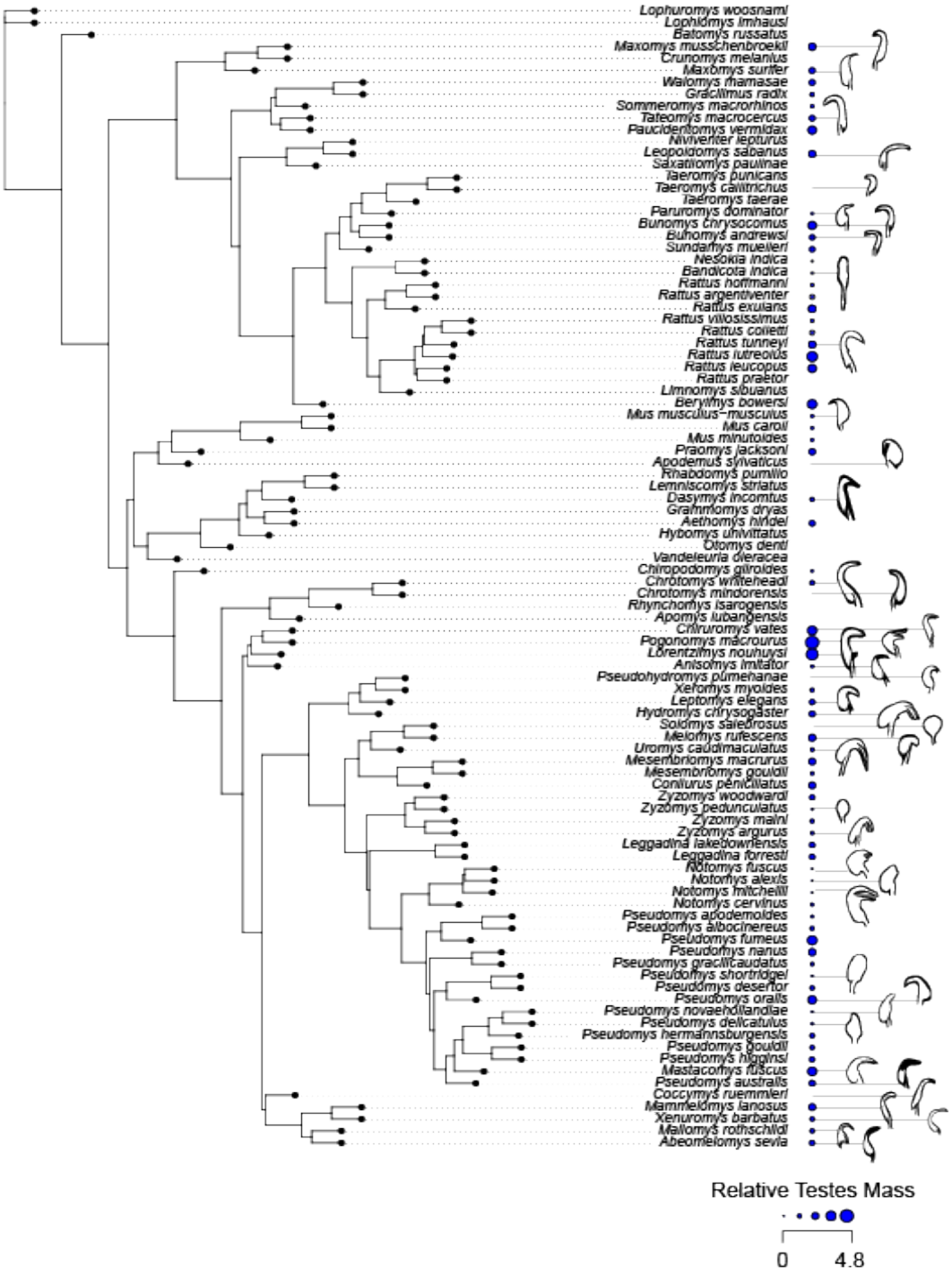

Figure 2: Rooted phylogeny of a subset of Murinae species. Blue circles indicate relative testes mass as a percent of body mass, with larger circles representing larger relative testes mass. Traces of sperm micrographs are shown for some species to show the diversity of sperm head morphology across murine rodents. Sperm head images are not to scale.

Sperm competition may underlie the divergence of some male reproductive traits

Some male reproductive traits appeared to evolve convergently across murines, independent of phylogenetic history (Figure 2). We wanted to test this quantitatively by estimating the role of phylogeny in predicting trait values (Pagel's ƛ; Pagel 1999). A ƛ of 0 indicates that a trait that is evolving independently of phylogeny, whereas a ƛ of 1 means that a trait is evolving according to a Brownian motion model, with more closely related taxa showing more similar trait values. Interestingly, most sperm morphology traits were largely explained by phylogeny (Table 1). For example, sperm head length, head width, apical hook angle, principal and end piece length, and tail length all had ƛ > 0.9. Most of these traits had ƛ significantly different from zero (likelihood ratio test FDR-corrected $P < 0.05$). However, ƛ for sperm head width was not significantly different from zero despite having a ƛ = 0.9 (likelihood ratio test FDR-corrected $P = 1$). Relative testes mass tended to evolve independently of phylogeny, with ƛ = 0.4 and not significantly different from zero (likelihood ratio test FDR-corrected $P = 0.33$, Table 1). Number of hooks or ventral processes had a low ƛ value (ƛ = 0.2) but was significantly different from zero (likelihood ratio test FDR-corrected $P = 0.01$).

Because relative testes mass is thought to be a proxy for the intensity of sexual selection (Harcourt *et al.* 1981), reproductive traits that are correlated with relative testes mass may be evolving in response to sexual selection (Pahl *et al.* 2018). We used phylogenetic logistic regression (binary traits; Ives and Garland 2009) or phylogenetic generalized least squares analysis to test if sperm morphology traits were correlated with relative testes mass in murines. Hook presences was significantly associated with relative testes mass, as species with larger testes tended to have at least one hook (FDR-corrected $P = 0.0023$; Figure 3). We found that apical hook length, apical hook angle, and ventral process angle were significantly correlated with relative

testes mass after controlling for phylogeny using a Brownian motion model (FDR-corrected P = 0.007, 0.002, 0.03; Figure 3; Table 2). These results are consistent with a previous study that found a strong positive correlation between relative testes mass and both apical hook angle and apical hook length in the Rattini, a tribe within Murinae (Pahl *et al.* 2018). However, traits related to sperm head size and sperm length were not significantly associated with relative testes mass (FDR-corrected P > 0.05; Table 2). Thus, sperm competition appeared to select for longer hooks and increased hook angles, but not other sperm morphology traits.

Table 1: Pagel's ƛ for phenotypic traits. Traits highlighted in gray had ƛ values significantly different from 0 (FDR-corrected P-value < 0.05), indicating traits that were evolving under a Brownian motion model and therefore showed evolutionary divergence that was largely explained by phylogeny. Sample sizes for each trait are shown in the "N" column.

Traits	N	Pagel's λ	FDR-corrected P-value
Head length (μm)	41	1.00	2.45E-07
Head width (μm)	36	0.90	1
Head area (μm^2)	22	0.55	1
Apical hook length (μm)	58	0.82	0.24
Apical hook angle (°)	56	1.00	3.26E-08
Ventral process length (μm)	36	0.62	0.02
Ventral process angle (°)	28	0.56	0.04
Principal and end piece length (μm)	40	1.00	5.98E-04
Tail length (μm)	42	1.00	5.98E-04
Number of hooks or ventral processes	53	0.24	0.01
Relative testes mass (% body mass)	64	0.43	0.33

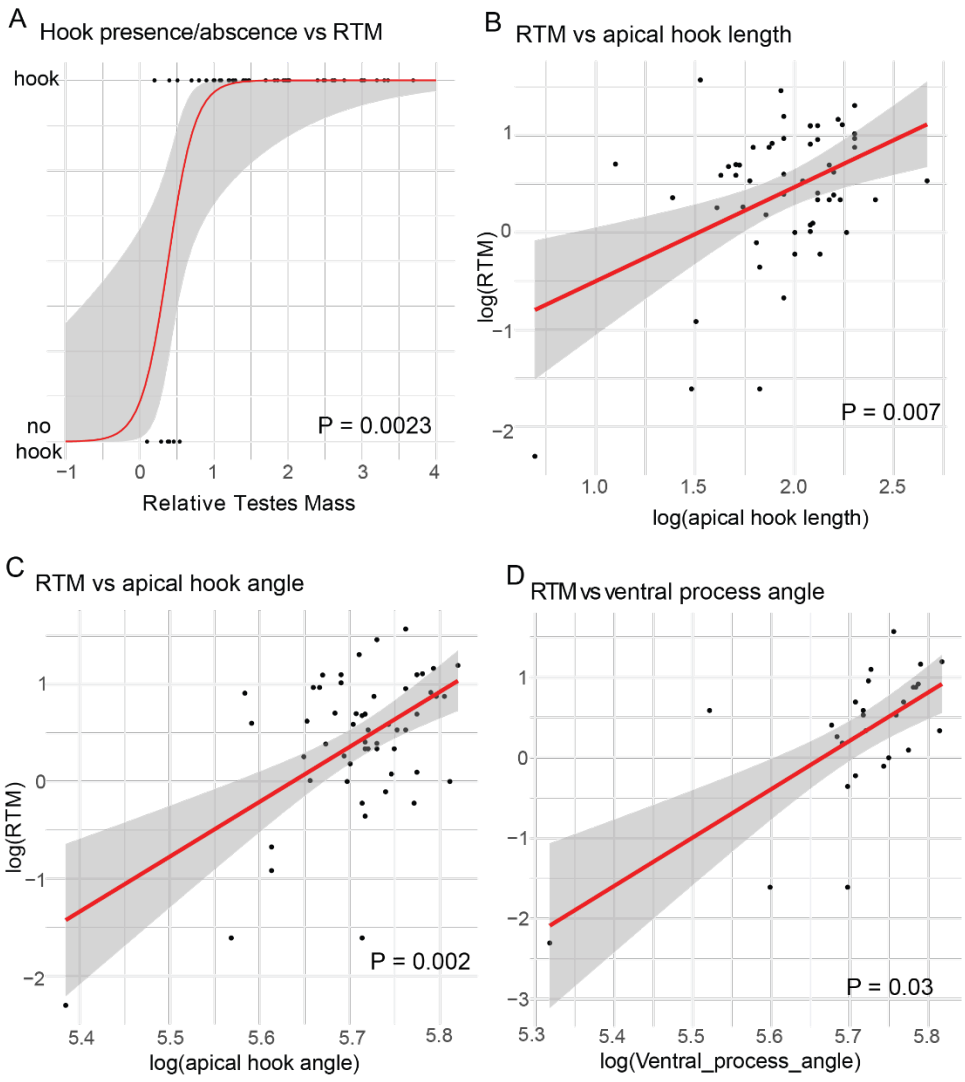

Figure 3: Correlations between relative testes mass (RTM) and sperm morphology traits. Red lines and gray areas show regressions and confidence intervals based on a logistic regression (A) or generalized linear models (B-D). P-values are based on phylogenetic logistic regression (A) or phylogenetic generalized least squares analyses using a Brownian motion model (B-D). Taxa with no hook were removed for (B and C), and taxa with no ventral processes were removed for (D).

Table 2: Phylogenetic generalized least squares results. The **"**Model" column indicates tests for different sperm morphology traits associated with relative testes mass (RTM). Rows highlighted in gray indicate traits that were significantly associated with relative testes mass after FDR-correction ($P < 0.05$).

Model	PGLS FDR-corrected P-value
Head length ~ RTM	0.543
Head width ~ RTM	0.076
Head area ~ RTM	0.384
Apical hook length ~ RTM	0.007
Apical hook angle ~ RTM	0.002
Ventral process length ~ RTM	0.063
Ventral process angle ~ RTM	0.030
Principal and end piece length ~ RTM	0.137
Tail length ~ RTM	0.116

Genes enriched in late spermatogenesis and seminal vesicles evolved more rapidly

We previously showed that genes enriched for expression in postmeiotic round spermatids had high rates of protein sequence divergence in *Mus* (i.e., high dN/dS; Kopania *et al.* 2022). To test if this result held across murines, we compared dN/dS across the full murine phylogeny (188 species) for genes induced in early meiotic and postmeiotic cell types in *Mus* based on expression data from Kopania *et al.* (2022). We used the concatenated dN/dS approach described in the methods to account for gene tree discordance. Genes induced during early meiosis in *Mus* had lower dN/dS than the genome-wide average in our murine dataset (Table 3; Figure 4A; genome-wide median dN/dS: 0.176, early meiosis median dN/dS: 0.157, FDR-corrected Wilcoxon rank-sum test $P << 0.0001$). In contrast, genes induced during postmeiotic spermatogenesis in *Mus* showed higher dN/dS than the genome-wide average (Table 3; Figure 4A; genome-wide median dN/dS: 0.176, postmeiotic median dN/dS: 0.194, FDR-corrected Wilcoxon rank-sum test $P << 0.0001$). These results suggest that more rapid protein sequence divergence in late spermatogenesis is a general pattern across murines, and that this pattern is consistent for both relatively closely related lineages (*Mus* only, Kopania *et al.* 2022) and at much deeper evolutionary time scales (Murinae).

We also estimated dN/dS for genes with testis-specific expression patterns in *Mus musculus*. These genes had an average dN/dS of 0.228 in our murine dataset, significantly higher than the genome-wide average and the median dN/dS for genes induced in postmeiotic spermatogenesis (Figure 4A; FDR-corrected Wilcoxon rank-sum test $P << 0.0001$ for all pairwise comparisons). Although these genes may not be testis-

specific in other species, their elevated dN/dS in murines suggest that their rapid evolution is not limited to *Mus*, and that they are subject to similar evolutionary forces across murines.

We also tested if the “faster late” pattern holds across a more nuanced dissection of spermatogenesis cell types using single-cell RNAseq data from Green *et al.* (2018). We compared dN/dS for marker genes associated with testis somatic cells, diploid spermatogonia, early meiosis (pre-leptotene), meiotic spermatocytes, postmeiotic spermatids, and elongating spermatids (Figure 5). Somatic and early spermatogenesis cell types tended to have lower dN/dS than the genome-wide average, while spermatocytes, spermatids, and elongating spermatids had higher dN/dS, supporting the faster late pattern (Table 3; Figure 4B; FDR-corrected Wilcoxon rank-sum test $P < 0.05$ for all cell types compared to genome-wide average).

In addition to the testes, other male reproductive tissues often express rapidly evolving genes. However, this pattern does not apply to all reproductive tissues, with some reproductive tissues tending to express highly conserved genes. One study in mice found that proteins in the seminal vesicles were particularly rapidly evolving, whereas those in other male reproductive tissues were more conserved than the genome-wide average (Dean *et al.* 2009). We showed that the rapid evolution of seminal vesicle genes was consistent across all murine rodents (Table 3; Figure 4C; Figure 5), as genes involved in the seminal vesicles had significantly higher dN/dS than average (FDR-corrected Wilcoxon rank-sum $P < 0.05$). Three tissues, the bulbourethral gland, coagulating gland, and dorsolateral prostate, had lower dN/dS than average (FDR-corrected Wilcoxon rank-sum $P < 0.05$), suggesting that proteins involved in these tissues may be highly conserved.

It is important to note that many coding loci did not pass our filters, and therefore many proteins associated with spermatogenesis cell types or male reproductive tissues were not included in this analysis. Most male reproductive genes excluded from our final dataset were filtered out because they did not have a confident 1:1 ortholog between mouse and rat. For most cell types or tissues, about 50-70% of genes in the expression enrichment lists we used were included in analyses, except for the seminal vesicles, from which only 40% of genes were included. Although there was high variation in the

number of genes associated with each tissue, our concatenated dN/dS approach should have eliminated any potential biases due to differences in numbers of genes.

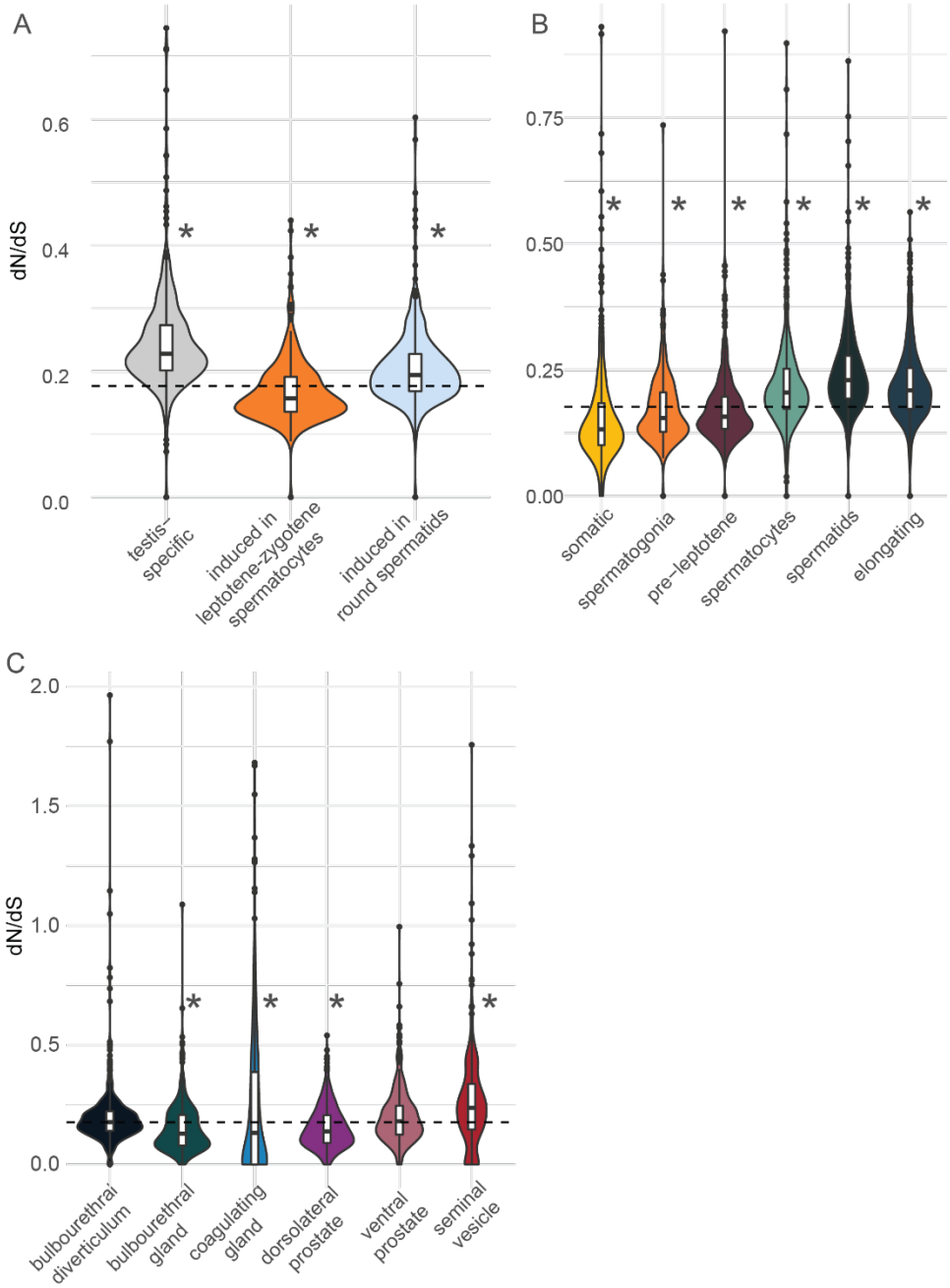

Figure 4: Violin plots of dN/dS for genes enriched for expression at different spermatogenesis stages or in different male reproductive tissues. Data are based on a concatenated dN/dS across all genes in a tissue or cell type calculated for each branch in the species tree. Therefore, variance in each violin plot is variance among branches, not among genes. (A) Genes testis-specific or induced for early meiotic or postmeiotic

expression in *Mus* (based on Kopania *et al.* 2022). (B) Genes enriched in different spermatogenesis cell types in *Mus musculus* based on single-cell expression data (Green *et al.* 2018). (C) Genes enriched in different male reproductive tissues based on proteomics data (Dean *et al.* 2009). For each plot, the dotted line represents the concatenated dN/dS for all coding loci in our dataset, averaged across branches. Asterisks indicate significant differences from the genome-wide average based on an FDR-corrected Wilcoxon rank-sum test ($P < 0.05$).

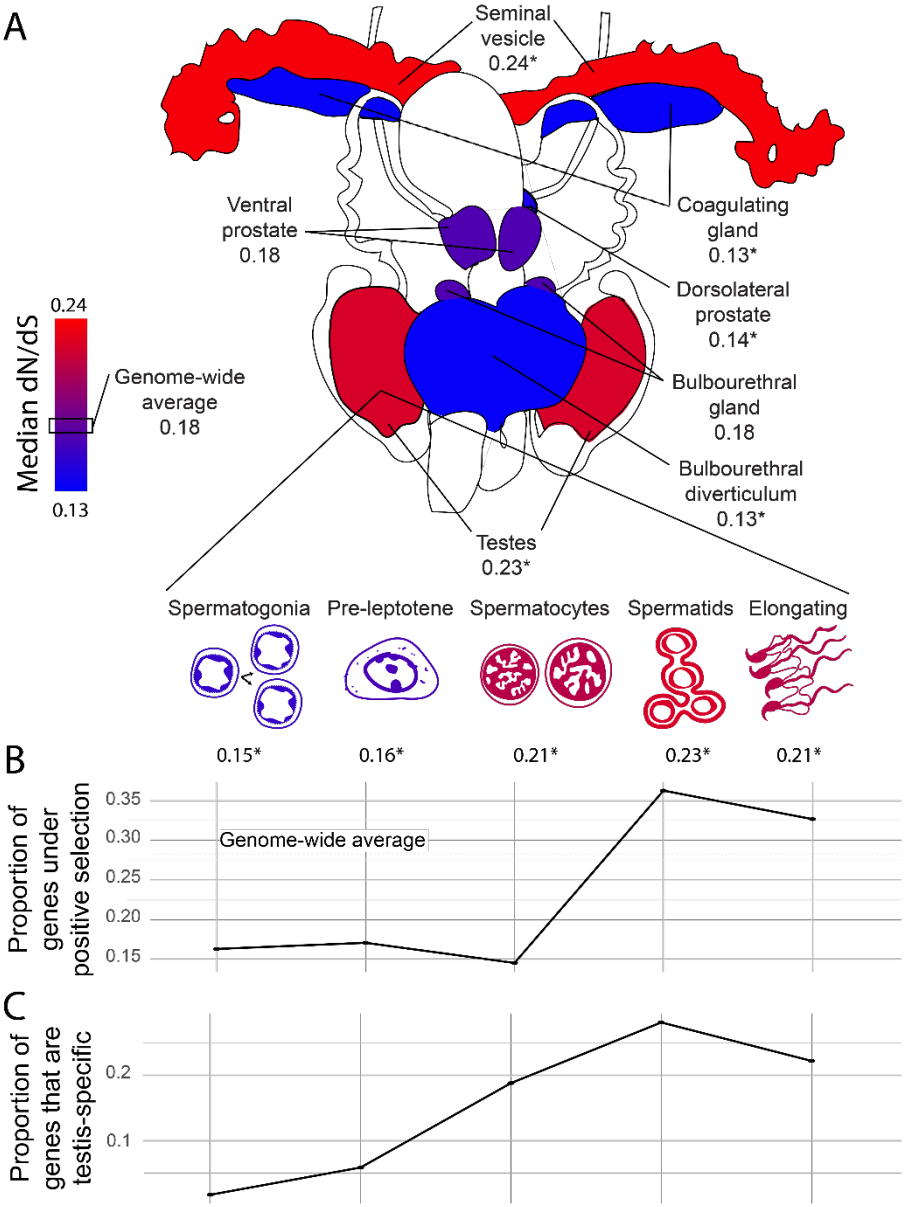

Figure 5: Molecular evolution by male reproductive tissue and spermatogenesis stage. In part (A), the top part of the figure shows the mouse male reproductive tract. The bottom part of the figure shows spermatogenesis cell types, with the earliest stages on the left to the latest stages on the right. Numbers below each label represent median dN/dS values for each tissue or cell type. Tissues and cell types are colored by their median dN/dS values, with warmer colors indicating higher dN/dS. The box on the scale indicates the median dN/dS (0.18) across all genes and branches in the murine species tree, and asterisks indicate significant differences from this median dN/dS. The median dN/dS for the testes is based on testis-specific genes from (Chalmel *et al.* 2007). The male reproductive tract was traced from (Dean *et al.* 2009), the pre-leptotene cell was traced from (Endo *et al.* 2015), and all other cell images were adapted from (Larson *et al.* 2018). Figure concept adapted from (Dean *et al.* 2009). (B) shows the proportion of genes enriched in each cell type with evidence for positive selection. The dotted line depicts the proportion of genes with evidence for positive selection out of all genes included in the test for selection. (C) shows the proportion of genes enriched in each cell type that are testis-specific.

Positive selection contributed to the rapid divergence of postmeiotic spermatogenesis genes

The rapid protein sequence evolution of genes enriched during late spermatogenesis and in seminal vesicles may result from relaxed purifying selection, more frequent positive selection, or both. We tested for positive selection on genes enriched in these cell types and tissues using a site test in PAML (M1a vs M2a; Yang 2007). For spermatogenesis cell types, proportionally fewer genes enriched in spermatogonia, pre-leptotene cells, and spermatocytes showed evidence for positive selection compared to the genome-wide average, whereas rapidly evolving round spermatids had a significantly higher proportion of genes with evidence for positive selection (Figure 5; Table 3; FDR-corrected $P < 0.05$). Genes enriched in somatic cells found in the testes also had proportionally fewer genes under positive selection (Table 3; FDR-corrected $P < 0.05$). Thus, higher rates of positive selection likely contributed to the rapid divergence

of genes enriched in postmeiotic round spermatids compared to other cell types in the testes.

For other male reproductive tissues, there were no significant differences in the intensity of positive selection acting on genes enriched in these tissues (Table 3; FDR-corrected $P > 0.05$). Relatively few genes (3-137 genes depending on the tissue) were both enriched in a particular tissue and included in our tests for selection, thus limiting power to detect significant differences in the proportion of genes under positive selection. Many genes predicted to be under positive selection based on previous studies, such as those encoding seminal vesicle secreted (SVS) proteins (Ramm *et al.* 2008), were excluded from selection tests because they are duplicated in some taxa (Dean *et al.* 2009), which can lead to false signatures of positive selection (Casola and Hahn 2009). Therefore, the lack of evidence for elevated rates of positive selection in any male reproductive tissues may reflect limitations of our dataset.

In addition to positive selection, relaxed purifying selection can result in rapid molecular evolution, and many male reproductive genes are highly tissue or cell type specific, leading to the prediction that they may be under relaxed purifying selection due to low pleiotropic constraint. Mid- to late-stage spermatogenesis cell types had proportionally more testis-specific genes compared to the genome-wide average (Table 3; FDR-corrected $P < 0.05$). Round spermatids, which had the highest average dN/dS value across enriched genes, also had the highest percent of enriched genes that were testis-specific (Figure 5; Table 3). Generally, a higher proportion of testis-specific genes were under positive selection compared to the genome-wide average. Testis-specific genes likely experience relaxed purifying selection because they are not constrained by functions in other tissues, and therefore testis-specific genes may be more free to diverge rapidly under positive selection (Murat *et al.* 2021). Round spermatids showed evidence for both elevated positive selection and testis-specificity compared to the genome wide average, and a relatively high proportion of testis-specific genes enriched in round spermatids were also evolving under positive selection (39.3%; Table 3). Relaxed purifying selection and positive selection may have acted together to allow the extremely rapid protein coding divergence of a subset of postmeiotic spermatogenesis genes.

Table 3: Summary of molecular evolution by spermatogenesis stage. Each row represents a different measure of molecular evolution, selection, or expression specificity. The first column shows the average values across all genes included in the full 188 taxa dataset. All other columns show averages for genes enriched in spermatogenesis cell types or male reproductive tissues, grouped by datasets from different sources. Gray boxes indicate significant differences from the average across all genes (FDR-corrected P < 0.05 based on a Wilcoxon rank sum test for dN/dS or a Pearson's chi-squared test for all other rows). LZ = leptotene-zygotene; RS = round spermatids; SO = somatic; SG = spermatogonia; PL = pre-leptotene; SC = spermatocytes; SD = spermatids; EL = elongating; BD = bulbourethral diverticulum; BG = bulbourethral gland; CG = coagulating gland; DP = dorsolateral prostate; VP = ventral prostate; SV = seminal vesicle

	All Genes	**Spermatogenesis cell types (Kopania *et al.* 2022)**		**Spermatogenesis cell types (Green *et al.* 2018)**						**Male reproductive tissues (Dean *et al.* 2009)**					
	All Genes	LZ	RS	SO	SG	PL	SC	SD	EL	BD	BG	CG	DP	VP	SV
Median dN/dS	0.18	0.16	0.19	0.13	0.15	0.16	0.21	0.23	0.21	0.13	0.18	0.13	0.14	0.18	0.24
Percent of genes with evidence for positive selection	28.2%	22.4%	32.2%	15.0%	16.3%	17.1%	14.5%	36.3%	32.7%	34.0%	20.4%	33.3%	28.2%	26.8%	25.0%
Proportion of testis-specific genes with evidence for positive selection	38.0%	25.0%	43.0%	50.0%	16.7%	21.7%	18.5%	39.3%	34.7%	NA	NA	NA	NA	NA	NA
Proportion of genes that are testis-specific	3.5%	1.6%	15.1%	3.4%	1.6%	5.8%	18.8%	28.2%	22.3%	NA	NA	NA	NA	NA	NA

Discussion

Murine rodents exhibit extremely diverse reproductive phenotypes (McLennan *et al.* 2017; Pahl *et al.* 2018; Peirce *et al.* 2018). Many of these traits are thought to vary in response to the intensity of sperm competition, but this had not been tested on a large dataset representing all of the major tribes within Murinae, or with a well-resolved phylogeny. Murines also show rapid evolution at the molecular level for genes expressed in reproductive tissues (Roycroft *et al.* 2021). Studies in the murine genus *Mus* have revealed heterogeneity in evolutionary rate across spermatogenesis cell types and male reproductive tissues (Dean *et al.* 2009; Larson *et al.* 2016; Kopania *et al.* 2022). However, these studies lacked the power to test for positive selection and identify the evolutionary forces underlying rapid protein sequence evolution. In this study, we inferred a phylogeny using exome capture data from 96 murine species representing the major tribes within Murinae. We used phenotypic data from these taxa and cell-type or tissue-specific expression data from *Mus* to investigate both the phenotypic and molecular evolution of male reproduction in murines.

Evolution of male reproductive phenotypes in murines

Murines have incredibly diverse reproductive traits (Breed *et al.* 2020). In this study, we focused on phenotypes related to two male reproductive traits: relative testes mass and sperm head morphology. We showed that relative testes mass evolved largely independently of phylogeny, with large testes evolving multiple times in murines (Figure 2, Table 1). Relative testes mass is often used as a proxy for the intensity of sexual selection, and relative testes mass is directly associated with the frequency of female multiple mating in primates (Harcourt *et al.* 1981) and house mice (Firman and Simmons 2008). However, the testis is a complex tissue with many cell types and many potential ways of increasing the rate of sperm production that may not result in an overall increase in testes size (Ramm and Schärer 2014). In addition to rates of sperm production, the testes are also involved in determining sperm morphology and the production of some hormones, so many selection pressures likely act on testis histology and size (Ramm and Schärer 2014). In murine rodents, most species are thought to mate multiply, but directly testing the intensity of sperm competition in rodents is

challenging and usually requires genotyping wild litters to test for multiple paternity, so it has only been done in a few species such as *Mus musculus* and *Rattus norvegicus* (Dean *et al.* 2006; Firman and Simmons 2008; Costa *et al.* 2016; Glass *et al.* 2016). Some behaviors such as paternal care are often used as proxies to infer monogamous mating systems, but these behavioral studies may be inconsistent with relative testes mass data (Cassaing *et al.* 2009). For example, *Mus spretus* shows some degree of paternal care in both laboratory and field studies, suggesting this species may be monogamous (Cassaing *et al.* 2009), but *M. spretus* has large relative testes mass compared to other rodents (Gomendio *et al.* 2006). Because it is not feasible to directly test the intensity of sperm competition in most murine species, we used relative testes mass as a proxy for the intensity of sperm competition, but with the caveat that the relationship between relative testes mass and sperm competition remains unclear in Murinae.

Sperm morphology is thought to be under strong selection (Pitnick *et al.* 2009), and postmating sexual selection can lead to the evolution of extreme sperm traits (Lüpold *et al.* 2016). Therefore, sperm morphological traits might be predicted to correlate with proxies of postmating sexual selection such as relative testes mass (Varea-Sánchez *et al.* 2016; Pahl *et al.* 2018). However, when we compared sperm morphology to relative testes mass, most sperm head characteristics were not correlated with relative testes mass after controlling for phylogeny (Table 2). Sperm morphological traits that are not correlated with relative testes mass may be evolving in response to different sexual selection pressures that do not select for an overall increase in sperm production, such as cryptic female choice (Firman *et al.* 2020; Higginson *et al.* 2012). We did see significant correlations between relative testes mass and apical hook length, apical hook angle, and ventral processes angle after controlling for phylogeny (Figure 3; Table 2), suggesting that these traits may be evolving in response to sperm competition.

These results are consistent with a previous study in Murinae (Immler *et al.* 2007) and in *Rattini,* a tribe within Murinae (Pahl *et al.* 2018). Apical hooks are a characteristic feature of sperm in many rodents, but their exact function is not known (Hook *et al.* 2021). In the wood mouse (*Apodemus sylvaticus*), sperm use the hook to

link up and form sperm trains with sperm from the same male, which facilitates faster movement through the reproductive tract (Moore *et al.* 2002). However, in other murine species such as *Mus musculus*, sperm trains are not common in ejaculates, and therefore they are probably not the main function of the hook (Firman and Simmons 2009). Sperm aggregates have been observed in *Mus musculus* and *Rattus norvegicus*, but they do not stay aggregated nearly as long as *Apodemus sylvaticus* sperm trains and they move slower than individual sperm (Immler *et al.* 2007). One alternative hypothesis in *Mus* is that the hook may allow sperm to attach to the oviduct epithelium to avoid being flushed out of the female reproductive tract between mating and oestrus (Firman and Simmons 2009). Hooks may interact with the female reproductive tract in other ways, such as interacting directly with the egg during fertilization (Hook *et al.* 2021). Although the exact function of the hook is unclear, sexual selection appears to play an important role in its evolution.

Variation in rates of molecular evolution across cell types and tissues

The rapid evolution of genes involved in reproduction is a widespread and well-characterized phenomenon (Swanson *et al.* 2001; Swanson and Vacquier 2002; Clark and Swanson 2005), but some reproductive cell types and tissues are also highly conserved (Dean *et al.* 2009; Finseth and Harrison 2018). Using expression data from *Mus,* we showed that testis-specific genes, late spermatogenesis genes, and seminal vesicle genes evolved rapidly (Figures 4 and 5; Table 3), confirming that results from previous studies in *Mus* extend to all murines (Dean *et al.* 2009; Kopania *et al.* 2022). In contrast, early spermatogenesis genes, the bulbourethral gland, coagulating gland (anterior prostate), and dorsolateral prostate showed significantly lower dN/dS than the genome-wide average (Figures 4 and 5; Table 3). We did not see a significantly lower dN/dS compared to the genome-wide average in the ventral prostate or bulbourethral diverticulum, different from the results in Dean *et al.* (2009). It may be that these genes were more conserved when calculating dN/dS based on *Mus* and *Rattus* but are more divergent across the whole murine phylogeny. It could also be that more conserved genes were excluded from our analysis, although this is unlikely because more divergent genes are more likely to be filtered out in our analysis pipeline.

We did not find evidence that rapid divergence in the seminal vesicles was due to positive selection (Figure 4C; Table 3), but this is likely because seminal vesicle *Svs* genes predicted to be under positive selection were excluded from our dataset. We filtered out *Svs* genes because they did not have a high confidence one-to-one ortholog between the mouse and rat reference genomes, likely because many *Svs* genes are duplicated (Dean *et al.* 2009), which can lead to false positive signatures of selection (Casola and Hahn 2009). In many taxa, proteins produced in the seminal vesicles make up parts of the ejaculate that directly interact with the female reproductive tract, and these genes tend to be under strong positive selection (Clark and Swanson 2005; Clark *et al.* 2006; Ramm *et al.* 2008). In mice, many seminal vesicle proteins are involved the formation of the copulatory plug (Ramm *et al.* 2008; Dean *et al.* 2009), and copulatory plugs increase male reproductive success, especially in the presence of sperm competition (Mangels *et al.* 2016; Lough-Stevens *et al.* 2020). The seminal vesicles are also enriched for genes involved in immune function in mice (Dean *et al.* 2009), and there is evidence for positive selection on immune function genes in primate ejaculates (Good *et al.* 2013). Immune system genes evolve rapidly in many taxa including Murinae (Roycroft *et al.* 2021), and these genes may explain the elevated dN/dS we observed in the seminal vesicles despite excluding the *Svs* genes from our study.

Rapid molecular evolution can result from relaxed purifying selection or positive selection, and these hypotheses are not mutually exclusive. Later stages of developmental processes are thought to experience relaxed developmental constraint (Abzhanov 2013), and the later stages of spermatogenesis tend to express genes that are highly specific (Green *et al.* 2018). Additionally, sperm elongate and take on their mature shape during these late stages, and sperm morphology is thought to be under strong sexual selection (Pitnick *et al.* 2009). Thus, we predicted that genes expressed primarily during late spermatogenesis would be evolving rapidly due to both relaxed constraint and positive selection. Our observations were consistent with these predictions, as we found greater protein sequence divergence in later spermatogenesis stages (Figure 4A, B; Figure 5) and evidence that late spermatogenesis genes were more likely to be evolving under positive selection (Figure 5; Table 3). Thus, positive selection probably contributed to some of the rapid evolution we observed in late

spermatogenesis genes. We also observed a significantly elevated proportion of testis-specific genes enriched in late spermatogenesis cell types (Figure 5; Table 3), so relaxed pleiotropic constraint due to high tissue specificity likely also contributed to this rapid divergence. Relaxed pleiotropic constraint and positive selection may have also interacted to allow for rapid divergence, as we saw a relatively high proportion of testes-specific genes under positive selection in late spermatogenesis.

Future Directions

Reproductive traits often evolve rapidly at both the phenotypic and molecular level, but few studies have directly connected phenotypic and molecular evolution for reproductive traits. Many studies have focused on the effects of individual genes rather than genome-wide patterns of rapid divergence (Ramm *et al.* 2008; Subrini and Turner 2021), and others that have tested for direct connections between mating system, phenotypic evolution, and molecular evolution have seen inconclusive results (Wong 2011; Claw *et al.* 2018). One approach to make these connections is to test for correlated shifts in molecular evolutionary rate associated with convergent phenotypes (Kowalczyk *et al.* 2019). Our dataset that includes convergent phenotypic changes in relative testes mass and sperm head morphology combined with exome sequence data provides a powerful framework for applying these tests. Future work will test for convergent evolutionary rate shifts to identify loci evolving rapidly in association with large testes, changes in sperm hook number, and changes in hook length and angle. These loci experiencing correlated rate shifts may be rapidly evolving due to relaxed purifying selection or positive selection. Our work so far showed that positive selection is a pervasive force underlying the rapid divergence of late spermatogenesis and seminal vesicle genes when testing for selection across all of Murinae. However, it remains unclear if the intensity of positive selection is similar across all taxa for these genes, or if selection is particularly strong in taxa thought to experience higher levels of sperm competition. Therefore, future work will use branch-site tests to identify bursts of episodic positive selection associated with convergent phenotypic traits (Kowalczyk *et al.* 2020). Connecting rapid phenotypic divergence to rapid molecular divergence and positive selection will provide novel insight into the tempo of male reproductive evolution

in murines and the evolutionary forces underlying the remarkable diversity of murine reproductive traits.

References

Abzhanov, A., 2013 von Baer's law for the ages: lost and found principles of developmental evolution. Trends in Genetics 29**:** 712-722.

Ahmed-Braimah, Y. H., R. L. Unckless and A. G. Clark, 2017 Evolutionary Dynamics of Male Reproductive Genes in the *Drosophila virilis* Subgroup. G3: Genes|Genomes|Genetics 7**:** 3145-3155.

Anisimova, M., J. P. Bielawski and Z. Yang, 2001 Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Molecular Biology and Evolution 18**:** 1585-1592.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin *et al.*, 2012 SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19**:** 455-477.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen *et al.*, 2013 Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22**:** 6018-6032.

Breed, W. G., 1997 Evolution of the Spermatozoon in Australasian Rodents. Australian Journal of Zoology 45**:** 459.

Breed, W. G., 2005 Evolution of the spermatozoon in muroid rodents. J Morphol 265**:** 271-290.

Breed, W. G., H. Hassan, M. Gonzalez, H. J. McLennan, C. M. Leigh *et al.*, 2019 Interspecific diversity of testes mass and sperm morphology in the Philippine chrotomyine rodents: implications for differences in breeding systems across the species. Reprod Fertil Dev 31**:** 705-711.

Breed, W. G., C. M. Leigh and E. J. Peirce, 2020 Reproductive biology of the mice and rats (family Muridae) in New Guinea—diversity and evolution. Records of the Australian Museum 72**:** 303-316.

Breed, W. G., and J. Taylor, 2000 Body mass, testes mass, and sperm size in murine rodents. Journal of Mammalogy 81**:** 758-768.

Cassaing, J., S. Cervera and F. Isaac, 2009 Laboratory and field evidence of paternal care in the Algerian mouse (*Mus spretus*). Journal of Ethology 28**:** 7.

Casola, C., and M. W. Hahn, 2009 Gene Conversion Among Paralogs Results in Moderate False Detection of Positive Selection Using Likelihood Methods. Journal of Molecular Evolution 68: 679-687.

Chalmel, F., A. D. Rolland, C. Niederhauser-Wiederkehr, S. S. W. Chung, P. Demougin *et al.*, 2007 The conserved transcriptome in human and rodent male gametogenesis. Proceedings of the National Academy of Sciences of the United States of America 104**:** 8346-8351.

Clark, N. L., and W. J. Swanson, 2005 Pervasive Adaptive Evolution in Primate Seminal Proteins. PLOS Genetics 1**:** e35.

Clark, N. L., J. E. Aagaard and W. J. Swanson, 2006 Evolution of reproductive proteins from animals and plants. Reproduction 131**:** 11-22.

Claw, K. G., R. D. George, M. J. MacCoss and W. J. Swanson, 2018 Quantitative evolutionary proteomics of seminal fluid from primates with different mating systems. BMC Genomics 19**:** 488.
Costa, F., J. L. Richardson, K. Dion, C. Mariani, A. C. Pertile *et al.*, 2016 Multiple Paternity in the Norway Rat, *Rattus norvegicus*, from Urban Slums in Salvador, Brazil. Journal of Heredity 107**:** 181-186.
Dean, M. D., K. G. Ardlie and M. W. Nachman, 2006 The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Molecular Ecology 15**:** 4141-4151.
Dean, M. D., N. L. Clark, G. D. Findlay, R. C. Karn, X. Yi *et al.*, 2009 Proteomics and Comparative Genomic Investigations Reveal Heterogeneity in Evolutionary Rate of Male Reproductive Proteins in Mice (*Mus domesticus*). Molecular Biology and Evolution 26**:** 1733-1743.
Dean, R., A. E. Wright, S. E. Marsh-Rollo, B. M. Nugent, S. H. Alonzo *et al.*, 2017 Sperm competition shapes gene expression and sequence evolution in the ocellated wrasse. Mol Ecol 26**:** 505-518.
Eddy, E. M., 2002 Male germ cell gene expression. Recent Prog Horm Res 57**:** 103-128.
Endo, T., K. A. Romer, E. L. Anderson, A. E. Baltus, D. G. de Rooij *et al.*, 2015 Periodic retinoic acid–STRA8 signaling intersects with periodic germ-cell competencies to regulate spermatogenesis. Proceedings of the National Academy of Sciences 112**:** E2347-E2356.
Finseth, F. R., and R. G. Harrison, 2018 Genes Integral to the Reproductive Function of Male Reproductive Tissues Drive Heterogeneity in Evolutionary Rates in Japanese Quail. G3: Genes|Genomes|Genetics 8**:** 39-51.
Firman, R. C., 2020 Of mice and women: advances in mammalian sperm competition with a focus on the female perspective. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 375**:** 20200082.
Firman, R. C., C. Gasparini, M. K. Manier and T. Pizzari, 2017 Postmating Female Control: 20 Years of Cryptic Female Choice. Trends in Ecology & Evolution.
Firman, R. C., and L. W. Simmons, 2008 The frequency of multiple paternity predicts variation in testes size among island populations of house mice. Journal of Evolutionary Biology 21**:** 1524-1533.
Firman, R. C., and L. W. Simmons, 2009 Sperm competition and the evolution of the sperm hook in house mice. Journal of Evolutionary Biology 22**:** 2505-2511.
Gibbs, R. A., G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren *et al.*, 2004 Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428**:** 493-521.
Glass, G. E., S. L. Klein, D. E. Norris and L. C. Gardner, 2016 Multiple Paternity in Urban Norway Rats: Extended Ranging for Mates. Vector-Borne and Zoonotic Diseases 16**:** 342-348.
Gomendio, M., J. Martin-Coello, C. Crespo, C. Magaña and E. R. S. Roldan, 2006 Sperm competition enhances functional capacity of mammalian spermatozoa. Proceedings of the National Academy of Sciences 103**:** 15113-15117.
Gómez Montoto, L., M. Varea Sánchez, M. Tourmente, J. Martín-Coello, J. J. Luque-Larena *et al.*, 2011 Sperm competition differentially affects swimming velocity

and size of spermatozoa from closely related muroid rodents: head first. Reproduction 142**:** 819-830.
Good, J. M., V. Wiebe, F. W. Albert, H. A. Burbano, M. Kircher *et al.*, 2013 Comparative Population Genomics of the Ejaculate in Humans and the Great Apes. Molecular Biology and Evolution 30**:** 964-976.
Green, C. D., Q. Ma, G. L. Manske, A. N. Shami, X. Zheng *et al.*, 2018 A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Developmental Cell.
Harcourt, A. H., P. H. Harvey, S. G. Larson and R. V. Short, 1981 Testis weight, body weight and breeding system in primates. Nature 293**:** 55.
Hibbins, M. S., M. J. S. Gibson and M. W. Hahn, 2020 Determining the probability of hemiplasy in the presence of incomplete lineage sorting and introgression. eLife 9**:** e63753.
Higginson, D. M., K. B. Miller, K. A. Segraves and S. Pitnick, 2012 Female reproductive tract form drives the evolution of complex sperm morphology. Proceedings of the National Academy of Sciences 109**:** 4538-4543.
Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano *et al.,* 2006 The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 34(Database issue):D590-8
Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh and L. S. Vinh, 2017 UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35**:** 518-522.
Hook, K. A., L. M. Wilke and H. S. Fisher, 2021 Apical Sperm Hook Morphology Is Linked to Sperm Swimming Performance and Sperm Aggregation in Peromyscus Mice. Cells 10**:** 2279.
Immler, S., H. D. M. Moore, W. G. Breed and T. R. Birkhead, 2007 By Hook or by Crook? Morphometry, Competition and Cooperation in Rodent Sperm. PLOS ONE 2**:** e170.
Ives, A. R., and T. Garland, Jr., 2009 Phylogenetic Logistic Regression for Binary Dependent Variables. Systematic Biology 59**:** 9-26.
Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde *et al.*, 2014 Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346**:** 1320-1331.
Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin, 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14**:** 587-589.
Katoh, K., K. Misawa, K. i. Kuma and T. Miyata, 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30**:** 3059-3066.
Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong *et al.*, 2011 Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477**:** 289-294.
Kopania, E. E. K., E. L. Larson, C. Callahan, S. Keeble and J. M. Good, 2022 Molecular Evolution across Mouse Spermatogenesis. Molecular Biology and Evolution.

Kosakovsky Pond, S. L., and S. D. W. Frost, 2005 Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution 22**:** 1208-1222.
Kosakovsky Pond, S. L., A. F. Y. Poon, R. Velazquez, S. Weaver, N. L. Hepler *et al.*, 2019 HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Molecular Biology and Evolution 37**:** 295-299.
Kowalczyk, A., R. Partha, N. L. Clark and M. Chikina, 2020 Pan-mammalian analysis of molecular constraints underlying extended lifespan. eLife 9**:** e51089.
Kowalczyk, A., W. K. Meyer, R. Partha, W. Mao, N. L. Clark *et al.*, 2019 RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35**:** 4815-4817.
Larson, E. L., E. E. K. Kopania and J. M. Good, 2018 Spermatogenesis and the Evolution of Mammalian Sex Chromosomes. Trends Genet 34**:** 722-732.
Larson, E. L., D. Vanderpool, S. Keeble, M. Zhou, B. A. J. Sarver *et al.*, 2016 Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. Genetics 203**:** 1841-1857.
Lough-Stevens, M., C. R. Ghione, M. Urness, A. Hobbs, C. M. Sweeney *et al.*, 2020 Male-derived copulatory plugs enhance implantation success in female Mus musculus. Biology of Reproduction 104**:** 684-694.
Lüpold, S., M. K. Manier, N. Puniamoorthy, C. Schoff, W. T. Starmer *et al.*, 2016 How sexual selection can drive the evolution of costly sperm ornamentation. Nature 533**:** 535.
Mangels, R., K. Tsung, K. Kwan and M. D. Dean, 2016 Copulatory plugs inhibit the reproductive success of rival males. Journal of Evolutionary Biology 29**:** 2289-2296.
Martin, S. H., J. Davey, C. Salazar and C. Jiggins, 2018 Recombination rate variation shapes barriers to introgression across butterfly genomes. bioRxiv.
McLennan, H. J., S. Lüpold, P. Smissen, K. C. Rowe and W. G. Breed, 2017 Greater sperm complexity in the Australasian old endemic rodents (Tribe: Hydromyini) is associated with increased levels of inter-male sperm competition. Reproduction, Fertility and Development 29**:** 921-930.
Mendes, F. K., and M. W. Hahn, 2016 Gene Tree Discordance Causes Apparent Substitution Rate Variation. Systematic Biology 65**:** 711-721.
Moore, H., K. Dvoráková, N. Jenkins and W. Breed, 2002 Exceptional sperm cooperation in the wood mouse. Nature 418**:** 174.
Murat, F., N. Mbengue, S. B. Winge, T. Trefzer, E. Leushkin *et al.*, 2021 The molecular evolution of spermatogenesis across mammals. bioRxiv**:** 2021.2011.2008.467712.
Murrell, B., S. Moola, A. Mabona, T. Weighill, D. Sheward *et al.*, 2013 FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Molecular Biology and Evolution 30**:** 1196-1205.
Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh, 2014 IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32**:** 268-274.
Pagel, M., 1999 Inferring the historical patterns of biological evolution. Nature 401**:** 877-884.

Pahl, T., H. J. McLennan, Y. Wang, A. S. Achmadi, K. C. Rowe *et al.*, 2018 Sperm morphology of the Rattini - are the interspecific differences due to variation in intensity of intermale sperm competition? Reprod Fertil Dev 30**:** 1434-1442.
Pease, J. B., D. C. Haak, M. W. Hahn and L. C. Moyle, 2016 Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLOS Biology 14**:** e1002379.
Peirce, E. J., H. J. McLennan, J. Tuke, C. M. Leigh and W. G. Breed, 2018 Evolution of the testis and spermatozoon in mice and rats (Subfamily Murinae) in the absence of sperm competition. Journal of Zoology 306**:** 58-68.
Pinheiro J, B. D., DebRoy S, Sarkar D, R Core Team, 2020 nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-148. <URL: https://CRAN.R-project.org/package=nlme>.
Pitnick, S., D. J. Hosken and T. R. Birkhead, 2009 Sperm morphological diversity, pp. 69-149 in *Sperm Biology*, edited by T. R. Birkhead, D. J. Hosken and S. Pitnick. Academic Press, London.
Ramm, S. A., P. L. Oliver, C. P. Ponting, P. Stockley and R. D. Emes, 2008 Sexual Selection and the Adaptive Evolution of Mammalian Ejaculate Proteins. Molecular Biology and Evolution 25**:** 207-219.
Ramm, S. A., and L. Schärer, 2014 The evolutionary ecology of testicular function: size isn't everything. Biological Reviews 89**:** 874-888.
Revell, L. J., 2012 phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3**:** 217-223.
Rowe, K. C., A. S. Achmadi and J. A. Esselstyn, 2016 A new genus and species of omnivorous rodent (Muridae: Murinae) from Sulawesi, nested within a clade of endemic carnivores. Journal of Mammalogy 97**:** 978-991.
Rowe, K. C., A. S. Achmadi, P.-H. Fabre, J. J. Schenk, S. J. Steppan *et al.*, 2019 Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. Journal of Biogeography 46**:** 2752-2768.
Roycroft, E., A. Achmadi, C. M. Callahan, J. A. Esselstyn, J. M. Good *et al.*, 2021 Molecular Evolution of Ecological Specialisation: Genomic Insights from the Diversification of Murine Rodents. Genome Biology and Evolution 13.
Shapiro, B., and M. Hofreiter, 2012 Ancient DNA: methods and protocols, pp.
Simmons, L.W., 2019. *Sperm competition and its evolutionary consequences in the insects*. Princeton University Press.
Simmons, L. W., and J. L. Fitzpatrick, 2012 Sperm wars and the evolution of male fertility. Reproduction 144**:** 519-534.
Subrini, J., and J. Turner, 2021 Y chromosome functions in mammalian spermatogenesis. eLife 10**:** e67345.
Swanson, W. J., and V. D. Vacquier, 2002 The rapid evolution of reproductive proteins. Nat Rev Genet 3**:** 137-144.
Swanson, W. J., Z. Yang, M. F. Wolfner and C. F. Aquadro, 2001 Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences 98**:** 2509-2514.
Thomas, G. W. C., and M. W. Hahn, 2019 Referee: Reference Assembly Quality Scores. Genome Biology and Evolution 11**:** 1483-1486.

Tung Ho, L. s., and C. Ané, 2014 A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Systematic Biology 63**:** 397-408.
Varea-Sánchez, M., M. Tourmente, M. Bastir and E. R. S. Roldan, 2016 Unraveling the Sperm Bauplan: Relationships Between Sperm Head Morphology and Sperm Function in Rodents. Biology of Reproduction 95.
Wong, A., 2011 The Molecular Evolution of Animal Reproductive Tract Proteins: What Have We Learned from Mating-System Comparisons? International Journal of Evolutionary Biology 2011**:** 908735.
Wong, A., 2014 Covariance between Testes Size and Substitution Rates in Primates. Molecular Biology and Evolution 31**:** 1432-1436.
Yang, Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24**:** 1586-1591.
Yin, J., C. Zhang and S. Mirarab, 2019 ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. Bioinformatics 35**:** 3961-3969.