THE INFLUENCE OF SEXUAL SELECTION AND SEXUAL CONFLICT ON THE EVOLUTION OF POST-INSEMINATION DYNAMICS

by

KATJA ROSE KASIMATIS

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2019

DISSERTATION APPROVAL PAGE

Student: Katja Rose Kasimatis

Title: The Influence of Sexual Selection and Sexual Conflict on the Evolution of Post-Insemination Reproductive Dynamics

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

William Cresko	Chairperson
Patrick C. Phillips	Advisor
Matthew Streisfeld	Core Member
Stevan J. Arnold	Core Member
Michael Harms	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019

DISSERTATION ABSTRACT

Katja Rose Kasimatis

Doctor of Philosophy

Department of Biology

June 2019

Title: The Influence of Sexual Selection and Sexual Conflict on the Evolution of Post-Insemination Dynamics

Sexual reproduction is a fundamental process that structures populations and modulates interactions between species. The reproductive process is shaped by selection acting on the variance in mating success. Additionally, conflict between the sexes over the mechanisms by which mating success is optimized effects reproduction. Selection can also act in a sex-specific manner outside of the reproductive process and drive a different class of sexual antagonisms. To understand how sexual conflict shapes evolution within and between the sexes, the action of selection must be connected to the lifecycle of an individual. Such a lifecycle-explicit framework allows for quantitative measurements of sex-specific selection, sexual conflict, and genetic load.

Here I connect the action of selection with the appropriate stage of the lifecycle to determine how conflict between the sexes contributes to genome evolution. Using theoretical approaches, I examine if sexually antagonistic viability selection can create genomic divergence between the sexes. I find that selection must be strong to generate measurable divergence, which produces a high genetic load. Additionally, I show that sampling variance can account for much of the signal attributed to sexually antagonistic selection in the literature.

Using experimental approaches, I manipulate sex-specific selection acting during the gametic phase to determine the molecular components of male fertilization success. I develop *Caenorhabditis* nematodes as a new model system for studying post-insemination reproductive interactions. Contrary to expectation, I find that nematode sperm proteins are hyper-conserved at the sequence level and rapidly evolving at the gene family level. This result suggests an alternative signature of sex-specific selection and conflict. Additionally, I develop a genetic tool for isolating sperm dynamics. This sterility induction system is the first external, non-toxic, reversible sterility induction system in animals.

Together my dissertation highlights how the genomic signatures of sexual selection and conflict are complex and require explicit empirical testing to validate both the phenotype and action of selection. Such complexity indicates that evolutionary systems biology approaches will be the most informative way to move the field forward and establish the importance of sexual conflict in shaping evolution.

This work includes published and unpublished coauthored material.

CURRICULUM VITAE

NAME OF AUTHOR: Katja Rose Kasimatis

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of California, Santa Cruz

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2019, University of Oregon
Bachelors of Science, Marine Biology, 2014, University of California, Santa Cruz

AREAS OF SPECIAL INTEREST:

Evolutionary Genetics
Population Genomics
Sexual Conflict
Gamete Biology

GRANTS, AWARDS, AND HONORS:

DeLill Nasser Award, Genetics Society of America, 2019

Travel Award, Society for the Study of Evolution, 2018

Special "OPPS" Travel and Research Award, University of Oregon, 2018

Achievement Rewards for College Scientists Scholar Award, ARCS Oregon Chapter, 2015-2018

Travel Award, American Genetics Association, 2016

Donald Wimber Travel Award, University of Oregon, 2016

Hill Fund Outstanding Graduate Teaching Fellow Award, University of Oregon, 2014-2015

Dan Kimble First Year Teaching Award, University of Oregon, 2014-2015

PUBLICATIONS:

Kasimatis, K.R., Moerdyk-Schauwecker, M.J., Timmermeyer, N. & Phillips, P.C. (2018) Proteomic and Evolutionary Analyses of Sperm Activation Identify Uncharacterized Genes in *Caenorhabditis* Nematodes. BMC Genomics 19: 1-13.

Kasimatis, K.R., Moerdyk-Schauwecker, M.J. & Phillips, P.C. (2018) Auxin-Mediated Sterility Induction System for Longevity and Mating Studies in *Caenorhabditis elegans*. G3 8: 2655-2662.

Kasimatis, K.R. & Phillips, P.C. (2018) Reproductive Interactions Facilitate Rapid Gene Family Evolution of a Hyper-Conserved Sperm Protein. G3 8: 353-362.

Borne, F., Kasimatis, K.R. & Phillips, P.C. (2017) Quantifying male and female pheromone-based mate choice in *Caenorhabditis* nematodes using a novel microfluidic technique. PLoS ONE 12: e0189679.

Kasimatis, K.R**.**, Nelson, T.C. & Phillips, P.C. (2017) The Genomic Signatures of Sexual Conflict. J. Heredity 108: 780-790.

Kasimatis, K. & Riginos, C. (2016) No simple trade-offs: relationships between egg size, clutch size, spawning mode, adult body size, and latitude in reef fishes. Coral Reefs 35: 387-397.

ACKNOWLEDGMENTS

I would like to thank to my advisor, Patrick Phillips, for his support, guidance, and mentorship. Dr. Phillips gave me the opportunity to be curious and explore many different projects for which I am grateful. I would also like to thank my committee members Stevan Arnold, William Cresko, Michael Harms, and Matthew Streisfeld for their guidance and advice.

I would like to acknowledge the support and guidance of the Phillips Lab members past and present. I would especially appreciate the constant support and mentorship of my office mate Stephen Banse. I would like to thank John Willis for teaching me my worm basics. I would like to thank Levi Moran for experimental evolution advice and career guidance. I am grateful to Dr. Harms and the Harms Lab for their feedback on the proteomic work and general biochemistry guidance.

I am grateful everyone who helped make this work possible. Megan Moerdyk-Schauwecker assisted with the transgenics and this work would not have been possible without her assistance. I would like to thank Erik Johnson, John Johnson, Ruben Lancaster, Christine Sedore, and Alex Smith for experimental assistance and Anastasia Teterina for data analysis advice. I am grateful for the administrative support of Arlene Deyo and Sara Nash – the Institute of Ecology and Evolution would not run without them. I would like to acknowledge Gennifer Merrihew and Michael MacCoss at the Genome Sciences Mass Spectrometry Center at the University of Washington for the mass spectrometry analysis. I would like to thank Matthew Rockman, Erich Schwarz, and John Wang for generously sharing unpublished genomic data. I would like to thank Göran Arnqvist, Stephen Banse, Matthew Hahn, Thomas Nelson, Christian Rödelsperger, Sean Stankowski, Willie

Swanson and multiple anonymous reviewers for constructive comments on my published work. I am grateful to Melissa Wilson and John Logsdon for the opportunity to contribute Chapter II as part of the symposium on the “Evolutionary Genomics of Sex”.

Funding for this dissertation work was provided by the National Institutes of Health grants R01 GM102511 and R01 AG49396 to Patrick C. Phillips and training grant T32 GM007413 to Katja R. Kasimatis. The ARCS Oregon Chapter generously supported me as a Chapter Scholar. I am grateful for conference and travel support from the American Genetics Association, Genetics Society of America, Graduate Evolutionary Biology and Ecology Students, Society for the Study of Evolution, and University of Oregon Department of Biology, and University of Oregon Graduate School.

Finally, I would like to thank my mother, Elaine Kasimatis, for being my mentor and role model.

In loving memory of Amand Kasimatis, Phillip Kasimatis, and Mr. Kitty Kasimatis.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Reproduction is fundamental. On the surface, this statement is simple. All organisms must pass their genetic material on to the next generation for the lineage to survive. In asexual organisms, clonal division makes this process straightforward. However, sexually reproducing organisms rely on the union of haploid gametes and therefore must interact with an individual of the opposite sex. The function of sexual reproduction is to promote recombination and segregation of alleles (Fisher 1930). The process of reproduction requires females and males to successfully find a mate, copulate, and fuse gametes. If the sexes invested equally in their gametes (i.e. isogamy) and offspring, this process would again be fairly trivial, at least from an evolutionary perspective. In fact, under these conditions, sexes as we define them would not need to exist and the biotic world would be very different. However, distinct genetic sexes do exist, despite the costly process of sexual reproduction (Lehtonen *et al.* 2012).

Males and females can broadly be defined by producing gametes of different types (i.e., anisogamy) (Parker 2014). The typical obvious difference is one of size, with eggs/ovules being defined as the large gamete and sperm/pollen as the small gamete. Anisogamy has two related consequences (Trivers 1972; Parker 2014; Lehtonen *et al.* 2016). First, the sexes must have the necessary developmental framework to produce sex-specific reproductive tissues and gametes. Second, at the critical step of fertilization the sexes contribute resources differently to the next generation. This differential investment creates the opportunity for selection to act in a sex-specific manner to optimize the

investment of each sex in the next generation and drive the evolution of further sexual dimorphism (Trivers 1972; Parker 2014).

Sexual selection and the mating phase of the lifecycle

Sexual dimorphism has been a puzzling concept since early naturalists first recognized it. In fact, Darwin himself first saw sexual dimorphism as a contradiction to the theory of natural selection (Darwin 1871). In *Decent of Man* (Darwin 1871), Darwin proposed a second selective force, namely sexual selection, to account for sexually dimorphic traits that should seemingly decrease viability. Specifically, Darwin proposed that sexual selection acts to increase reproductive fitness in a sex-specific manner. Many mechanisms by which sexual selection can act both within and between the sexes have been proposed, including: Fisherian sexual selection (Fisher 1930), chase away sexual selection (Holland and Rice 1998), sensory bias (Ryan *et al.* 1990), sexy sons (Weatherhead and Robertson 1979), and indicator mechanisms (Zahavi 1975). These theories have been thoroughly reviewed in Arnold (Arnold 1983), Andersson (1994), Arnqvist and Rowe (2005) Andersson and Simmons (2006), and Kuijper *et al*. (Kuijper *et al.* 2012). An overarching comment across the different theoretical treatments of sexual selection is that they have major discrepancies over the explicit inclusion of transmission and the treatment of the genetic basis of traits (discussed in Arnold 1983). Additionally, some models, such as the "good genes" model, confound natural and sexual selection, while parental care based models confound selection across generations. These modeling shortcuts challenge the ability to analyze fitness effects and heritable evolutionary response to sexual selection (Arnold 1983). In response to these confounds, field has generally moved to dynamics models of selection in a lifecycle-structured framework

(Arnold and Wade 1984). In the case of sexual selection, this means specifically focusing on the fitness effects of variation in mating success.

Each generation new zygotes are formed, survive to be reproductive adults, and reproduce. This sequence constitutes the lifecycle of an individual. The lifecycle can be broadly partitioned into three different phases: individuals have a survival phase (zygote to reproductive adult) in which they are subject to natural selection, a reproductive phase (reproductive adult to gametes transferred) in which they are subject to sexual selection, and a gametic phase in which haploid gametes are subject to gametic selection (Fig. 1.1A). The length of each phase will vary by life history. Some organisms exist predominantly in the diploid survival phase, while others spend more time in the haploid gametic phase. The mating phase itself is almost always very short. Variation in the length of these phases affects how total fitness is determined. Selection during the reproductive and gametic phases, by definition, views the sexes independently and therefore acts in a sex-specific manner. However, even during the survival phase selection can act in a sex-specific manner (Fig. 1.1B).

Sex-specific selection can have profound effects because female and male function is derived largely from the same genomic content. This shared genomic basis is strictly true for the autosomal genome, and in some cases substantially holds true for even sex chromosomes (when they exist). Consequently, at the phenotypic level females and males are distinct while at the genotypic level they are not, making the sexes an extreme form of polyphenism (Grath and Parsch 2016; Kasimatis *et al.* 2017). This mismatch across the genotype and phenotype map necessitates that: i) the autosomal genome can recognize the sexual environment (i.e., the genetic sex of an individual) and

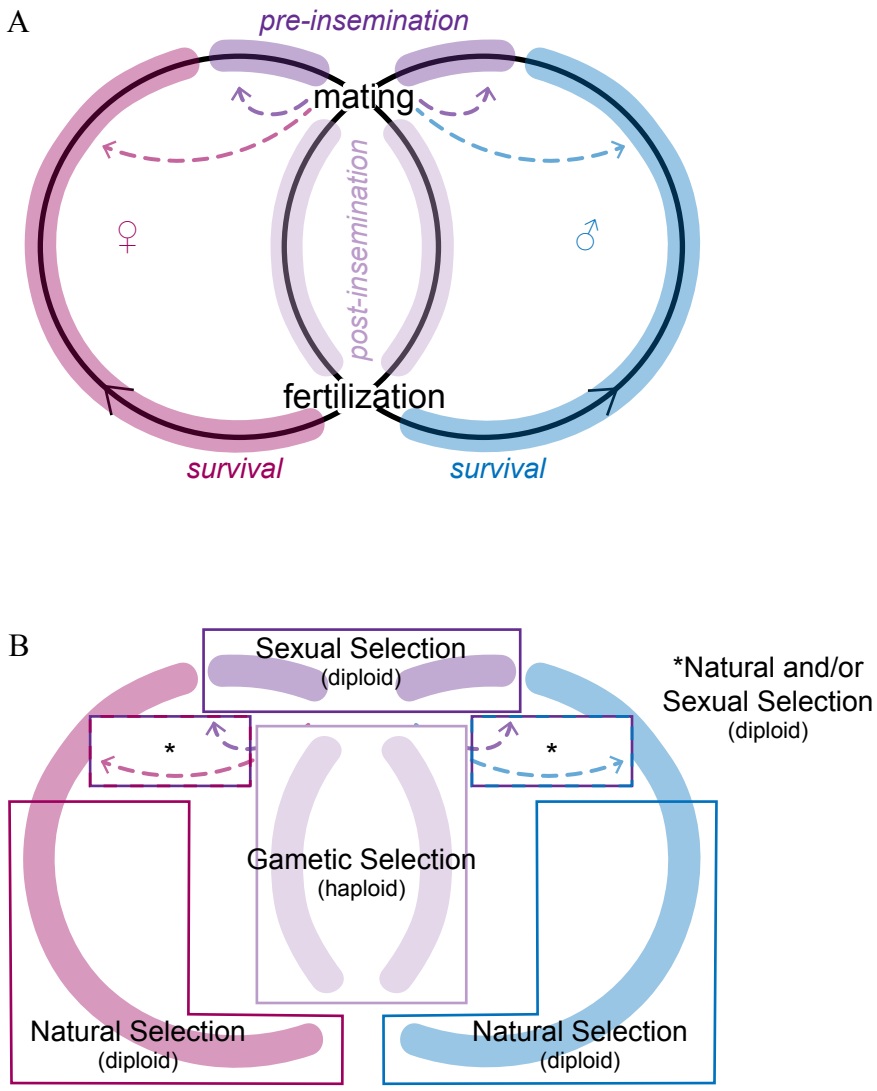


Figure 1.1. The lifecycle of an individual. A) A generation begins when gametes fuse and a zygote is formed. That zygote is then subject to natural selection while it grows to be a reproductive adult. Females are shown in pink and males in blue. Only during the reproductive process are the sexes intimately linked (shown in purple). Reproduction occurs over two stages. First, pre-insemination mate choice interactions and mate competition culminate in a mating event. After mating females and males can either return to a survival state (shown by the pink and blue dashed lines, respectively) or immediately return to a mating state (shown by the purple dashed lines). Then during post-insemination, gametes must interact for a successful fertilization event to occur. Fertilization starts the next generation. B) The action of selection can be parsed by the stage of the lifecycle. During the survival and resource allocation phase, natural selection acts on diploid individuals. During mating sexual selection acts, again on diploid individuals. Those individuals that transition back to the lifecycle (i.e. do not die) can be shaped by either natural selection of sexual selection. Finally, gametic selection acts on haploid egg and sperm cells prior to fertilization.

ii) the translation between genotype and phenotype often be sex-dependent. From an evolutionary perspective, this translation mismatch generates the opportunity for antagonism between the sexes. Specifically, since selection views the sexes as distinct "populations" it will act in a sex-specific manner to optimize the fitness within each sex. However, the common genetic basis tethers the evolutionary responses of the sexes, which prevents either sex from reaching its fitness optimum and leads to a decrease in overall population fitness (Arnqvist 2004). This antagonistic interaction is referred to as sexual conflict (though sexual antagonism is often used interchangeably).

History of sexual conflict

Sexual conflict was originally defined by Parker (1979) as conflict between the evolutionary interests of the individuals of the two sexes occurring during reproduction. The concept was heavily influenced by studies of *Drosophila* mating conducted by Bateman (1948) and subsequent theoretical work developed by Trivers (1972). Together this body of work solidified the concept that reproduction in not harmonious between the sexes, such that an increase in the reproductive success of one partner does not always translate to increased reproductive success in the other partner (see discussion in Arnqvist and Rowe 2005). This idea of non-harmonious reproduction established the field of sexual conflict. The foundational idea of the field was that intersexual conflict during the reproductive phase could structure populations, drive mating system evolution, and generate divergence leading to speciation (Parker 1979; Holland and Rice 1998; Arnqvist and Rowe 2005; Gavrilets 2014). However, the original definition of sexual conflict is somewhat vague and lacks the specific clarity needed to study these goals (see Arnqvist and Rowe 2005). Do the evolutionary interests of the sexes really differ? How are

evolutionary interests quantified? Additionally, while this game theory-based definition is applicable at the phenotypic level, it is hard to translate to a mechanistic basis at the genetic level. To create a more operational definition of conflict, Rice took a population genetics perspective and classified sexual conflict as a specific type of intergenomic conflict, coining the terms interlocus and intralocus sexual conflict (Rice and Holland 1997; Rice 1998).

Now, at least 17 distinct definitions of sexual conflict appear to exist in the literature (Table 1.1). These range from phenotypically-centered definitions to genotypically-centered definitions, with some interpretations being so broad as to define sexual conflict as any sex-by-genotype interaction (Rice and Chippindale 2001). As new definitions are added, the connection to reproduction has become tenuous, such that the antagonism between the sexes is no longer sexual conflict *senu stricto* (Kasimatis *et al.* 2017). Moreover, the link between sexual conflict and the action of selection varies dramatically across definitions, with some considering that sexual selection drives sexual conflict (Arnqvist and Rowe 2005), others that sexual conflict drives sexually antagonistic selection (Parker 2006), and a small set that consider no direct link between these processes (Kokko and Jennions 2014). The myriad of definitions make this field confusing, especially for non-specialists, which leads to even further muddling of the concept and vagueness in the field. Additionally, the different definitions imply very different empirical frameworks for identifying sexual conflict (Table 1.1). In fact, the group of definitions based on opposing evolutionary interests between the sexes do not translate to an empirically testable hypothesis. Together these problems hinder cohesive forward movement that would address the foundational goals within the field.

Table 1.1. Example definitions of sexual conflict, building on Arnqvist and Rowe (2005) (see Table 7.1). Definitions are divided into six general classes based on the empirical method by which that form of conflict could be detected.

Definition	Detection	Reference
Reproductive Interests and Strategies		
Conflict between the evolutionary interests of the individuals of the two sexes	Measure the interests of sexes	Parker (1979)
Different evolutionary interests of the two sexes	Measure the interests of sexes	Parker and Partridge (1998)
Each parent's fitness is generally maximized if it invests less and the other parent invests more than would maximize the other parent's fitness	Measure reproductive investment relative to maximum?	Lessels (Arnqvist and Rowe 2005)
When the genetic interests of males and females diverge	Measure the interests of sexes	Chapman et al. (2003)
Differences in the evolutionary interests of males and females that occur over traits related to courtship, mating, and fertilization through to parental investment	Measure the interests of sexes	Chapman (2006)
The sexes maximize their fitness via different, and often mutually incompatible, strategies	Measure the strategies of sexes	Adler and Bonduriansky (2014)
Phenotypic Covariance		
The ratio of the number of offspring that a sex optimally produces to the number of offspring the limiting sex optimally produces	Measure covariance between mating rate and offspring	Trivers (1972)
When traits favored by reproductive competition within one sex are costly for individuals of the other sex, when expressed either phenotypically or as a result of male-female interactions	Measure covariance between reproductive traits between the sexes	Parker (1984); Rice (1998)
Sex difference in the covariance between promiscuity and offspring numbers	Measure covariance between mating rate and offspring	Shuster and Wade (2003)
When the slopes of any relation between either (i) a single trait expressed in both sexes or (ii) an outcome of male-female interactions on one hand, and fitness on the other, differ in sign in the two sexes	Measure covariance between trait and fitness	Gavrilets et al. (2001)
Genetic Covariance		
The fitness of the alleles at the A locus depend on the identity of the alleles at the B locus in the opponent	Measure genetic covariance	Rice and Holland (1997)
Genotype-by-sex interaction for fitness	Measure covariance between genotype and sex	Rice and Chippindale 2001 (2001)

Table 1.1 ***continued***

Definition	Detection	Reference
Reproductive Trait Optima		
When an increase in the average frequency in one sex translates into a reduction in the average fitness in the other sex	Measure frequency-dependent fitness	Pizzari and Snook (2003)
Differences in the roles played by the sexes in the process of reproduction, which in turn leads to differences between the sexes in the costs and benefits of mating and reproduction	Measure fitness optima in reproductive traits of each sex	Gavrilets (2014)
Expression of a trait increases the fitness of one sex while reducing the fitness of the other sex	Measure fitness optima in traits of each sex	Edward et al. (2015)
Males and females have incompatible fitness optima related to courtship, mating, fertilization, offspring provisioning, and parental care	Measure fitness optima in reproductive traits of each sex	Furness et al. (2015)
The negative fitness consequences generated as a result of different reproductive trait optima between the sexes	Measure fitness optima in reproductive traits of each sex	Kasimatis et al. (2017)
Sex-Specific Selection		
Divergent reproductive strategies of the sexes generate different selective pressures on many traits	Measure sex-specific selection	Bonduriansky and Chenoweth (2009)
Selection acts in opposing directions in males and females	Measure sex-specific selection	Mank (2017)
Other		
Cost-free "tool" that allows individuals of one sex to alter individuals of the other sex in a costly manner	Measure between sex manipulation	Kokko and Jennions (2014)
Genes disagree about the consequences that a phenotypic change will have for their carrier's and/or social partner's reproductive success	Measure the interests of genes	Gardner and Ubeda (2017)

Ideally, a scientific definition of biological process should to lead to clear hypothesis tests within a reproducible empirical framework. The difficulty that consistently appears to arise in definitions of sexual conflict stems from what should actually be measured: is sexual conflict a phenotypic property or a genomic property? Following its original usage, sexual conflict occurs at the phenotypic level, with a clear focus on interactions between the sexes (i.e. mating dynamics) and how the abiotic environment shapes those interactions. However, studies over the last decade have moved away from mating interactions to focus more broadly on how genomic pleiotropy can generate a signal of sexual conflict (Fry 2010; Connallon *et al.* 2010; Patten *et al.* 2010; Connallon and Clark 2012; 2013). Additionally, the translation of a shared genome into sexually dimorphic phenotypes sets the stage for genetic constraints and between-sex covariance to occur, which could also drive sexual antagonism. The distinction between the phenotypic context and genomic consequences of sexual conflict can best be clarified by partitioning antagonistic effects based on the action of selection and the phase of the lifecycle during which selection is occurring (Fig. 1.1).

Connecting sexual conflict to the lifecycle

Sexual antagonism/conflict is not itself a selective force, because it does not act on a variance in fitness. Rather, sexual antagonism/conflict results from fitness effects generated at a particular phase of the lifecycle and can generate between sex variance in fitness (Fig. 1.2). Many conflicts between the sexes can arise as a result of natural selection generated by variation in survival due to sex-specific attributes, such as being the female or male with highest fitness for a given environment. For example, increased predation pressure on male peacocks due to their exaggerated tails results in sex-specific

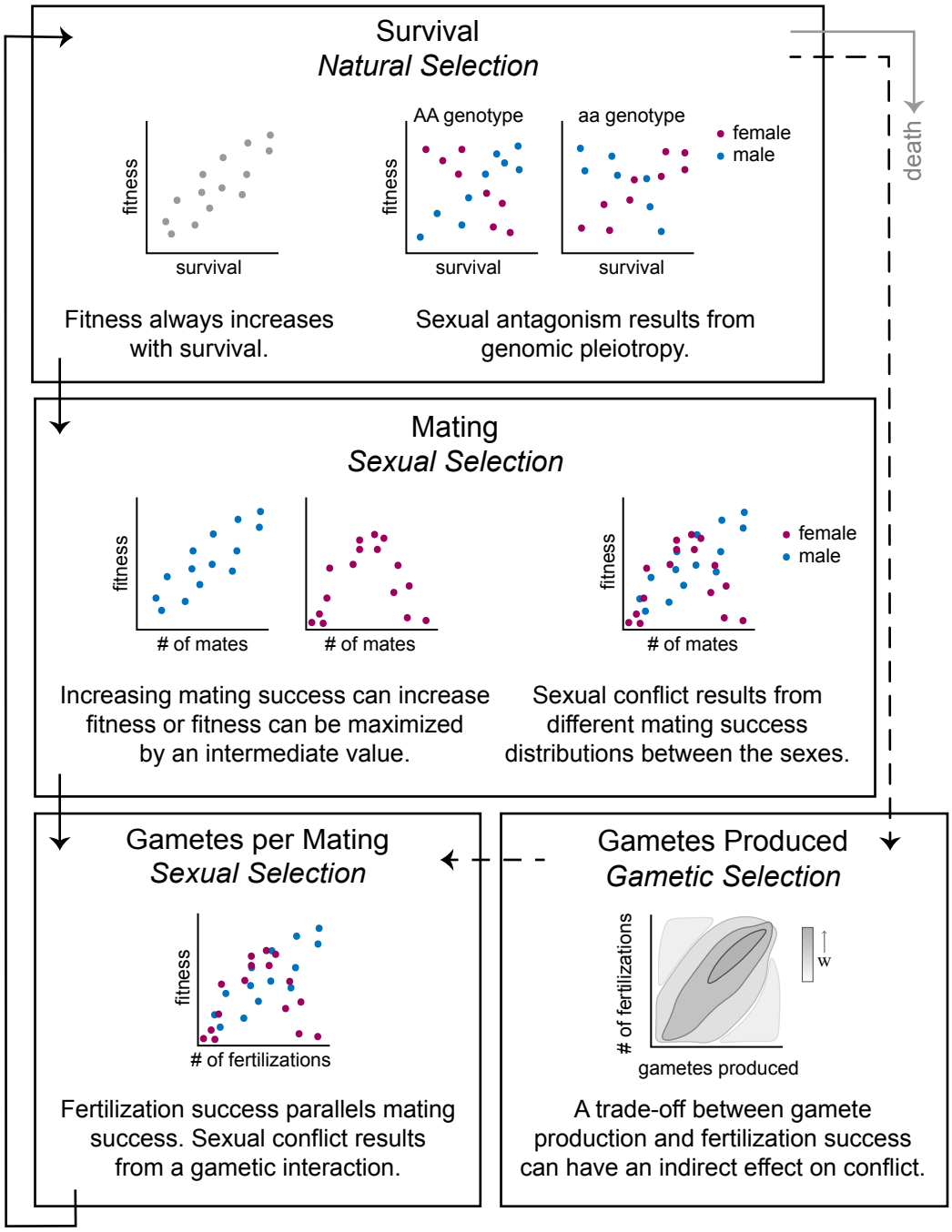


Figure 1.2. The the lifecycle is detailed in Figure 1.1. During the survival phase, natural selection will always act to increase fitness. If selection occurs in a sex-specific manner, an antagonism between the sexes can arise due to genomic pleiotropy. This antagonism can be resolved by decoupling the genetic architecture of the trait between the sexes. During mating, sexual selection can drive sexual conflict between the sexes if the traits optimizing mating success between the sexes have different fitness values. The conflict is a consequence of mating interactions and the resolution is constrained by genetic covariance between the sexes. During post-insemination interactions, gametes are subject to selection. Variance in fertilization success can sex-specific selection, which parallels sexual selection and sexual conflict during pre-insemination. Gametic processes are also influenced by total gamete production, which is subject to natural selection and sexual selection.

selection and is the direct result of sexual differentiation. Expression of genes leading to exaggerated tails in females would be deleterious and selected against. However, the fitness differences are not generated directly via an interaction during mating. In fact, sexual antagonism resulting from sex-specific natural selection does not require the sexes to interact at all. Under natural selection, fitness has a positive relationship with survival. Here selection can truly view females and males independently, such that traits can exhibit sex-specific survivorship (for example Czorlich *et al.* 2018). The source for such an antagonism is the shared genomic basis of traits. In our example above, if tail development could be uncoupled between the sexes the antagonism would no longer exist. Thus, sexual antagonisms occurring during the survival phase are a result of genomic pleiotropy (Fig. 1.2). The genetic basis for these antagonisms suggests that bottom-up experiments may be a powerful way of identifying candidate loci (Kasimatis *et al.* 2017). Additionally, since the source of antagonism is a sexually antagonistic pleiotropy these conflicts can be resolved by creating local sex-specific genetic architecture (Kirkpatrick 2010) or recruiting such loci to the sex chromosome (van Doorn and Kirkpatrick 2007).

During the reproductive phase females and males are intimately linked, which distinguishes this stage of the lifecycle from all others. Sexual reproduction necessitates that females and males share gametes. This process can occur via a direct interaction as seen in internally fertilizing organisms, an indirect interaction mediated through a pollinator as seen in plants, or an indirect interaction mediated by an external cue as seen in broadcast spawning/pollinating organisms. The mode of sexual reproduction will influence the action of sexual selection and the forms of conflict (Furness *et al.* 2015).

Importantly though, sexual selection acts independently in each sex to optimize mating success (Arnold 1983). This interaction of sex-specific selection and the requirement of a successful mating provides the opportunity for sexual conflict (Fig. 1.2). If all females and all males optimized mating success in exactly the same manner (i.e. no variance in mating success), then sexual conflict would not exist in the population. This scenario can be observed in strict monogamy conditions (Pitnick *et al.* 2001). However, as sexual selection pushes females and males in different directions of trait space, the potential for conflict occurs. In turn, sexual conflict increases the variance in mating success and therefore the opportunity for sexual selection is also increased (Arnold 1994; Jones 2009). Thus, a positive feedback loop between sexual selection and conflict is generated and is unique to sexual conflict *senu stricto*.

The evolutionary response between the sexes due to conflict will depend on the genetic basis of the traits involved. Under a shared genetic basis (i.e. antagonistic pleiotropy) the genetic basis constrains the evolutionary response, as is seen for sexually antagonistic traits. However, unlike sexual antagonism, under sexual conflict uncoupling the genetic architecture will not resolve the conflict as long as the antagonistic mating interaction continues. A polygenic basis for conflict traits is the foundation for antagonistic coevolution between the sexes (Holland and Rice 1998). Here the driver of conflict is still the interactions between the sexes not the genetic basis itself. Again, mating interactions maintain sexual conflict regardless of the genetic basis, which distinguishes sexual conflict from sexual antagonism. In fact, sexual conflict can even occur if traits are sex-limited and the genetic basis is completely uncoupled between the sexes. Thus, sexual conflict cannot be resolved at the genomic level. It can only be

resolved if the mating dynamics change or the environment shifts. The empirical measurement of sexual conflict must then be phenotypic.

During the gametic phase selection acts on haploid cells, which distinguishes this phase of the lifecycle from the other phases (Fig. 1.1). As with mating interactions, the mode of reproduction – internal fertilization, external fertilization, or pollinator mediated – will strongly shape the action of selection. Gametic selection can act on variation in fertilization success, paralleling sexual selection (Kuijper *et al.* 2012). Here interactions among gametes generate the potential for sexual conflict (Fig. 1.2). However, gamete production is also influenced by natural selection. This process can indirectly contribute to sexual conflict by generating within-sex variance that impacts fertilization success. Alternatively, if total gamete production and total reproductive success do not align, then sexual antagonism can occur (Fig. 1.2). Natural selection can also shape gametic processes through selection on gene drivers, which can negatively impact one sex and again generate sexual antagonism (McLaughlin and Malik 2017).

Disentangling the actions of gametic selection can be challenging due to the complexity of this phase of the lifecycle, which in part is why this phase is understudied. There are two critical questions that will move our understanding of sexual antagonism/conflict during this stage forward. First, what is the actual phenotype under selection? This question is particularly important for distinguishing if selection is acting on a fertilization interaction or gamete production *per se*. Second, are the gamete expressing genes based from on the haploid genome? This gametic expression can introduce additional sources of genomic conflict and impact the efficacy of selection. More work focusing on this stage of the lifecycle is needed as selection and conflict

during this phase have a high potential to be potent drivers of evolution by dictating fertilization success.

Does the distinction between natural selection driven conflicts and sexual selection driven conflicts matter? We believe it does for the following reasons. The action of sex-specific selection and phase in the lifecycle in which selection occurs will determine the stability and resolution of sexual conflicts within a population. Our cohesive, lifecycle-grounded framework allows for: i) explicit models and equations of the postulated processes, and ii) defines explicit empirical tests of competing hypotheses and/or a means of measuring relevant parameters.

Dissertation outline

Reproducible hypothesis testing on the causes and consequences of sexual conflict requires a lifecycle-explicit theory and empirical studies. The research described in my dissertation aims to strength our theoretical understanding of sexual antagonism and explicitly address the action of selection during the gametic phase of the lifecycle. This work integrates proteomic techniques, molecular evolution, and cutting-edge genomic tools to effectively address the genomics of reproductive success. I take advantage of the tools and tractability of *Caenorhabditis* nematodes to develop a new model system for post-insemination mating interactions. *Caenorhabditis* nematodes are an excellent system for studying the genomics of reproductive success for multiple reasons. First, multiple mating systems present within the genus creating natural variation in the strength of sexual selection (Kiontke *et al.* 2011). Second, nematodes have a unique sperm biology characterized by amoeboid-like crawling sperm cells (Justine 2002), which allows us to address gametic selection and interactions in a novel phenotypic environment.

Additionally, natural variation in sperm size (Vielle *et al.* 2016) suggests that there is also natural variation in sperm competitive ability (LaMunyon and Ward 2002). Finally, nematodes are small, easy to culture, can be cryopreserved, and have a wealth of genetic and genomic tools all of which allow for experiments to be done in a novel way, unfeasible in other systems (Brenner 1974; Kenyon 1988; Lee *et al.* 2017).

Chapter II describes genomic signatures for detecting sexual conflict (Kasimatis *et al.* 2017). In addition to myself, Thomas C. Nelson and Patrick C. Phillips contributed significantly to this published work. Sexual conflict represents an extreme form of an environment-dependent fitness effect. In this way, many of the predictions from environment-dependent selection can be used to formulate expected patterns of genome evolution under sexual conflict. However, the pleiotropic and transmission constraints inherent to having alleles move between sex-specific backgrounds from generation to generation further modulate the anticipated signatures of selection. This chapter outlines methods for detecting candidate sexual conflict loci both across and within populations and highlights the need to integrate genotype, phenotype, and functional information to truly distinguish sexual conflict from other forms of sexual differentiation.

Chapter III investigates the ability of sex-specific selection to created divergence between the sexes at a single locus. In addition to myself, Peter L. Ralph and Patrick C. Phillips contributed significantly to this unpublished work. While sexually antagonistic selection might favor different alleles within males and females, segregation randomly reassorts alleles at autosomal loci between sexes each generation. This process of homogenization during transmission thus prevents between-sex allelic divergence generated by sexually antagonistic selection from accumulating across multiple

generations. However, recent empirical studies have reported high male-female F_{ST} statistics. In this chapter I use a population genetic model coupled with individual-based simulations to evaluate whether these observations could plausibly be produced by sexually antagonistic selection.

Chapter IV describes the evolutionary history of a critical nematode sperm protein (Kasimatis and Phillips 2018). In addition to myself, Patrick C. Phillips contributed significantly to this published work. The major sperm protein (MSP) is unique to the phylum Nematoda and is required for proper sperm locomotion and fertilization. In this chapter, I annotate the major sperm protein gene family and analyze their molecular evolution in 10 representative species across Nematoda. This gene family does not conform to the standard expectation for the evolution of reproductive proteins. However, the molecular evolution of the MSP gene family is nonetheless consistent with the widely repeatable observation that reproductive proteins evolve rapidly (Swanson and Vacquier 2002), though in terms of gene structure, copy number, and genomic organization.

Chapter V investigates the unique sub-cellular sperm membranous organelles in nematodes (Kasimatis *et al.* 2018b). In addition to myself, Megan J. Moerdyk-Schauwecker, Nadine Timmermeyer, and Patrick C. Phillips contributed significantly to this published work. Nematode sperm contain subcellular vesicles known as membranous organelles that are necessary for male fertility, yet play a still unknown role in overall sperm function. In this chapter, I take a novel proteomic approach to characterize the functional protein complement of membranous organelles in two *Caenorhabditis* species: *C. elegans* and *C. remanei*. I identify distinct protein compositions between membranous organelles and the activated sperm body. Two particularly interesting and undescribed

gene families which localize to the membranous organelles are examined using molecular evolution analyses and CRISPR-based functional tests.

Chapter VI describes the development of an inducible sterility system for *C. elegans* (Kasimatis *et al.* 2018a). In addition to myself, Megan J. Moerdyk-Schauwecker and Patrick C. Phillips contributed significantly to this published work. Precisely controlling fertilization has proved a major challenge across model systems. Using the auxin-inducible degradation system targeting the *spe-44* gene within the nematode *C. elegans*, I designed a means of externally inducing spermatogenesis arrest. This chapter shows that exposure to auxin during larval development induces both hermaphrodite self-sterility and male sterility. Moreover, male sterility can be reversed upon cessation of auxin exposure. This sterility induction system has multiple applications in the fields of spermatogenesis, mating systems evolution, and aging.

Taken together, these chapters exemplify what can be learned about sexual antagonism and sexual conflict when using a lifecycle-explicit framework. Importantly, approaches at the genotypic and phenotypic levels must be used in concert to truly identify the source of sexual conflict and driving action of selection.

CHAPTER II

GENOMIC SIGNATURES OF SEXUAL CONFLICT

This chapter was published in volume 108 of the Journal of Heredity in 2017. Thomas C. Nelson and Patrick C. Phillips are co-authors on this publication. Patrick C. Phillips and I developed the ideas. Thomas C. Nelson and I co-wrote the manuscript. I was the principle investigator for this work.

The citation for this publication is as follows:

Kasimatis, K. R., T. C. Nelson, and P. C. Phillips. 2017. Genomic Signatures of Sexual Conflict. J. Hered. 108:780–790.

INTRODUCTION

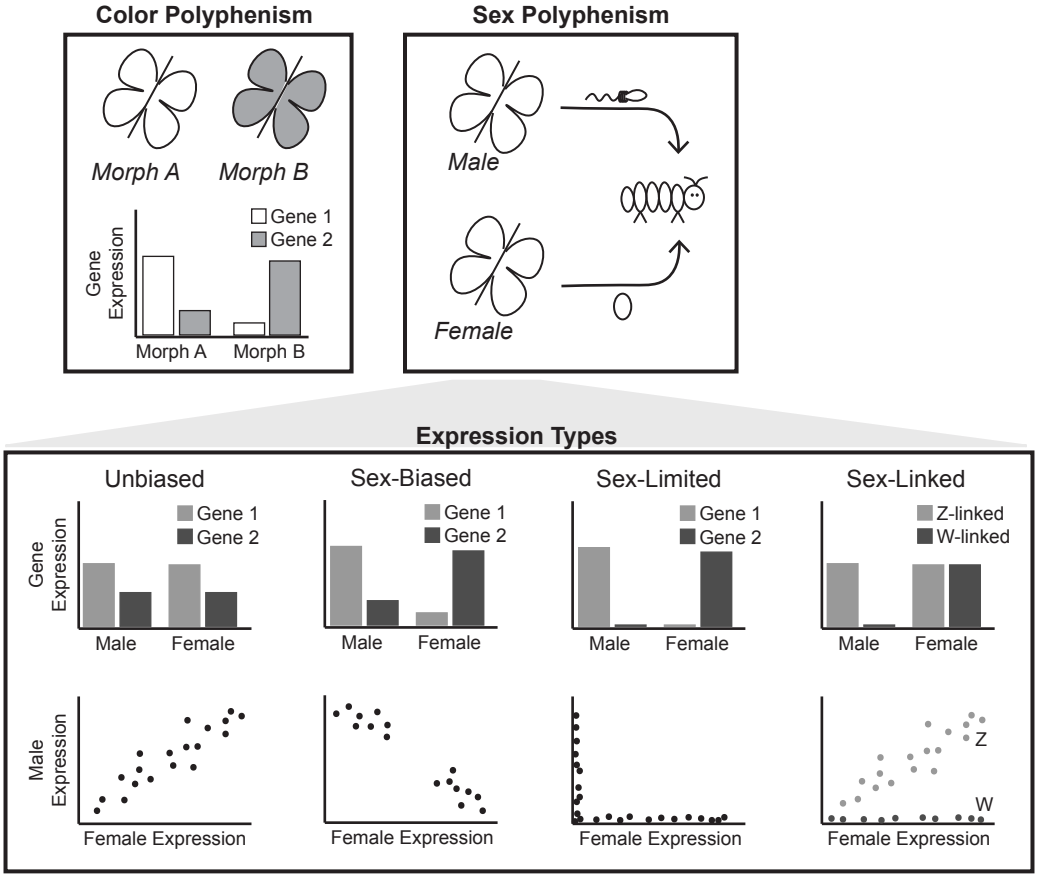
Reproducing optimally is in the evolutionary interest of both sexes, yet rarely is that optimum realized. Since sexual selection acts differentially on each sex, reproductive trait values are often skewed in the favor of one sex to the detriment of the other. This asymmetry in the realization of each sex's reproductive interests due to conflicts over mating and parental efforts is classically described as sexual conflict (Trivers 1972; Parker 1979). Focusing within the process of mating, sexual conflict can be more specifically described as the negative fitness consequences generated as a result of different reproductive trait optima between the sexes. For example, Bateman (1948) first characterized mating rate as a sexual conflict trait by showing that fecundity measured as a function of mating success differs between males and females in *Drosophila*

melanogaster, such that the fitness of males increases monotonically with increased mating rate, whereas females have an optimum mating rate beyond which total reproductive fitness decreases. Following Bateman, studies across a variety of taxa have shown that polyandrous females incur a cost of mating in the form of decreased fecundity or lifespan relative to monogamous females (Fowler and Partridge 1989; Rice 1996; Watson *et al.* 1998; Pitnick *et al.* 2001). Sexual conflict also occurs as part of post-insemination male ejaculate-female reproductive tract dynamics (Hollis *et al.* 2015; McDonough *et al.* 2016; Wilburn and Swanson 2016). In particular, seminal fluid proteins can alter female physiology and behavior and contribute to decreased female lifespan (Poiani 2006; Chapman 2011; Sirot *et al.* 2015). Additionally, sexual conflict can occur at the point of sperm-egg fusion due to fertilization rate differences (Swanson *et al.* 2001; Clark *et al.* 2009; Pujolar and Pogson 2011). Collectively, these studies demonstrate the fitness costs of sexual selection pushing males and females in opposing directions of phenotype space.

While sexual conflict is defined at the phenotypic level, the potential for conflict is a genomic property. The alleles that promote the optimal reproductive success of a given sex may have a positive, neutral, or negative impact on the other sex. In the case of negative effects, the subset of sexually differentiated genes with opposing fitness effects in males and females can broadly be categorized as intergenomic conflicts, of which sexual conflict is a special case (Rice and Holland 1997). If males and females had separate genomes, selection could optimize each for male- and female-specific attributes. However, males and females share a common genome and therefore can be viewed as representing an extreme form of polyphenism (Box 2.1). As selection acts to optimize the

Box 2.1. Sex as a Polyphenism

A polyphenism is an association of a genome with two or more discrete phenotypes across different environments (Suzuki and Nijhout 2006; Simpson et al. 2011). Therefore, polyphenisms are an extreme and discrete form of phenotypic plasticity. For example, in a butterfly color polyphenism, there is an association between the gene expression driving each phenotype and the season in which that phenotype is seen. As shown in the figure below, gene expression is regulated such that it is optimized for each environment and any incorrect expression signifies an environmental mismatch. Similarly, the sexes represent a polyphenism: gene expression is associated with a male or female phenotype based on the sexual environment (Mank 2017; Reuter et al. 2017). However, unlike other polyphenisms, the sexes must interact each generation to reproduce. Males and females are likely to require different genetic functions simply to operate as separate sexually differentiated individuals. These differences by themselves do not necessarily constitute conflict. Only when expression selected for a function in one sexual background has a negative effect on the other does conflict arise. Similarly, not all potentially antagonistic differentiation generated by sex-specific function can be said to be sexual conflict. Gene expression that leads to sex-specific fitness effects is classified as sexual conflict only when those functional effects influence the mating and fertilization success of the individual expressing those genes.



Box 2.1 *continued*

Males and females express a range of unbiased, sex-biased, sex-limited, and sex-linked genes that allow each sex to function properly. While unbiased gene expression may capture the early stages of sexual conflict, the sex-related forms of expression may represent varying degrees of sexual conflict (Connallon and Knowles 2005; Grath and Parsch 2016). Sex-biased genes have skewed expression patterns such that when the expression levels for male and female genes are plotted against each other (as shown below), a negative correlation exists. These patterns show an association with a sex at a given time, but are not necessarily a fixed property of a gene and can change based on the external environment (Grath and Parsch 2016). Such a correlation is expected for intergenomic conflicts such as sexual conflict, though not exclusive to this class. Further, genes that are inherently sexually dimorphic in expression, such as those expressed in gonadal tissues, have a negative intersexual correlation despite there being no necessary link to a mating interaction over which there is conflict. Thus, just as gene expression can be associated with a physical environment, so too are sex-biased genes associated with a sexual environment. Sex-limited expression represents an extreme form of sex-biased expression, where there is no correlation between male and female gene expression levels (as shown below). Such an expression state can represent a resolution to genomic conflicts, including sexual conflict. Similarly, by linking gene expression with the heterogametic sex chromosome, sexual conflict can be resolved through sex-limited expression. Here the simplest expression profile is shown with sex-limited expression of the heterogametic chromosome (W) and complete dosage compensation of the homogametic chromosome (Z) between the sexes. However, dosage compensation can be widely variable among taxa and therefore the homogametic sex chromosome may show a range of expression patterns (Mank 2009; Bachtrog et al. 2011). In all cases, caution must be taken in inferring past sexual conflict as other evolutionary dynamics, such as anisogamy, parental care, and imprinting, can also result in the evolution of sex-limited expression.

reproductive success of one sex, there is a response in the other sex, creating a genomic tug-of-war such that different alleles within the same genome are favored depending on the sexual environment in which they find themselves. Yet in each generation sex unites the genome and maintains the opportunity for sexual conflict. The resulting functional and evolutionary consequences depend on the balance between pleiotropy and linkage of the genes underlying the conflict trait. Male and female beneficial alleles contributing to a conflict trait can exist at different loci, resulting in interlocus sexual conflict (Rice and Holland 1997; Parker 2006). A single locus can also exhibit sexually antagonistic pleiotropy, whereby different alleles have opposing effects on male and female fitness. If this pleiotropy affects the optimal reproductive success of each sex, then the genetic basis of conflict is referred to as intralocus sexual conflict (Parker 1979; 2006).

Sexual conflict studies have predominantly focused on characterizing conflict traits, so naturally the search for the underlying genes has followed a forward genetics approach (Wilkinson *et al.* 2015). Thus far, a handful of genes underlying post-insemination conflict traits have been identified in this manner (Chapman *et al.* 1995; Swanson and Vacquier 1997; Wigby and Chapman 2005; Clark *et al.* 2009). Here the conflict traits are identified during a stage of mating in which the interaction driving sexual conflict is explicit and in a manner in which these traits can then be linked with the genome. However, since such studies necessarily exist on a gene-by-gene basis, they do not address the broader role of sexual conflict in genome evolution.

Reverse genetic approaches are gaining popularity for identifying sexually antagonistic loci. For instance, one potential approach is to focus on very rapidly evolving genes, as these are often related to reproduction (Clark *et al.* 2006) or immunity

(Sackton *et al.* 2007). This logic, however, is somewhat circular: reproductive proteins are considered rapidly evolving primarily because a number of independent studies have reported them to be so. A more common approach is to quantify gene expression differences between the sexes (Box 2.1). Although such studies have certainly identified numerous transcriptional sexual dimorphisms (Yang *et al.* 2006; Innocenti and Morrow 2010; Baker *et al.* 2011; Viguerie *et al.* 2012; Bohne *et al.* 2014), sexual differentiation does not equate to sexual conflict. The genes that enhance male survival at the expense of female survival represent an intersexual trade-off (sexual antagonism), but differential gene expression by itself cannot be taken as evidence for sexual conflict, *sensu stricto* (Box 2.1). Therefore, the sex-specific effects of genes alone cannot be taken as primary evidence that those genes underlie a conflict trait.

The genomics revolution has transformed both our conceptual understanding of how evolutionary processes affect the genome as well as our ability to detect the signatures of these processes. Genomic data have the potential to address key questions about the evolutionary importance of sexual conflict, including: How many loci are involved in sexual conflict in natural systems? How many loci contribute to a particular conflict trait? How important is sexual conflict relative to other evolutionary processes in shaping genome evolution? Currently we lack a null expectation to accurately assess genomic data and distinguish true sexual conflict from signatures of sexual differentiation and intergenomic conflict. Perhaps more important, as we will argue, is a general absence of a direct link between genomic variation and reproductive function in many approaches. Such a relationship is critical for distinguishing evolution generated by sexual conflict from more general forms of sexual differentiation. In the following

sections, we synthesize our understanding of sexual conflict as a driver of genomic evolution with a specific focus on hypotheses of genomic signatures of sexual conflict, methods for their detection, and the functional information needed to tie genomic evolution to conflict traits.

GENOMICS OF SEXUAL CONFLICT: SELECTION TOWARD MULTIPLE AND MOVING FITNESS OPTIMA

Interlocus Sexual Conflict: Coevolution and Linkage Disequilibrium

Interlocus sexual conflict, in which multiple loci contribute to the conflict trait, creates an intersexual genetic covariance and drives coevolution between the sexes (Gavrilets and Waxman 2002; Gavrilets and Hayashi 2006; Hayashi *et al.* 2007). These dynamics are analogous to those proposed by the Red Queen Hypothesis, where biotic interactions drive rapid, continuous evolutionary change (Van Valen 1973). Such interactions cause various modes of coevolution – fluctuating, escalatory, and chase-away – based on the action of selection (Brockhurst *et al.* 2014). In the case of sexual conflict, antagonistic male-female interactions are the self-driving force behind coevolution and the coevolutionary trajectory of the sexes can follow any of these Red Queen modes. Specifically, fluctuating Red Queen dynamics result from frequency-dependent selection, such that the sexes coevolve in a time-lagged, matching fashion. Here cyclic coevolution is predicted as long as genetic variation for the conflict trait is maintained (Haygood 2004). Conversely, escalatory Red Queen dynamics are characterized by an "arms race" between the sexes, whereby directional selection drives continuous coevolution (Gavrilets 2000; Gavrilets *et al.* 2001; Gavrilets and Waxman 2002; Hayashi *et al.* 2007;

Pennell *et al.* 2016). Similarly, chase-away Red Queen dynamics are also driven by directional selection and therefore can lead to continuous coevolution. However, here the driving antagonistic interaction is male exploitation countered by female resistance, such that males must overcome the ever-increasing female resistance threshold to mate (Holland and Rice 1998). Chase-away dynamics in particular create the opportunity for allelic diversification and assortative mating based on genotype fitness (Gavrilets and Waxman 2002; Gavrilets and Hayashi 2006; Hayashi *et al.* 2007). Unlike traditional Red Queen scenarios, such as host-parasite interactions, interlocus sexual conflict occurs within a shared genome and thus the evolutionary dynamics are shaped by the degree of sex-specific expression and pleiotropy at each locus contributing to the conflict trait.

When interlocus sexual conflict persists over long timescales, recurrent positive selection can result in accelerated protein evolution. Using a comparative molecular evolution framework based on this expectation, studies have investigated the evolutionary dynamics of interlocus sexual conflict by estimating the ratio of nonsynonymous to synonymous substitutions (ω-ratio) for genes assumed to be involved in sexual conflict so as to determine if the proteins are adaptively coevolving (Yang 1998; Clark and Aquadro 2010). This approach has been particularly well-suited for studying potential sperm-egg fusion conflicts in broadcast spawning marine invertebrates (reviewed in Swanson and Vacquier 2002; Clark *et al.* 2006; Vacquier and Swanson 2011). Here polyspermy (the fertilization of a single egg by multiple sperm) drives coevolution between sperm and egg recognition proteins with signatures of escalatory (Clark *et al.* 2009) and chase-away Red Queen dynamics (Levitan 2006; Manier and Palumbi 2008; Pujolar and Pogson 2011; Sunday and Hart 2013). Elevated rates of

evolution in seminal fluid proteins from *Drosophila* provide another example in which the rate of divergence of fertilization success related proteins suggests a role in sexual conflict (Begun *et al.* 2000; Wolfner 2002; Wagstaff 2005). These studies exemplify some of the most powerful information we have on the Red Queen Hypothesis on a macroevolutionary scale.

However, rapid evolution by itself cannot be taken as sole evidence for directional selection via sexual conflict. With a view toward seeking to understand broad patterns of genome evolution, it should be noted that the rapid turnover of amino acid residues may result from adaptive evolution at multiple amino acid residues, but this expectation need not be the case: reduced effective population sizes at loci under selection reduce the efficacy of purifying selection, and thus the observed amino acid substitutions may instead be effectively neutral or mildly deleterious (Smith and Haigh 1974; Vitti *et al.* 2013; Dapper and Wade 2016). This may be especially important for the evolution of genes on sex chromosomes, which have smaller effective population sizes than autosomes because of their hemizygous state (Bachtrog *et al.* 2011). However, sex-specific selection can cause deviations from the expected effective population sizes of sex chromosomes due to the faster-X effect (Vicoso and Charlesworth 2009; Mank *et al.* 2010), making sex-linked genes a potentially interesting subset of genes for such macroevolutionary interlocus sexual conflict studies. Moreover, divergence between populations must be driven by Red Queen dynamics occurring within populations and can therefore be more precisely described by taking advantage of within-population polymorphism at the whole-genome scale (Wilkinson *et al.* 2015).

Interlocus sexual conflict impacts genomic variation within a population through two related mechanisms. First, Red Queen dynamics affect patterns of polymorphism and haplotype structure. In particular, escalatory and chase-away Red Queen dynamics have the potential to produce classical signatures of persistent directional selection, namely selective sweeps (Fig. 2.1). Specifically, selective sweeps reduce nucleotide variation near the locus under selection because the advantageous allele only exists on limited number of genetic backgrounds. Any rare variants physically linked to the selected allele will also hitchhike to higher frequency and skew the site frequency spectrum (SFS) toward intermediate and high-frequency derived variants, creating "U-shaped" spectra (Fig. 2.1) (Smith and Haigh 1974; Nielsen *et al.* 2005). Until recombination can break the linkage between the selected allele and nearby variants, an in-progress or recently completed sweep will also impact local haplotype structure: selection reduces the total number of observed haplotypes and extends the length of the haplotype containing the selected allele (Sabeti *et al.* 2002; 2007). Under fluctuating Red Queen dynamics, frequency-dependent selection may further affect patterns of variation as selective sweeps remain incomplete and even reverse direction. Like positive selection, frequency-dependent selection can skew the SFS toward high-frequency derived variants (Fig. 2.1) (Huerta-Sanchez *et al.* 2008). If genetic variation is sampled while selected alleles are at intermediate frequency, however, frequency-dependent selection may produce excesses of intermediate frequency polymorphisms that are physically linked to the sites under selection (Charlesworth 2006). This type of constant selection will also maintain multiple distinct haplotypes associated with alternative alleles. Depending on the strength of selection and local recombination rates, this perturbation of haplotype variation may be

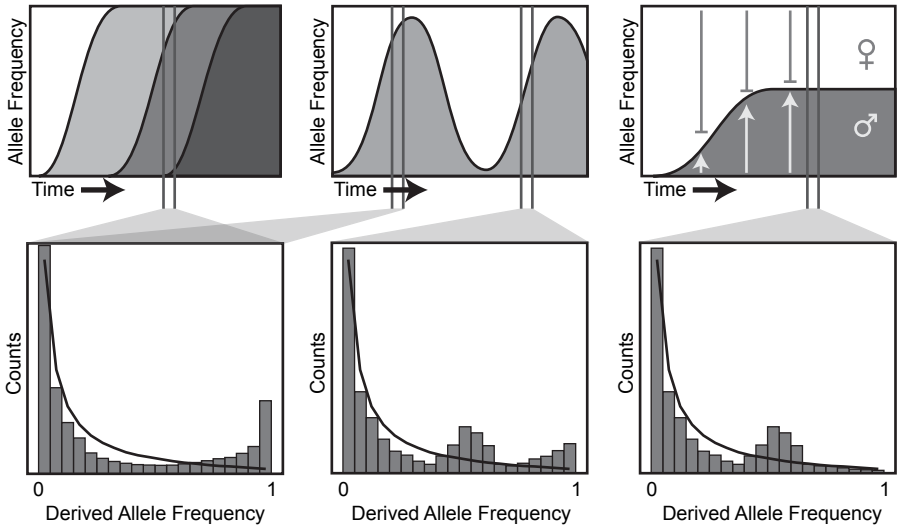
	Interlocus Sexual Conflict		Intralocus Sexual Conflict
Model of Selection	Red Queen: Escalatory & Chase-Away	Red Queen: Fluctuating	Sexually Antagonstic Pleiotropy
Action of Selection	Recurrent Positive	Frequency-Dependent	Balancing
Effect on Site Frequency Spectrum			
Effect on Linkage	• Long-range linkage disequilibrium maintained by assortative mating based on fitness		• Negligible
Detection Between Populations	• Estimate ration of nonsynonymous to synonymous substitutions (ω-ratio) to test for adaptive coevolution (e.g. PAML, HyPhy)		• Increased coalescence times within either or both populations.
Detection Within a Population	• Within-locus: Test for skewed SFS. Outlier tests for common statistics (π, Tajima's D); likelihood-based tests (e.g. SweepFinder). • Among interacting loci: Long-rang LD between female/male beneficial alleles at separate loci.		• F_{ST} scans for sexually differentiated SNPs. • Associations between genotype and mating success.
Confirmation	• Estimate the correlation in evolutionary rate between loci to prioritize candidate pairs for experimental biology.		• Test fitness effects sex-matched and sex-mismatched genotypes.

Figure 2.1. The effects of sexual conflict on polymorphism and linkage and methods for their detection. Interlocus and intralocus sexual conflict represent specific instances of broader categories of selection. Trajectories of conflict-associated alleles are shown through time, with unique alleles shown in shades of gray. Representative shifts in the site frequency spectrum surrounding conflict alleles are shown below their respective trajectories. Solid lines represent the neutral expectation based on Watterson (1975). Effects on linkage, and methods for detection are summarized below and in the text.

extensive enough to be detectable as a signature of selection. Note, however, that these variant patterns are not unique to sexual conflict and thus distinguishing between natural selection and sexual conflict will be a challenge without additional functional information.

A second and perhaps more specific effect of interlocus sexual conflict is the generation of linkage disequilibrium among male- and female-beneficial loci due to positive assortative mating generated by fertilization success (Gavrilets and Waxman 2002; Tomaiuolo and Levitan 2010; Patten *et al.* 2010). Unlike the patterns of sequence variation discussed above, this pattern is the direct result of a shared genome between the sexes. Such a pattern of linkage disequilibrium is especially the case with chase-away or escalatory Red Queen dynamics, as only high fitness males can overcome female resistance and successfully fertilize high fitness females. Offspring from these matings therefore contain both male- and female-beneficial alleles at a higher frequency than would be expected under random mating and, if male- and female-beneficial loci occur on the same chromosome, sex-beneficial alleles will initially be found in repulsion – female- and male-beneficial alleles will reside on maternal and paternal chromosomes, respectively. Recombination in the offspring will promote positive linkage disequilibrium by creating physical linkage between sex-beneficial alleles. This pattern of linkage disequilibrium should in principle be detectable within populations, although we are unaware of any studies that have used an approach based on this prediction to study sexual conflict.

The process of linking sex-beneficial alleles is also thought to underlie the evolution of sex chromosomes (reviewed in Bachtrog *et al.* 2011; Mank *et al.* 2014;

Kirkpatrick 2017). Sex chromosomes are predicted to be hotspots of sexual conflict (Rice 1984) due to their transmission dynamics and their different effective population sizes relative to autosomes. In particular, the build-up of sexually antagonistic loci causes recombination suppression between the sex chromosomes and creates patterns of antagonistic divergence over time (Charlesworth 1991; Wright *et al.* 2016). By taking advantage of high quality genomic data, recent studies across multiple taxa support this accumulation of sexually antagonistic loci on sex chromosomes (Nam *et al.* 2015; Lucotte *et al.* 2016; Wright *et al.* 2017). Thus, a qualitatively different approach to studying interlocus sexual conflict would be to focus on sex chromosome – particularly neo-sex chromosome – evolution and the accumulation of sexually antagonistic loci. However, despite this *a priori* expectation, potential conflict loci must still be linked with a conflict trait as not all genes physically linked to the sex determining locus will necessarily be sexually antagonistic.

For a variety of reasons, the loci involved in interlocus sexual conflict are also likely to demonstrate sex-biased or sex-limited expression. For example, as in classic cases of interlocus sexual conflict, if the genes involved are expressed only in the reproductive tract, then they will by definition have sex-limited expression (Swanson and Vacquier 1997; Kamei 2003; Findlay *et al.* 2014). In general, sex-limited expression has the potential to alleviate the negative pleiotropic effects of a sexual environment-gene expression mismatch (Box 1). However, the resolution of intralocus sexual conflict, which often results in sex-limited expression, has the potential to generate or exaggerate interlocus sexual conflict (Pennell and Morrow 2013; Berger and Martinossi-Allibert 2016; Pennell *et al.* 2016). Additionally, genes whose expression is already sex-limited,

such has genes functioning in the development of the reproductive tract or secondary sexual characteristics, can still be involved in interlocus sexual conflict even though these genes are already evolving in a sex-specific manner. Indeed, if mating dynamics promote conflict between the sexes, then genes already involved in reproduction and whose expression is sex-limited may be in fact preferentially recruited into conflict.

The detection of candidate loci involved in interlocus sexual conflict can take advantage of existing tools developed to interrogate patterns of genomic variation and gene expression, and distinguishing sexual conflict from other potential causes of within-genome antagonisms should ideally involve both approaches (Fig. 2.1). In particular, incorporating sex-biased or sex-limited gene expression data could help distinguish candidate interlocus sexual conflict loci from loci under positive natural selection (Cheng and Kirkpatrick 2016). Multiple algorithms are available that identify signatures of selection from genome-wide single nucleotide polymorphism (SNP) data, generated either through whole-genome shotgun sequencing or reduced-representation approaches (Nielsen 2005; Schrider and Kern 2016). To scan for sex-biased or sex-limited expression across the entire transcriptome, RNA-seq studies are becoming commonplace, even in nonmodel organisms. Since linkage disequilibrium between alleles at multiple interacting loci may be an important signature of interlocus sexual conflict, it will be beneficial to use sequencing protocols that retain genotype – and ideally haplotype – information for all individuals in a population. This presents a problem especially for studies using small organisms or very large sample sizes, where pooling of multiple individuals is common. Luckily, DNA library preparation methods are rapidly improving for small amounts of starting DNA, and molecular barcoding now allows multiplexing with tens of thousands

of unique barcodes. Certainly, these methods will continue to improve in the coming years.

Intralocus Sexual Conflict: Pleiotropy and Balancing Selection

Intralocus sexual conflict is a specific case of antagonistic pleiotropy, with the pleiotropic effects being dependent on shared effects across different sexual environments (Box 1). Thus, intralocus sexual conflict is similar to an environment-dependent fitness trade-off, where sex is the environment. Such trade-offs will create a sex-by-genotype interaction and would be expected to drive balancing selection (Connallon and Clark 2014). Further, this form of antagonistic pleiotropy should prevent alleles with differential sex-specific fitness effects from reaching fixation, thus maintaining sexual conflict (Arnqvist 2011). However, few predictions and little empirical evidence exists on the genomic consequences of intralocus sexual conflict.

Effects of intralocus sexual conflict on genomic variation should be similar to those predicted from more classical environment-dependent fitness tradeoffs that produce balanced polymorphisms (Charlesworth *et al.* 1997; Lenormand 2002; Charlesworth *et al.* 2003). Over long timescales, balancing selection can extend the expected residence time of an allele within a population far beyond that for neutral or positively selected sites, leading to increased polymorphism in linked genomic regions. Persistent balancing selection should therefore be detectable in between-population or between-species comparisons. Ideally, studies of intralocus conflict would focus on ongoing conflict within a mating population, when the phenotypic and fitness consequences of conflict can be directly measured. But this level of study may make the genomic signatures of sexual conflict more difficult to detect.

In a general scenario of an environment-dependent fitness tradeoff, selection generates genetic divergence among chromosomes carrying alternative alleles (Lenormand 2002). At the locus under selection, maintenance of variation depends on selection overriding the force of migration between environments. In flanking genomic regions, divergence extends out from the selected locus as a function of these forces and the recombination rate. In other words, larger selection coefficients and lower migration and recombination rates result in more extensive divergence among chromosomes.

A key difference between intralocus sexual conflict and other models of environment-dependent fitness tradeoffs arises when we consider the migration rate. In most, limited migration among environments allows for the selective maintenance of alternative alleles and produces signatures of divergence at and around selected loci (Slatkin 1987; Lenormand 2002; Yeaman and Whitlock 2011; Roesti *et al.* 2015). Under sexual conflict, however, the different environments are males and females. Thus, complete outcrossing among selective environments in each generation enforces an extremely high effective "migration" rate. Nevertheless, recent models suggest that intralocus sexual conflict can promote and maintain divergence between recombining sex chromosomes when conflict loci are in linkage disequilibrium with the sex-determining region (Kirkpatrick and Guerrero 2014). Genetic variation can also be maintained on autosomal loci if sexually antagonistic selection coupled with assortative mating is sufficiently strong (Arnqvist 2011). However, a more general understanding of how intralocus sexual conflict impacts genomic variation is currently lacking. The shared genome between the sexes likely provides a potent barrier to allelic divergence, as recombination will rapidly reduce divergence among chromosomes carrying alternative

alleles. Even if intralocus sexual conflict remains unresolved for long periods of time, differentiation among chromosomes carrying alternative alleles is likely to be extremely localized unless recombination rates are very low. Therefore, while a single instance of intralocus sexual conflict may be a weak force structuring genomic variation, opposing selection in males and females may still result in allele frequency differences between the sexes at a conflict locus. Open questions remain about how many loci are involved in these conflicts and what their combined effects on genomic variation may be.

Genomic Detection of Sex-Specific Selection

Given that the sexes share alleles, the detection of candidate loci involved in intralocus sexual conflict is somewhat more complicated than for interlocus conflict. Methods based on the site frequency spectrum may be able to detect conflict loci through the expected excess of intermediate frequency polymorphisms (Fig. 2.1). But as noted above, this effect will be highly localized unless conflict alleles are very young. The skew in the SFS will also closely resemble other forms of balancing selection, making it difficult to identify sexual conflict as the causative process. Recently, genome scan approaches have been adapted to find sexually differentiated SNPs by identifying allele frequency differences between the sexes (Lucotte *et al.* 2016; Cheng and Kirkpatrick 2016; Flanagan and Jones 2017). These approaches are appealing because they use established population genetic statistics, such as F_{ST}, to identify sexually antagonistic loci. Specifically, a male-female F_{ST} measures the change in allele frequency due to opposing selection between the sexes. This signature can be caused either by sexual conflict over reproductive fitness or sex-specific viability effects. In the latter case, sexually differentiated loci could be linked to sexual conflict in that they affect the optimal

reproductive interest of each sex. However, sex-specific viability effects can also result simply from natural selection acting differentially on males and females. A male-female F_{ST} alone cannot distinguish between these selective processes and therefore candidate genes would still need functional verification. Coupling these regions of differentiation to sex-specific expression does not necessarily distinguish among these possibilities as sexual differentiation does not equal to sexual conflict. Since the sexes for the most part share the same genome and reproduce each generation, any allele frequency difference between males and females only reflects a single generation of selection, regardless of the source of selection. This single generation limitation makes the feasibility of a male-female genome scan suspect.

To investigate this possibility more deeply, we developed a population genetic model to quantify the change in allele frequency at a single locus after a single generation of sexually antagonistic selection. Specifically, we examined the change in frequency of an allele separately within males and females to account for the sex-specific effects of selection. In particular, we focused on a male-beneficial, female-antagonistic allele. Using these sex-conditional allele frequency changes, we calculated the difference in allele frequency between the sexes (similar to calculating a Fisher's exact test) and male-female F_{ST}. The complete model is outlined in Box 2.2.

Overall, we find that a single generation of selection is not sufficient to change the frequency of an allele such that there is a distinct genomic signature unless selection is unrealistically strong. Although both the difference in allele frequency between the sexes and male-female F_{ST} increase as a function of selection (Fig. 2.2), the actual change in allele frequency is very small. For example, under an additive effects scenario, strong

Box 2.2. Population Genetic Model of Intralocus Sexual Conflict

Can whole-genome scans identify sexually antagonistic loci by comparing allele frequency differences within males and females? We consider this possibility by developing a simple population genetic model of the change in allele frequency between the sexes due to a single generation of sexually antagonistic selection. Specifically, we use a framework that allows us to estimate sex-conditional allele frequency changes such that selection can act differentially on an allele depending on the sexual environment (following Kidwell et al. 1977). This approach is actually fairly straightforward because transmission dynamics do not have to be considered. Instead, only the within-generation changes in genotype frequencies within male- and female-specific pools need to be tracked.

Consider a locus segregating for alleles A and a that are sex-specific beneficial: allele A is female-beneficial and allele a is male-beneficial. We represent sexually antagonistic selection as the cost of having the allele favored in the other sex. The relative fitnesses of the genotypes AA, Aa, and aa in males are thus $1 - s_f :: 1 - h_1 s_f :: 1$, where s_f is the cost of a male possessing the female-favorable allele and h_1 is the dominance coefficient in males. The relative fitnesses of the genotypes AA, Aa, and aa in females are similarly $1 :: 1 - h_2 s_m :: 1 - s_m$, where s_m is the cost of having the male allele and h_2 is the dominance coefficient in females.

Since sexual selection is typically stronger in males, we quantified the sex-conditional change in allele frequency of the male-beneficial allele. We can describe the frequency of the male-beneficial allele in males after a single generation of selection (q_m) as the relative frequency of a-containing male genotypes after selection divided by mean fitness of males the generation before selection:

$$q_m = \frac{q^2 + pq(1 - h_1 s_f)}{p^2(1 - s_f) + 2pq(1 - h_1 s_f) + q^2}.$$

Here q is the starting frequency of a, and $p = 1 - q$. Similarly, the frequency of the male-beneficial allele in females after a single generation of selection (q) is the frequency of the allele in females divided by the mean fitness of females:

$$q_f = \frac{q^2(1 - s_m) + pq(1 - h_2 s_m)}{p^2 + 2pq(1 - h_2 s_m) + q^2(1 - s_m)}.$$

The magnitude of change in allele frequency is simply the difference in allele frequencies between the sexes, such that:

$$\Delta q = q_m - q_f.$$

This can be translated into the familiar statistic (Wright 1931; Cheng and Kirkpatrick 2016) as:

$$F_{ST} = \frac{\Delta q^2}{4pq}.$$

As outlined in the main text, we considered the effects of dominance for five intralocus sexual conflict scenarios: additive beneficial and conflict allele effects ($h_1 = h_2 = 0.5$), conflict allele dominance ($h_1 = h_2 = 1$), beneficial allele dominance ($h_1 = h_2 = 0$), female-beneficial allele dominance ($h_1 = 1, h_2 = 0$), and male-beneficial allele dominance ($h_1 = 0, h_2 = 1$) (Fig. 2.2 & 2.3). For simplicity, we used a symmetrical cost of selection between the sexes ($s_m = s_f$) when looking at dominance effects. We also examined the effects of asymmetrical selection costs between the sexes, focusing on the additive dominance case ($h_1 = h_2 = 0.5$). A haploid version of the model is qualitatively similar to the diploid additive case (data not shown).

sexually antagonistic selection (s = 0.1) gives a virtually undetectable male-female F_{ST} (F_{ST} = 0.0007). These values are qualitatively consistent with a similar model and empirical estimates of sex-specific F_{ST} in humans and *Drosophila* produced by Cheng and Kirkpatrick (2016). Conflict allele or beneficial allele dominance yield outcomes that are qualitatively similar as those for an additive effects scenario, although sex-specific dominance can have qualitative and quantitative effects (Fig. 2.3). Under a scenario such as female-beneficial allele dominance, the maximum male-female F_{ST} with strong selection is approximately 1.5 times larger than under additive effects. Even so, this male-female F_{ST} still has a negligible impact on single locus detection (F_{ST} = 0.001).

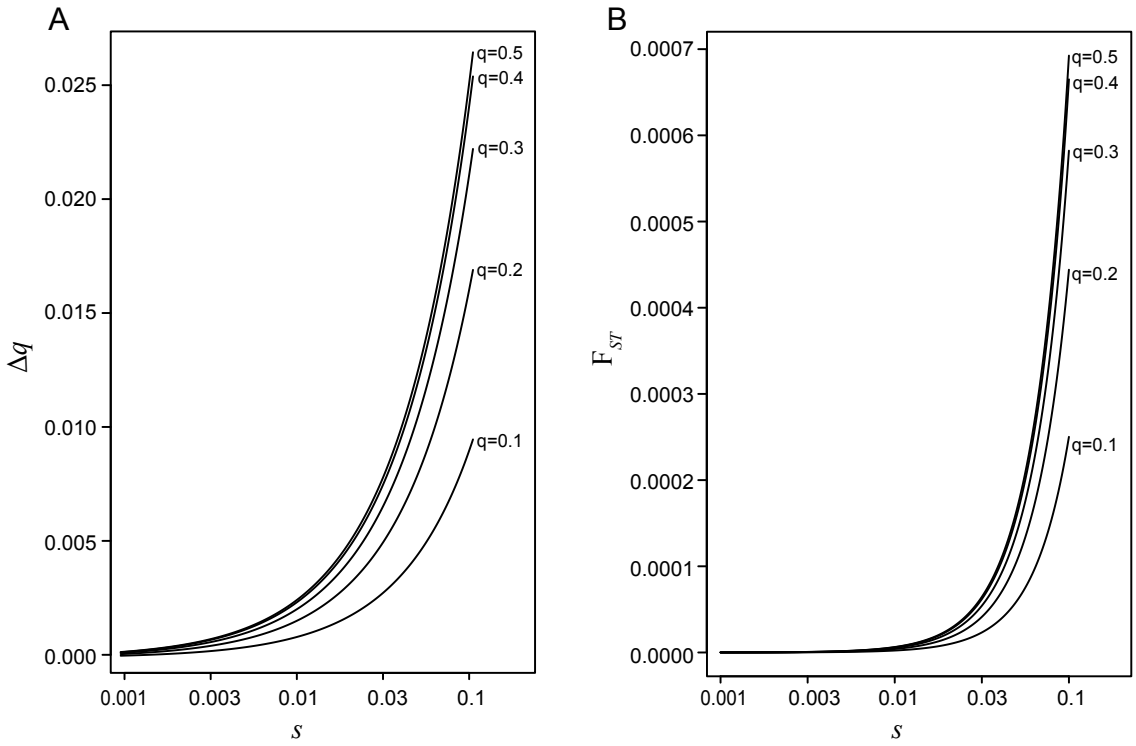


Figure 2.2. The change in the male-beneficial allele frequency with additive beneficial and conflict allele effects ($h_1 = h_2 = 0.5$) after a single generation of sexually antagonistic selection, where the cost of selection is equal between the sexes ($s_m = s_f = s$). A) The difference between male and female allele frequencies (Δq) as a function of selection for different initial values of q. B) Male-female F_{ST} as a function of selection for different values of q. For both response measures, the change in allele frequency increases as the cost of selection increases. The maximum change in allele frequency is seen when the male and female beneficial alleles start at the same frequency (q = 0.5).

Overall there appears to be very little power to detect sex-specific differentiation within a generation, even when selection is strong (also see Cheng and Kirkpatrick 2016).

Since sexual selection acts differentially between the sexes, we also examined the effects of asymmetrical fitness costs between the sexes. Selection had to be at least an order of magnitude different between the sexes to detect any quantitative differences male-female F_{ST}. The biological relevance of this high cost of selection in natural populations is questionable, particularly when selection is inherently linked with reproductive success. For instance, a selection cost of $s = 0.1$ suggests that 10% of individuals die each generation or do not contribute to the next generation due to their sexual context alone. These results suggest that the genetic load created by sexual conflict

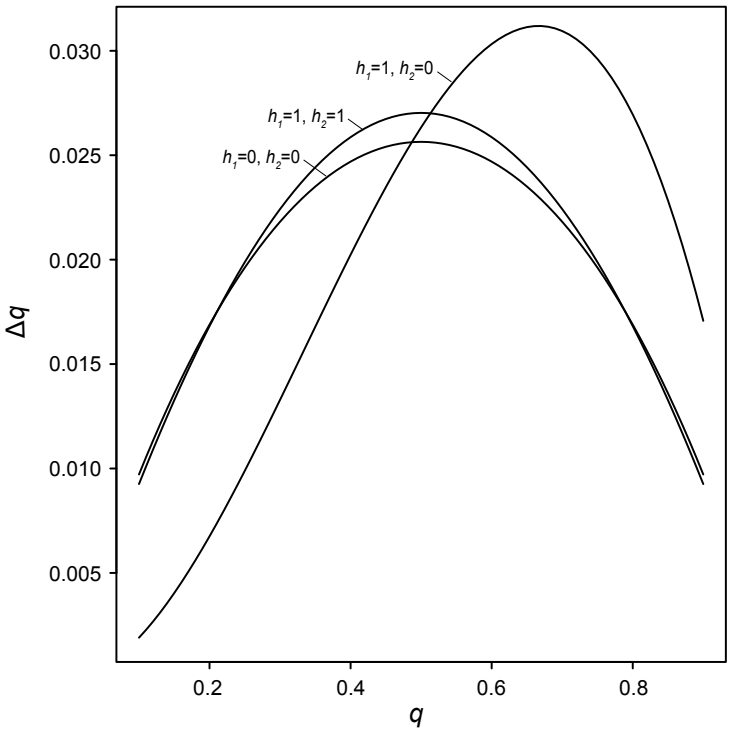


Figure 2.3. The change in the male-beneficial allele frequency due to a single generation of sexually antagonistic selection is dependent on the dominance relationships between the sexes. When there is conflict allele ($h_1 = h_2 = 1$) or beneficial allele ($h_1 = h_2 = 0$) dominance in both sexes, the difference in allele frequency is maximized when the male and female beneficial alleles start at the same frequency ($q = 0.5$). Additionally, there is a diminishing response to selection, such that the difference in allele frequency between the sexes forms a concave surface. Sex-specific beneficial allele dominance changes the shape and magnitude of the selection response curve. For example, when the female-beneficial allele is dominant ($h_1 = 1$, $h_2 = 0$), the response surface is shifted and stretched. Here the cost of selection is equal between the sexes ($s_m = s_f = 0.1$).

has the potential to be quite large. Similarly, if selection across the sexes is indeed this strong, then the opportunity to resolve conflict should also be strong and we would therefore expect natural selection for alleles that counter sexual antagonistic selection. However, if multiple loci are contributing to a conflict trait, then the average cost of selection could potentially be lower (Cheng and Kirkpatrick 2016). Alternatively, this single generation framework may not capture the full effects of sex-specific selection. Our model assumes there is uniform representation of alleles within the gamete pool at the start of each generation. However, selection on certain gamete types, could skew the gamete pool such that only a subset of individuals of each sex contribute their alleles to the next generation (Kidwell *et al.* 1977; Arnqvist 2011). Such dynamics have the potential to drive male-female differentiation at a locus and warrant further exploration.

From a practical perspective, detecting such small F_{ST} values is likely to be unrealistic. Nevertheless, recent studies using human data suggest that such sex-specific signatures of selection can in fact be detected at the whole genome level (Lucotte *et al.* 2016). Similarly, data from pipefish found remarkably high male-female F_{ST} values (Flanagan and Jones 2017). The ability to detect male-female differentiation using a genome scan suggests that such strong selection costs may be more prevalent than believed (though the mechanism of selection is unknown). If true, our view that the sexes share largely the same genome must be fundamentally altered. However, these male-female tests of differentiation are subject to false positives from many sources. Extremely large male and female sample sizes would be required not only to detect such F_{ST} values, but also to prevent false positives resulting simply from random sampling effects. Additionally, when possible, corrections must be made for linkage to the sex

chromosomes and, in particular, the sex determining region as these can also drive spurious male-female differentiation. Such a correction may be difficult in systems where the sex determining region or even sex chromosome is unknown, thus leading to false positive values. We suggest that a genome-wide permutation test (as in Churchill and Doerge 1994) should be performed to determine the null expectation and distinguish false positives from true signatures of male-female divergence. While genome scan approaches may be appealing for the moment, we suggest caution in interpreting their results until further theoretical work is completed, so that we more fully understand the biology underlying the statistic being measured and are able to distinguish false positives from signatures of sexually antagonistic SNPs.

Synthesis: Linking Genomic Signatures with Conflict Traits

While we can readily identify sexually differential expression patterns, not all of these necessarily correspond to sexual conflict. Since sexual conflict is defined at the phenotypic level, ultimately, we must link candidate loci identified via genetic and genomic signatures with their functional role within conflict traits. The most successful examples to date involve systems that are *prima facia* involved in reproductive interactions, such as sperm and seminal fluid proteins (reviewed in Swanson and Vacquier 2002; Clark *et al.* 2006; Vacquier and Swanson 2011). Expanding this framework to more general phenotypes will be a challenge. Several approaches can be taken to narrow down and prioritize these candidate loci (reviewed in Findlay and Swanson 2010). For instance, automated function-prediction software can verify if candidate loci can realistically be expected to be involved in a mating interaction.

Unfortunately, such programs tend to be somewhat generic and have limitations to the extent of functional information that they provide (Friedberg 2006). Therefore, they should be treated as a coarse pass over data, but not the only method via which sexual conflict is assigned. Coupling divergence methods with expression patterns is perhaps more informative (Harrison *et al.* 2015; Cheng and Kirkpatrick 2016), but still suffers from the potential weakness of conflating correlation and causation in the context of sexual conflict.

Evolutionary rate covariation (ERC) is another method that is particularly useful for prioritizing candidate interlocus sexual conflict genes (Clark and Aquadro 2010; Clark *et al.* 2012; Wolfe and Clark 2015). This method uses the ratio of nonsynonymous to synonymous substitutions to determine if the evolutionary rate of two proteins is correlated. Such a correlation is expected if proteins are physically interacting and subject to intermolecular coevolution (Clark *et al.* 2006; Clark and Aquadro 2010), if proteins are functionally related and therefore experiencing similar selective pressures (Clark and Aquadro 2010), or if gene expression patterns covary (Fraser *et al.* 2004; Hakes *et al.* 2007). By comparing the evolutionary rates for positively selected loci, potential Red Queen coevolutionary dynamics can be identified (Clark *et al.* 2009). While potentially powerful, this method does require a comparative genomics framework with sufficient dense phylogenetic signal.

In the end, true verification of sexual conflict necessitates experimental biology. Sexual conflict does not create a unique genomic signature – other forms of environment-specific selection, including sex itself as an environmental context, can produce the signatures outlined above. Classic molecular genetics approaches are likely to be

necessary to determine the function of candidate conflict-related loci. Such approaches, once wildly unrealistic for most systems, are increasingly coming into reach with the advent of general purpose genome editing approaches (Bono *et al.* 2015). Only when coupled with reproductive experiments will screening for a conflict trait and the mechanism of selection be possible. This complete link between genotype, phenotype, and function is necessary to describe the complete sexual conflict pathway and should be the standard to which we strive.

CONCLUSION

Reproduction is a fundamental biological process, yet the complexity of this process creates the potential for antagonistic interactions to become more pronounced. Sexual conflict, in its original usage, defines the negative fitness consequences generated by the process of mating and fertilization. This requirement of a male-female reproductive interaction distinguishes sexual conflict from other forms of antagonistic selection. We urge that a precise definition of sexual conflict be used when studying evolutionary dynamics. Such specificity will become increasingly important as we move from vague descriptions of possible evolutionary patterns to identifying specific genetic loci at genomic scales. Despite an emerging ability to detect potential loci involved in sexual conflict, these signatures are largely indistinguishable from those caused by natural selection and intergenomic conflicts in general. To firmly move the field into the genomics era, we need to couple an analysis of evolutionary patterns to the sex-specific functional context of putative conflict-related loci. That the field is now poised to

capitalize on these approaches promises many exciting developments and novel insights into the evolution of sexual conflict in the very near future.

BRIDGE

In Chapter II I outlined methods for identifying genomic signatures of sexual conflict. However, most of these signatures are not unique to this form of selection, with the exception of within-locus male-female divergence. In Chapter III, I expand on the model presented in Chapter II. I develop a population genetic model of sexually antagonistic selection that is inclusive of transmission to investigate the strength of selection required to detect genomic divergence between the sexes within a generation and the population fitness cost generated by this process.

CHAPTER III

LIMITS TO GENOMIC DIVERGENCE UNDER SEXUALLY ANTAGONISTIC SELECTION

Peter L. Ralph and Patrick C. Phillips are co-authors on this paper. Patrick C. Phillips and I developed the model. Peter L. Ralph and I performed the simulations. I am principle investigator for the work. I wrote the manuscript.

The preprint and supplementary material for this paper can be found at:
Kasimatis, K.R. *et al.* Limits to genomic divergence under sexually antagonistic selection. *bioRxiv* doi: https://doi.org/10.1101/591610

INTRODUCTION

Females and males use largely the same genome to produce distinct phenotypes and behaviors. This ubiquitous phenomenon requires an association between dimorphic phenotypes and their sexual environment (Kasimatis *et al.* 2017; Mank 2017b). Genes residing on a sex chromosome have a physical link to sex determination. Particularly, on heteromorphic sex chromosomes, the lack of recombination allows for selection to act in a sex-specific manner to optimize beneficial genes within each sex (Rice 1984, 1987; Charlesworth and Charlesworth 1980). Conversely, the shared genetic basis of autosomal genes prevents such sex-specific optimization of fitness. When autosomal-based traits have different optimal fitness values in each sex, then selection acts in a sexually antagonistic manner to push females and males in opposing directions in phenotype space (Rice and Holland 1997; Bondurianksy and Chenoweth 2009). However, recombination and meiotic segregation uncouple beneficial alleles from their sexual environment every generation, preventing the resolution of

antagonism via the creation of separate female and male genomic pools. This homogenization process tethers together the evolutionary responses of the sexes and creates an inherent intersexual genomic conflict (reviewed in Bonduriansky and Chenoweth 2009; Kasimatis *et al.* 2017).

Identifying sexually antagonistic loci – particularly using reverse genomics approaches – has proved challenging. Initial studies calculated differentation between females and males using WrightâĂŹs fixation index (F_{ST}), and interpreted high values as evidence of sexually antagonistic selection. Empirical data from multiple taxonomic groups (Lucotte *et al.* 2016; Flanagan and Jones 2017; Wright *et al.* 2018; Dutoit *et al.* 2018) suggest that hundreds to thousands of SNPs have elevated male-female autosomal divergence with outliers exceeding $F_{ST} = 0.01$ (Lucotte *et al.* 2016; Wright *et al.* 2018) and even approaching $F_{ST} = 0.2$ (Flanagan and Jones 2017). Taken at face value, both the number of sexually antagonistic alleles and the degree of divergence are striking. However, these results are difficult to evaluate as they suggest that there must be quite strong selection within each sex to drive such high divergence within a generation (see Cheng and Kirkpatrick 2016).

Several different processes could in principle generate divergence (or apparent divergence) between the sexes. First, sex biases in chromosome segregation through associations with the sex determining region could distort allele frequencies between the sexes. Over time, this segregation distortion can contribute to the generation of neo-sex chromosomes (Jaenike 2003; Kozielska *et al.* 2010) – particularly heteromorphic sex chromosomes – leading to sex-specific differentiation in the trivial sense that the locus is completely absent in one sex. Second, gametic selection resulting in a sex-specific fertilization bias could also distort allele frequencies (Joseph and Kirkpatrick 2004). Both of these processes occur during the gametic phase of the lifecycle and have long been recognized for their potential ability to distort segregation ratios within the sexes (reviewed in Immler and Otto

2018). In contrast, sexually antagonistic viability selection that occurs post-fertilization is a fundamentally different mechanism because there is no direct co-segregation of sex with the alleles under selection. Previous work on sexually antagonistic viability selection has largely focused on its potential role in maintaining genetic variation due to the sex-specific pleiotropic effects of the locus. In particular, Kidwell *et al.* (1977) laid out a framework for analyzing sexual antagonism that has widely been used in the field (Arnqvist 2011; Connallon *et al.* 2010; Connallon and Clark 2011; Patten and Haig 2009; Fry 2010). A little appreciated feature of the Kidwell model is that it tracks allele frequencies (rather than diploid genotype frequencies) in adults from each generation to the next. Although the model incorporates diploid selection, this sampling paradigm is sufficient because the "random union of gametes" model of mating only requires allele frequencies to generate diploid genotype frequencies in the next generation. However, this model simplification prevents the inclusion of other models of mating, such as assortative mating among genotypes.

In this paper, we will first build a model of sexually antagonistic viability selection, segregation, and transmission, extending the model of Kidwell *et al.* (1977) to include assortative mating. We use this model to evaluate how much between-sex differentiation is produced across a range of selection, dominance, and assortative mating parameters. Second, we use these results to evaluate the claims that the observed between-sex allelic differentiation is caused by sexually antagonistic viability selection. We then use simulation to test the conclusions of our deterministic model, as well as the role of sampling variance in generating loci with high between-sex differentiation. Both our single locus model and individual-based simulations with antagonistic loci distributed genome-wide indicate that antagonistic selection must be remarkably strong to produce non-negligible divergence between the sexes. Instead, simulations indicate that sampling variance in allele frequency is much more likely to account for extreme between-sex divergence and must therefore be

explicitly included in any analyses of putative signatures of male-female divergence.

METHODS

Model

Consider an autosomal locus in which are found two alleles: one male-beneficial (A_1) and one female-beneficial (A_2). Sexual antagonism results in a fitness cost to individuals carrying the allele favored in the other sex (Kidwell *et al.* 1977; Bodmer 1965). The life cycle is shown in Figure 3.1. Each generation begins with zygotic frequencies equal in each sex, but then genotype-dependent survival results in different genotype frequencies in each sex at time of mating. The relative fitnesses of genotypes A_1A_1, A_1A_2, and A_2A_2 in females are $1 :: 1 - h_f s_f :: 1 - s_f$, where s_f is the cost of a female having the male-favorable allele and h_f is the dominance coefficient in females. Writing the frequencies of the three genotypes in zygotes as $p_{11}(t)$, $p_{12}(t)$, and $p_{22}(t)$ at the start of generation t, the genotype frequencies in females after selection will then be proportional to $p_{11}(t)$, $p_{12}(t)(1 - h_f s_f)$, and $p_{22}(t)(1 - s_f)$, respectively. Similarly, the relative fitnesses of the genotypes A_1A_1, A_1A_2, and A_2A_2 in males are $1 - s_m :: 1 - h_m s_m :: 1$, and the genotype frequencies in males after selection are proportional to $p_{11}(t)(1 - s_m)$, $p_{12}(t)(1 - h_m s_m)$, and $p_{22}(t)$, respectively.

Therefore, the frequency of the female-beneficial allele in females post-selection, which we denote $p_f(t)$, is

$$p_f(t) = \frac{p_{11}(t) + \frac{1}{2}p_{12}(t)(1 - h_f s_f)}{p_{11}(t) + p_{12}(t)(1 - h_f s_f) + p_{22}(t)(1 - s_f)}. \quad (1)$$

The same quantity for males is:

$$p_m(t) = \frac{p_{11}(t)(1 - s_m) + \frac{1}{2}p_{12}(t)(1 - h_m s_m)}{p_{11}(t)(1 - s_m) + p_{12}(t)(1 - h_m s_m) + p_{22}(t)}. \quad (2)$$

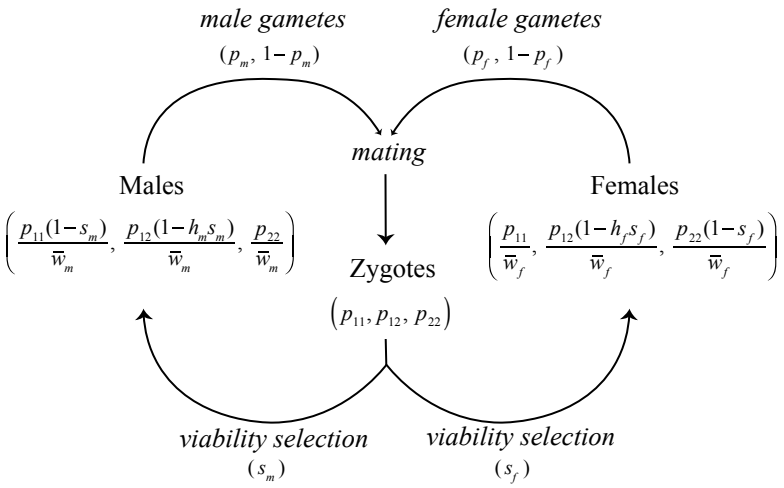


Figure 3.1: Lifecycle of the model. Zygotes are subject to sexually antagonistic viability selection (s_m and s_f), perturbing allele frequencies in adults in a sex-specific manner. Sex-specific adult allele frequencies are given in Equations 1 and 2, where $\bar{w}_f = \mathbf{w_f} \cdot \mathbf{p_t}$ and $\bar{w}_m = \mathbf{w_m} \cdot \mathbf{p_t}$. Surviving adults produce gametes of each allele type in frequencies corresponding to Equations 1 and 2. At this time meiotic segregation breaks the association between the locus and sex. Females and males mate with frequencies proportional to the mate choice matrix ($\mathbf{M}$) to produce the zygote pool in the next generation. Kidwell *et al.* (1977) gives the recursion for the allele frequencies in gametes (p_m, p_f), under the assumption of random, genotype-independent mating.

In a deterministic model of non-overlapping generations without gamete-specific selection, the genotype frequencies in the next generation are determined by the frequency of gametes joining from each of the nine possible mating combinations weighted according to mate choice. We parameterize mate choice using a matrix whose rows are indexed by male genotypes and columns by female genotypes, such that M_{ij} is the frequency of pairings of male genotype i with female genotype j relative to that expected under random mating.

We focus on three common mating scenarios by structuring the mate choice matrix as:

$$\mathbf{M} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_2 & m_1 & m_2 \\ m_3 & m_2 & m_1 \end{bmatrix}.$$

Under random mating, each pairing occurs with equal likelihood ($m_1 = m_2 = m_3$). Positive assortative mating by genotype occurs when females and males with the same genotype mate more frequently than those with different genotypes ($m_1 > m_2 = m_3$). Conversely, disassortative mating by genotype – or positive assortative mating by fitness – occurs when A_1A_1 individuals mate with A_2A_2 individuals ($m_1 = m_2 < m_3$).

The genotype frequencies in the next generation can be concisely calculated with some matrix algebra. Let $\mathbf{w_f} = (1, 1 - h_f s_f, 1 - s_m)$ and $\mathbf{w_m} = (1 - s_m, 1 - h_m s_m, 1)$ be the vectors of relative fitnesses in females and males respectively. Then, define the 3×3 matrix of fitness-weighted mate pairings, $\mathbf{F}$, so that for each pair of genotypes a and b, the entry $\mathbf{F}_{ab} = w_m(a) M_{ab} w_f(b)$. In other words, $\mathbf{F} = \mathrm{diag}(w_m) M \,\mathrm{diag}(w_f)$, where $\mathrm{diag}(w_m)$ denotes the matrix with $\mathbf{w_m}$ on the diagonal and zeros elsewhere. Finally, define $\beta = \mathrm{diag}(1, 1/2, 0)$ and $\gamma = \mathrm{diag}(0, 1/2, 1)$. Then, the vector of frequencies of each genotype among zygotes (before selection) in the next generation can be calculated using the current frequencies as a weighted sum over possible mating pairs:

$$\begin{aligned} p_{11}(t+1) &= \frac{\mathbf{p(t)^T} \beta \mathbf{F} \beta \mathbf{p(t)}}{\mathbf{p(t)^T F p(t)}} \\ p_{22}(t+1) &= \frac{\mathbf{p(t)^T} \gamma \mathbf{F} \gamma \mathbf{p(t)}}{\mathbf{p(t)^T F p(t)}} \\ p_{12}(t+1) &= 1 - p_{11}(t+1) - p_{22}(t+1). \end{aligned} \tag{3}$$

Here, $\mathbf{p(t)} = (p_{11}(t), p_{12}(t), p_{22}(t))$ is the column vector of genotype frequencies and $\mathbf{p(t)}^T$ is its transpose. This set of equations can be derived by noting that the relative frequencies of A_1A_1, A_1A_2, and A_2A_2 genotypes produced in the next generation are $\mathbf{p(t)}^T \beta \mathbf{F} \beta \mathbf{p(t)}$,

$\mathbf{p(t)}^T(\beta\mathbf{F}\gamma + \gamma\mathbf{F}\beta)\mathbf{p(t)}$, and $\mathbf{p(t)}^T\gamma\mathbf{F}\gamma\mathbf{p(t)}$, respectively; since $\beta + \gamma = I$, the identity matrix, these sum to $\mathbf{p(t)^T F p(t)}$, the denominator in equations (3).

We then used Mathematica v11.1.1.0 (Wolfram Research, Inc.) to find the equilibria of this system and determine stability of those equilibria. The complete notebook is provided in File S1.

Within-generation statistics

Sex-specific viability selection creates differences in allele frequencies between the sexes each generation. We can therefore quantify the effects of sexually antagonistic selection using the male-female F_{ST} statistic, which we calculate as the squared difference in allele frequencies between sexes, normalized by the total heterozygosity across sexes (Cheng and Kirkpatrick 2016; Wright 1951):

$$F_{ST} = \frac{(p_m - p_f)^2}{4(p(t)_{11} + p(t)_{12}/2)(p(t)_{22} + p(t)_{12}/2)}. \tag{4}$$

Sex-specific selection creates divergence between the sexes by increasing the frequency of the beneficial allele in each sex. Therefore, at the population level, this opposing action of section skews genotype frequencies away from Hardy-Weinberg equilibrium. The degree of inbreeding within the population due to sex-specific effects can be quantified using Wright's F_{IS} statistic (Wright 1951):

$$F_{IS} = \frac{p(t)_{12}}{2(p(t)_{11} + p(t)_{12}/2)(p(t)_{22} + p(t)_{12}/2)} - 1.$$

A population fitness cost due to sexual antagonism (i.e., genetic load), is generated each generation. Within each sex, the genetic load is the difference between the maximum possible fitness and the mean fitness (Haldane 1957, 1937). The population's average genetic load (L) is the average of the loads for each sex (assuming an equal sex ratio), which is given by:

$$L = 1 - \frac{\bar{w}_m + \bar{w}_f}{2}, \tag{5}$$

where $\bar{w}_m = p_{11}(t)(1 - s_m) + p_{12}(t)(1 - h_m s_m) + p_{22}(t)$ and $\bar{w}_f = p_{11}(t) + p_{12}(t)(1 - h_f s_f) + p_{22}(t)(1 - s_f)$.

Simulations

We used R (R Core Team 2018) (File S2) to simulate allele frequency dynamics at a single locus in a population subject to selection and drift. During viability selection each generation, each individual survived with probability equal to their (sex- and genotype-dependant) fitness. Then, the genotype frequencies within each sex were multiplied to give the matrix of relative frequencies of possible mating pairs, which was further weighted by the mate choice matrix. To generate the next generation, a fixed number of mating pairs are sampled from this distribution, and offspring are produced by random choice of parental alleles.

We also implemented simulations with sexually antagonistic selection acting at many loci, genome-wide with SLiM v3.1, an evolution simulation framework (Haller and Messer 2019) (recipes in File S3). Individuals each had a genome of 100 Mb, a uniform recombination rate of 10^{-8}, and a mutation rate of 10^{-10}. All mutations are sexually antagonistic (we do not simulate neutral variation): each new mutations was beneficial in a randomly chosen sex and detrimental in the other, with selection coefficients drawn independently for each sex from a Gaussian distribution with mean zero and standard deviation 0.01. Each mutation also had dominance coefficients drawn independently for each sex from a uniform distribution between 0 and 1. The model had overlapping generations: each time step, first viability selection occurred (with probability of survival equal to fitness), followed by reproduction by random mating. The number of new offspring was chosen so that the population size fluctuated around 10,000 diploids, and simulations were run for 1,000 time steps. For a neutral comparison, we also simulated from the same scenario but with no fitness effects. We ran 5 independent simulations of each scenario (i.e., neutral and sexually antagonistic).

After the final generation, genetic load and male-female F_{ST} at each locus were calculated. F_{ST} values were calculated both using all individuals within the population as well as using smaller subsamples of 100 individuals and 50 individuals with equal numbers of each sex. Subsample sizes were chosen to reflect sample sizes currently used in the literature. Male-female F_{ST} values within the subsamples were calculated using the modification of Wright's derivation (Eq. 4) as well as using Weir and Cockerham's F_{ST} estimator (Weir and Cockerham 1984; Bhatia *et al.* 2013) to examine the impact of the statistic used on the distribution of F_{ST} values. Under equal female and male subsampling, Weir and Cockerham's F_{ST} is equivalent to Hudson's F_{ST} (Bhatia *et al.* 2013).

RESULTS

We first examine the conditions under which our model supports a stable polymorphism, and then examine the degree of between-sex divergence and genetic load expected under both equilibrium and non-equilibrium (selective sweep) conditions. Finally, we verify these results using simulations, which also provide an opportunity to explore the effects of statistical sampling on inferences of sex-specific differentiation from genomic samples.

Transmission dynamics at a sexually antagonistic locus

Maintenance of polymorphism requires symmetric selection between the sexes under random mating: We will quantify the strength and degree of asymmetry between the sex-specific allelic effects using the overall strength (s) and the ratio of selection coefficients (α), so that $s_m = s$ and $s_f = \alpha s$. The full solution for the maintenance of polymorphism under arbitrary patterns of dominance can be solved by setting $p(t+1) = p(t)$ in the recursion equations above (Equations (3); File S1). Under general conditions, this system yields a fifth-order polynomial that does not readily generate a closed form solution in symbolic form, although the equilibria can be easily found numerically. Symbolic solutions are

possible under some specific conditions.

Assuming random mating and additivity of allelic effects ($h_m = h_f = 0.5$), the frequency of the A_1 allele at equilibrium (denoted $\hat{p}_{A1}$) can be expressed in terms of the strength of selection and asymmetry in selection:

$$\hat{p}_{A1} = \frac{1}{2} - \frac{1-\alpha}{2s\alpha}. \tag{6}$$

When selection is equally antagonistic across the sexes, an equilibrium frequency of $\hat{p}_{A1} = 0.5$ is always predicted. This theoretical solution is well supported by the stochastic simulations as well (Fig. 3.2A-B). The bounds on the non-trivial equilibrium frequency can be found by setting $\hat{p}_{A1}$ to zero or one. By solving these equations for α in terms of the strength of selection (s), we find that for the equilibrium to be stable, α and s must satisfy the condition:

$$\frac{1}{1+s} < \alpha < \frac{1}{1-s}. \tag{7}$$

These bounds can also be found by calculating the Jacobian matrix for the full set of transition equations (File S1) and agree with those identified by Kidwell *et al.* (1977). In general, the equilibrium conditions describe an expanding envelope in parameter space that allows more asymmetry in the pattern of antagonistic selection as the absolute strength of selection increases (Fig. 3.3A & B). To a first order approximation in s, equation (7) shows that the equilibrium is stable only if asymmetry is not larger than the strength of selection, such that $|\alpha - 1| < s$, as shown in Fig. 3.3A. Thus, when selection is weak or moderate, the maintenance of a polymorphism requires approximately equal selection between the sexes. However, the permissible degree of asymmetry increases with the strength of selection (Fig. 3.3B). For example, when $s \geq 0.4$ a stable polymorphism can be maintained so long as the asymmetry in fitness ($|1 - \alpha|$) is less than 50%. Selection coefficients of this magnitude mean mortality rates of 40% or higher each generation due to a single incorrect sexually

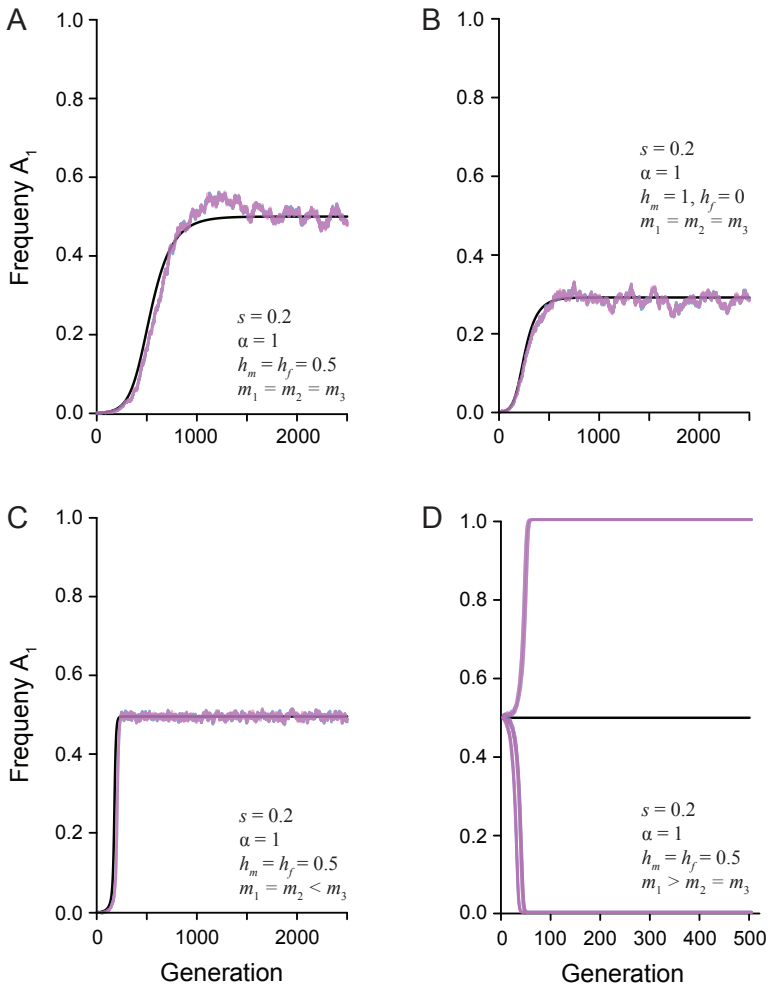


Figure 3.2: The change in the frequency of a newly derived sexually antagonistic allele (A_1) over time. The black line represents the predicted allele frequency from the recursion equation. The overlaid pink and blue lines represent the simulated population (N = 20,000) of females and males, respectively. The strength of selection (s), ratio of selection between the sexes (α), dominance relationship (h_m and h_f), and mate choice coefficients (m_1, m_2, and m_3) are given in each panel. A) Random mating with additive dominance and symmetric selection between the sexes maintains a stable polymorphism. B) Random mating with complete male dominance and symmetric selection between the sexes maintains a stable polymorphism. C) Assortative mating by fitness with additive dominance maintain a stable polymorphism. D) Assortative mating by genotype with additive dominance has an unstable equilibrium. Multiple simulated populations show how drift will quickly lead to fixation or loss of the A_1 allele.

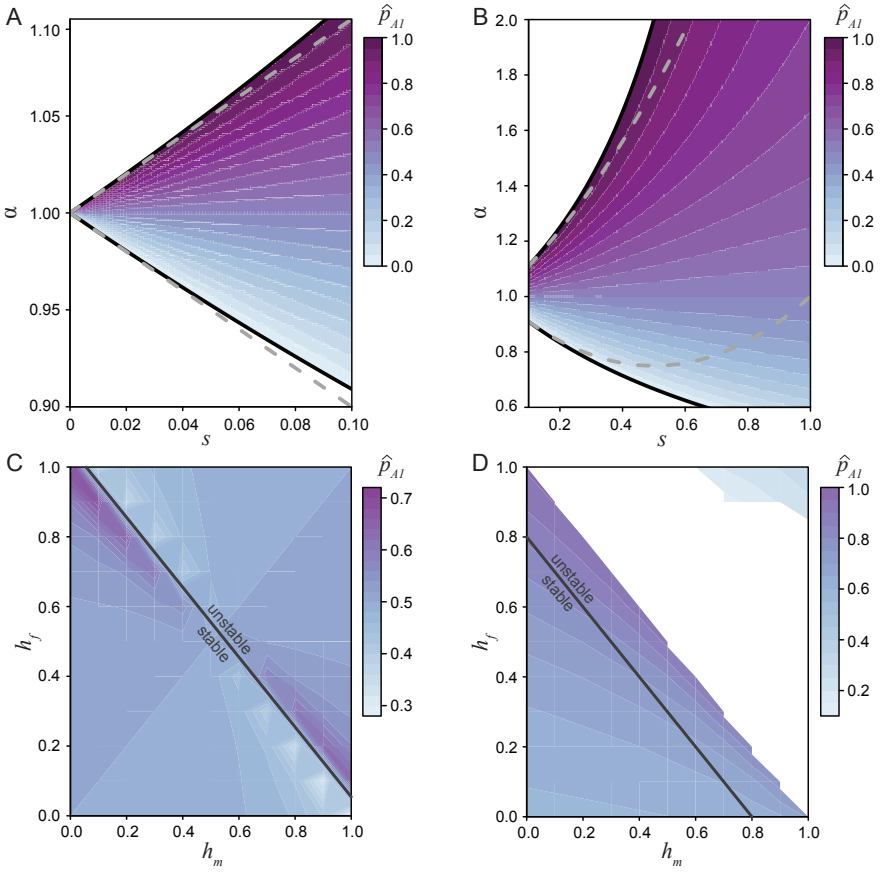


Figure 3.3: The equilibrium space for the A_1 allele under differing selection and dominance conditions. A) The equilibrium space at an additive locus ($\hat{p}_{A1}$, Equation 10), when selection is weak and related between the sexes by the ratio α. Here the equilibrium space is symmetric around $\alpha = 1$ and confined to approximately equal selection between the sexes. The solid black line represent the permissible bounds on α (7) and the dashed gray line represents the first order Taylor series approximation. B) The equilibrium space at an additive locus increases as the strength of selection increases. The solid black line represents the bounds on α and the dashed gray line represents the second order Taylor series approximation. C) The equilibrium space across all dominance conditions when selection is equal between the sexes ($s = 0.1, \alpha = 1$). When the dominance coefficients between the sexes sum to no greater than one ($h_m + h_f \leq 1$), then the equilibrium is stable. However, when the sum is greater than one the equilibrium is unstable. D) Strong, asymmetric selection ($s = 0.4, \alpha = 1.5$) narrows the equilibrium space and range of stable conditions ($h_m + h_f \leq 0.8$).

antagonistic allele, which seems biologically implausible. Therefore, under additivity, any stable antagonistic polymorphisms must have approximately equal fitness effects in the two sexes, while less balanced antagonistic loci will quickly be fixed or lost.

On the other hand, if dominance is allowed to vary between the sexes but selection is equally antagonistic across the sexes ($\alpha = 1$), there is always a single real, non-trivial equilibrium (Fig. 3.3C), whose stability depends on the sum of the dominance coefficients between the sexes. When $h_m + h_f \leq 1$ the equilibrium is stable (File S1). This stability boundary makes sense as the mean fitness of homozygous individuals is lower than that of heterozygous individuals (assuming equal sex ratios):

$$1 - \frac{s}{2} \leq 1 - \frac{s(h_m + h_f)}{2}.$$

In other words, the equilibrium remains stable if the deleterious effects of dominance in one sex do not outweigh the benefits in the other sex. Interestingly, weak selection at a locus with sex-beneficial dominance ($h_m = h_f = 0$) can maintain a stable polymorphism despite greater asymmetry in selection than can an additive model (File S1). This expansion of the stability region is likely a result of heterozygotes being shielded from antagonistic selection and suggests that modifying dominance can act to maintain sexual antagonism at a locus. Conversely, when $1 < h_m + h_f \leq 2$, dominance favors the deleterious allele in each sex, pushing the population to an unstable state and leading to the fixation of the less costly allele. Allowing for asymmetry in the strength of selection narrows the equilibrium space and reduces the range of dominance coefficients resulting in stability (Fig. 3.3D).

Assortative mating by fitness expands the polymorphism space: Under positive assortative mating by fitness, high fitness matings occur between disparate genotypes and therefore produce an excess of heterozygotes each generation. Under this mating dynamic (with $m_3 > m_2 = m_1$), up to three real non-trivial equilibria can exist depending on the selection and dominance parameters (File S1). However, as with random mating, at most one equi-

librium is stable. When selection is symmetrically antagonistic across the sexes ($\alpha = 1$), an A_1 allele frequency of approximately 0.5 is always predicted, regardless of dominance. This prediction is borne out by the single locus simulation results, which further show that assortative mating by fitness tends to make the stable equilibrium more robust to the effects of genetic drift (Fig. 3.2C). Increasing the asymmetry of selection can introduce an additional unstable equilibrium, and increasing the strength of sex-deleterious dominance (towards $h_m = h_f = 1$) can introduce a second unstable equilibrium (File S1). These theoretical predictions agree with previous simulations of assortative mating (Arnqvist 2011). As with random mating, the relationship between the strength and asymmetry in selection is the critical factor in determining when equilibria are stable. Specifically, when the asymmetry in selection is sufficiently large, fixation of the more favored allele is expected. Fixation only tends to occur under unrealistically large viability costs, however, and so the predominant outcome of assortative mating by fitness is the maintenance of heterozygotes and an expansion of the equilibrium space relative to random mating.

Assortative mating by genotype leads to fixation: In contrast to assortative mating by fitness, if assortative mating is by genotype ($m_1 > m_2 = m_3$), there is only a single non-trivial equilibrium (File S1). This equilibrium is always unstable, regardless of dominance, as shown by the leading eigenvalue of the Jacobian matrix. Figure 3.2D shows allele frequency trajectories that start at this unstable equilibrium rapidly go to loss or fixation (with the choice determined by random genetic drift). Thus, these mating dynamics shrink the equilibrium space and lead to the loss of the weaker antagonistic allele.

Male-female divergence is exceptionally low

A number of studies have observed high mean male-female divergences (measured by F_{ST}). For instance, Dutoit *et al.* (2018) found a mean male-female $F_{ST} = 0.0016$ across genes with male-biased expression in a sample of 43 flycatchers of each sex. Wright *et al.* (2018) found

a larger average male-female F_{ST} value of 0.03 across sex-biased genes in transcriptomes of 11 male and four female Trinidadian guppies. Similarly, Flanagan and Jones (2017) identified 473 genome-wide outliers having male-female F_{ST} values above roughly 0.05 in a RADseq study of 171 male and 57 female gulf pipefish. Finally, Lucotte *et al.* (2016) found an average male-female F_{ST} of 0.067 across autosomal SNPs in the human HAPMAP data that showed significant nonzero male-female F_{ST} in all 11 populations (with around 100 samples of both sexes per population). Our previous work showed that selection within a single generation at an additive locus must be strong to generate a male-female $F_{ST} > 0.01$ (Kasimatis *et al.* 2017). The model we study here allows us to estimate the strength of antagonistic selection required to produce male-female F_{ST} values as large as these, both at stably polymorphic loci and at loci undergoing a selective sweep.

When selection and dominance coefficients are chosen such that a stable equilibrium is maintained, divergence between the sexes tends to be exceptionally low (Fig. 3.4A). For example, a 10% viability cost ($s = 0.1$) results in a between-sex F_{ST} value of 0.0007 at equilibrium (assuming an additive locus and random mating). An equilibrium male-female F_{ST} value of 0.0016 (as in flycatchers) requires at least a 15% viability cost within each sex ($s = 0.15$, $\alpha = 1$). To produce equilibrium F_{ST} values an order of magnitude larger (as reported for the largest loci in the other taxa) requires a 30-65% viability cost ($s = 0.30 - 0.65$, $\alpha = 0.8 - 2.0$). For these values to be a product of viability selection, the field would need to have overlooked as much as 50% genotype-dependant mortality (or infertility) for each sex every generation, which seems implausible in these taxa.

Greater divergence can be generated across a broader range of selection values when an antagonistic locus transiently sweeps through a population. Here a viability cost of 10% produces higher divergence than at equilibrium, although divergence is still low in absolute terms ($F_{ST} < 0.002$ across dominance values, under random mating). Again, at

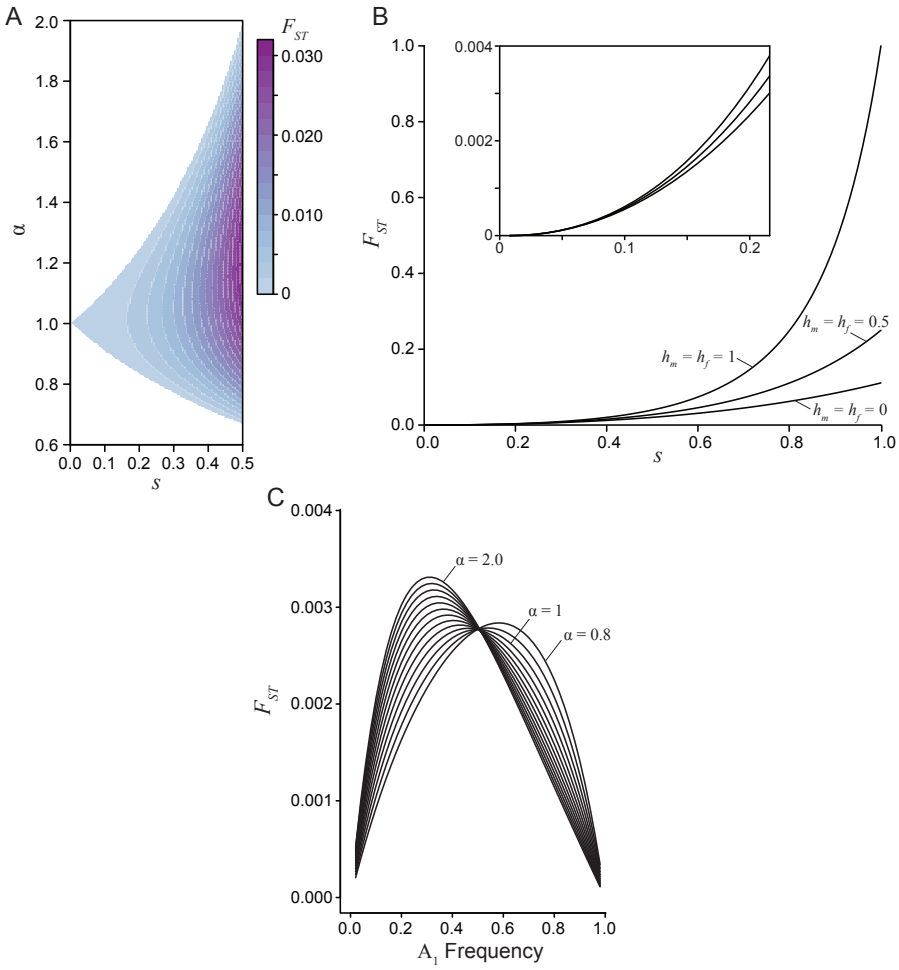


Figure 3.4: Divergence between the sexes. A) Male-female F_{ST} for an additive locus ($h_m = h_f = 0.5$) at equilibrium, where the strength of selection between the sexes is related by the ratio α. B) Male-female F_{ST} as a function of selection for three dominance regimes: sex-specific beneficial ($h_m = h_f = 0$), additive ($h_m = h_f = 0.5$), and deleterious ($h_m = h_f = 1$). Sex-specific beneficial dominance always results in the lowest divergence between the sexes. The inset graph highlights the similarly low divergence values generated under weak and moderately weak selection. C) Male-female F_{ST} at a sex-beneficial locus ($h_m = h_f = 0$) as a function of A_1 allele frequency for varying degrees of asymmetry in selection ($0.8 \leq \alpha \leq 2$) with a fixed mean selection coefficient ($0.5(s_m + s_f) = 0.2$). When $\alpha = 1$ represents a stable equilibrium state, while the other curves are not at equilibrium.

least a 30% viability cost would be required to produce F_{ST} values above 0.05. Sex-specific beneficial dominance ($h_m = 0, h_f = 0$) is expected to generate the lowest levels of between-sex divergence, while sex-specific deleterious dominance ($h_m = 1$, $h_f = 1$) yields the greatest levels divergence, though such a scenario seems biologically unstable (Fig. 3.4B). Importantly, under weak selection dominance has only a negligible effect on divergence. In fact, varying dominance does not generate quantitative changes in F_{ST} unless selection is remarkably strong ($s > 0.5$). Rather, asymmetry in selection seems a more important driver of divergence in non-equilibrium populations, as this asymmetry is precisely the factor that moves populations away from equilibrium conditions to a state in which the least costly allele sweeps to fixation. Across the range of α values with a fixed mean strength of selection between the sexes, divergence slightly increases as asymmetry between the sexes increases (Fig. 3.4C). However, the male-female F_{ST} values are of the same magnitude despite strong asymmetry when the mean strength of selection in confined in this manner. When the strength of selection varies independently between the sexes, increasing α yields much greater divergence, though this result is confounded by overall stronger selection in one sex. Overall, substantial divergence between the sexes still requires strong selection in non-equilibrium populations.

Sexual antagonism generates a substantial genetic load

Since male and female fitness are each maximal under fixation for different alleles at an antagonistic locus, sexually antagonistic selection generates a genetic load within the population at both a polymorphic equilibrium and during a selective sweep. At equilibrium under random mating, the load is maximized if the strengths of selection in each sex are equal (Fig. 3.5A), and dominance has little to no effect. Importantly, across strengths of selection up to $s = 0.5$, the load generated at equilibrium exceeds F_{ST} between males and females by nearly a factor of 10 (Fig. 3.5B). For example, a 10% viability cost ($s = 0.1$)

results in a reduction of population fitness up to 5%, with a maximum F_{ST} value of 0.0007. The load produced by a single antagonistic locus with F_{ST} equal to the **mean** male-female F_{ST} reported in human HAPMAP data (Lucotte *et al.* 2016) would exceed 20% (Fig. 3.5B). This relationship indicates that even weak selection driving low – and probably undetectable – levels of divergence can generate a substantial fitness reduction due to the sex-specific nature of selection.

An alternative way to examine load is by quantifying the excess of heterozygosity due to sexually antagonistic selection, using the F_{IS} statistic (the inbreeding coefficient). Here sex-specific selection creates homozygous pools of each sex within a generation, which leads to an excess of heterzygotes at the start of the next generation. Under weak selection, these departures from Hardy-Weinberg equilibrium are of similar magnitude as male-female F_{ST} (Fig. 3.5C). However, under strong sexual antagonism – such as that required to generate the empirically observed divergence values – F_{IS} can approach 10%.

Antagonistic loci that do not have a polymorphic equilibrium tend to produce even greater load while sweeping. Unless selection was very weak ($s < 0.05$), load tended to exceeded 10%. Under strong, asymmetric selection load can approach 70% during a sweep. Additionally, the fitness cost of sexual antagonism remains after an allele fixes. The load generated during a sweep is affected by dominance, with additive loci generating loads that are intermediate to the other dominance scenarios. Beneficial dominance within each sex can apparently resolve some of the underlying antagonism by shielding selection on heterozygotes and therefore reducing the load. In contrast, sex-specific deleterious dominance generated the greatest load.

Genome-wide antagonistic selection also produces low divergence

Our analytical results are based on a single-locus model, yet empirical studies report averages across large numbers of loci. To complement the single-locus theory, we quantified the

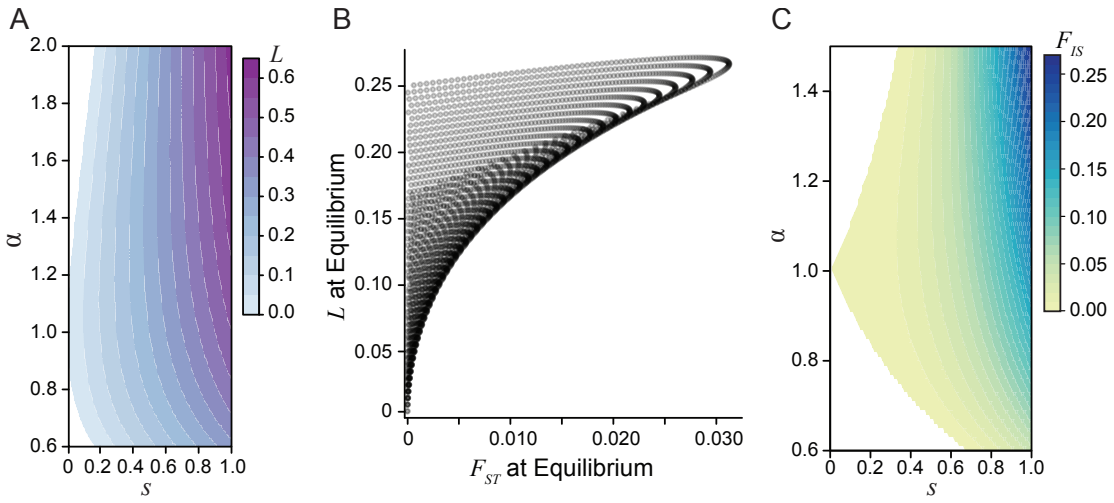


Figure 3.5: The genetic load created by sexually antagonistic selection. A) The genetic load generated at equilibrium for an additive locus across strengths (s) and asymmetries (α) of selection. B) A comparison of male-female divergence and genetic load for a locus at equilibrium across a gradient of selection coefficients with varying asymmetry. The load generated at a locus exceeds the degree of divergence between the sexes. Each curve corresponds to a different fixed strength of selection from $s = 0$ to $s = 0.5$ and each point along the curves corresponds to a different value of α from 0.6 to 2. C) The population inbreeding coefficient F_{IS} for an additive locus ($h_m = h_f = 0.5$) at equilibrium. The excess of heterzygous individuals in the population represents the departures from Hardy-Weinberg equilibrium due to sex-specific selection.

effects of sexually antagonistic selection throughout the genome using individual-based simulations in SLiM Haller and Messer (2019). Simulations in which every new mutation was sexually antagonistic in a population of 10,000 individuals resulted in a mean male-female F_{ST} of 0.00005 and a between-replicate standard deviation of 0.0001, consistent with the single-locus theory (since s was around 0.01). However, entirely neutral simulations (equal mutation rates but no selection) resulted in the same mean and SD of male-female F_{ST} values. Both the sexually antagonistic and neutral simulations averaged around 1,400 SNPs after 1,000 generations of evolution. Although qualitatively similar, the distribution of male-

female F_{ST} values across loci was statistically significantly different between the neutral and sexually antagonistic simulations (Kolmogorov-Smirnov test: $D = 0.11$, $p < 0.001$; Fig. 3.6A). However, this difference in distributions was driven by the larger number of intermediate frequency alleles in the sexually antagonistic simulations. In particular, neutral simulations across all five replicates had only two SNPs with a frequency above 10%, while the sexually antagonistic simulations had over 200 SNPs with a frequency above 10%. Despite there being true differences between the neutral and sexually antagonistic simulations, the male-female divergences observed were still exceptionally low. In fact, neither model had any loci with male-female F_{ST} greater than 0.001 (Fig. 3.6A). Sexually antagonistic simulations had an average 21% decrease in population fitness ($L = 0.21 \pm 0.02$) after 1000 generations of evolution, again consistent with single-locus calculations. Even the minimum load observed under the sexually antagonistic scenario corresponded to a 18% decrease in population fitness.

Sampling variance can generate spurious signals of male-female divergence

Both the single locus model and genome-wide simulations indicate that, while theoretically possible, we would need strong sexually antagonistic viability selection to maintain high divergence between the sexes. Alternatively, the large observed F_{ST} statistics might be due to sampling variance. The empirical studies we cite have relatively small sample sizes (N = 15-200). The male-female F_{ST} values we reported above from simulation were calculated from the *entire* population. To evaluate the effect of sampling, we calculated male-female F_{ST} values from random samples of individuals in our SLiM simulations of two sizes: 100 individuals (50 females and 50 males) and 50 individuals (25 females and 25 males). This subsampling produced dramatically higher male-female F_{ST} values under both the neutral (100 individuals, mean $\pm$ standard deviation across replicates: $F_{ST} = 0.005 \pm 0.004$; 50 individuals: $F_{ST} = 0.01 \pm 0.007$) and sexually antagonistic (100 individuals: $F_{ST} =$

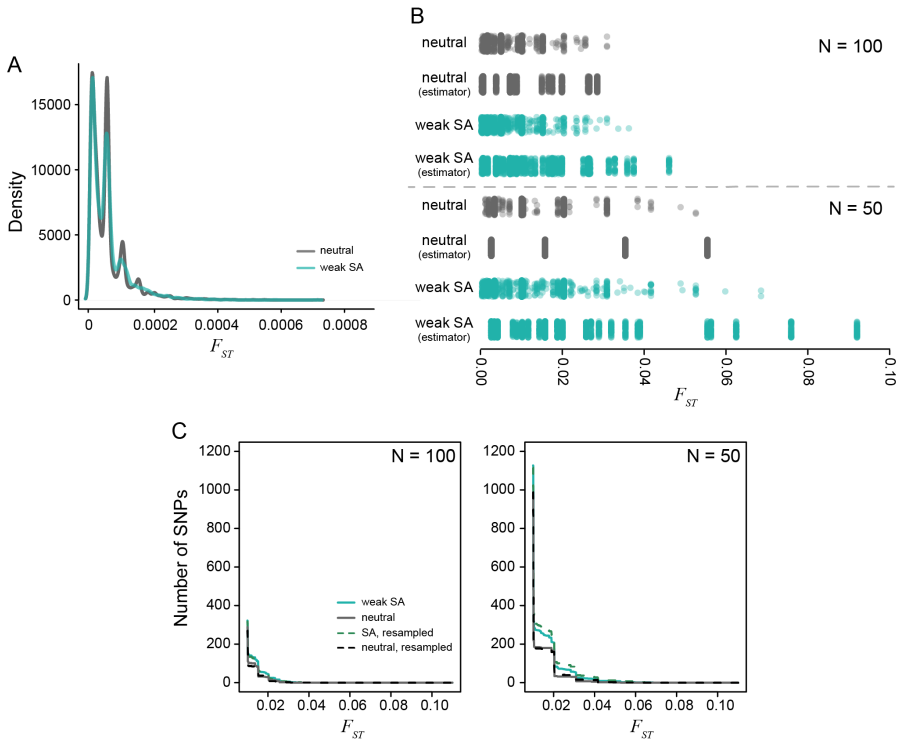


Figure 3.6: The distribution of per locus F_{ST} values generated from simulated populations after 1000 generations of evolution. A) The density of male-female F_{ST} values for a population of 10,000 individuals is centered around F_{ST} = 0.0005. The neutral (gray) and sexually antagonistic (teal) simulations were similar but statistically significantly different. B) The distribution of male-female F_{ST} values when subsampling the full populations to either 100 individuals or 50 individuals with equal sex ratios. Divergence values were calculated using the theoretical derivation (Eq. 4) and Weir and Cockerham's F_{ST} (noted as estimator). The sexually antagonistic simulations are significantly different from the neutral simulations due to an increased sampling variance in the sexually antagonistic scenario. Values less than $2e-4$ were excluded from the plot. C) Cumulative distribution curves of per-locus male-female F_{ST} values, both between random samples from the two sexes (solid lines) and between sets of individuals chosen randomly independently of sex (dotted lines). Male-female F_{ST} distributions differed between neutral (grey) and antagonstic (teal) simulations but were not higher for between-sex comparisons, showing that higher F_{ST} values in the antagonistic simulation was not directly due to selection.

0.005 ± 0.004; 50 individuals: $F_{ST} = 0.01 \pm 0.008$) simulations (Fig. 3.6B). There was a significant difference in the distribution of male-female F_{ST} values between the neutral and sexually antagonistic simulations both when subsampling at 100 individuals (Kolmogorov-Smirnov test: $D = 0.054$, $p < 0.01$) and 50 individuals ($D = 0.096$, $p < 0.001$). However, there was no correlation between the F_{ST} values calculated from the full population and those obtained from samples of either 100 individuals ($r = 0.003$) or the 50 individual subset ($r = -0.012$). This lack of correlation also holds true for the neutral model (100 individuals: $r = 0.034$; 50 individuals: $r = 0.025$).

Empirical studies often use estimators of F_{ST}, such as Weir and Cockerham's F_{ST} or Hudson's F_{ST}, to account for population size and sample allele frequencies (rather than population allele frequencies). Therefore, in addition to using Wright's derivation (Eq. 4, we calculated male-female F_{ST} values in subsampled individuals using Weir and Cockerham's derivation (Weir and Cockerham 1984; Bhatia *et al.* 2013) (which is equivalent to Hudson's F_{ST} under our sampling conditions). We found that using an estimator of F_{ST} did not account for the sampling variance (Fig. 3.6B). Specifically, the tail of the distribution of male-female F_{ST} values was qualitatively between statistics.

Although there were more high male-female F_{ST} sites in samples from the sexually antagonistic simulations (Fig. 3.6B), this did not seem to be a direct result of selection, but rather due to the fact that there are many more intermediate frequency alleles in the sexually antagonistic simulations because of balanced polymorphisms. To test this hypothesis, we calculated F_{ST} between two random samples of size 50 drawn from each simulation *independently* of sex, and also between random samples of size 25. If the enrichment of high between-sex F_{ST} sites in the antagonistic simulations are in fact due to the difference in allele frequency distribution rather than the direct result of selection, then the enrichment should persist even in these samples drawn after randomizing sex. Indeed this enrichment

persists, as shown in Fig. 3.6C. Thus, the higher number of intermediate frequency sites in the sexually antagonistic model creates a higher sampling variance of F_{ST}, as expected based on theory (Jakobsson *et al.* 2013). In particular, the tail of the F_{ST} distributions show many higher values, such that an increase by two orders of magnitude relative to the full population was observed (Fig. 3.6B). These results suggest that separating signals of weak antagonistic selection from sampling noise will be extremely difficult.

DISCUSSION

Sexually antagonistic viability selection creates allelic divergence between the sexes because the proportions of each genotype that die before reproduction differs between the sexes. This between-sex divergence for non-sex-linked elements is created anew each generation because chromosomal segregation re-assorts autosomal associations across the sexes during sexual reproduction. An emerging trend in sexual antagonism research is the use of male-female genomic comparisons to identify sexually antagonistic loci. These recent studies identified hundreds of sexually divergent autosomal loci with mean divergence between the sexes in the range of 2-7% (Lucotte *et al.* 2016; Flanagan and Jones 2017; Wright *et al.* 2018). Taken as reported, these studies suggest the extent and strength of sexually antagonistic selection is far greater than might be anticipated. To assess these claims, we used a population genetic model to determine the magnitude of divergence generated by sexually antagonistic viability selection, the strength of selection required to drive such divergence, and the population fitness costs generated by this process.

Although sexual antagonism has been a topic of particular interest over the last few decades (Arnqvist and Rowe 2005), some of the early investigations of sex-specific selection were largely motivated as part of a general attempt to elucidate all possible means by which the large amounts of segregating polymorphisms observed within natural populations could

be maintained (Lewontin 1974). In this context, Kidwell *et al.* (1977) focused on how strong sex-specific selection ($s > 0.5$) could maintain a polymorphism at an autosomal locus when alleles had opposing effects in the sexes. Our analysis agrees with Kidwell *et al.* (1977), but we find that maintenance of such polymorphisms would create a substantial genetic load. Additionally, with weaker selection, the parameter space allowing a stable polymorphism becomes quite narrow. This refinement of the Kidwell model highlights the necessity for considering biologically relevant conditions – as similarly discussed by Smith and Hoekstra (1980) – particularly when theory is is informing signatures of selection within the genome. An important contribution of this model is the explicit inclusion of transmission, which allows for non-random mate choice – a potentially important underlying component of sexual conflict (Arnqvist and Rowe 2005) – to be considered. Assortative mating can indeed have a large impact on the conditions for the maintenance of polymorphism. Supporting previous simulations (Arnqvist 2011), we found positive assortative mating by fitness maintained polymorphisms. In particular, we show that the combination of asymmetrical selection between the sexes and deleterious dominance conditions expanded the equilibrium space relative to random mating. However, such deleterious sex-specific dominance would likely be selected against, suggesting that the strength of selection is the more relevant parameter in natural populations.

While the maintenance of polymorphism may have been a primary motivation for previous work, a goal of modern genomics is to use specific signals of genomic differentiation to identify the loci underlying sexually antagonistic genetic effects (Mank 2017a; Kasimatis *et al.* 2017). Building on our previous work (Kasimatis *et al.* 2017) allowed us to consider the expected degree of between-sex divergence both when an antagonistic polymorphism is maintained at equilibrium in the population, and when no such stable equilibrium exists, so one of the two alleles sweeps towards fixation to the detriment of one sex. Our

model and accompanying simulations highlight several potential limitations of detecting sex-specific differentiation in empirical studies.

First, detectable quantitative divergence between the sexes requires exceptionally strong sexually antagonistic selection. Previous work indicates that F_{ST} values between populations is of order s^2 (Charlesworth and Charlesworth 2010), however, we find that this underestimates the strength of selection when measuring a male-female F_{ST}. Even a 10% viability cost in each sex resulted in between-sex F_{ST} values of less than 0.001 (Fig. 3.4), a signal that is unlikely to be distinguishable from noise without sampling many thousands of individuals within each sex. Critically, to achieve divergence values greater than 0.03 – such as in humans – would require a 30% to 60% viability cost in each sex under our model. These remarkably high sex-specific mortality rates are, to the best of our knowledge, not observed in nature (see Singh and Punzalan 2018) and would presumably be fairly evident in observations of within-generation population biology. (However, exceptionally high fecundity animals might withstand such high sex-specific mortality (see Williams 1975).) Strong sex-specific gametic selection is more plausible than viability selection on adults, but such high levels of genotype-dependent gamete "mortality" in these organisms still do not seem consistent with empirical observations. Additionally, gametic selection requires an epistatic association between the autosomal locus and the sex determining region, which seems implausible across so many loci.

Second, asymmetry in the strength of selection between the sexes is critical in determining the degree of divergence generated. When the strength of selection is weak and approximately the same between the sexes, polymorphisms may be stably maintained, but between-sex divergence is small. However, there is no *a priori* reason to expect that antagonistic mutations should be perfectly symmetrical in their effects and therefore that polymorphic loci should be stable over time. Alleles with more asymmetric effects will often

sweep, producing larger but transient between-sex divergences, although again only under moderate to strong selection. Here, we found that dominance has little quantitative effect on male-female divergence, particularly when selection is weak. In general, understanding what the distribution of sex-specific effects underlying antagonistic selection looks like will provide important information on the potential for sexually antagonistic loci to contribute to genetic variation and genome evolution.

Third, regardless of the allele-frequency dynamics, the genetic load created by sexually antagonistic selection is substantial (also see Cheng and Kirkpatrick 2016). Even weak selection generates a measurable decrease in population mean fitness. Interestingly, sex-specific beneficial dominance can mitigate load to some extent and potentially provides an opportunity for alleles that modify these dominance relationships to invade. This overdominance-like scenario would be expected to generate a form of cryptic genomic conflict and could potentially lead to the persistence of antagonism. Overall, however, our single locus results indicate that male-female allelic divergence is extremely difficult to generate and that the fitness costs of unresolved antagonism are considerable.

Thus, the theoretical predictions from our single locus model seem at odds with the empirical patterns reported to date. Taken as true measurements of sexually antagonistic selection, the empirical data could be described by two, non-exclusive genomic patterns. Divergent loci could either be stable polymorphisms or could be arising and sweeping to fixation through a constant genomic churn of antagonistic interactions. Our results show that either of these explanations require an exceptionally high genetic load. Again, there is currently no indication that mortality occurs in such a high, sex-specific manner, particularly in some of the vertebrate species that have been examined.

Individual-based simulations with many linked selected loci genome-wide recapitulate the predictions of the single-locus model, finding again that even in this more com-

plex situation, weak selection can only produce very low levels of divergence. Most importantly, however, we found that estimating male-female F_{ST} from samples of the sizes used in the literature (hundreds or less) produced distributions with larger means and longer tails, even in the complete absence of antagonistic selection. Even in simulations with antagonistic selection, any high divergence values were a result of random sampling noise, and did not correlate with the true divergence values or strength of selection. These simulations highlight the sensitivity of F_{ST} statistics to sampling variance, which is a major obstacle for identification of antagonistic loci from sex-specific differentiation. Additionally, using an estimator of F_{ST} (rather than the theoretical parameter) did not account for these sampling effects. Most existing empirical studies have not taken these effects fully into account. Our simulations are not intended to be comprehensive, but demonstrate that sampling variance can be more important than selection itself in driving high estimates of divergence, and highlight the need for proper sampling theory.

At the very least, studies reporting high male-female F_{ST} values should compare these values to empirical distributions found by random permutation of sex labels, as done by Dutoit *et al.* (2018) and correcting for multiple comparisons as done by Cheng and Kirkpatrick (2016). Additionally, connecting significant SNPs to a phenotype such as sex-biased expression (Cheng and Kirkpatrick 2016) should help clarify the action of selection. However, population substructure may remain a concern, since such a permutation test does not account for cryptic correlations with sex. For instance, suppose that the sampled population is composed of a mixture of two diverged subpopulations, and that the sex and admixture coefficients of the sampled individuals are correlated. (The samples in Dutoit *et al.* (2018) were all taken from a single island, so this seems unlikely to explain their results.) Other estimation issues beyond sampling variance may well play a role in the large observed male-female F_{ST} values. Reads from the sex chromosome that are wrongly

aligned to an autosome, particularly in the heterogametic sex, have the potential to generate spurious F_{ST} peaks, an issue that may affect some classes of genes – such as those with sex-biased expression – more than others. Furthermore, existing studies report large numbers of loci with high average F_{ST}. Should we interpret this as evidence of antagonistic selection across many loci simultaneously, or at just one a few loci that affect others through linkage? This is not clear, because each generation's sex-specific selection on a single antagonistic allele will also cause between-sex frequency differences at other loci to the extent they are in linkage disequilibrium with the locus under selection. More work is needed to quantify this effect so that they can be included in analyses of natural populations. Many of the loci currently identified as being caused by sex-specific antagonistic selection seem likely to be spurious signals resulting from poor statistical inference. While we believe sexually antagonistic selection does contribute to genomic evolution, we strongly caution against the use and over-interpretation of male-female F_{ST} statistics until better sampling theory is developed.

BRIDGE

In Chapter III I investigated the claims that sexually antagonistic selection can create significant genomic divergence between the sexes at a locus. Contrary to previous claims, the results show that selection must be unrealistically strong to generate such a pattern and that the identified signals of divergence are more likely stochastic noise due to sampling variance. These results indicate that this bottom-up, genome-scanning method is perhaps not appropriate for identifying sexual conflict loci. An alternative approach is to identify candidate classes of genes based on their function and then analyze their evolutionary history. In the next chapter, I choose the Major Sperm Protein as a critical nematode sperm protein for male reproductive success. Chapter IV investigates the Major Sperm Protein gene family for evidence of sexual selection and a role in post-insemination sexual conflict.

CHAPTER IV

RAPID GENE FAMILY EVOLUTION OF A NEMATODE SPERM PROTEIN DESPITE SEQUENCE HYPER-CONSERVATION

This chapter was published in volume 8 of the journal G3 in 2018. Patrick C. Phillips is a co-author on this publication. Patrick C. Phillips and I developed the ideas. I performed all analyses. Patrick C. Phillips was the principle investigator for the work. I wrote the manuscript.

The full supplementary material for this publication can be found at: https://doi.org/10.1534/g3.117.300281

The citation for this publication is as follows:

Kasimatis, K. R., and P. C. Phillips. 2018. Rapid Gene Family Evolution of a Nematode Sperm Protein Despite Sequence Hyper-conservation. G3 8:353–362.

INTRODUCTION

Post-insemination reproductive tract dynamics are fundamentally important for determining an individual's reproductive success. In animals with internal fertilization, the male ejaculate must interact with the female reproductive tract and ovum as well as potentially needing to outcompete the sperm of other males. Just as pre-insemination processes are shaped by sexual selection, so too are post-insemination interactions. However, the dynamics of the latter case are predominantly driven by molecular

interactions, as opposed to behavioral ones, and therefore the appropriate unit of evolutionary analysis is the molecular evolution of the reproductive proteome (McDonough *et al.* 2016; Wilburn and Swanson 2016). Studies across a wide range of vertebrate and invertebrate taxa have consistently shown that reproductive proteins have an elevated ratio of non-synonymous to synonymous substitutions (dN/dS) relative to non-reproductive proteins (Swanson and Vacquier 2002; Clark *et al.* 2006; Vacquier and Swanson 2011; Mordhorst *et al.* 2015). In fact, sperm-specific, seminal fluid, and egg-specific proteins evolve at astonishingly rapid rates, often the fastest observed within a given genome. Within these reproductive categories, evolutionary rates differ based on sex and functional protein class. Specifically, male reproductive proteins evolve more rapidly than their female counterparts (Drosophila 12 Genomes Consortium *et al.* 2007; Harrison *et al.* 2015) and, within male proteins, seminal fluid proteins show the strongest signals of positive selection (Begun *et al.* 2000; Wagstaff 2005; Drosophila 12 Genomes Consortium *et al.* 2007; Findlay *et al.* 2009; Walters and Harrison 2010; Dean *et al.* 2011). These rapid evolutionary rates in males are often attributed to sexual selection in the form of sperm competition (Dhole and Servedio 2014). However, male reproductive proteins are involved in a variety of roles including sperm motility, antimicrobial response, oxidative protection, sperm capacitation, and immunity modulation in addition to modifying female behavior and physiology (Poiani 2006; Perry and Rowe 2015). Such a diversity of functions suggests that pleiotropic trade-offs may be common and that these signatures of protein evolution may in fact be driven by multiple selective pressures (Poiani 2006; Good *et al.* 2013; Dapper and Wade 2016).

The standard approach to studying reproductive proteins is gene-based: the sequence evolution of a gene of interest is analyzed across multiple species. While this approach provides valuable information, it does not capture the full effects of selection across the levels of genomic organization. In particular, gene families are highly dynamic in their genomic organization, gene copy number, and transcriptional architecture (Demuth and Hahn 2009; Innan and Kondrashov 2010; Schrider and Hahn 2010), creating an additional source of variation upon which selection can act (Perry *et al.* 2007; Xue *et al.* 2008; Conrad *et al.* 2010). For example, positive selection can drive gene family expansion through selection for divergent gene copies or maintain neutrally duplicated genes (Innan and Kondrashov 2010). Therefore, to fully understand the evolutionary history of a gene, both genic and genomic approaches are necessary to capture the multiple levels of genomic organization.

Nematodes are an excellent system for taking a genomic-based approach to reproductive protein evolution and addressing standing questions on the pleiotropic trade-offs influencing their evolution. First, multiple annotated reference genomes exist (Blaxter and Koutsovoulos 2015), which allows gene families to be analyzed for both structure and organization. Additionally, nematodes exhibit variation in life-history (Blaxter and Koutsovoulos 2015), including the presence of multiple mating systems (Felix *et al.* 2014) – gonochoristic and self-fertilizing hermaphroditic – creating variation in the mechanisms influencing mating and sperm dynamics. Finally, nematodes have a unique sperm biology characterized by large, crawling sperm (Justine 2002). The most abundant protein is the major sperm protein (MSP) (Klass and Hirsh 1981; Burke and Ward 1983). This multi-gene family has almost exclusively been described biochemically

(Burke and Ward 1983; Smith and Ward 1998; Haaf *et al.* 1998; Baker *et al.* 2002). Specifically, MSP is a dimeric molecule that polymerizes to form branching filaments, which form the pseudopod of the cell and are used to crawl in a tread-milling fashion (Burke and Ward 1983; Bottino *et al.* 2002; del Castillo-Olivares and Smith 2008). These filaments are structurally similar to actin filaments and, in fact, MSP replaces the function of actin in sperm cells (Nelson *et al.* 1982). In addition to its role in locomotion, studies in *Caenorhabditis elegans* have shown that MSP has pleiotropic effects, namely acting as an oocyte signaling molecule (Miller *et al.* 2001). Despite their central role in fertilization, MSP genes have not been rigorously annotated outside of *C. elegans* nor has the molecular evolution of this gene family been characterized.

Here, using a novel annotation of the large MSP gene family across ten different species, combined with rate-based tests and an analysis of synteny, we show that MSPs display a remarkable combination of nearly complete sequence conservation at the individual sequence level contrasted with extensive lineage-specific evolution of the gene family within species. Thus, nematode MSPs appear to be yet another example of the rapid evolution of reproductive proteins, but in this case, this pattern emerges only when the entire genomic context of the gene family is taken into account.

MATERIALS AND METHODS

MSP gene annotations

The *Caenorhabditis elegans* major sperm protein (MSP) gene family (PRJNA13758) was used as the reference sequence for annotations. The *C. elegans* genome is a high-quality whole-genome assembly (CEGMA: 100% complete, 0% partial; BUSCO 98% complete,

n = 982) (Howe *et al.* 2017) with well-curated annotations (Lee *et al.* 2017) and therefore we are confident using the annotated MSP genes as our query dataset. Thirty-one MSP genes have been identified, predominately using biochemical and molecular genetic techniques (Burke and Ward 1983). Note that the gene sequence for *msp-32* is markedly different from the other *C. elegans* MSP genes in overall length, so we verified the predicted sequence using PCR amplification of the gene from the standard N2 lab reference strain and Sanger sequencing.

MSP genes were annotated in the genomes of nine species: *C. sp. 34* (PRJDB5687), *C. briggsae* (PRJNA10731), *C. remanei* (PRJNA248909), *C. angaria* (PRJNA51225), *Pristionchus pacificus* (PRJNA12644), *Strongyloides stercoralis* (PRJEB528), *Ascaris suum* (PRJNA62057), *Wuchereria bancrofti* (PRJNA275548), and *Trichinella spiralis* (PRJNA257433). Annotations were made using custom blast searches in Geneious v9.1.5 (Kearse *et al.* 2012). Blast searches were conducted using all 31 *C. elegans* MSP gene copies based on nucleotide sequence (Megablast) for *Caenorhabditis* species and amino acid sequence (tblastn) for the other species. Results were hand-curated to ensure accuracy in assignment and predicted gene annotations. Specifically, all blast results were checked to ensure the hit corresponded to a true gene (i.e. contained a start and stop codon) and contained a MSP domain (Tarr and Scott 2005). A total of 121 genes were annotated across the nine species. The predicted gene annotation was edited in 5 genes due to a mis-called start or stop codon or a mis-called intron splice site.

MSP secondary structure was predicted using the Phyre^2 server (Mezulis *et al.* 2015). Structural models and residue mapping were visualized using the PyMOL Molecular Graphics System v1.8 (Schrödinger, LLC).

Evolutionary rate tests

The MSP gene sequences were aligned using ClustalW (Thompson *et al.* 1994). Amino acid divergence of the global sequence alignments was calculated for all pairwise gene combinations within a species. Because the unusual nature of evolution in this gene family precluded orthology assignments across family members, we also calculated the distribution of pairwise divergences relative to the *C. elegans* reference rather than attempting to estimate phylogeny-based measures of the average rate of evolutionary change, such as ω (Yang 2007). Unrooted maximum likelihood phylogenies were constructed in PhyML based on sequence alignments of all genes across all species (Guindon and Gascuel 2003). To corroborate that the MSP genes on Chromosome II form species-specific clades based on chromosome-level clustering, we calculated the approximate likelihood of the inferred topology relative to the next most likely tree without species-specific clades (File S1) (Anisimova and Gascuel 2006). The test was run against five independently-inferred, randomized phylogenies to avoid being caught in a local maximum.

To determine if nucleotide sequence identities were higher within genomic clusters than between clusters, we conducted a permutation analysis of pairwise sequence identity by randomizing the order genes throughout the genome and computing the difference in mean nucleotide sequence identity of the randomly re-assigned clusters using clusters of the same size of those observed within the genome. This allowed us to create a null distribution in which the hypothesis that sequence identity did not depend on genomic location was true (difference between measures equal to zero). This distribution was generated from a total of 10^5 permutations, and the probability of rejecting the null

hypothesis calculated by examining how often the randomized comparisons equaled or exceeded the observed difference among the actual clusters.

Synteny analyses

Synteny of the MSP genes within *Caenorhabditis* was analyzed using species with high-quality whole-genome assemblies: *C. elegans*, *C. sp. 34*, *C. briggsae*, and *C. remanei*. The *C. elegans* MSP genes form three gene clusters: one on Chromosomes II and two on Chromosome IV. Additional genes falling within these clusters that were able to serve as syntenic chromosome anchors were identified using the UCSC Genome Browser (Kent *et al.* 2002). The Chromosome II gene anchors were highly conserved across these species and were located on the chromosome or scaffolds to which MSP genes also mapped (Table S1). The Chromosome IV gene anchors displayed more variation in the location to which they mapped across species and had little to no overlap with the MSP genes annotated in these species (Table S2). Therefore, only the MSP genes that mapped to Chromosome II were included in the synteny analyses and all the other MSP genes were categorized as unique to their given species.

Gene dosage analyses

To determine if gene copy number was correlated with gene dosage, we performed a linear model of copy number versus sperm size within R v3.2.1 (R Core Development Team 2015). Sperm size – given as spermatid diameter – was obtained from estimates provided in the literature: *C. elegans* (Vielle *et al.* 2016), *C. sp. 34* (Woodruff *et al.* 2017), *C. remanei* (Vielle *et al.* 2016), *C. briggsae* (Vielle *et al.* 2016), *C. angaria* (Vielle *et al.* 2016), *P. pacificus* (Rudel *et al.* 2005), *A. suum* (Theriot 1996).

As an additional test of the possible influence of gene dosage and gene-family diversity, gene expression patterns were analyzed for *C. elegans* using median expression within larval stage 4 males, as assembled within WormBase (Lee *et al.* 2017). We fit a linear model to determine if either i) chromosome-level clusters or ii) isopeptide subfamilies predicted MSP expression patterns.

RESULTS

MSP gene family annotation

We annotated MSP genes in nine representative species across Nematoda using the 31 *C. elegans* MSP gene copies as a reference (Fig. 4.1). Species were chosen from four of the five major nematode clades (Blaxter and Koutsovoulos 2015) based on the availability of high quality whole-genome assemblies. We sampled five species from the *Caenorhabditis* genus to capture variation across different mating systems and to provide the context for fine-scale genomic analysis. For each of the species chosen, we blasted each of the *C. elegans* MSP genes against the reference genome. We annotated MSP genes in eight of the nine species. Interestingly, we were unable to annotate any MSP genes in *Trichinella spiralis* (Clade I). The amino acid sequence identity of potential *T. spiralis* orthologs to the *C. elegans* gene family was at most 37.5% identical (T01_10172) with no identifiable MSP domain motifs, so we expanded the blast search to include all the MSP genes annotated in the other eight species. Again, we did not find amino acid sequence identity greater than 39.2% (exon 3 of T01_1333 to *Pristionchus pacificus*). The genus *Trichinella* is reported to have crawling sperm (Justine 2002) and therefore the complete lack of MSP genes seems unlikely. If very few gene copies are

present, as may be the case due to the global decrease in genes in the lineage leading to *T. spiralis* (Markov *et al.* 2014), then the sequence could simply be missing from the genomic information available, despite the high quality of the genome (CEGMA: 96.8% complete and 0.0% partial; BUSCO: 87.4% complete for n = 982). In contrast to the apparent lack of MSP in *T. spiralis*, we identified four MSP genes in *Ascaris suum*, contrary to biochemical based reports of a single gene with two isoforms (King *et al.* 1992).

	Copy N°	Mode CDS Length	Exons	Intron Position	Median Residue Changes: within	to CE
C. elegans ♂♀	31	384 (n=29)	1	–	1 (1, 2)	
C. sp. 34	31	384 (n=31)	2	83/84	0 (0, 2)	4 (3, 5)
C. briggsae ♂♀	18	384 (n=15)	2	83/84	2 (1, 3)	3 (2, 6)
C. remanei	25	384 (n=21)	2[a]	83/84	2 (1, 4)	4 (2, 4)
C. angaria	9	384 (n=7)	1	–	2 (1, 4)	5 (4, 7)
P. pacificus	29	384 (n=24)	3[b]	33/34; 83/84	2 (1, 12)	14 (13, 16)
S. sterocoralis	3	384 (n=3)	1	–	3 (2, 3.5)	13 (12, 14)
A. suum	4	384 (n=3)	2	33/34	9 (4, 9)	21 (20, 26)
W. bancrofti	1	384 (n=1)	2	33/34	0	21 (21, 22)
T. spiralis	0	–	–	–	–	–

[a]Except for CRE_17766 and CRE_09979 which have 1 exon.
[b]Except for PPA_02009 which has 2 exons, PPA_32400 and PPA_25646 which have 4 exons, and PPA_33077 and PPA_20094 which have 5 exons.

Figure 4.1. The evolution of the major sperm protein (MSP) gene family across Nematoda (species tree from Blaxter and Koutsovoulos 2015). For each species, the number of gene copies, coding sequence length (given as the mode), number of exons, amino acid residues between which the intron(s) is located, and sequence divergence estimates are given. Sequence divergence is given as the median number of pairwise amino acid residue changes within MSP genes copies of each species as well as the pairwise divergence between the copies of each species and the 31 *C. elegans* (CE) reference MSP genes. The lower and upper quartiles of the pairwise divergences are given in parentheses. Species the from basal nematode clades have fewer MSP gene copies relative to Clade V species. However, there is a high degree of sequence conservation across all species. The estimated evolutionary divergence time within *Caenorhabditis* is tens of millions of years, while the common ancestor between *C. elegans* and Clade III is estimated to have diverged over a 500 million years ago (Blaxter 2009).

In the nematode genomes with clearly identifiable MSP genes, copy number ranged from 1 to 31 (Fig. 4.1). Gene copy number appears to have dramatically increased in the Clade V nematodes. This copy number increase may be a general pattern across Clade V species (see Markov *et al.* 2014) or could potentially be an artifact of the genomes available. Currently only high quality genomes exist for parasitic species for non-Clade V nematodes, while Clade V genomes all come from free-living species. Parasitism can lead to reductions in genome size (Hunt *et al.* 2016) and, while there is no specific evidence for overall genome reduction in these nematodes, fewer coding genes are annotated in these parasitic species relative to free-living ones (Howe *et al.* 2017). Alternatively, increases in gene copy number are often associated with selection for increasing gene dosage (Ohno 1970). If true, sperm size and MSP gene copy number would be predicted to be positively correlated, as larger cells would require more protein to move (Burke and Ward 1983). In contrast, we did not find a correlation between sperm diameter and gene copy number ($F_{1,5} = 0.80$, $p = 0.41$). Nor was there an apparent trend between mating system (hermaphroditic or gonochoristic) and gene copy number.

Coding sequence length was conserved across the phylum (mode CDS length = 384nt for 134 of 152 gene copies annotated). The number of exons, however, varied between species, though within a species the number of exons and the intron splice site was conserved (except for five genes in *P. pacificus* and two genes in *C. remanei*, Fig. 4.1). A parsimonious model of intron evolution suggests an ancestral gene state of two exons with a single, short intron toward the beginning the gene. In the lineage leading to Clade V there appears to have been a gain of a second intron toward the end of the gene with a secondary loss of the ancestral intron position within the lineage leading to

Caenorhabditis. In *P. pacificus*, MSP genes had both a greater number of exons and more variability in the number of exons than seen in the other species, consistent with previous studies (Rödelsperger *et al.* 2013). Three of the species sampled – *Strongyloides stercoralis*, *C. angaria*, and *C. elegans* – showed independent losses of introns in all gene copies.

The MSP amino acid sequence is hyper-conserved

Given the two very different functions of the MSP during post-insemination dynamics – locomotion and signaling – we expected to see patterns that might reflect evolutionary divergence of protein function. The median amino acid divergence between MSP gene copies within a species was less than 2.5% for all species except *A. suum*, which had a median within species divergence of 7% (Fig. 4.1). These low within-species divergences suggested that the MSP amino acid sequence has been highly conserved within individual lineages. Comparisons of sequence divergence across the phylum revealed that the median pairwise divergence for each species compared to *C. elegans* ranged from 2.3% to 16.5%, with sequence divergence increasing with evolutionary distance. In particular, the maximum median amino acid divergence (16.5%) was seen between *C. elegans* and both Clade III representatives, representing over a billion years of total evolutionary divergence time (Blaxter 2009). This extremely low level of sequence divergence is comparable to known highly conserved, ancient gene families such as actin (Mills *et al.* 2001), histone (Pehrson and Fuji 1998; Malik and Henikoff 2003), and ubiquitin (Sharp and Li 1987; Tan *et al.* 1993). For example, mouse and human actin homologs have 79% to 88% sequence identity (Mills *et al.* 2001). In comparison, the degree of genomic divergence between mouse and human is roughly similar to that between *C. elegans* and

C. briggsae (Kiontke 2005), which have a mean MSP sequence identity of 95%. In order to perform a direct evolutionary rate comparison to determine the extent of MSP sequence conservation, we calculated the amino acid divergence for the actin gene family within *Caenorhabditis*. The median pairwise divergence of actin paralogs across *Caenorhabditis* species ranged from 0.8% to 1.1% (Table S3). These actin divergence values are very comparable to those seen within MSP gene copies of each of the *Caenorhabditis* species (median within species pairwise divergence range: 0% to 1.5%), while divergence among species was slightly higher (median pairwise divergence to *C. elegans* range: 2.3% to 3.9%). Converse to what is seen in actin, MSP sequence conservation appears to be stronger within a species than divergence among species, potentially due to the young age of paralogs or strong within-species constraint. Overall, within *Caenorhabditis*, the MSP genes appear to evolve at a rate similar to actin, making this one of the most highly conserved gene families known.

The low within species amino acid divergence of MSPs in the *Caenorhabditis* species is primarily caused by multiple genes having invariant protein sequences. These protein sequence identities allowed us to group MSP genes into species-specific subfamilies based on isopeptide sequence (Fig. S1; Table S4). Even after grouping redundant sequences, most subfamilies had no more than five amino acids residues that were different from the *C. elegans* reference (Fig. 4.2). Further, the majority of amino acid changes at any given residue occurred in only a single subfamily rather than across all subfamilies of a species (Fig. 4.2; Fig. S1). Three residues in particular (15G, 16T, and 80F) appear less constrained than the rest of the amino acid sequence (Fig. 4.2).

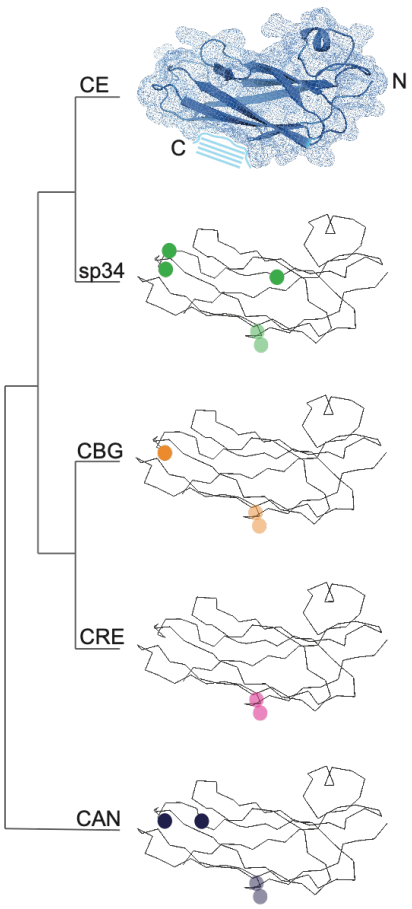


Figure 4.2. The major sperm protein amino acid sequence is highly conserved across *Caenorhabditis*. A space-filling molecule highlighting secondary structure is shown for *C. elegans* (CE). The N-terminus (N), C-terminus (C), and dimer interface (highlighted with stripes) are shown. Ribbon structures are shown for *C. sp. 34* (sp34), *C. briggsae* (CBG), *C. remanei* (CRE), and *C. angaria* (CAN). Circles mark amino acid changes relative to *C. elegans* present in three or more species-specific isopeptide subfamilies (Table S4). Transparency is used for the residues on the back side of the molecule. The residue changes highlighted are consistent across species and do not fall in predicted binding domains. Protein structures were obtained from the Phyre[2] server (Mezulis *et al.* 2015) using the published MSP crystal structure (Baker *et al.* 2002).

These residues are not involved in protein folding or filament formation (Haaf *et al.* 1998; Baker *et al.* 2002; del Castillo-Olivares and Smith 2008), suggesting they likely do not affect locomotion. Additionally, there were very few amino acid changes in the end of the protein sequence (residues 109-127). These residues have been shown to be essential for both filament formation (del Castillo-Olivares and Smith 2008) and stimulating oocyte release (Miller *et al.* 2001), highlighting the strong functional constraint on the amino acid sequence. A noticeable exception to this strong whole protein sequence conservation was seen for four genes each comprising a unique subfamily. These subfamilies had a diverged end located at either the N-terminus (CE-13) or C-terminus (sp34-4, CBG-1, and CBG-2) (Fig. S1). These diverged termini range

from 19 to 113 amino acids and have no predicted secondary structure. Given the high degree of sequence similarity in the rest of the protein, these additional domains are unexpected and may represent MSP proteins with functions outside of locomotion, although their actual function is currently unknown.

Lineage-specific MSP gene family evolution within *Caenorhabditis*

We took advantage of the high MSP copy number within *Caenorhabditis* to explore the evolutionary history of the MSP gene family from a genomic perspective. Due to the high degree of sequence conservation, we could not rely on traditional sequence-based approaches (such as Yang 2007) to infer evolutionary homology. Therefore, we instead took a synteny-based approach coupled with phylogenetic relationships structured by synonymous variation to examine orthology. Specifically, if the MSP gene family was a large, ancestral family, we expected to see: 1) conservation of synteny across species and 2) phylogenetic clustering of orthologous gene copies from each species into monophyletic clades.

Chromosome II was the only genomic location in which *C. elegans*, *C. sp. 34*, *C. briggsae*, and *C. remanei* had overlapping occupancy of MSP genes (Table S1 & S2). *Caenorhabditis angaria* was not included due to incomplete genome assembly in this region. We used a conserved set of 12 genes on Chromosome II, spanning the *C. elegans* Chromosome II MSP gene cluster, to provide a genomic scaffold against which to evaluate the local evolution of MSP genes (Table S1). The gene anchors were conserved and syntenic between *C. elegans* and *C. sp. 34* (Fig. 4.3). The order of the anchors was also conserved in *C. brigssae* and *C. remanei* but in an inverted orientation. Importantly,

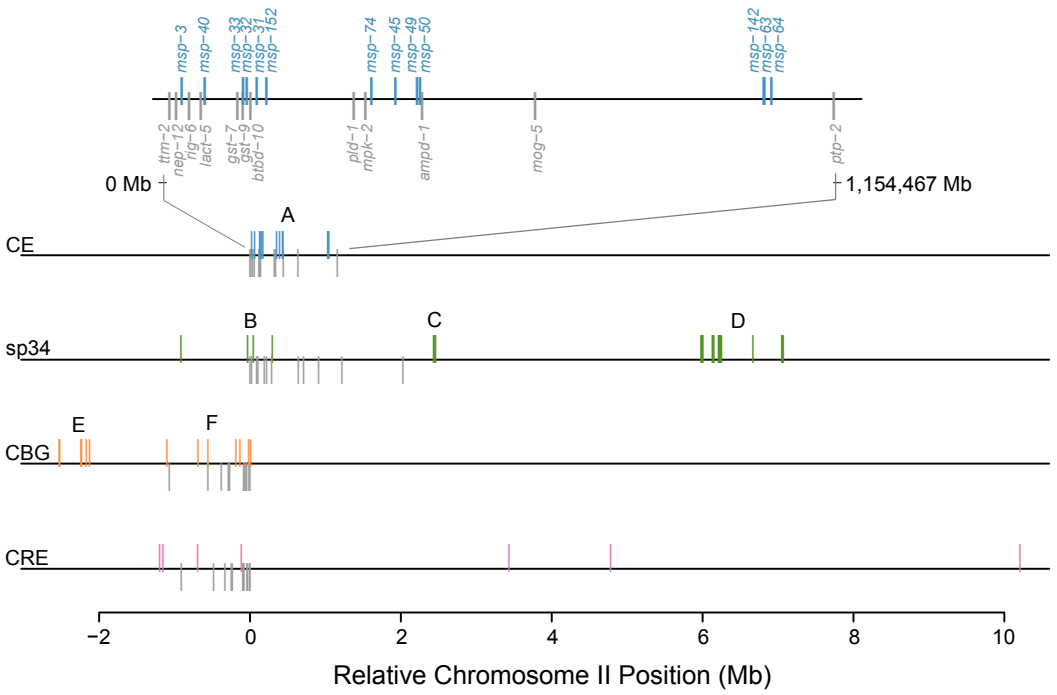


Figure 4.3. MSP genes are not syntenic across *C. elegans* (CE), *C. sp. 34* (sp34), *C. briggsae* (CBG), and *C. remanei* (CRE). The majority of MSP genes map to Chromosome II. The syntenic region is defined around the *C. elegans* gene anchors (shown as a gray down-strike, Table S1). The x-axis is given as relative Chromosome II position, which was defined by setting the first gene anchor (*ttm-2*) as the origin. These anchors are conserved and syntenic across species, although they are in an inverted orientation in *C. briggsae* and *C. remanei*. The MSP genes in *C. sp. 34* (green upstrike), *C. briggsae* (orange upstrike), and *C. remanei* (pink upstrike) do not fall within the gene anchors, but rather form non-syntenic clusters across the chromosome. The MSP gene cluster labels correspond to the phylogenetic clades labeled in Figure 4.4.

the MSP genes form separate gene clusters across the chromosome that are distinct within each species, with little overlap relative to the gene anchors. Additionally, all the species had MSP gene clusters on Chromosome II occupying regions in which MSP genes are completely absent in *C. elegans* and each within-species gene cluster occupied a unique region of Chromosome II.

Despite homology of MSP genes and some overlap of genes with the syntenic Chromosome II anchors, phylogenetic analysis did not show one-to-one MSP orthologs across species. Rather, phylogenetic structuring of Chromosome II MSP genes mirrored the physical grouping of genes, such that monophyletic clades corresponded to each

species-specific MSP chromosome-level gene cluster (log-likelihood of species-specific clades = -3910.59, Fig. 4.4). Indeed, local monophyletic structure within clusters was maintained when all gene copies in the genome were included in the analysis (Fig. S2). Further, phylogenetic analysis of only the MSP clusters overlapping with the Chromosome II syntenic anchors (Fig. 4.4, clusters: CE-A, sp34-B, CBG-F) reinforced species-specific monophyly (data not shown). The strict structuring of predominantly synonymous nucleotide variation within gene clusters is contrary to an expectation of local syntenic identity by descent and lacks concordance with known species relationships. Instead, gene sequence history appears to track genes through cluster-specific gene conversion via non-homologous DNA repair (Innan and Kondrashov 2010). The role of gene conversion appears particularly strong when examining within species pairwise nucleotide sequences identities across the whole genome (Fig. 4.5). As seen in *C. elegans*, *C. briggsae*, and *C. sp. 34*, nucleotide variation – and, in particular, synonymous variation – is more similar within genomic clusters than between clusters ($p < 0.001$). Additionally, within gene clusters, the physical proximity of genes appears correlated with sequence identity – as seen for the *C. elegans* chromosome II cluster (Fig. 4.5A) – further supporting the action of gene conversion. This pattern of unique, non-syntenic gene clustering at both the physical chromosome and evolutionary history levels does not support the expectation of an ancestral, preserved gene family. Rather, such a pattern is reflective of a model in which gene copy variation is generated by lineage-specific duplications, with sequence identity enforced within tandem duplicates by cluster-specific gene conversion.

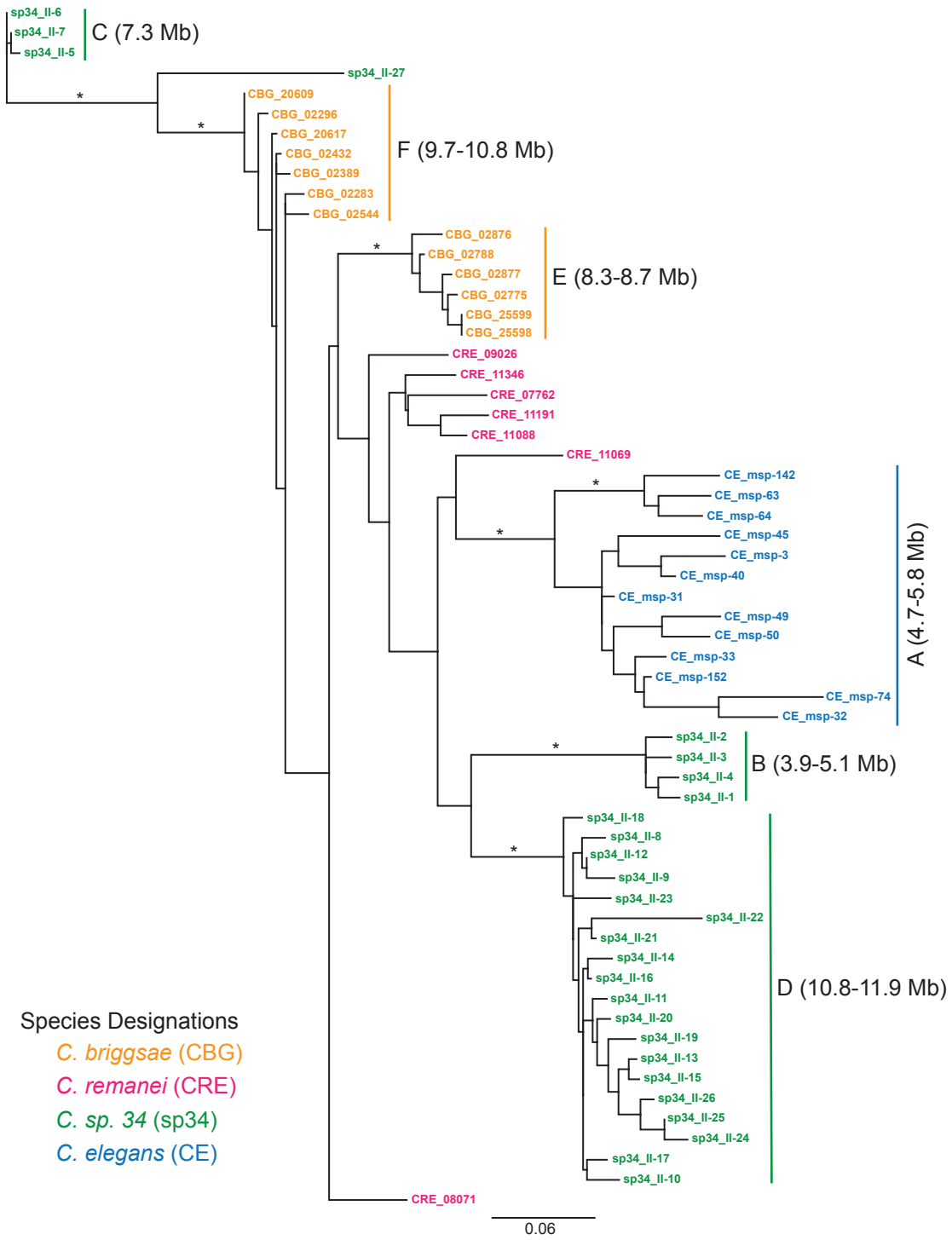


Figure 4.4. Maximum likelihood phylogeny of the MSP genes that map to Chromosome II. The labeled monophyletic clades correspond to distinct chromosome-level MSP gene clusters (shown in Figure 3). Asterisks denote bootstrap values greater than 80%. Approximate likelihood ratio analysis supports this topology as the best representation of the evolutionary relationships (mean log-likelihood = -3910.59 for five independently-inferred phylogenies). The structuring of phylogenetic variation based on gene clusters and complete lack of recapitulation of species relationships suggests the MSP genes are not orthologous.

Patterns of expression do not explain gene-family evolution within *C. elegans*

Within *C. elegans*, we were able to directly assess MSP gene expression and examine the relationship between expression and genomic organization and sequence hyper-conservation. Specifically, using RNA expression data, we examined if chromosome-level clustering or isopeptide subfamily designation were correlated with gene expression patterns. Gene expression differences between chromosome-level clusters were marginally significant ($F_{2,28} = 4.99$, $p = 0.014$) with cluster IV-β having the highest mean expression and IV-α the lowest (mean expression and standard error for Chromosome II: 2,343.4 ± 423 FPKM, Chromosome IV-α: 1,263.3 ± 276 FPKM, Chromosome IV-β: 3,150.5 ± 406). Perhaps more importantly, expression within an individual cluster could range by an order of magnitude in adjacent genes. Gene expression differences among isopeptide subfamilies were also marginally significantly different ($F_{12,18} = 2.36$, $p = 0.048$). Interestingly, *msp-32* – a diverged terminus MSP – had the lowest expression, though again the functional implications require more targeted information.

DISCUSSION

Male reproductive proteins have come to be synonymous with rapid evolution driven by sperm competition and antagonistic male-female coevolution (Swanson and Vacquier 2002; Wilburn and Swanson 2016). Here, combining custom annotation of MSP genes with genic and genomic analyses, we investigated the evolutionary history of the MSP gene family across the phylum Nematoda. The MSP is arguably the most important nematode sperm protein and, given our knowledge of sperm protein evolution in other systems, we expected to see signatures of positive selection. However, MSPs do not

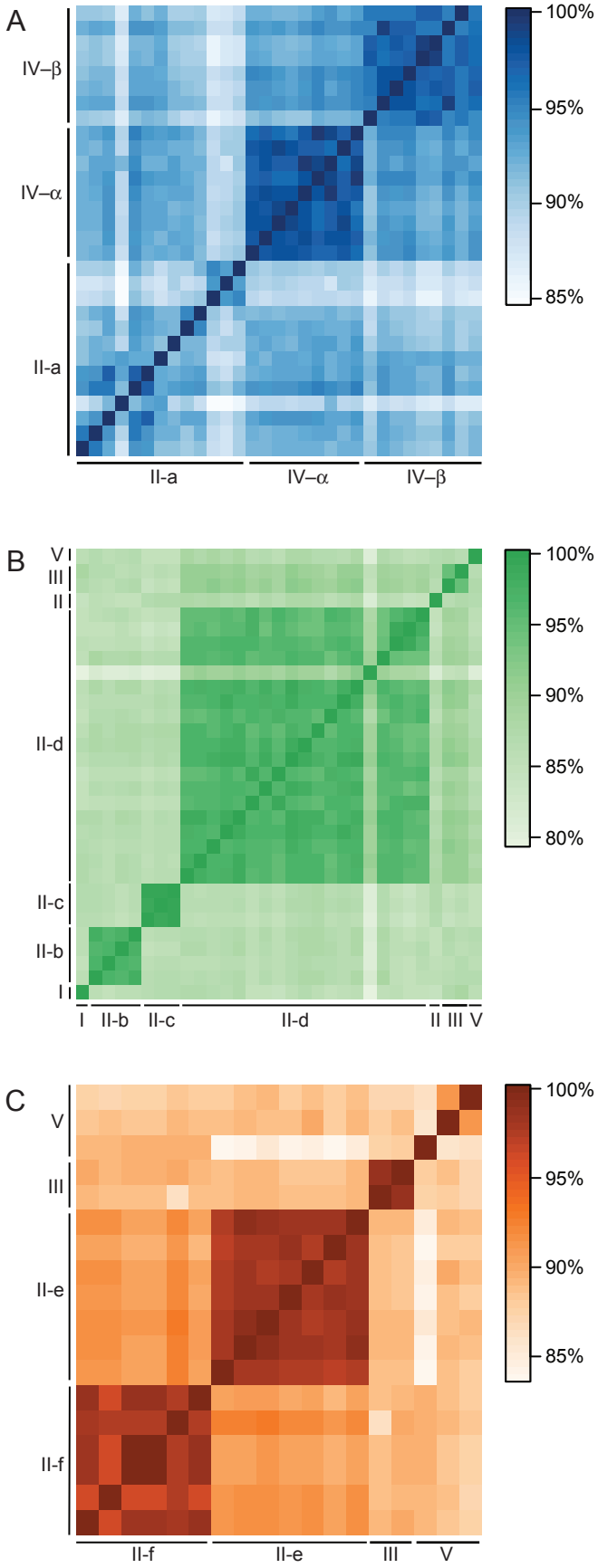


Figure 4.5. Nucleotide sequence identity for A) *C. elegans*, B) *C. sp. 34*, and C) *C. briggsae*. Each square represents the percent sequence identity between a gene pair. The genes are ordered increasing along each chromosome from I to V, as applicable. The majority of MSPs map to chromosome II in all species. These MSPs are labeled based on their chromosome-level cluster (II-a through -f), corresponding to the labels in Figures 3 and 4. Overall, genes are more similar within these chromosome-level clusters than between clusters in all three species ($p < 0.0001$).

conform to this standard expectation. Rather, these genes show a degree of hyper-conservation that is observed in fundamental eukaryotic proteins, such as actin. Specifically, greater than 83.5% amino acid sequence identity has been maintained for the more than 500 million years since these groups shared a common ancestor, making MSP genes some of the most conserved genomic elements yet identified.

The high degree of constraint observed is potentially reflective of the pleiotropic tradeoffs to which MSP genes are subject. Biochemical studies of MSP have identified that much of the protein is important for proper dimerization and filament formation (Haaf *et al.* 1998; Baker *et al.* 2002; del Castillo-Olivares and Smith 2008). Further, nonsynonymous mutations at these interaction sites results in incorrect or loss of filament formation (del Castillo-Olivares and Smith 2008). Such strong functional constraint likely results in equally strong purifying selection, as mutations of this sort could effectively poison a cell through the loss of locomotory function and therefore prevent fertilization from being achieved. Thus, given these structural dependencies and their fundamental role in the most basic attribute of fitness that is fertilization, it is perhaps not surprising that MSPs are highly constrained (albeit at nearly every single amino acid). However, the MSP also acts as an oocyte signaling molecule. Here we would predict to see sexual selection resulting from male-female dynamics drive sequence divergence of gene copies. Four genes had a diverged terminus, possibly reflective of such neofunctionalization, and further functional characterization of these genes is warranted. Nevertheless, within a species MSP copies are essentially identical, suggesting that strong pleiotropic tradeoffs can hinder evolution driven by intersexual interactions.

While well studied in other contexts, gene family dynamics are still underappreciated in reproductive protein studies. We found evidence of extensive MSP gene family evolution within *Caenorhabditis* in the face of the strong pleiotropic constraint on gene sequence variation. Two alternative models can explain the emergence of dynamic gene family variation across a genus. First, a large set of paralogs could be derived from a common ancestor with subsequent differentiation within each lineage. Alternatively, there could be lineage-specific evolution, such that the gene copies arose after branching from a common ancestor and are therefore unique to each lineage. Our data best support a lineage-specific model of gene family evolution, whereby the MSP gene family evolves through independent gene translocations followed by tandem duplication and cluster conservation via gene conversion (Fig. 4.3 & 4.4). Three lines of evidence indicate this model of evolution: synteny analysis, phylogenetic structuring of synonymous variation, and intron evolution. MSP genes form distinct, species-specific clusters across the genome that are highly variable in both the number of genes present and the physical length of chromosome occupied. If clusters of MSP genes were preserved from an ancestral family and subsequently translocated as clusters throughout the genome, we would expect to see proportional spacing of MSP genes through clusters with simultaneous translocation of linked genes. Instead, syntenic analysis provides no evidence of gene hitchhiking within clusters. Rather, these data support independent movement of single genes throughout the genome. Following a translocation event is a pattern of tandem gene duplication, which is supported by the phylogenetic grouping of gene clusters based on synonymous nucleotide variation. Further, there is a lack of recapitulation of known species relationships within the gene trees, again suggesting

independent duplication events. These phylogenetic patterns also suggest strong gene conversion within MSP gene clusters as the mechanism by which sequence identity is maintained. Gene conversion was particularly evident in the extremely high sequence similarity of synonymous variation within genomic clusters, while more variation was measured between clusters (Fig. 4.5). While gene conversion can mask signals of orthology, we do not believe this to be the case. In particular, the patterns of intron loss observed are not consistent with the maintenance of ancestral paralogs, as it is highly unlikely that a conserved family would lose all introns simultaneously across the genome. Rather MSP genes appear to have a highly dynamic nature that is independent within each *Caenorhabditis* species. While this pattern of sequence conservation and gene family evolution is not unique to the MSP family (Perry *et al.* 2007; see Sackton *et al.* 2007; Gao and Zhu 2016; Lee *et al.* 2016), the degree of copy number variation and genomic re-organization seen for the MSP family is more extensive than previously observed.

Lineage-specific duplications have been quantified on a broad-scale across Nematoda and are believed to be related to dosage constraints (Markov *et al.* 2014; Baskaran *et al.* 2015) and life-history transitions (Baskaran *et al.* 2017). However, the mechanism driving this rapid lineage-specific evolution within a single genus is still somewhat unclear. Gene families can be positively selected for diversification of gene copies, which is clearly not the case for the MSP gene family since the amino acid sequence is highly conserved both within and between species (Fig. 4.1). Positive selection can also act to change the transcriptional architecture of a family and thereby effect gene dosage (Innan and Kondrashov 2010). Again, this mechanism does not appear

to drive MSP gene family evolution, as gene copy number is decoupled from sperm size. Transcriptional architecture may, however, play a role through the subfunctionalization of MSP gene expression. In particular, copy number could be correlated with expression level if all genes copies were not equally expressed. In such a scenario, stabilizing selection could act on protein expression level, with gene copy number neutrally evolving. For example, in *Pristionchius* nematodes gene expression, in general, is not correlated with lineage-specific duplication events, suggesting subfunctionalization of copy variants may be common (Baskaran and Rödelsperger 2015). While we annotated multiple MSP genes in each genome, there is currently little to no information outside of *C. elegans* as to whether all gene copies are expressed. While expression data from *C. elegans* show a marginal association chromosome-level clusters, there is a high degree of variance in expression both within and between clusters. Thus, these existing whole-worm, single developmental stage transcriptome data are too limited to draw any strong conclusions. Important future studies should examine if there is differential expression of copies throughout spermatogenesis and sperm activation. Such a quantitative study of the transcription and translation of MSP genes would be valuable, though challenging due to sequence hyper-conservation.

This neutral model of gene copy expansion seems likely to drive chromosome-level cluster expansion. However, it does not particularly explain the translocation of genes throughout the genome. A distinguishing feature of MSPs is their involvement in reproduction and particularly their function as an oocyte signaling molecule. If pleiotropy constrains the MSP sequence from coevolving with its female receptor, then positive selection may act instead on the gene family to counter any female coevolutionary

response. Here gene conversion could act not only to preserve MSP-MSP interactions, but also to transfer any compensatory mutations due to male-female coevolution to other duplicates (Scienski *et al.* 2015). Adaptive evolution has been shown to drive copy number variation in *C. elegans* on short time scales (Farslow *et al.* 2015) and may explain the dynamic movement of MSP genes throughout the genome of individual lineages, though a direct test of this hypothesis would be challenging. Our study highlights the necessity of using whole-genome data when probing the evolutionary history of a gene. Although the pattern of sequence evolution seen for this reproductive protein is unusual, MSP genes are consistent with a broader perspective in which reproductive interactions are capable of driving rapid evolution at the genome- as well as the sequence-level.

BRIDGE

In Chapter IV I demonstrated that the major sperm protein gene family is hyper-conserved at the sequence level while evolving in a rapid, lineage-specific manner at the gene family level. These results are unexpected, perhaps due to the unique sperm biology of nematodes. Additionally, they indicate that multiple levels of genomic organization must be considered to achieve the complete evolutionary history of a gene. Following this conclusion, in Chapter V characterize and describe the evolutionary histories of several newly identified nematode-specific gene families. I proteomically characterize a unique sperm sub-cellular organelle that is critical for nematode male fertility and examine the molecular evolution of its component proteins along with functionally characterizing a newly named gene family.

CHAPTER V

PROTEOMIC AND EVOLUTIONARY ANALYSES OF SPERM ACTIVATION IDENTIFY UNCHARACTERIZED GENES IN *Caenorhabditis* NEMATODES

This chapter was published in volume 19 of the journal BMC Genomics in 2018. Megan J. Moerdyk-Schauwecker, Nadine Timmermeyer, and Patrick C. Phillips are co-authors on this publication. Patrick C. Phillips and I designed the study. Nadine Timmermeyer designed the original version of The Shredder, which was further refined by me. I collected samples for mass spectrometry analysis and analyzed the data. Mass spectrometry was performed by the Genome Sciences Mass Spectrometry Center and the University of Washington. Megan J. Moerdyk-Schauwecker and I created strain PX623. I performed the fecundity analyses. Patrick C. Phillips was the principle investigator for the work. I wrote the manuscript.

The full supplementary material for this publication can be found at: https://doi.org/10.1186/s12864-018-4980-7

The citation for this publication is as follows:

Kasimatis, K. R., M. J. Moerdyk-Schauwecker, N. Timmermeyer, and P. C. Phillips. 2018. Proteomic and evolutionary analyses of sperm activation identify uncharacterized genes in Caenorhabditis nematodes. BMC Genomics 19:593.

BACKGROUND

Despite coming in a wide variety of morphologies, sperm exhibit three key cellular traits that are widely conserved across metazoans (reviewed in Dunbar and O'Rand 1991; Tarín and Cano 2012). First, it appears all sperm undergo a histone-to-protamine chromatin condensation (Eirín-López *et al.* 2005). Second, the vast majority of sperm swim using a flagellum coupled to an actin/myosin cytoskeleton (Morrow 1999). Third, most sperm contain an acrosome or acrosome-like membrane domain that aids in sperm-egg recognition and fusion (Tanphaichitr *et al.* 2015). In contrast to other animals, the phylum Nematoda has a distinctly different sperm morphology and molecular biology (Nelson *et al.* 1982). Namely, nematodes have large, amoeboid-like sperm cells that use non-actin mediated locomotion (Nelson *et al.* 1982). While other species with aflagellate sperm rely on passive diffusion for locomotion (Morrow 1999; Tarín and Cano 2012), nematodes use Major Sperm Protein (MSP)-mediated motility to crawl (Nelson and Ward 1980; Nelson *et al.* 1982). Nematode sperm also lack an acrosome (Nelson *et al.* 1982), and membrane remodeling during spermiogenesis (sperm activation) is instead largely driven by membranous organelles (Ward *et al.* 1983). Both the use of MSP-mediated motility and the presence of membranous organelles are critical components of nematode sperm biology that are unique to and conserved across this ancient phylum.

Perhaps not surprisingly, these two unique components of nematode sperm interact with one another throughout spermatogenesis. Membranous organelles are membrane bound vesicles derived from the Golgi that are found throughout the dividing cell (Ward *et al.* 1983). During spermatogenesis membranous organelles and MSP associate to form fibrous body membranous organelles. As spermatogenesis concludes,

these fibrous body membranous organelles dissociate and the membranous organelles migrate to the cell periphery while the MSP remains distributed throughout the cytoplasm (Fig. 5.1A) (L'Hernault 2006). During spermiogenesis MSP forms branching filaments, which structure the pseudopod of motile sperm (Burke and Ward 1983; Bottino *et al.* 2002). Meanwhile, the membranous organelles remain associated with the cell body, fusing with the cell membrane to create cup-like structures reminiscent of secretory vesicles (Fig. 5.1A) (Nelson and Ward 1980; Ward *et al.* 1983). Unlike an acrosome reaction, however, the membranous organelles fuse prior to any contact with an oocyte. The role of membranous organelles and the function of these fusion events remains unknown, largely because of the challenge of studying subcellular components in single gametes. Nevertheless, mutant screens targeting faulty spermatogenesis have shown that incorrect membranous organelle fusion results in sterility (Achanzar and Ward 1997; Chatterjee *et al.* 2005; Washington and Ward 2006) and therefore that these organelles must play an important functional role within sperm. One hypothesis for membranous organelle function is that the increased membrane surface area and incorporation of additional proteins is important for membrane microdomain remodeling and fluidity (Roberts and Ward 1982; Xu and Sternberg 2003). Since membranous organelles release their contents into the extracellular space, they may have an additional function as a source of seminal fluid proteins and therefore be involved in post-insemination reproductive tract dynamics. However, without information on the composition of membranous organelles, determining the full functional role of their fusion is a challenge.

Here we take a novel approach that co-opts sperm activation events to proteomically characterize membranous organelles within two *Caenorhabditis* species.

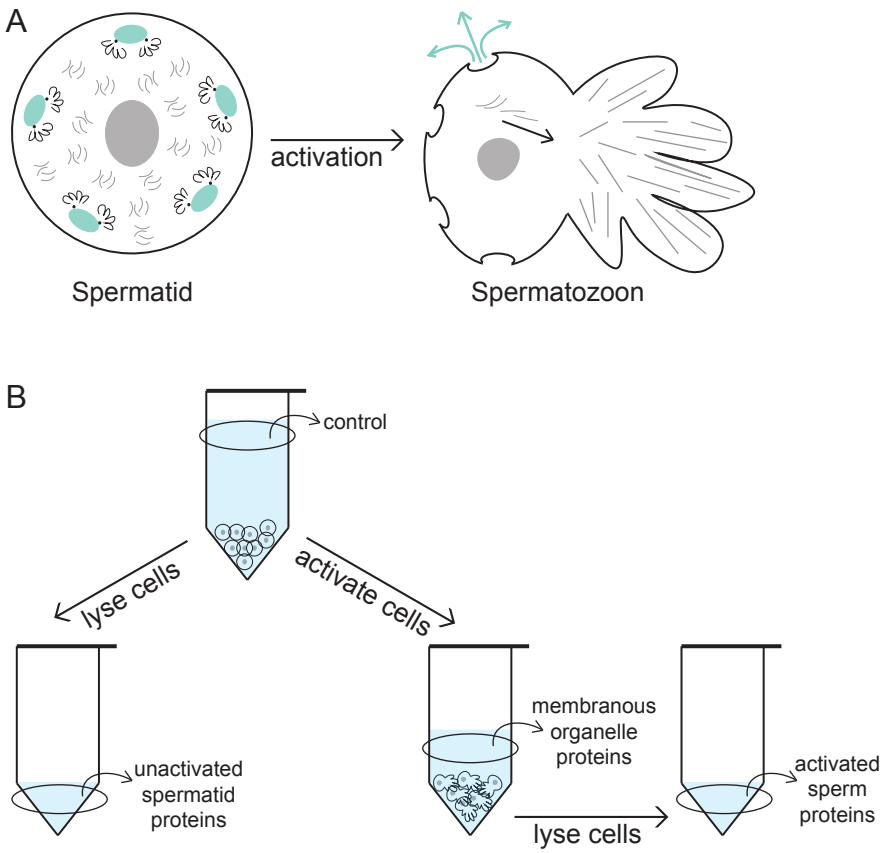


Figure 5.1. Spermiogenesis in nematodes. A) In un-activated spermatids, membranous organelles (shown in teal) migrate to the cell periphery, while Major Sperm Protein (shown in gray) is distributed throughout the cell. Upon sperm activation, Major Sperm Protein forms the pseudopod of the cell and is used to crawl, while the membranous organelles fuse with the cell membrane and release their contents into the extracellular space. B) Diagram of the sperm collection processes. Un-activated spermatid proteins were collected by concentrating spermatids collected using microfluidic dissection (see Fig. 2) and lysed to release proteins. For the activated proteome, un-activated spermatids were first collected using a male crushing technique and then concentrated. The supernatant before sperm activation represents a control for cell lysis. Spermatids were activated *in vitro* by changing the intracellular pH. The supernatant after activation represents the proteins released during membranous organelle fusion. The activated sperm cells were lysed and the membranes pelleted. The supernatant after cell lysis represents the proteins associated with the activated sperm body.

We identify two particularly interesting gene families—the Nematode-Specific Peptide family, group D and Nematode-Specific Peptide family, group F—that are previously undescribed and use evolutionary analysis and genomic knockouts to more directly probe their function.

RESULTS

Proteomic characterization of spermiogenesis in *C. elegans*

Un-activated spermatids were collected from males using a novel microfluidic dissection technique. This male dissection technique utilizes a custom microfluidic device with a fine glass needle to slice through the cuticle and testis of males to release stored spermatids (Fig. 5.2). The un-activated spermatids were lysed to characterize non-membrane-bound sperm proteins (Fig. 5.1B). The un-activated spermatid proteome was dominated by the MSP, confirming that pure sperm cell samples were being collected (Additional File 1). The most abundant proteins, however, were from the Nematode-Specific Peptide family, group D (NSPD), which comprised approximately 50% of the total protein abundance. Since mass spectrometry identified a single peptide motif for these proteins, NSPD abundance was described at the gene family level. The NSPD family is uncharacterized, but has been previously shown to exhibit a pattern of male-enriched expression (Lee *et al.* 2017). Actin proteins were also identified at < 1% abundance, which is comparable to previous biochemical estimates (Nelson *et al.* 1982). While relatively few total protein calls were made, fully one third of the un-activated spermatid proteome is previously uncharacterized in biological function.

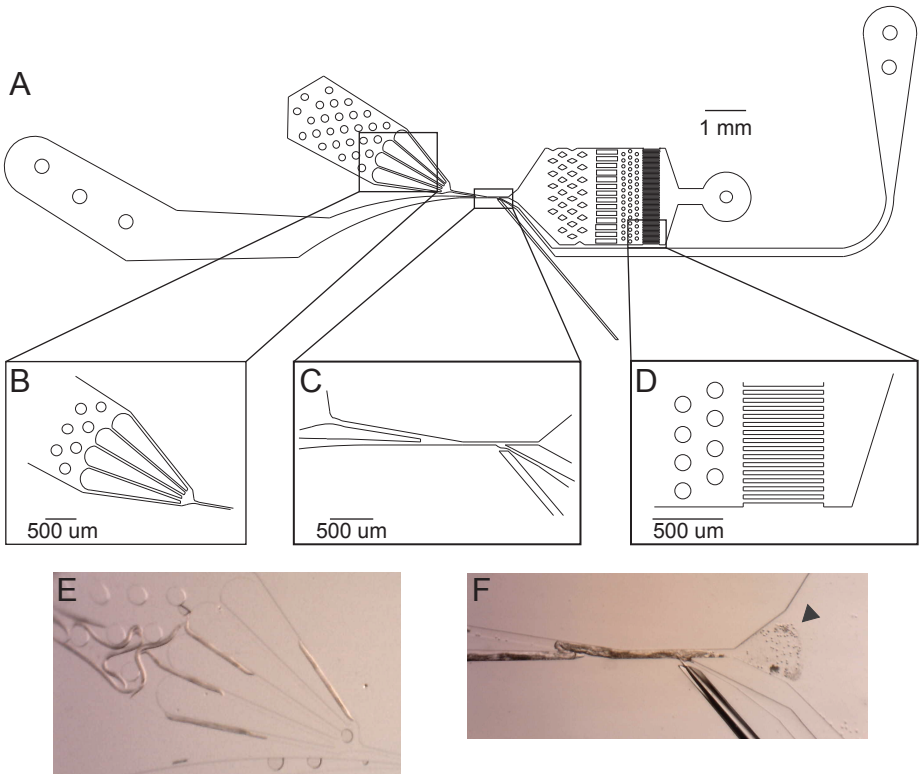


Figure 5.2. Schematic of The Shredder. A) The Shredder is a microfluidic dissection device with a single worm loading arena, a needle insertion, a sperm filtration and collection arena, and two flush channels. B) The male loading arena. The bifurcating design sequentially loads males into the dissection channel. C) The male dissection channel. Males are pushed into the channel from the loading arena and sperm cells are flushed out the right. The needle channel is separated from the male dissection channel by a thin filament of PDMS, which creates a water-tight seal around the needle. D) The sperm filter (10 um) prevents collection of non-sperm components. E) Males in the loading arena for sequential loading into the dissection channel. F) Dissected male and released spermatids (indicated by the triangle) for collection.

To isolate soluble proteins within the membranous organelle from those associated with the sperm body, we took advantage of natural membranous organelle-membrane fusion during sperm activation. Since this analysis required a higher-throughput, un-activated spermatids were collected using a male crushing technique (modified from Klass and Hirsh 1981; Miller 2006). This method squeezes the testis out of males to release spermatids. Spermatids were then activated *in vitro* by changing the intracellular pH (Ward *et al.* 1983) and the proteomes of the membranous organelle secretions and activated sperm fractions were collected via centrifugation (Fig. 5.1B). Again, the MSP was in high abundance, though now identified in both the membranous organelle and activated sperm proteomes (Fig. 5.3). Interestingly, our data reveal three previously unannotated genes (Y59E9AR.7, Y59H11AM.1, and ZK1248.4) as MSPs based on high nucleotide sequence identity and presence of the MSP domain (Kasimatis and Phillips 2018). Overall, 62% of the proteins identified in in the un-activated spermatid proteome were also identified in either the membranous organelle or activated sperm proteome. The lack of one-to-one correspondence between the un-activated proteome and the two activated components is unsurprising given the low total number of proteins identified and the pseudo-quantitative nature of shotgun proteomics. Nevertheless, all the proteins identified were previously found in the un-activated spermatid proteome collected by Ma *et al.* (2014).

The proteins released from the membranous organelle during activation were distinct from those remaining in the activated sperm (Fig. 5.3A). Seventeen proteins were unique to the membranous organelle proteome, including the NSPD family, which comprised 10% of the total membranous organelle protein abundance (Fig. 5.3B). The

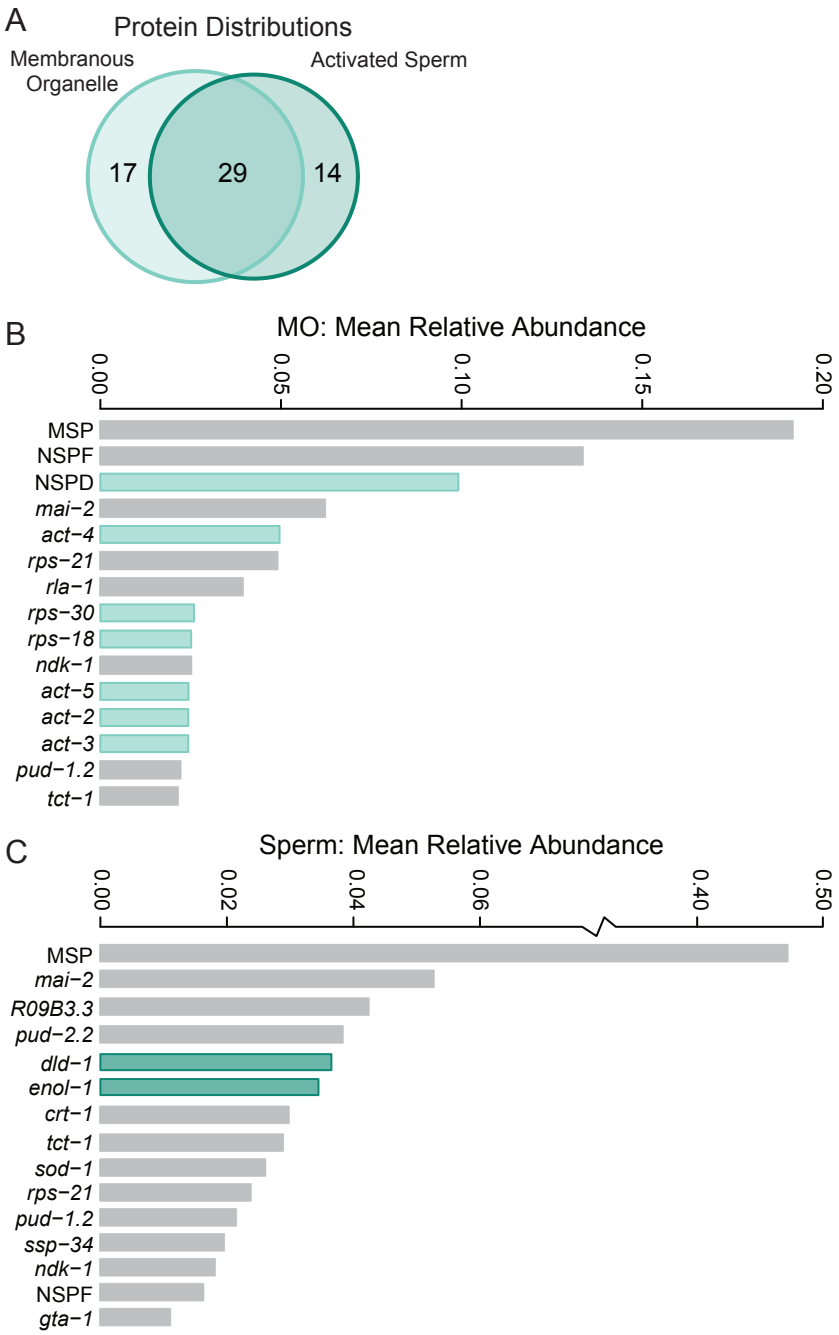


Figure 5.3. Proteomic characterization of the membranous organelle and activated sperm proteomes in *C. elegans*. A) The two proteomes were distinct, with 17 proteins found only in membranous organelles and 14 proteins found only in activated sperm. B) The 15 most abundant proteins identified in the membranous organelles. Proteins unique to membranous organelles (highlighted in teal) include the Nematode-Specific Peptide family, group D (NSPD) as well as several housekeeping gene families. C) The 15 most abundant proteins identified in activated sperm. The proteins unique to activated sperm (highlighted in teal) are predominantly involved in energy production. Protein abundance is shown as the relative mean normalized spectrum abundance frequency.

actin gene family was also unique to the membranous organelle, as were several other housekeeping-related gene families. Within the activated sperm proteome, we identified 14 unique proteins, the majority of which were involved in energy production (Fig. 5.3C). Of noticeable interest were the genes F34D6.7, F34D6.8, and F34D6.9, which again were described using a single abundance measure due to identical mass spectrometry peptide sequence identification. These genes were in fact the most abundant membranous organelle protein after MSP, with a ten-fold greater abundance in membranous organelles than in activated sperm (Fig. 5.3B-C). The F34D6.7, F34D6.8, and F34D6.9 genes in *C. elegans*, display male-specific expression (Lee *et al.* 2017), consistent with our observations. They are organized distinctly from other genes in this region as an array and have a nucleotide sequence similarity of 93.9%. Given their genomic organization, sequence similarity, and co-localization of expression, these genes appear to be a small gene family that originated via tandem duplication. Additionally, an amino acid blast search of these F34D6 sequences in NCBI reveals that they are nematode-specific. Thus, they comprise a newly identified Nematode-Specific Peptide family, which we designate as NSP group F (NSPF).

Proteome composition is largely conserved between species

Spermatids were also collected from the obligate outcrossing nematode *C. remanei.* To compare proteome composition between divergent species, we condensed all protein calls to the gene family level. Within *C. remanei*, we identified 64 gene families in the membranous organelle proteome and 94 gene families within the activated sperm proteome, with 51 families being shared between the proteomes (Additional File 2). Of all the proteins identified, eight did not have an annotated *C. elegans* ortholog. However,

a BLAST search against the *C. elegans* genome indicates that three of these genes (CRE18007, CRE13415, CRE00499) may have unannotated orthologs. Of the remaining unique genes, three appear to be paralogs (CRE12049, CRE30219, CRE30221), suggesting a potential *C. remanei*-specific sperm protein family. A total of 34 gene families were identified in both *C. elegans* and *C. remanei*, capturing the majority of highly abundant genes identified. However, more proteins of low abundance were identified in *C. remanei*. Three gene families – NSPD, Actin, and Ribosomal Proteins, Large subunit – unique to the membranous organelle proteome in *C. elegans* were identified in low abundance within activated sperm in *C. remanei*, potentially because of differential success in activating *C. remanei* sperm *in vitro* (Additional File 2). Two noticeable differences between species were the presence of histone proteins and the absence of NSPF orthologs in *C. remanei*.

Evolutionary analysis of membranous organelle proteins

Proteomic analysis identified NSPD and NSPF proteins as being highly abundant and localized their expression to the membranous organelle. Yet no information exists about the molecular or biological function of these genes. To better understand the nature of these gene families, we analyzed their evolutionary history across the *Elegans* supergroup within *Caenorhabditis*. We made custom annotations of these gene families in 11 species using the annotated *C. elegans* genes (ten NSPD and three NSPF) as the query dataset. Our sampling included the three lineage transitions to self-fertilizing hermaphroditism (Braendle and Felix 2006; Kiontke *et al.* 2011) and the single lineage transition to sperm gigantism (Woodruff *et al.* 2017) found within this supergroup.

Across all 12 species we identified 69 NSPD homologs (Additional File 3). The NSPD gene family ranged from three to ten gene copies, with *C. elegans* having the highest copy number and *C. kamaaina* having the lowest (Fig. 5.4). Coding sequence length was largely conserved between paralogs, but differed across species. Sequence length differences were particularly driven by a 24-30 base pair region in the middle of the gene containing repeating of asparagine and glycine amino acids, which tended to be the same length within a species, but differed across species (Additional File 4).

	Copy #	Location	Mode CDS (bp)	Mean ID (%)	ω[A]
nigoni	7	II, III, V	216 (n = 7)	90.4	
briggsae ♀	4[B]	II, III, IV	219 (n = 3)	86.9	0.23
sinica	7[B]	7 scaffolds	219 (n = 4)	86.0	
remanei	7[B]	II, X, 19, 42	219 (n = 4)	84.6	
latens	5	6, 11, 12, 71	219 (n = 3)	95.2	0.37
sp. 33	4	--	219, 222 (n = 2)	90.6	
wallacei	6	1, 3, 6	219 (n = 3)	95.2	
tropicalis ♀	5	593, 629, 630	219 (n = 5)	91.1	0.07
doughertyi	6	6 scaffolds	219 (n = 3)	89.6	
sp. 34 ☞	5	II, III	204 (n = 3)	81.3	0.09
elegans ♀	10	I, II, III, IV	216 (n = 6)	92.1	
kamaaina	3	596, 1050	213 (n = 2)	95.3	

[A]Calculated using reduced sequence alignments.
[B]Contains one gene with additional C-terminus sequence.

Figure 5.4. The evolution of the Nematode-Specific Peptide family, group D (NSPD) across the *Caenorhabditis Elegans* Supergroup. Listed for each species are: the number of gene copies annotated, the genomic location (Roman numerals represent chromosome level assemblies and numbers represent scaffolds), the mode coding sequence length in base pairs (n = number of gene copies of said length), the mean amino acid sequence identity between paralogs, and the alignment-wide estimate of the ratio of non-synonymous to synonymous substitutions (ω). The complete gene annotation list is provided in Additional File 3 and the sequence alignments are given in Additional File 4.

Despite these species-specific repeats, amino acid sequence identity between paralogs was high, ranging from 81.3-95.3%. No secondary structure was predicted for these genes and in fact they were biochemically categorized as being 73% intrinsically disordered due to low sequence complexity and amino acid composition biases (Dunker *et al.* 2002; Wright and Dyson 2015).

The NSPD genes were broadly distributed across the genome, occurring as single copies on multiple chromosomes or scaffolds in each species (Additional File 3). This seemingly independent arrangement of individual genes throughout the genome precluded a robust syntentic analysis. Additionally, phylogenetic analysis showed NSPD genes predominantly cluster within species and thus they do not convey a strong signal of ancestral gene orthology (Additional File 5). Since orthologous genes could not be assigned, the protein coding sequences were analyzed within the four monophyletic clades represented. Even within these shorter evolutionary timescales, orthologous genes were not readily apparent, again suggesting species-specific evolution at the gene family level. To assess variation in evolutionary rate across the gene family, we estimated a single, alignment-wide ratio of non-synonymous to synonymous substitutions (ω) using reduced sequence alignments. Specifically, we removed the species-specific amino acid repeats in the middle of the gene, which were highly sensitive to alignment parameters. The ω-values varied widely from 0.07 to 0.37 with the more recently derived clades having higher values (Fig. 5.4), although none indicate a strong signal of positive selection. Rather, these genes seem to be weakly constrained outside of the species-specific repeats, which was unexpected given their disordered nature.

We identified and annotated 22 NSPF orthologs in ten species (Additional File 3). Like the NSPD family, the NSPF genes do not have a predicted secondary structure and are 40% intrinsically disordered. They are, however, biochemically predicted to be signaling peptides (mean signal peptide score = 0.9) with a predicted cleavage site between amino acid residues 20 and 21 (Additional File 6). No genes were located within *C. sp. 34* genome (which is very well assembled). Nine species had two gene copies, while *C. doughertyi* has a single copy and, as mentioned, *C. elegans* has three annotated copies. Examination of 249 sequenced *C. elegans* natural isolates (Cook *et al.* 2017) suggests that *nspf-2* arose through a duplication of *nspf-1* as, while all copies of *nspf-1* align to the same position, there is variation in the intergenic space across the isolates. This duplication appears fixed within the *C. elegans* lineage—though one strain (CB4856) has a premature stop codon—and sequence identity is high between duplicates. Additionally, the *C. elegans* NSPF gene family has translocated to Chromosome II while the other species show conserved synteny to Chromosome IV (Fig. 5.5). Using syntenic relationships coupled with gene orientation and phylogenetic clustering, we were able to assign gene orthology within the family (Additional File 7). Within these orthologous groups, species relationships were largely recapitulated with ω-values of 0.53 and 0.26 for the *nspf-1* and *nspf-3* orthologs, respectively. However, when the *C. elegans* lineage was excluded, the ω-values sharply decreased to 0.15 for the *nspf-1* and 0.17 for the *nspf-3* orthologs, indicating a pattern of sequence constraint (Fig. 5.6). We explicitly tested if the *C. elegans* lineage was evolving at a different rate than the other lineages. Indeed, the *nspf-1* (ω = 1.1, C.I. of ω = 0.78 – 1.5, -2Δ*ln* = 5.11) and to a lesser extent the *nspf-3* (ω = 0.57, C.I. of ω = 0.34 – 0.87, -2Δ*ln* = 2.34) *C. elegans* lineages showed some evidence of

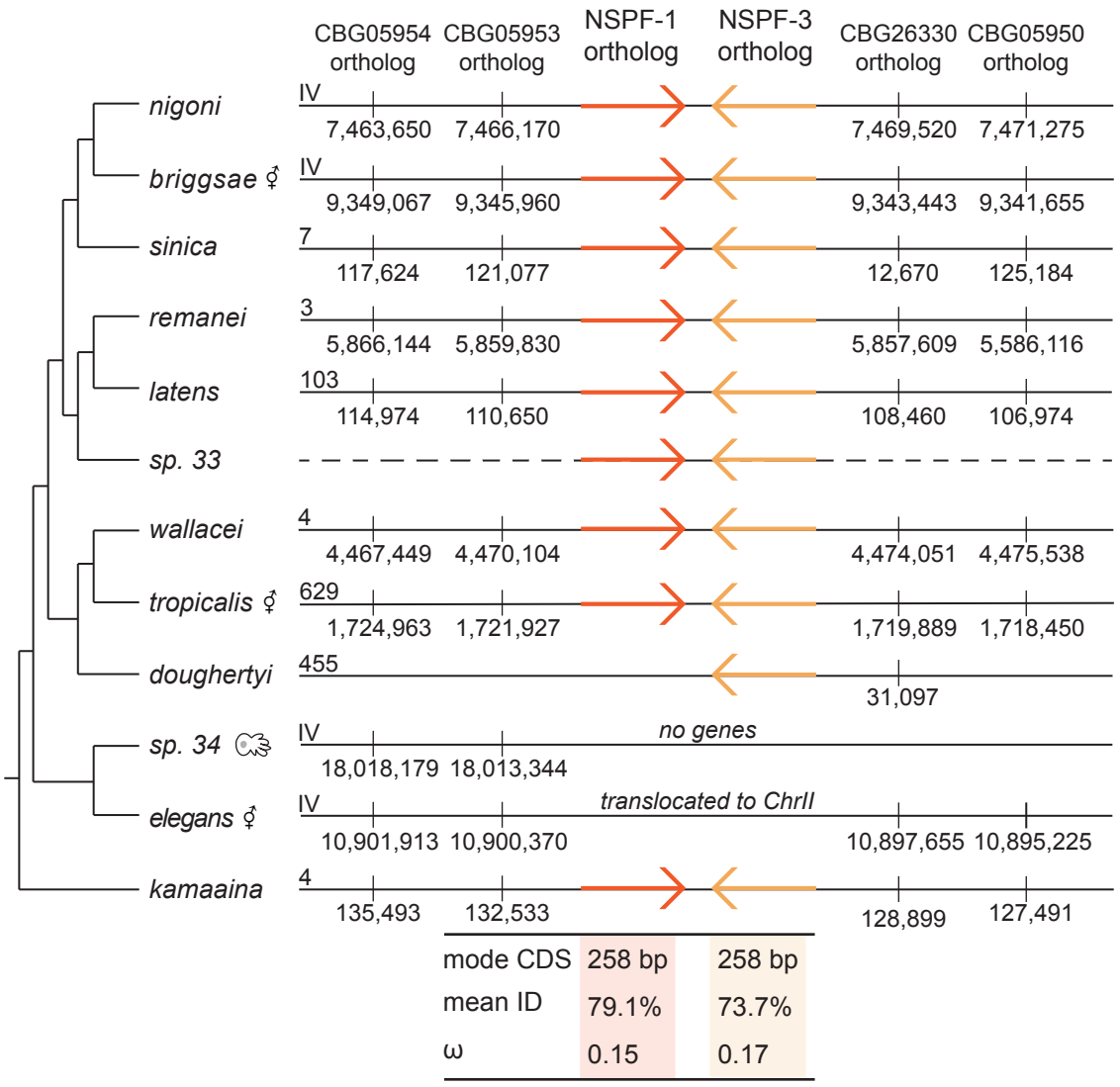


Figure 5.5. The evolution of the Nematode-Specific Peptide family, group F (NSPF) across the *Elegans* Supergroup. The orthologous *nspf-1* and *nspf-3* genes are shown in orange on the chromosome or scaffold to which they locate. The Chromosome IV gene anchors used to determine synteny are shown. For each orthologous group the mode coding sequence length (in base pairs), the mean amino acid sequence identity, and the alignment-wide estimate of the ratio of non-synonymous to synonymous substitutions (ω) are shown. The *C. elegans* orthologs are excluded from the mean identity and ω estimates as they show distinctly different patterns of evolution. The complete gene annotation list is provided in Additional File 3 and the sequences alignments are given in Additional File 6.

positive selection, although the differences in the likelihoods of the two models were not statistical significant.

Functional analysis of the NSPF gene family

Given the high abundance of the NSPF protein, the conserved nature of these genes, and their potential as signaling peptides, we hypothesized these genes could be important for male fertility either during spermatogenesis or in sperm competition. Using CRISPR, we knocked out the three NSPF genes in the *C. elegans* standard laboratory strain (N2) to directly test the function of this gene family. We quantified male reproductive success, by allowing single males to mate with an excess of females over a 24 hour period. Very little difference in progeny production was observed between knockout and wildtype males ($t = -0.81$, $df = 26$, $p = 0.42$; Fig. 5.6A). Given the size of our experiment and the large sampling variance in individual fecundity, we would have been able to detect a difference between backgrounds of 24% with 80% power, so we possibly missed some effects if they were particularly subtle. We also measured the role of these genes in male competitive success, finding again that knocking out these genes had no effect on male fertility (Fig. 5.6B).

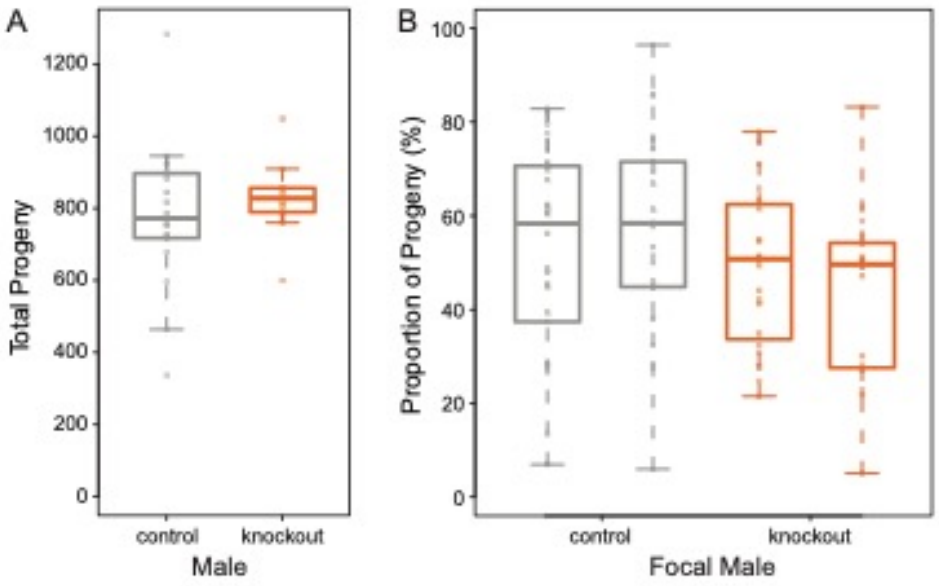


Figure 5.6. Functional assays of the NSPF gene family in *C. elegans* male fertility. A) In a non-competitive sperm setting, knockout males (orange) do not produce significantly fewer progeny than control males (gray) when given an excess of females with which to mate (t = -0.81, df = 26.0, p = 0.42). B) In a competitive sperm setting, knockout males (orange) do not produce significantly fewer progeny than control wildtype males (z = -0.12, p = 0.90) nor do they produce a significant deviation from 50% of the total progeny production (proportions test: $\chi^2 = 1.27$, df = 1, p = 0.26, C.I. of progeny produced = 27.4 – 55.9%). All fecundity data are provided in Additional File 7.

In fact, knockout males were no worse competitors than wildtype males (z = -0.12, p = 0.90) and produced roughly 50% of the progeny measured (proportions test: $\chi^2 = 1.27$, df = 1, p = 0.26, C.I. of progeny produced = 27.4 – 55.9%). Overall, then, despite is prevalence within the sperm membranous organelle, the NSPF gene family does not appear to play an important role in male fertilization success.

DISCUSSION

We used a proteomic approach coupled with molecular evolution analyses and direct functional assays to characterize the composition and role of membranous organelles in nematode sperm. Our approach capitalized upon the natural sperm activation process to

accurately isolate secreted membranous organelle proteins for the first time. This proteome set captures the most abundant proteins found in sperm and shows that the composition of the membranous organelle proteome is seemingly distinct from that of the activated sperm body. Since the complete proteomes were likely not identified, the abundance values presented are relative and therefore direct comparisons across samples is misleading. Nevertheless, interesting and uncharacterized gene families were identified as some of the most abundant proteins sampled. Unsurprisingly, the most abundant protein in activated sperm was the major sperm protein (MSP). Interestingly, MSPs were also the most abundant proteins in the membranous organelle. Since MSP proteins are important not only for motility, but also for oocyte signaling (Miller *et al.* 2001), identifying them as an abundant membranous organelle component implicates membranous organelle fusion as an additional method by which free-floating MSP is added to the seminal fluid (see Kosinski *et al.* 2005). There are 31 annotated MSP gene copies in *C. elegans*, with potentially more uncharacterized copies as seen here, and as of yet we do not know if some of them might be subfunctionally located within different parts of the sperm (Kasimatis and Phillips 2018). We also found that sperm proteome composition was largely conserved between *C. elegans* and *C. remanei*, particularly within the activated sperm itself. This is the first investigation of the proteome of a gonochoristic nematode. Although similarity is the rule, we did identify several *C. remanei* proteins lacking *C. elegans* orthologs, which are potentially a unique sperm family and warrant future molecular characterization, including determining if they are gonochoristic-specific genes.

Two gene families identified in the membranous organelles are particularly notable. First, the NSPD gene family was unique to the membranous organelle. This previously uncharacterized gene family shows high sequence similarity between paralogs and low levels of divergence between species. The high degree of similarity between paralogs is particularly interesting as these genes are not organized as a single cluster and therefore sequence similarity is likely not maintained through non-homologous DNA repair (i.e., gene conversion) (Chen *et al.* 2007). Additionally, NSPDs lack secondary structure and are in fact predicted to be intrinsically disordered. This lack of divergence coupled with little biochemical constraint is unusual and suggests NSPD function requires a specific amino acid sequence along its entire length. However, not all regions of the gene appear to be under the same constraint, as evidenced by the short species-specific repeating motif, although the functional relevance of this motif remains unknown. The pattern of seemingly independent gene copy number expansion and genomic organization despite sequence constraint observed here is strikingly similar to the evolutionary pattern we previously observed in the MSP gene family (Kasimatis and Phillips 2018), and suggests lineage-specific gene family evolution rather than preservation of an ancestral gene family structure.

The newly defined NSPF family showed enriched expression in the membranous organelle, as well as sequence conservation across the clade. While the degree of gene family evolution was far more limited, the duplication of *nspf-2* in *C. elegans* isolates combined with apparent gene losses in *C. sp. 34* and *C. doughertyi* suggest that this family is not completely static. The *C. elegans* lineage, in particular, appears to be evolving differently from the rest of the genus, including changes in copy number and

genomic organization. Despite their predicted signaling function, we found no compelling evidence that these genes are involved in male reproductive success, though a subtle fertility difference could have been swamped out by the high individual variance in fecundity. These null results suggest that this family could be redundant as is supported by apparent species-specific gene losses, although if true we might expect to see greater sequence divergence across the genus due to genetic drift. Alternatively, this family may play a role in female post-mating physiological response or male re-mating behavior and not on male fertility *per se*.

One noticeable difference between these nematode-specific gene families is the lack of a signal peptide in NSPD genes, which is puzzling given that membranous organelles are Golgi-derived vesicles and thus proteins are presumably loaded through ER-Golgi signaling pathways. One possibility is that proteins produced in very high abundance—such as NSPDs and MSPs—could passively leak from the ER into membranous organelles (Alberts *et al.* 2002). Alternatively, transporters on the surface of membranous organelles could actively or passively transport proteins into the vesicle (Beer and Wehman 2017). An entirely different explanation for identifying non-signaling proteins in the secreted proteomes is that activation releases other exosomes similar to the budding MSP vesicles previously shown in fully activated sperm (Kosinski *et al.* 2005). However, such exosomes have not yet been identified during spermiogenesis itself. These questions of packaging warrant future studies tagging the NSPD proteins, though such an endeavor may prove challenging given their high sequence similarity, short size, and disordered nature.

While these data represent a foundation for membranous organelle molecular biology, no clear functional role for the soluble proteins within this subcellular component stands out. Nevertheless, two non-exclusive hypotheses suggest themselves. First, membranous organelles may serve as a contributor to the overall composition of the seminal fluid (although perhaps a minor contributor). The presence of MSP within the organelles supports this hypothesis. Future studies that track where membranous organelle proteins are found after activation—at the female vulva opening, in the spermatheca, or possibly transferred back to the male cloaca—will be valuable in verifying this hypothesis. Alternatively, the membranous organelle could be more important during spermatid stasis and establishing membrane fluidity upon activation (Roberts and Ward 1982). Here, membrane fusion is the more critical functional component, and the release of membranous organelle contents would then represent an incidental "trash dump" as sperm cells move on to the next phase of their life cycle. The presence of actin exclusively in the membranous organelle supports this hypothesis, as activated sperm function is known to be actin-independent. Additionally, the null functional data for the NSPF family support this "trash dump" hypothesis. Both hypotheses warrant continued investigation to further understand the functional role of this unique sperm component.

CONCLUSIONS

Overall, our findings of sequence conservation over such long evolutionary time periods are contrary to observations within many other organisms, where elevated signals of positive selection are detected in seminal fluid proteins (Swanson and Vacquier 2002;

Clark *et al.* 2006; Mordhorst *et al.* 2015). From an evolutionary perspective, then, patterns of evolution in secreted membranous organelle proteins do not match expectations for typical seminal fluid proteins. However, this pattern of sequence conservation coupled with lineage-specific gene family evolution observed here has also been previously identified for the MSP gene family (Kasimatis and Phillips 2018). There thus appears to be a "nematode sperm protein evolution syndrome" in which structural rearrangements and copy number variants are a more prevalent mechanism of genetic evolution than sequence divergence *per se*. Such a pattern could potentially be due to the conserved and unique sperm biology in nematodes, especially the biochemistry of locomotion. These results further support the need for taking a holistic approach when understanding the evolutionary history of genes.

METHODS

Sperm collection

Worm culture and strains: Sperm were collected from *Caenorhabditis elegans* (standard laboratory strain N2 and strain JK574: *fog-2*(q71) V on the N2 background) and *C. remanei* (strain EM464). The *fog-2* mutation blocks *C. elegans* hermaphrodite self-sperm production, resulting in a functionally male-female population, thereby increasing the ease with which males could be collected. All strains were raised on NGM-agar plates seeded with OP50 *Escherichia coli* bacteria and raised at 20ºC (Brenner 1974). Synchronized cultures of larval stage 1 animals were produced through hypochlorite treatment (Kenyon 1988). Males sourced for microfluidic dissection were isolated from females starting as young adults (44 hours post-larval stage 1) for 24 hours

to build up their stored spermatid supply. Males sourced for testis crushing were maintained on mixed sex plates at population densities of approximately 1,000 animals until the second day of adulthood (62 hours post-larval stage 1).

Microfluidic-based sperm collection: The Shredder (final design: v5.0; Additional File 8) was designed using CAD software (Vectorworks 2013 SP5, Nemetschek Vectorworks, Inc) to function as a precise method of dissecting the male testis. The design has a single worm inlet that sequentially pushes males past a glass dissection needle, which slices through the cuticle, punctures the testis, and releases stored spermatids (Fig. 5.2). Two additional liquid channels flush males out of the dissection channel and flush sperm through a filtration system into the sperm outlet. Single layer devices were fabricated from polydimethylsiloxane (PDMS) using soft lithography (Qin *et al.* 2010) and bonded to a glass microscopy slide following exposure to air plasma. Dissection needles were made using a laser micropipette puller (Sutter Instrument P-2000) and inserted into each device following bonding.

A single Shredder could be used once to dissect up to 20 males. Each device was first flushed with 20 mM ammonium bicarbonate (pH 7.8), after which 20 virgin males were loaded into the worm inlet. The collected spermatids were concentrated by centrifugation (500 rcf for 15 minutes) and then lysed in liquid nitrogen. The cell membranes were pelleted, leaving the spermatid proteins in the supernatant for collection. A total of four pooled *C. elegans* replicates (259 males) and five pooled *C. remanei* replicates (265 males) formed the un-activated spermatid proteome for each species.

Testis-crushing sperm collection: To increase the amount of protein collected, particularly the membranous organelle protein contribution, we also used a male crushing technique to collect spermatids (modified from Klass and Hirsh 1981; Miller 2006). Males were raised in mixed sex populations and size separated from females on the second day of adulthood. This developmental time point was optimal for maximizing the difference in diameter between the sexes and minimizing progeny. The sexes were separated using Nitex nylon filters (35 um grid for *C. elegans* and 30 um grid for *C. remanei*) with an average male purity of 91%. The filtration set-up was kept within a sterilized box to reduce external contamination.

Males were pelleted and plated between two 6" x 6", silane-coated (tridecafluoro-1,1,2,2-tetrahydrooctle-1-trichlorosilane) plexiglass squares. The plexiglass was then placed between two 6" x 6" x 1" wooden blocks. A heavy-duty bench vise was used to apply pressure to males, releasing the testis and spermatids. Spermatids were washed off the plexiglass using 20 mM ammonium bicarbonate (pH 5.6) onto a 10 um grid Nitex nylon filter. This filter size was large enough to let spermatids freely pass, but not adult carcasses or eggs. Spermatids were concentrated by centrifugation and the supernatant collected (Fig. 5.1B). Supernatant collected before sperm activation was used to control for proteins released by cell lysis. No protein was measured in the pre-sperm activation supernatant. Spermatids were activated *in vitro* by adding 100 uL of 70mM triethanolamine (TEA) to the pelleted volume (Ward *et al.* 1983) and were left to activate on a chilled block for 15 minutes. Our ability to activate sperm was verified by microscopy. The supernatant was collected to provide the membranous organelle proteome (Fig. 1B). The remaining activated cells were lysed as before and the proteins

were collected as the activated sperm proteome. Six pooled replicates for *C. elegans* (maximum 19,075 males) and four pooled replicates for *C. remanei* (maximum 13,400 males) formed the membranous organelle and activated sperm proteomes for each species.

Proteomic characterization of sperm

Tandem mass spectrometry: The proteomes were prepared and characterized by the Genome Science Mass Spectrometry Center at the University of Washington. Samples were denatured and digested according to standard protocols (Merrihew *et al.* 2008) and then analyzed on a Thermo Velos-Pro mass spectrometer coupled with a Thermo Easy nano-LC. Analytical replicates were run for each sample. MS/MS data were analyzed using the Comet database search algorithm (Eng *et al.* 2013) with either the *C. elegans* (PRJNA13758) or *C. remanei* (PRJNA53967) reference protein database. Peptide q-values and posterior error probabilities were calculated using Percolator (Käll *et al.* 2008). Peptides were assembled into protein identification using ID picker (Zhang *et al.* 2007) with a 1% false discovery rate cutoff.

Proteomic data analysis: Raw MS/MS information for each proteome was processed so as to include the minimum number of proteins that account for the observed peptides (i.e. parsimonious proteins) and filtered to exclude non-nematode proteins. Additionally, we combined isoform calls into a single gene and condensed four classes of genes (MSP family, NSPD family, SAMS family, F34D6 family) to the gene family level because of identical peptide coverage and high overall sequence similarity of paralogs. Overall, then, our final datasets were the most conservative representation of our data. We then calculated the relative normalized spectrum abundance frequency (measured NSAF

divided by the total worm NSAF) for each protein. The two runs were combined by taking the mean relative NSAF of each protein.

Biological functions for each protein were assigned using WormBase when possible (Lee *et al.* 2017). The composition of the membranous organelle and activated sperm proteomes were compared to determine which proteins were shared and which were unique to a given proteome. Since the *C. remanei* genome is not as well functionally annotated, *C. elegans* orthologous gene families were assigned to characterize biological function. Proteome composition between species was compared at the gene family level. All analyses were performed using the R statistical language (R Core Team 2015).

Evolutionary analysis of the membranous organelle

Gene annotations: We used the well-annotated *C. elegans* reference genome (PRJNA13758: CEGMA: 100% complete, 0% partial; BUSCO 98% complete, n = 982) to compile our query dataset for the NSPD and NSPF (genes F34D6.7, F34D6.8, and F34D6.9) gene families. Genes were annotated in 11 species across the *Caenorhabditis Elegans* supergroup: *C. sp. 33* (from J. Wang), *C. sp. 34* (PRJDB5687), *C. briggsae* (PRJNA10731), *C. doughertyi* (PRJEB11002), *C. kamaaina* (QG2077_v1), *C. latens* (PX534_v1), *C. nigoni* (PRJNA384657), *C. remanei* (PRJNA248909), *C. sinica* (PRJNA194557), *C. tropicalis* (PRJNA53597), and *C. wallacei* (from E. Schwarz). Annotations were generated using custom amino acid blast (tblastn) searches in Geneious v10.2.3 (Kearse *et al.* 2012). Blast results were hand-curated for accuracy. In particular, five NSPF sequence motifs found to be conserved between *C. elegans* and *C. briggsae*

were used as markers during annotation. We annotated a total of 59 NSPD genes and 19 NSPF family genes (Additional File 3) in the 11 species.

The *Caenorhabditis* Natural Diversity Resource (Cook *et al.* 2017) was used to probe the duplication and translocation of the NSPF family across the 249 isotypes identified from whole genome sequencing of 429 natural isolates. The NSPF gene region (II: 2,687,625 – 2,690,180) was extracted using SAMTOOLS. Coverage was calculated and those positions with less then 3x coverage were masked. A consensus sequence for each isotype was created. These sequences were aligned using ClustalW (Thompson *et al.* 1994) in Geneious.

Synteny of the NSPF family was analyzed to determine gene orthology. The *C. elegans* NSPF family formed a cluster on Chromosome II, however, the *C. briggsae* NSPF family formed a cluster on Chromosome IV. Therefore, additional genes surrounding both the *C. elegans* and *C. briggsae* clusters were identified using the UCSC Genome Browser (Kearse *et al.* 2012). These genes served as syntenic Chromosome II and IV anchors, respectively, following the approach outlined in Kasimatis and Phillips (2018). The NSPD family was spread across more than half the chromosomes in *C. elegans* and *C. briggsae*, precluding rigorous syntenic analysis.

Secondary structure was predicted using the Phyre^2 server (Mezulis *et al.* 2015). Biochemical predications about protein structure and function were made using the Predictors of Natural Disordered Regions Server (Dunker *et al.* 2002) and the SignalP Server (Petersen *et al.* 2011).

Evolutionary rate tests: The gene sequences for the NSPF and NSPD families were aligned using ClustalW. Amino acid sequence identity was calculated for all pairwise

gene combinations within a species as well as across the clade. Unrooted maximum likelihood phylogenies were constructed in PhyML (Guindon and Gascuel 2003) of orthologous genes for the NSPF family. Since orthology could not be assigned within the NSPD family, phylogenies were constructed based on monophyletic species trios. Alignment-wide estimates of the non-synonymous to synonymous substitution ratio (ω-ratio) were calculated using HyPhy (Pond *et al.* 2005) under a GTR mutation model. Selection within the NSPF family was estimated across the genus for orthologous genes. Additionally, orthologous genes were analyzed using a branch-site framework in the package BS-REL (Kosakovsky Pond *et al.* 2011) within HyPhy to determine if the *C. elegans* branch in particular was evolving differently than the rest of the gene family. The NSPD family was analyzed using reduced alignments of all genes within monophyletic species triplets. Reduced alignments were constructed by removing the species-specific repeating amino acid motifs (~8 residues) in the middle of the gene. Here sequence alignment was highly dependent on the gap/extension penalty, thereby potentially confounding evolutionary inference.

Functional verification of the NSPF gene family

Strain generation by CRISPR/Cas9: Guide sequences were chosen using the CRISPRdirect (Naito *et al.* 2015), MIT CRISPR Design (http://crispr.mit.edu) and Sequence Scan for CRISPR (Xu *et al.* 2015) tools. For deletion of the *nspf-1*, *nspf-2*, and *nspf-3* genes, cr:tracrRNAs (Synthego) targeting the sequences CAGAGCCCATAATTCAAAGACGG and AGATGAGATTCTAATCAGGTAGG were annealed and pre-incubated with Cas9 (PNA Bio) in accordance with the manufacturer protocol. Young adult N2 individuals were injected in the gonad with a final mix

consisting of 1.7 µM of each cr:tracrRNA, 1.65 µg/µl Cas9 and 50 ng/µl of the oligonucleotide repair template (5'-GTAAGAATACAATTTTTCTTTGTGACTTACCGTCTGGTAGGGTGGCAGATCAGTGTTCAGAAGGAAGTGA-3'), along with an additional cr:tracrRNA and oligonucleotide repair template to allow for screening by *dpy-10* co-conversion (see Paix *et al.* 2015). Individuals from broods containing Roller or Dumpy individuals were screened for the deletion by PCR and confirmed by Sanger sequencing. Individuals with confirmed deletions were then crossed to males with the *him-5* mutation (strain CB4088: *him-5*(e1490) on the N2 background). The *him-5* mutation increases the frequency of X chromosome non-disjunction events during meiosis, resulting in roughly 30% male progeny from self-fertilizing hermaphrodites (Hodgkin *et al.* 1979). Five generation of backcrossing were done to purge potential off-target CRISPR affects. The resulting strain, PX623, (fx*Df*1 II; *him-5*(e1490) V) was used for functional analyses of the NSPF genes.

Fertility assays

We assayed the fertility of knockout males in both non-competitive and competitive sperm environments. To assess the overall reproductive success of knockout males, we mated a single knockout male with three wildtype, virgin females (strain JK574) for 24 hours. As a control, wildtype males (strain JK574) were mated to wildtype females following the same male to female ratio. Matings were done on small NGM-agar plates (35 mm diameter) seeded with 10 uL OP50 *E. coli*. After 24 hours, each male was removed and the females were transferred to a new plate to continue laying eggs. Females were transferred to new plates every 24 hours until progeny production ceased.

The total number of progeny was counted as a measure of each male's reproductive success (Additional File 7). To measure competitive ability, individual wildtype, virgin females (strain JK574) were mated with a knockout male and an RFP marked male (strain PX626: fx*Is*2[$P_{hsp\text{-}16.41}$::PEEL-1::*tbb-2* 3' UTR, $P_{rpl\text{-}28}$::mKate2::*unc-54* 3'UTR, $P_{rps\text{-}0}$::HgrR::*unc-54* 3'UTR, I: 2851040]; *fog-2*(q71) V). Again as a control, virgin females were mated to a wildtype male and an RFP marked male. Worms were mated overnight on small NGM-agar plates seeded with 10 uL OP50 *E. coli* and then the males were removed. Progeny were collected over the next 24 hours, counted, and screened for the number of RFP positive progeny. Two independent biological replicates of the competitive assay were performed (Additional File 9).

The fertility data were analyzed using R, with the significance of non-competitive reproductive success evaluated using Welch's Two Sample t-test and an analysis of the power of the comparison computed using the package *pwr* (Champely 2017). Male sperm competitive success was analyzed using a generalized linear model framework with random effects and a Poisson distribution within the package *lme4* (Bates *et al.* 2015). An equality of proportions test was performed for the competitive sperm assay to determine if wildtype and knockout males sired half of the total progeny.

BRIDGE

In Chapter V I defined the sperm sub-cellular membranous organelle proteome and demonstrated that gene family expansion and reorganization are the predominant evolutionary patterns for nematode-specific sperm proteins. Together Chapters IV and V indicate a nematode sperm evolutionary syndrome where sequence evolution is constrained and gene family evolution is dynamic. However, to truly understand both the function and action of selection on sperm and putative seminal fluid proteins in nematodes, a finely tuned method for controlling sperm production and transfer to females is required. In Chapter VI I develop an external, non-toxic sterility induction system for *C. elegans*, with applications for spermatogenesis, mating, and longevity studies.

CHAPTER VI

AUXIN-MEDIATED STERILITY INDUCTION SYSTEM FOR LONGEVITY AND MATING STUDIES IN *Caenorhabditis elegans*

This chapter was published in volume 8 of the journal G3 in 2018. Megan J. Moerdyk-Schauwecker and Patrick C. Phillips are co-authors on this publication. Megan J. Moerdyk-Schauwecker and I designed the technology. Megan J. Moerdyk-Schauwecker and I created strain PX627 and PX629. I collected and analyzed the data. Patrick C. Phillips was the principle investigator for the work. Patrick C. Phillips and I wrote the manuscript.

The full supplementary material for this publication can be found at:
https://doi.org/10.1534/g3.118.200278

The citation for this publication is as follows:
Kasimatis, K. R., M. J. Moerdyk-Schauwecker, and P. C. Phillips. 2018. Auxin-Mediated Sterility Induction System for Longevity and Mating Studies in Caenorhabditis elegans. G3 8:2655–2662.

INTRODUCTION

Sexual reproduction is among the most fundamental of biological processes, and the ability to control the means and mode of sexual reproduction provides a powerful tool for studying a wide variety of important questions. First and foremost, interactions within

and between the sexes are mediated by fertilization. Thus, precise control of fertilization allows for the nature of reproductive interactions to be directly manipulated, addressing questions regarding potentially antagonistic interactions between members of the same sex (e.g., sperm competition) (Karr and Pitnick 1999; Edward *et al.* 2015), between members of the opposite sex (e.g., sexual conflict) (Arnqvist and Rowe 2005), and between parents and offspring (e.g., parent-offspring conflict) (Trivers 1972). In addition, reproduction itself is a critical field of study. For example, investment in offspring is often thought to represent a trade-off with other aspects of an individual's life history, including overall lifespan (Stearns 1989; Schluter *et al.* 1991). Directly manipulating the dynamics of reproduction allows these trade-offs to be specifically assessed. More prosaically, some experiments, such as longevity studies, require the separation of parents and offspring and in some systems this separation is best accomplished by simply not allowing the adults to reproduce in the first place (Park *et al.* 2017).

Currently there are few techniques available to control reproduction short of direct physical manipulation and/or separation of the sexes. In these cases, mostly in model organisms, chemical interventions or genetic mutations can be used to induce sterility in one of the sexes. For example, sterility induction mechanisms—both genetic and chemical—are common in the agricultural industry as a method of preventing cross-pollination (Kempe and Gils 2011). In *Drosophila*, several genetic mutations can be used to generate either female (Schüpbach and Wieschaus 1991; Volpe *et al.* 2001) or male sterility (Castrillon *et al.* 1993). However, since these mutations tend to be recessive, they must be maintained over a balancer chromosome or in a heterozygous population, making them manually intensive to use. Vertebrate models offer many more challenges

to reproductive control, and therefore few sterility induction approaches exist in these systems (see Hsu *et al.* 2009). *Caenorhabditis elegans* is a major model system for genetics, development, neurobiology, and aging. Within *C. elegans*, a limited number of sterility mutants are available (L'Hernault 2006; Nishimura and L'Hernault 2010; Ellis and Stanfield 2014) and can be maintained by mating hermaphrodites to males. In some cases, temperature sensitive sterility mutants (Hirsh and Vanderslice 1976; Ward and Miwa 1978) exist. While these mutants can be an effective tool, by necessity they require a temperature shift, which can affect lifespan (Park *et al.* 2017). Lifespan can also be affected due to the pleiotropic effects of reproductive genes (Murakami and Johnson 1996). An alternative scheme is to prevent progeny production in hermaphrodites using chemical treatments (Mitchell *et al.* 1979), though these techniques are manually intensive and not conducive to high-throughput assays. Further, chemical intervention can potentially generate unaccounted for fitness effects, which not only confound the biological interpretation of results but also make reproducibility a challenge.

An ideal sterility system would be inducible, driven by an external treatment, and, when possible, reversible. To the best of our knowledge such an approach does not exist, even within model organisms. To address this need, we used the non-toxic, non-native auxin inducible degradation (AID) system (Nishimura *et al.* 2009) coupled with knowledge of a critical spermatogenesis gene to create an external sterility induction system in *C. elegans*. We show that this system induces hermaphrodite self-sterility and complete, but reversible sterility of males. This method has broad applications in nematode biology, including studies of aging, gametogenesis, and mating systems evolution.

Constructing an inducible spermatogenesis arrest

The AID system (Nishimura *et al.* 2009; Zhang *et al.* 2015) was chosen as the optimal method for an external sterility induction system in *C. elegans*, as auxin is non-native, non-toxic, and cost-effective. The auxin hormone regulates gene expression in *Arabidopsis thaliana* by activating the F-Box transport inhibitor response 1 (TIR1) protein – the substrate recognition component of a Skp1-Cullin-F-box E3 ubiquitin ligase complex which ubiquitinates degron-tagged proteins for degradation by the proteasome (Tan *et al.* 2007; Nishimura *et al.* 2009). This system has been co-opted as an inducible genetic mechanism in a variety of organisms by degron-tagging a protein of interest and choosing a promoter to drive TIR1 expression in the necessary cell type (Kanke *et al.* 2011; Zhang *et al.* 2015; Trost *et al.* 2016; Natsume *et al.* 2016). We targeted a necessary spermatogenesis gene *spe-44*, causing a spermatogenesis arrest and therefore sterility. Specifically, *spe-44* is one of eleven sperm-specific transcription factors (Reinke 2003) and is predicted to have hundreds of downstream targets, including the critical Major Sperm Protein (Kulkarni *et al.* 2012). Constitutive TIR1 expression was driven using the germline promoter of *pie-1*, which is one of few genes known to have strong sperm expression in hermaphrodites and males (Merritt *et al.* 2008). These three components—auxin, $P_{pie\text{-}1}$*::TIR1*, *spe-44::degron*—generate a fully controllable sterility induction system in *C. elegans.*

MATERIALS AND METHODS

Molecular biology

Guide sequences were chosen using the tools CRISPRdirect (Naito *et al.* 2015), MIT CRISPR Design (http://crispr.mit.edu) and Sequence Scan for CRISPR (Xu *et al.* 2015). For the TIR1 insertion, a guide targeting the sequence GAAATCGCCGACTTGCGAGGAGG near the ttTi4348 MosSCI site was inserted into pDD162 (Dickinson *et al.* 2013) using the Q5 site-directed mutagenesis kit (NEB) to create pMS18. This insertion region was previously shown to be permissive for germline expression (Frøkjær-Jensen *et al.* 2012). The plasmid pMS30 was created by Gibson assembly using the NEBuilder HiFI Kit (NEB) and included: homology arms amplified from N2 genomic DNA, the *pie-1* promoter amplified from pCM1.127 (Addgene #21384) (Merritt *et al.* 2008), the *C. elegans* optimized *AtTIR1::mRuby* fusion and *unc-54* terminator amplified from pLZ31 (Addgene #71720) (Zhang *et al.* 2015), and the self-excising drug selection cassette (SEC) amplified from pDD282 (Addgene #66823) (Dickinson *et al.* 2013). The plasmid backbone was also derived from pDD282. An 11 bp segment of the genomic DNA sequence was omitted from the homology arms to prevent re-cutting. All plasmid assembly junctions were confirmed by Sanger sequencing. Sequencing showed that pMS30 contained a single nucleotide substitution in one of the LoxP sites of the SEC. However, this substitution did not notably impact SEC removal.

The degron::3X-FLAG tag utilized asymmetric homology arms (Richardson *et al.* 2016) for *spe-44* insertion and contained appropriate silent sites to prevent re-cutting. The insert was synthesized as a GeneArt String (ThermoFisher) and amplified by PCR prior to injecting.

Strain generation by CRISPR/Cas9

The *P_{pie-1}::TIR1::mRuby* construct was injecting into the gonad of young adult hermaphrodites (standard laboratory strain N2) using a mixture of 50 ng/µl pMS18, 10 ng/µl pMS30 and 2.5 ng/µl pCFJ421 (Addgene #34876) (Frøkjær-Jensen *et al.* 2012). Screening and removal of the SEC was done following Dickinson *et al.* (2013). Presence of the insertion and removal of the SEC was confirmed by PCR and Sanger sequencing.

To degron tag *spe-44*, a cr:tracrRNA (Synthego) targeting the sequence ATTGAATATGACTAGGTCCTGG near the C-terminus of *spe-44* was annealed and pre-incubated with Cas9 (PNA Bio) in accordance with manufacturer protocol. A mix of 1.7 µM cr:tracrRNA, 1.65 µg/µl Cas9 (PNA Bio), and 80 ng/µl of the PCR repair template, was then injected into the gonad of young adult N2 hermaphrodites containing the *P_{pie-1}::TIR1::mRuby* construct. Included in the injection mix was an additional cr:tracrRNA and oligonucleotide repair template, allowing for screening through *dpy-10* co-conversion (Paix *et al.* 2015). Progeny from broods containing individuals with a Dumpy or Roller phenotype were then screened for the *spe-44::degron* insertion by PCR and confirmed by Sanger sequencing.

Confirmed double mutants were backcrossed 5 times to N2 to create the final strain PX627 (fx*Is*1[*P_{pie-1}::TIR1::mRuby*, I:2851009]; *spe-44*(fx110[*spe-44*::*degron*]). This strain was crossed to strain CB4088, to create the male-rich strain PX629 (fx*Is*1[*P_{pie-1}::TIR1::mRuby*, I:2851009]; *spe-44*(fx110[*spe-44*::*degron*]) IV; *him-5* (e1490) V).

Worm culture and strains

The *C. elegans* strains PX627, PX629, N2, and JK574 (*fog-2*(q71) V) were maintained on NGM-agar plates seeded with OP50 *Escherichia coli* at 20ºC (Brenner 1974). The

fog-2 mutation blocks self-sperm production in hermaphrodites, making them functionally female. Synchronized cultures of larval stage 1 (L1) animals were obtained through hypochloride treatment of gravid adults (Kenyon 1988). To induce sterility, worms were transferred to NGM-agar plates containing 1 mM indole-3-acetic acid (Auxin, Alfa Aesar) following Zhang *et al.* (2015). Zhang *et al.* (2015) showed this auxin concentration to be non-toxic to adults with no larval development defects or fecundity effects. Auxin plates were stored in the dark at 4ºC to prevent compound degradation.

Auxin exposure assays were carried out on small plates (35 mm) seeded with 100 µL *E. coli* and a sample size of 130 hermaphrodites per developmental stage and 100 males per stage. Developmental stages were scored using the known growth rate of animals at 20°C (Byerly *et al.* 1976). Animals were considered fertile if at least one viable progeny was produced. Adult male developmental exposure assays were done by plating synchronized L1 PX629 worms on NGM-agar plates until day 1 of adulthood. Males were then transferred to small auxin plates seeded with 10 µL *E. coli* along with two virgin females (strain JK547). Males were transferred to new virgin females twice a day until no fertilized eggs were seen on plates. Sterility induction was analyzed using a general linear model (GLM) with a binomial distribution in the R statistical language (R Core Team 2015). Male sterility recovery experiments were done by plating synchronized L1 animals on auxin plates and leaving worms on auxin until day 1 or day 2 of adulthood. Males were then transferred to small NGM plates seeded with 10 µL *E. coli* and given three virgin females (strain JK574) with which to mate. Plates were monitored until fertilized eggs appeared. Experiments within a given replicate set were conducted contemporaneously.

The baseline fertility for PX627 hermaphrodites was determined by counting the total number of progeny produced and compared to wild-type hermaphrodites (strain N2). Twenty hermaphrodites of each strain were maintained on small NGM-agar plates seeded with 10 µL *E. coli* until they used all their self-sperm. Self-progeny data were analyzed using a t-test in R. For the hermaphrodite self-sterility mating experiments, synchronized L1 PX627 worms were plated onto auxin plates and removed three hours into adulthood. Virgin females (strain JK574) were used as a control. Individual pseudo-females were mated with two males (strain JK574) overnight on small NGM-agar plates seeded with 10 µL *E. coli*, after which males were removed. All the progeny laid over the subsequent 24 hours were counted. Two independent biological replicates were done with 23 to 35 pseudo-females in each treatment. Mating data were analyzed using a GLM framework with random effects and a Poisson distribution using the *lme4* v.1.13 package (Bates *et al.* 2015) in R.

Lifespan assays

Lifespan data were collected using automated lifespan machines following Stroustrup *et al.* (2013). Briefly, worms were synchronized by letting day 2 adults (strains PX627 and N2) lay eggs over a two hour time period. Auxin self-sterility was achieved by allowing PX627 hermaphrodites to lay directly on auxin plates or by transferring larval stage 4 (L4) progeny to auxin plates. Both self-sterility treatments were transferred to NGM-agar plates on day 1 of adulthood. As a control, egg lays for both PX627 and N2 hermaphrodites were done on NGM-agar plates. At day 1 of adulthood, these animals were transferred to small plates containing 51 µM 5-fluoro-2'-deoxyuridine (FUdR, VCI

America) to inhibit reproduction (Mitchell *et al.* 1979). Control worms were transferred to fresh FUdR plates 24 hours later.

On day 5 of adulthood, all worms were transferred onto medium scanner plates (60 mm) with sealable lids to minimize dehydration. NGM-agar scanner plates contained 40 mM potassium phosphate buffer (pH 6.0), 1mM magnesium sulfate, and 5 mg/mL cholesterol, along with 100 mg/mL nystatin to prevent fungal growth while on the automated lifespan system. Control plates also included 51 μM FUdR. All scanner plates were seeded with 200 μL *E. coli*. A total of 35 to 60 adult hermaphrodites were transferred to each plate with four technical replicates of each treatment. The 16 plates were randomly arranged on a modified Epson v700 scanner in a temperature controlled 20°C room and held in place by rubber mat. Plates were imaged approximately every hour for twenty days across two independent biological replicates.

Images were analyzed using the Worm Browser software developed with the automated lifespan system (Stroustrup *et al.* 2013). This process includes specifying the location of individual plates on the scanner, detecting individual worms, and analyzing worm movement. The resulting data are time of death calls for each individual worm based on the cessation of movement. All plates were hand annotated to ensure that non-worm objects were excluded. Additionally, the time of death calls for the first and last 10% of worms on each plate were checked as these time points are more error prone. The final lifespans were calculated using the egg lay as day zero.

To analyze the influence of our sterilization approach on longevity, we used a mix-model survival analysis as outlined in Lucanic *et al.* (2017). Longevity effects were evaluated using both a mixed-model Cox Proportional Hazard (CPH) model (Therneau *et*

al. 2012) using the *coxme* v.2.2-5 package {Vienna:_Bx9fvNW}, as well as via GLM using the *lme4* package in R. In each case, the coxme and GLM approaches yield equivalent results and so only the *coxme* results are presented as they represent the more comprehensive analytical framework for these data. Using the automated lifespan machine, a small subset of individuals initially placed on a plate are missing and presumed lost over the course of an assay. Such individuals would normally be classified as "censored" in normal survivorship analysis. However, because mortality is determined retrospectively when an individual ceased to move, the moment of loss of such individuals cannot be determined and so they must simply be classified as missing rather than censored at a given time point. For these analyses, the environmental treatment within which each individual was raised (FUdR, auxin) and the genotype of the individual (wild-type or $P_{pie\text{-}1}$*::TIR1::mRuby; spe-44::degron*) were treated as fixed effects, while replicate and plate (nested within replicate) were treated as random effects. Specific *a priori* hypotheses about effects of FUdR and genetic background were tested via contrast coefficients using the *mcp* procedure of the multcomp procedure in R (Hothorn *et al.* 2008).

RESULTS

Self-sterility induction in hermaphrodites

Caenorhabditis elegans hermaphrodites are protandrous, such that they produce several hundred sperm cells during their final larval stage and then switch to oocyte production for the remainder of their lifespan (Hirsh *et al.* 1976). Tagging the *spe-44* gene resulted in a slight reduction in progeny production (~7%) relative to wild-type hermaphrodites,

likely due to problems during spermatogenesis resulting from the degron tag. This trend was most notable on day 2 of adulthood with wild-type hermaphrodites laying significantly more progeny ($t = 3.55$, d.f. = 37, $p < 0.01$; Fig. 6.1). However, while overall life time reproductive success was marginally different between wild-type and *spe-44::degron* hermaphrodites, it was not significantly so (mean ± sd: N2 = 311.7 ± 32, PX627 = 289.3 ± 39, $t = 1.99$, d.f. = 37, $p = 0.054$).

We examined the necessary and sufficient windows of auxin exposure during hermaphrodite development to induce self-sterility. To prevent sperm production, hermaphrodites must be exposed to auxin during their larval development (Fig. 6.2A). When systematically analyzing exposure starting at the L4 stage—the developmental stage during which sperm are produced— through the first 30, 60, or 90 minutes of adulthood, all were hermaphrodites self-sterile (n = 50 per exposure time).

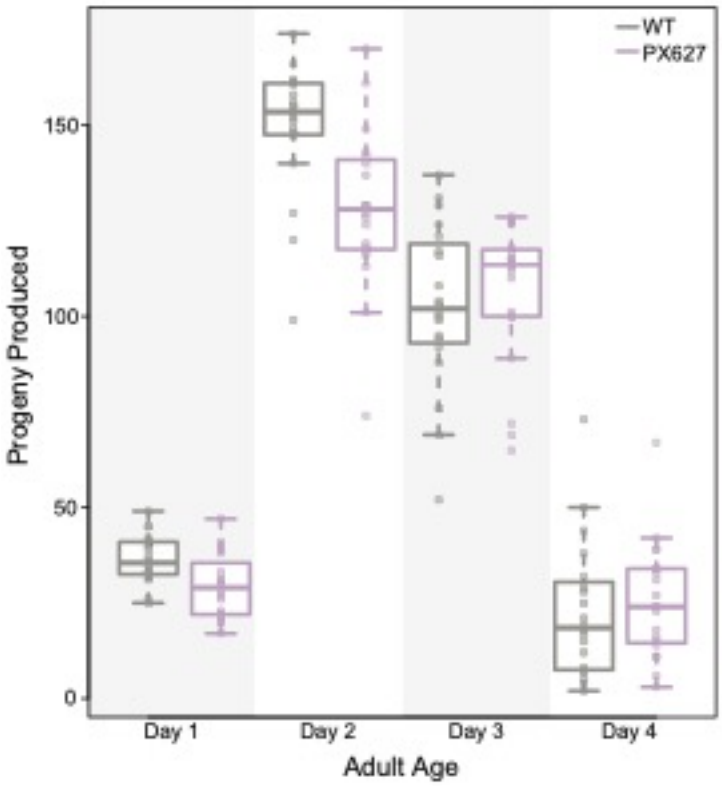


Figure 6.1. Baseline fecundity for *spe-44::degron* hermaphrodites. Progeny production per day is shown for wild-type hermaphrodite (gray) and PX627 hermaphrodite (purple). While overall reproductive success is not significantly different ($t = 1.99$, d.f. = 37, $p = 0.054$), wild-type hermaphrodites laid significantly more progeny on day 2 of adulthood ($t = 3.55$, d.f. = 37, $p < 0.01$) than PX627 hermaphrodites.

In fact, the L4 window alone was both necessary and sufficient to drive self-sterility (n = 50). Self-sterile hermaphrodites continued to lay unfertilized oocytes throughout their adult life, as is characteristic of certain classes of spermatogenesis mutants (L'Hernault 2006). Adult exposure to auxin had no effect on progeny production (Fig. 6.2A).

Despite being self-sterile, hermaphrodite oogenesis was unaffected. In particular, when mated to a male, auxin-treated hermaphrodites had comparable progeny counts to hermaphrodites made functionally female through the *fog-2* mutation (Fig. 6.3). Interestingly, these mated self-sterile hermaphrodites were highly consistent in the number of progeny they produced.

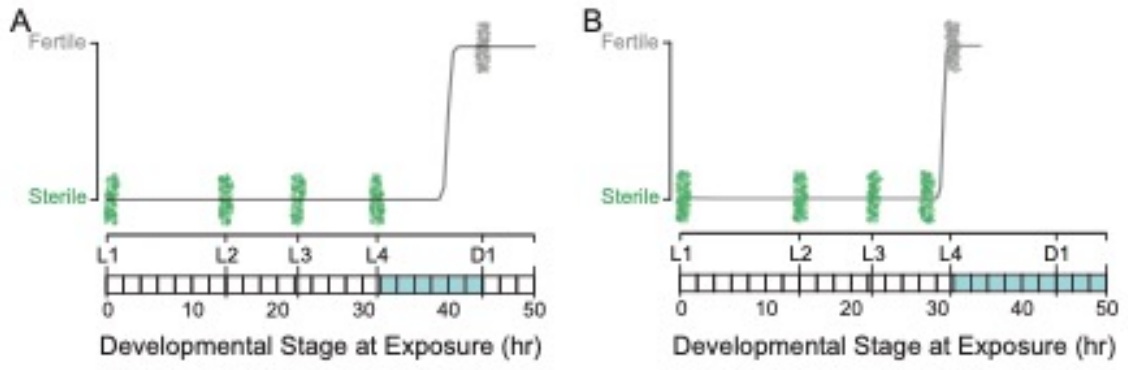


Figure 6.2. Auxin exposure induces hermaphrodite self-sterility and male sterility. A) Hermaphrodites were exposed to auxin starting at each of the four larval stages (L1 – L4) and first day of adulthood (D1). Spermatogenesis (highlighted in blue) occurs during L4 and continues approximately 30 minutes into adulthood (Hirsh *et al.* 1976). Each point represents an individual screened (n = 130 per stage). When exposed to auxin during larval development, all hermaphrodites were self-sterile. However, adults exposed to auxin were fully fertile. The logistic regression is shown in black (residual deviance < 0.0001). B) Males were exposed to auxin starting at each of the four larval stages (L1 – L4). Spermatogenesis (highlighted in blue) begins during L4 and continues throughout adulthood (L'Hernault 2006). Each point represents an individual screened (n = 100 per stage). Males exposed to auxin prior to L4 were sterile, however, males exposed to auxin at L4 produced a low number of progeny. The logistic regression is shown in black (residual deviance < 0.0001).

However, control *fog-2* females laid significantly more progeny than self-sterile hermaphrodites ($z = -2.78$, $p < 0.01$), potentially due to adaptation to obligate outcrossing within the laboratory strain (Teotonio *et al.* 2012; Palopoli *et al.* 2015), although it is equally possible that there are subtle partial spermiogenesis effects at play in the knock-down lines.

Inducible sterility of males is reversible within a single generation

We tested the sterility induction of males using a male-enriched *C. elegans* strain. Like hermaphrodites, males begin spermatogenesis during L4, however they continue producing sperm throughout adulthood, whereas hermaphrodites do not (L'Hernault 2006). We examined the window of auxin exposure during larval male development sufficient to induce sterility. Interestingly, L4 exposure alone was not sufficient to induce complete sterility, as these males still produced a low number of progeny. Rather males had to be exposed to auxin at least 2 hours prior to the L3/L4 molt (Fig. 6.2B). To measure the sterility induction onset at adulthood, males were raised on standard NGM plates and exposed to auxin starting at day 1 of adulthood. Within 24 hours of auxin exposure, no progeny were observed from male-virgin female matings, indicating that males were fully sterile ($n = 44$).

To determine if sterility in males could be reversed following consistent exposure to auxin during larval development, males were transferred from auxin to standard NGM plates at day 1 and day 2 of adulthood. Day 1 adult males began to recover their fertility within approximately 12 hours and all males were fully fertile within 24 hours ($n = 30$ of 30). Day 2 adult males, however, had a much slower recovery period and not all males became fertile ($n = 16$ of 30).

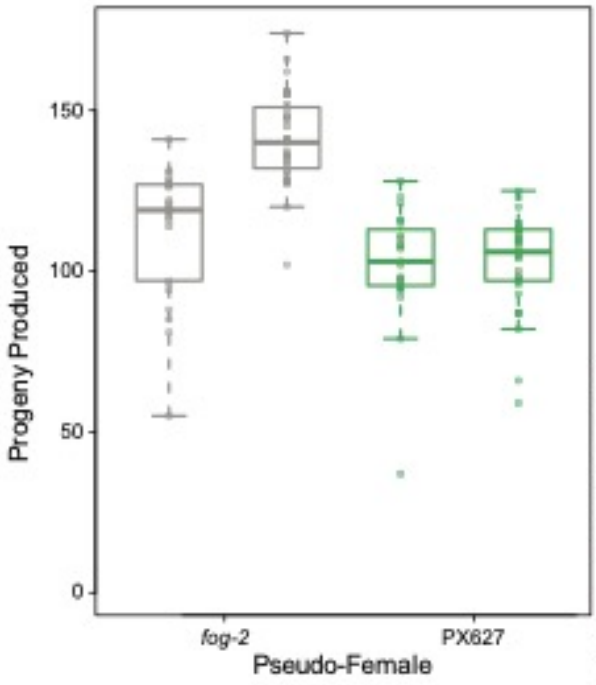


Figure 6.3 Self-sterile PX627 hermaphrodites (green) can recover their fertility when mated with a wild-type male, as compared to *fog-2* functional females (gray). Each bar within a genotypic set represents an independent replicate. While self-sterile hermaphrodites produced fewer progeny than *fog-2* females ($z = -2.78$, $p < 0.01$), their progeny production was invariable across replicates ($t = -0.097$, d.f. = 40.73, $p = 0.92$).

Hermaphrodite self-sterility induction as a tool for aging research

Lifespan assays in *C. elegans* are complicated by the difficult and relatively labor intensive process of separating individuals of an aging cohort from their offspring. A variety of approaches to address this problem are used in the literature, with treatment of adults by the pyrimidine analog FUdR (a chemotherapy agent) being the most widely used. FUdR interferes with DNA synthesis, thereby preventing the production of viable offspring. The sterility induction system developed here allows hermaphrodites to be treated during larval development in order to induce self-sterilization and to then be transferred to whatever media type is necessitated by a given experiment, such as plates treated with bioactive compounds (see Lucanic *et al.* 2017). We examined this potential use by contrasting longevities in wild-type (N2) and $P_{pie\text{-}1}$*::TIR1::mRuby; spe-44::degron* (PX627) adults living on plates containing FUdR with those of PX627 individuals reared on auxin plates either throughout the entire larval development period or during the L4

stage alone before being transferred to standard NGM plates for the remainder of their lives.

The survivorship curves of individuals sterilized via either FUdR or auxin were similar to one another, although they differed slightly in quantitative details (Fig. 6.4). A comparison of adult wild-type and PX627 individuals raised on FUdR in the absence of auxin yielded highly similar survivorship profiles and median lifespans (N2_FUdR = 18.1 days, PX627_FUdR = 17.5 days; CPH contrast: $z = 2.17$, $p = 0.0996$). PX627 individuals sterilized with FUdR also displayed quite similar overall longevity profiles, with individuals raised on auxin for their entire larval periods displaying nearly identical median lifespans to those treated with FUdR (whole larval period PX627_auxin = 17.7 days). However, auxin-exposed worms tended to display a lower rate of mortality late in life, yielding an overall significant difference between these treatments (CPH contrast: $z = 3.04$, $p = 0.0089$). The largest difference in lifespan was observed in PX627 individuals exposed to auxin during only the L4 stage of development, which had both longer median and maximum lifespans than matched FUdR treated individuals (L4 PX627_auxin = 19.2; CPH contrast: $z = 5.76$, $p < 0.0001$). Additionally, replicate trials from individuals treated only during the L4 stage tended to display more error variance (total variance attributable to replicate + plate effects) than the other experimental treatments (12% versus 2-4%, respectively).

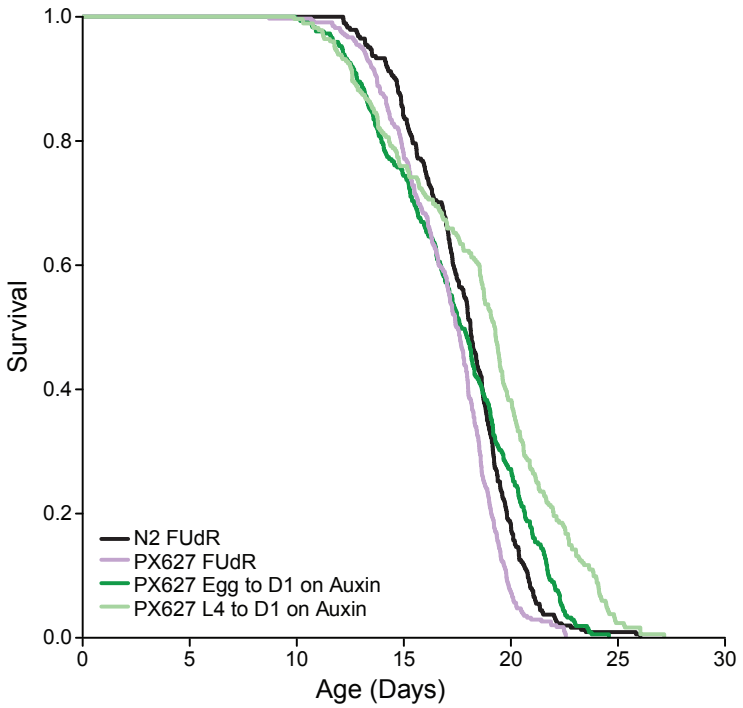


Figure 6.4. Lifespan curves comparing FUdR sterility to auxin-induced self-sterility. Wild-type (N2, black) and (PX627, purple) adults were FUdR treated. PX627 individuals were exposed to auxin from egg to day 1 of adulthood (dark green) or during the L4 stage alone (light green). Each survivorship curve represents six to eight pooled replicates each with over 100 individuals. The survivorship profiles were very similar across treatments and genetic backgrounds, though the PX627 L4 auxin treatment showed a quantitatively distinct profile.

DISCUSSION

Sexual reproduction integrates multiple processes across an organism's life, including the generation of gametes, the act of finding and securing mates, and the production of offspring. Each of these steps has an associated cost (Lehtonen *et al.* 2012). On top of these direct consequences, antagonistic interactions between the sexes during the process of mating as well as conflicts between parents and offspring can further exacerbate reproductive costs. Additionally, the interplay between reproduction and other major life history processes, such as aging and stress response, can add additional fitness trade-offs

(Adler and Bonduriansky 2014). However, precise manipulation and quantification of reproductive trade-offs in an experimental setting has proved challenging.

Sterility induction system

Using the AID system, we designed an external, non-toxic spermatogenesis arrest in *C. elegans*, resulting in hermaphrodite self-sterility and reversible male sterility. Hermaphrodite self-sterility could be induced through auxin exposure during the spermatogenesis developmental window alone. However, auxin exposure throughout larval development also induced complete self-sterility and had no noticeable effects on development (also see Zhang *et al.* 2015). Since this continued larval exposure required very little manual intervention, it is the preferred method for sterility induction, rather than multiple transfers of individuals on and off auxin during late larval development. While tagging the critical spermatogenesis gene *spe-44* gave complete self-sterility, the degron tag itself seems to create some inherent sperm loss. This observation is perhaps unsurprising given the wide-ranging role of this transcription factor. Specifically, while wild-type progeny production peaks on day 2, tagged hermaphrodites had more consistent progeny production over days 2 and 3 again likely due an overall decrease in the number of sperm produced. Despite the degron tag effect, overall progeny production was not significantly different from wild-type hermaphrodites in the absence of auxin and self-sterile hermaphrodites recovered their full fertility when mated with a wild-type male.

In males, auxin exposure during larval spermatogenesis initiation alone was not sufficient to induce sterility, but rather had to occur within the L3 stage. This earlier exposure window corresponds to the earlier expression profile of *spe-44* relative to other

spermatogenesis genes observed by Kulkarni *et al.* (2012). However, males could be fully sterilized using auxin exposure throughout larval development or early adult development. Moreover, males could recovery their fertility, though in an age-dependent manner. The inability of all day 2 males to recover their fertility could be due to a decrease in the transcription level of *spe-44* over time or reduced mating behavior.

Our sterility induction system has a broad range of applications within the fields of spermatogenesis, sperm competition, and mating systems evolution. The temporal control over male sperm production allows for an increased understanding of sperm dynamics, including the rate at which sperm are produced and the amount of sperm stored. Additionally, this temporal control could be co-opted for precise studies of sperm competitive behavior under multiple mating scenarios. A particularly interesting application of this system is the study of mating systems evolution. For example, a genetically identical population could be simultaneously evolved under hermaphroditic and obligate male-female mating regimes. Alternatively, populations could be evolved to switch between mating regimes to better understand the genomic implications of these transitions.

A new approach for aging studies

C. elegans is one of the premiere model systems for studying the biology of aging. The first life-extending mutations were discovered in *C. elegans* (Friedman and Johnson 1988; Kenyon *et al.* 1993) and since then this system has been used in hundreds of studies to investigate a wide variety of questions in aging research (reviewed in Park *et al.* 2017). In particular, a number of studies have shown that the reproductive state of an individual, especially those controlled by germline-soma signaling systems, can have

important consequences for longevity (Shi and Murphy 2014; Angeles-Albores *et al.* 2017). From a practical standpoint, reproduction can greatly complicate longevity assays in nematodes. Since the age at first reproduction is much shorter than median lifespan, there is the potential for several generations to be living on a plate at the same time, even if one starts with an initial age synchronized cohort. In most longevity studies this problem is solved either by manually removing ("picking") adults to fresh media every day, which is very labor intensive and prone to error, or using a chemical means to sterilize reproductive adults. The most common sterilization technique involves the use of FUdR, which disrupts DNA replication in proliferating tissues such as the germline. Actively poisoning a subject while trying to accurately track their health and lifespan is obviously less than ideal.

The sterility induction system developed here provides an ideal alternative to existing chemical sterilization approaches in *C. elegans* and other nematodes. First, worms only need to be exposed to auxin during their larval development and can then be transferred to regular media as adults. This exposure window cuts down on expense as well as the need to constantly replenish an environmental toxin throughout adult life. Additionally, unlike FUdR treatment, auxin sterilization has no impact on oogenesis—a major cellular process throughout hermaphrodite adulthood. Further, there is no concern about potential interactions between the sterilization agent and other external treatments such as food quality or chemical interventions (Lucanic *et al.* 2017). Overall, we find that longevity trajectories of self-sterilized individuals are very similar to individuals raised on FUdR (Fig. 6.4). The only substantive difference that we observed was in individuals that had only been exposed to auxin during the L4 stage that immediately precedes sexual

maturity. These individuals lived longer and displayed more variable outcomes than those that were exposed to auxin throughout their entire larval period. Potentially, even those these individuals are sterile, there may be some progression through spermatogenesis that has lifespan ramifications. This observation will require further study and provides an opportunity for deeper investigation of the relationship between reproduction and lifespan.

Overall, inducible sterility implemented during larval development followed by a transfer to standard media appears to be a viable, non-toxic, and more natural means of conducting long-term longevity studies with *C. elegans.* The only major disadvantage to this approach is the dependence upon the genetic background that we have constructed here or reconstructing the required degron system components in other genetic backgrounds. While these components may limit the applicability in some genetic studies, there is a very large advantage for direct environmental and/or chemical manipulation studies. Further, new aging related mutant screens might be initiated using the presented *spe-44*::*degron* hermaphrodites as the parental strain.

Conclusion

Our targeted approach of a critical spermatogenesis gene and the potential applications should in principle be transferable to other systems where auxin-induction is viable, such as *Drosophila* and zebrafish. Additionally, many other types of cell-specific arrests should be targetable using the auxin-inducible system. Now firmly within the era of CRISPR/Cas9 transgenics, targeted, external induction systems, such as the method presented here, are possible. When coupled with the power of automated assays and next-

generation sequencing techniques, the field is poised to gain a wealth of information previously unattainable.

CHAPTER VII

CONCLUSION

Sexual reproduction is a near ubiquitous process that structures populations and contributes to the diversity of life. Despite its importance, the reproductive phase of the lifecycle is a complex process requiring individuals to find a mate of the opposite sex, share gametes, and fuse gametes. When the sexes contribute differently to mate acquisition and/or offspring survival, selection can act independently on females and males independently and therefore select on reproductive success in a sex-specific manner. However, the sexes must interact during reproduction and any differences in sex-specific fitness will create a conflict between the sexes. Additionally, since selection acts on a common genome, the evolutionary responses of the sexes are tethered. From an evolutionary perspective, this tethering can sustain mating conflicts between the sexes, but is ultimately not the source of conflict. Alternatively, during the survival phase of the lifecycle, sexually antagonistic pleiotropy can generate viability antagonisms between the sexes with no interaction between the sexes required. Therefore, to understand sexual conflict the source of the conflict must be identified (genomic or phenotypic) and associated with the correct phase of the lifecycle.

Historically conflict between the sexes has been studied at the phenotypic level by identifying mating conflicts and their relation to population fitness (Bateman 1948; Parker 1979). More recent work has focused on the genetic basis of sexual conflict with an emphasis on unbiased genome scans for antagonistic loci (Mank 2017; Kasimatis *et al.* 2017). These genome-scan studies have as of yet not made a connection between mating

phenotypes and fitness. However, very few studies have made the complete link from genotype to phenotype to fitness (except see Chapman *et al.* 1995). To accomplish this challenge, the field needs foundational theory, experimental tools, and genomic resources. In my dissertation, I expand our theoretical understanding of sexual antagonism and develop a new model system for studying sexual selection and sexual conflict across the genotype-phenotype map. This works allows the sexual conflict field to capitalize on systems biology approaches to make predictions as to how antagonistic mating interactions can drive divergence within and between populations.

In Chapters II and III I examined the genomic consequences of sexual conflict. Importantly, despite sexual conflict being a driver of genome evolution, there are no genomic signatures exclusive to this action of selection. Moreover, sex-biased genes can show signals of sexual conflict without involving any inherent interaction between the sexes. This lack of clear genomic signal of sexual conflict necessitates linking candidate loci with phenotypes and fitness effects to determine both the action of selection as well as the mode of antagonism. In particular, the single locus model presented in Chapter III suggests that sexually antagonistic loci should be transient in the genome, suggesting the potential for a Red-Queen type churn of antagonistic loci as suggested by Rice (Rice and Holland 1997; 1998). Such a pattern will be hard to capture using current methods and sampling approaches. If the field wishes to use genome scans for identifying sexual antagonism, then larger genomic datasets in non-human systems must be collected. While a wealth of clinical information is available on humans, directly linking genotypes with phenotypes is challenging in this setting as direct experiments cannot be conducted. Additionally, evolution of the human lineage is different from many other species and in

therefore unlikely to provide a general paradigm of sexual conflict as a driver of genome evolution. The field is not currently in a strong position to move forward analyzing genomic data – rather we need to critically assess our goals, revisit our theoretical foundations, and build a correct statistical framework to make this work meaningful (see Lewontin 1991).

The theory I developed in Chapter III also suggests that invasion of modifiers should be selected for to mediate conflict. Specifically, changing the dominance relationships between the sexes should reduce the load and act as a temporary resolution state. Recent work has identified sex-specific beneficial dominance underlying traits with antagonistic effects between the sexes (Czorlich *et al.* 2018), indicating more robust theory in is needed. Such sex-specific modifiers of dominance could create a cryptic form of divergence between populations that may contribute to reproductive isolation, supporting previous hypotheses on sexual conflict as a driver of speciation (Rice 1998; Gavrilets 2014).

Until genomic signatures can provide better candidate lists, the focus of the field should remain on identifying conflict traits and then linking them with their underlying basis. One area well-suited for such work is post-insemination reproductive interactions. Here interactions are between cells and proteins making the link between phenotype and genotype more straightforward. In internally fertilizing species such interactions are contained within the female reproductive tract making them a challenge to study. In fact, most of what is known about post-insemination comes from *Drosophila*. Thus, widely accepted hypotheses in the field come from an albeit vast body of work in one genus and largely one species. In particular, it is widely accepted that sperm proteins evolve rapidly

due to sperm competition (i.e. sexual selection) and male-female interactions (i.e. sexual conflict) (Begun *et al.* 2000; Swanson and Vacquier 2002; Clark *et al.* 2006; Vacquier and Swanson 2011). In Chapters IV, V, and VI, I develop nematodes as a model system for studying post-insemination dynamics with an emphasis on male fertility. Nematodes have dramatically different sperm biology, which allows us to assess the ubiquity of existing hypothesis on the molecular evolution of sperm proteins. Additionally, the wealth of genetic and genomic information in worms allows for experimental verification of candidate conflict phenotypes.

Contrary to previous studies on the evolution of sperm proteins, Chapters IV and V show that nematode sperm proteins are hyper-conserved at the amino acid sequence. However, gene family composition and organization appear highly dynamic. This work highlights the need to address multiple levels of genomic organization to understand the history of a gene. Moreover, this pattern suggests that genome scan methods alone may be insufficient in some contexts to identify sexual conflict loci. Additionally, the dramatically different sperm biology of nematodes highlights that physiological differences during post-insemination can potentially have a large effect on the molecular evolution the proteins involved. Therefore, we need to expand away from studying reproduction within a single organism. Given the advances in cellular techniques and proteomic technology such work is now feasible in non-model organisms, potentially in the field, to capture the true nature of sperm competition and how this process relates to male fertility.

Studying the relationship between pre- and post-insemination reproductive in determining male fertility success is another challenge in the field that *Drosophila*

research has provided direction on. Here studies indicate that pre-insemination interactions determine male fertilization success (Pischedda and Rice 2012). However, this work is not based on strong hypothesis-testing. In Chapter VI, I design an external, inducible sterility system for *C. elegans*. The sterility induction capitalizes on the auxin inducible degron system (Nishimura *et al.* 2009; Zhang *et al.* 2015) to prevent males from producing functional sperm and hermaphrodites from producing self-sperm. This method has a wide variety of applications., such as: i) teasing apart selection on male pre- and post-insemination success, ii) studying aging, and iii) examining the evolution of mating systems. This work more broadly supports auxin-induction as an excellent mechanism for fine temporal and tissue-specific control over cellular functions.

One particular experimental framework that emerges from the sterility induction systems is isolation of pre-insemination and post-insemination reproductive stages. Specifically, using an experimental evolution approach, I am currently testing the relative contributions of pre-insemination mating success and post-insemination fertility success to overall male reproductive fitness. The experimental framework is briefly outlined as follows. To isolate gametic selection, sterility is induced in adult males, which prevents isolates the sperm transferred to females when males were very young adults. This mid-life sterility induction allows me to assess sperm storage and longevity within the female. Alternatively, after sterility is induced fully fertile males can be introduced to the population to create a source of sperm competition. Now sterile males must rely on their previously transferred sperm to outcompete incoming sperm from the competitors. This sterility and competition paradigm directly isolates post-insemination competition and selects for sperm defensive capability. Finally, a population where sterility is not induced,

but competitors are added represents the fully suite of pre- and post-insemination competition dynamics. Together this framework represents the type of work needed to connect the action of sex-specific selection across the lifecycle.

The research presented in this dissertation used a combination of theoretical, proteomic, and genomic approaches to understand the dynamic nature of sexual conflict. The fields of sexual selection and sexual conflict have rich histories, yet they are now at a turning point as we move forward in the genomics and systems era. Importantly, foundational questions in the field can now be experimentally tested. This work will require uniting approaches at the genotypic and phenotypic levels. Only then can we fully understand the importance of sexual conflict in structuring populations and driving mating interactions.

REFERENCES CITED

Achanzar W. E., Ward S., 1997 A nematode gene required for sperm vesicle fusion. J. Cell. Sci. **110**: 1073–1081.

Adler M. I., Bonduriansky R., 2014 Sexual Conflict, Life Span, and Aging. Cold Spring Harbor Perspectives in Biology **6**: a017566–a017566.

Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P., 2002 Transport from the ER through the Golgi Apparatus. In: Molecular Biology of the Cell. 4th edition, Garland Science.

Andersson M., 1994 *Sexual Selection*. Princeton University Press, New York.

Andersson M., Simmons L. W., 2006 Sexual selection and mate choice. Trends Ecol Evol **21**: 296–302.

Angeles-Albores D., Leighton D. H. W., Tsou T., Khaw T. H., Antoshechkin I., Sternberg P. W., 2017 The *Caenorhabditis elegans* Female State: Decoupling the Transcriptomic Effects of Aging and Sperm-Status. G3 **7**: 2969–2977.

Anisimova M., Gascuel O., 2006 Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. Systematic Biology **55**: 539–552.

Arnold S. J., 1983 Sexual selection: the interface of theory and empiricism. Mate choice.

Arnold S. J., 1994 Bateman's principles and the measurement of sexual selection in plants and animals. Am Nat **144:** S126–S149.

Arnold S. J., Wade M. J., 1984 On the measurement of natural and sexual selection: theory. Evolution **38**: 709–719.

Arnqvist G., 2004 Sexual conflict and sexual selection: lost in the chase. Evolution **58**: 1383–8– discussion 1389–93.

Arnqvist G., 2011 Assortative mating by fitness and sexually antagonistic genetic variation. Evolution **65**: 2111–2116.

Arnqvist G., Rowe L., 2005 *Sexual conflict*. Princeton University Press, Princeton.

Bachtrog D., Kirkpatrick M., Mank J. E., McDaniel S. F., Pires J. C., Rice W., Valenzuela N., 2011 Are all sex chromosomes created equal? Trends in Genetics **27**: 350–357.

Baker A. M. E., Roberts T. M., Stewart M., 2002 2.6Å Resolution Crystal Structure of Helices of the Motile Major Sperm Protein (MSP) of *Caenorhabditis elegans*. J Mol Biol **319**: 491–499.

Baker D. A., Nolan T., Fischer B., Pinder A., Crisanti A., Russell S., 2011 A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, *Anopheles gambiae*. BMC Genomics **12**: 296.

Baskaran P., Rödelsperger C., 2015 Microevolution of Duplications and Deletions and Their Impact on Gene Expression in the Nematode Pristionchus pacificus. PLoS ONE **10**: e0131136.

Baskaran P., Jaleta T. G., Streit A., Rödelsperger C., 2017 Duplications and Positive Selection Drive the Evolution of Parasitism-Associated Gene Families in the Nematode *Strongyloides papillosus*. Genome Biol Evol **9**: 790–801.

Baskaran P., Rödelsperger C., Prabh N., Serobyan V., Markov G. V., Hirsekorn A., Dieterich C., 2015 Ancient gene duplications have shaped developmental stage-specific expression in *Pristionchus pacificus*. BMC Evol Biol **15**: 1–12.

Bateman A. J., 1948 Intra-sexual selection in *Drosophila*. Heredity **2**: 349–368.

Bates D., Mächler M., Bolker B., Walker S., 2015 Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Soft. **67**: 1–48.

Beer K. B., Wehman A. M., 2017 Mechanisms and functions of extracellular vesicle release in vivo-What we can learn from flies and worms. Cell Adh Migr **11**: 135–150.

Begun D. J., Whitley P., Todd B. L., Waldrip-Dail H. M., Clark A. G., 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics **156**: 1879–1888.

Berger D., Martinossi-Allibert I., 2016 Intralocus sexual conflict and the tragedy of the commons in seed beetles. Am Nat **188**: E98–E112.

Bhatia, G., N. Patterson, Sankararaman S., Price, A. L., 2013 Estimating and interpreting FST: the impact of rare variants. Genome Res. **23:**1514–1521.

Blaxter M., 2009 Nematodes (Nematoda). In:Hedges SB, Kumar S (Eds.), *The Timetree of Life*, The Timetree of Life, pp. 247–250.

Blaxter M., Koutsovoulos G., 2015 The evolution of parasitism in Nematoda. Parasitology **142 Suppl 1**: S26–39.

Bodmer W. F., 1965 Differential Fertility in Population Genetics Models. Genetics **51**: 411–424.

Bohne A., Sengstag T., Salzburger W., 2014 Comparative Transcriptomics in East African Cichlids Reveals Sex- and Species-Specific Expression and New Candidates for Sex Differentiation in Fishes. Genome Biol Evol **6**: 2567–2585.

Bonduriansky R., Chenoweth S. F., 2009 Intralocus sexual conflict. Trends Ecol Evol: 1–9.

Bono J. M., Olesnicky E. C., Matzkin L. M., 2015 Connecting genotypes, phenotypes and fitness: harnessing the power of CRISPR/Cas9 genome editing. Mol Ecol **24**: 3810–3822.

Bottino D., Mogilner A., Roberts T., Stewart M., Oster G., 2002 How nematode sperm crawl. J. Cell. Sci. **115**: 367–384.

Braendle C., Felix M.-A., 2006 Sex Determination: Ways to Evolve a Hermaphrodite. Curr Biol **16**: R468–R471.

Brenner S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77**: 71–94.

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

Burke D. J., Ward S., 1983 Identification of a large multigene family encoding the major sperm protein of *Caenorhabditis elegans*. J Mol Biol **171**: 1–29.

Byerly L., Cassada R. C., Russell R. L., 1976 The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. Dev Biol **51**: 23–33.

Castrillon D. H., Gönczy P., Alexander S., Rawson R., Eberhart C. G., Viswanathan S., DiNardo S., Wasserman S. A., 1993 Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single P element mutagenesis. Genetics **135**: 489–505.

Champely S., 2017 *pwr: Basic Functions for Power Analysis*. R package version 1.2-1.

Chapman T., 2006 Evolutionary Conflicts of Interest between Males and Females. Curr Biol **16**: R744–R754.

Chapman T., 2011 Seminal fluid-mediated fitness traits in *Drosophila*. Heredity **87**: 511–521.

Chapman T., Arnqvist G., Bangham J., Rowe L., 2003 Sexual conflict. Trends Ecol Evol **18**: 41–47.

Chapman T., Liddle L. F., Kalb J. M., Wolfner M. F., Partridge L., 1995 Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature **373**: 241–244.

Charlesworth B., 1991 The evolution of sex chromosomes. Science **251**: 1030–1033.

Charlesworth B., Charlesworth D., Barton N. H., 2003 The Effects of Genetic and Geographic Structure on Neutral Variation. Annu. Rev. Ecol. Evol. Syst. **34**: 99–125.

Charlesworth B., Nordborg M., Charlesworth D., 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70**: 155–174.

Charlesworth D., 2006 Balancing selection and its effects on sequences in nearby genome regions. PLoS Genetics **2**: e64.

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

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

Chatterjee I., Richmond A., Putiri E., Shakes D. C., Singson A., 2005 The *Caenorhabditis elegans spe-38* gene encodes a novel four-pass integral membrane protein required for sperm function at fertilization. Development **132**: 2795–2808.

Chen J.-M., Cooper D. N., Chuzhanova N., Férec C., Patrinos G. P., 2007 Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet **8**: 762–775.

Cheng C., Kirkpatrick M., 2016 Sex-specific selection and sex-biased gene expression in humans and flies. PLoS Genetics **12**: e1006170–18.

Churchill G. A., Doerge R. W., 1994 Empirical threshold values for quantitative trait mapping. Genetics **138**: 963–971.

Clark N. L., Aquadro C. F., 2010 A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Mol Biol Evol **27**: 1152–1161.

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

Clark N. L., Alani E., Aquadro C. F., 2012 Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res. **22**: 714–720.

Clark N. L., Gasper J., Sekino M., Springer S. A., Aquadro C. F., Swanson W. J., 2009 Coevolution of interacting fertilization proteins. PLoS Genetics **5**: e1000570.

Connallon T., Chenoweth S. F., 2019 Dominance reversals and the maintenance of genetic variation for fitness. PLoS Biol **17**: e3000118.

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

Connallon T., Clark A. G., 2012 A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. Genetics **190**: 1477–1489.

Connallon T., Clark A. G., 2013 Antagonistic versus nonantagonistic models of balancing selection: characterizing the relative timescales and hitchhiking effects of partial selective sweeps. Evolution **67**: 908–917.

Connallon T., Clark A. G., 2014 Balancing selection in species with separate sexes: insights from Fisher's geometric model. Genetics **197**: 991–1006.

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

Connallon T., Cox R. M., Calsbeek R., 2010 Fitness consequences of sex-specific selection. Evolution **64**: 1671–1682.

Connallon T., Sharma S., Olito C., 2019 Evolutionary consequences of sex-specific selection in variable environments: four simple models reveal diverse evolutionary outcomes. Am Nat **193**: 93–105.

Conrad D. F., Pinto D., Redon R., Feuk L., Gokcumen O., Zhang Y., Aerts J., Andrews T. D., Barnes C., Campbell P., Fitzgerald T., Hu M., Ihm C. H., Kristiansson K., MacArthur D. G., MacDonald J. R., Onyiah I., Pang A. W. C., Robson S., Stirrups K., Valsesia A., Walter K., Wei J., Consortium W. T. C. C., Tyler-Smith C., Carter N. P., Lee C., Scherer S. W., Hurles M. E., 2010 Origins and functional impact of copy number variation in the human genome. Nature **464**: 704–712.

Cook D. E., Zdraljevic S., Roberts J. P., Andersen E. C., 2017 CeNDR, the *Caenorhabditis elegans* natural diversity resource. Nucleic Acids Res **45**: D650–D657.

Czorlich Y., Aykanat T., Erkinaro J., Orell P., Primmer C. R., 2018 Rapid sex-specific evolution of age at maturity is shaped by genetic architecture in Atlantic salmon. Nat Ecol Evol **97**: 5693–11.

Dapper A. L., Wade M. J., 2016 The evolution of sperm competition genes: the effect of mating system on levels of genetic variation within and between species. Evolution **70**: 502–511.

Darwin C., 1871 *The Descent of Man and Selection in Relation to Sex*. Murray, London.

Dean M. D., Findlay G. D., Hoopmann M. R., Wu C. C., MacCoss M. J., Swanson W. J., Nachman M. W., 2011 Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. BMC Genomics **12**: 306.

del Castillo-Olivares A., Smith H. E., 2008 Critical contact residues that mediate polymerization of nematode major sperm protein. J Cell Biochem **104**: 477–487.

Demuth J. P., Hahn M. W., 2009 The life and death of gene families. BioEssays **31**: 29–39.

Dhole S., Servedio M. R., 2014 Sperm competition and the evolution of seminal fluid composition. Evolution **68**: 3008–3019.

Dickinson D. J., Ward J. D., Reiner D. J., Goldstein B., 2013 Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. Nat Meth **10**: 1028–1034.

Drosophila 12 Genomes Consortium, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature **450**: 203–218.

Dunbar B. S., O'Rand M. (Eds.), 1991 *A Comparative Overview of Mammalian Fertilization*. Plenum Press, New York.

Dunker A. K., Brown C. J., Lawson J. D., Iakoucheva L. M., Obradović Z., 2002 Intrinsic disorder and protein function. Biochemistry **41**: 6573–6582.

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

Edward D. A., Stockley P., Hosken D. J., 2015 Sexual conflict and sperm competition. Cold Spring Harbor Perspectives in Biology **7**: a017707.

Eirín-López J. M., Frehlick L. J., Ausió J., 2005 Protamines, in the footsteps of linker histone evolution. J Biol Chem **281**: 1–4.

Ellis R. E., Stanfield G. M., 2014 The regulation of spermatogenesis and sperm function in nematodes. Sem Cell Dev Biol **29**: 17–30.

Farslow J. C., Lipinski K. J., Packard L. B., Edgley M. L., Taylor J., Flibotte S., Moerman D. G., Katju V., Bergthorsson U., 2015 Rapid increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. BMC Genomics **16**: 1044.

Felix M.-A., Braendle C., Cutter A. D., 2014 A streamlined system for species diagnosis in *Caenorhabditis* (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. PLoS ONE **9**: e94723.

Findlay G. D., Swanson W. J., 2010 Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays **32**: 26–36.

Findlay G. D., MacCoss M. J., Swanson W. J., 2009 Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. Genome Res **19**: 886–896.

Findlay G. D., Sitnik J. L., Wang W., Aquadro C. F., Clark N. L., Wolfner M. F., 2014 Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. PLoS Genetics **10**: e1004108–16.

Fisher R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Flanagan S. P., Jones A. G., 2017 Genome-wide selection components analysis in a fish with male pregnancy. Evolution **71**: 1096–1105.

Fowler K., Partridge L., 1989 A cost of mating in female fruitflies. Nature **338**: 760–761.

Fraser H. B., Hirsh A. E., Wall D. P., Eisen M. B., 2004 Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci **101**: 9033–9038.

Friedberg I., 2006 Automated protein function prediction--the genomic challenge. Briefings in Bioinformatics **7**: 225–242.

Friedman D. B., Johnson T. E., 1988 A mutation in the age-1 gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. Genetics **118**: 75–86.

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

Frøkjær-Jensen C., Davis M. W., Ailion M., Jorgensen E. M., 2012 Improved Mos1-mediated transgenesis in *C. elegans*. Nat Meth **9**: 117–118.

Furness A. I., Morrison K. R., Orr T. J., Arendt J. D., Reznick D. N., 2015 Reproductive mode and the shifting arenas of evolutionary conflict. Ann. N.Y. Acad. Sci. **1360**: 75–100.

Gao B., Zhu S., 2016 The drosomycin multigene family: three-disulfide variants from *Drosophila takahashii* possess antibacterial activity. Nature **6**: 32175.

Gardner A., Úbeda F., 2017 The meaning of intragenomic conflict. Nat Ecol Evol. **1**: 1807–1815.

Gavrilets S., 2000 Rapid evolution of reproductive barriers driven by sexual conflict. Nature **403**: 886–889.

Gavrilets S., 2014 Is Sexual conflict an “engine of speciation?” Cold Spring Harbor Perspectives in Biology **6**: a017723–a017723.

Gavrilets S., Hayashi T. I., 2006 The dynamics of two- and three-way sexual conflicts over mating. Phil. Trans. R. Soc. B **361**: 345–354.

Gavrilets S., Waxman D., 2002 Sympatric speciation by sexual conflict. Proc Natl Acad Sci **99**: 10533–10538.

Gavrilets S., Arnqvist G., Friberg U., 2001 The evolution of female mate choice by sexual conflict. Proc. Royal Soc B **268**: 531–539.

Good J. M., Wiebe V., Albert F. W., Burbano H. A., Kircher M., Green R. E., Halbwax M., Andre C., Atencia R., Fischer A., Paabo S., 2013 Comparative population genomics of the ejaculate in humans and the great apes. Mol Biol Evol **30**: 964–976.

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

Guindon S., Gascuel O., 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. System Biol **52**: 696-704.

Haaf A., LeClaire L., Roberts G., Kent H. M., Roberts T. M., Stewart M., Neuhaus D., 1998 Solution structure of the motile major sperm protein (MSP) of *Ascaris suum* - evidence for two manganese binding sites and the possible role of divalent cations in filament formation. J Mol Biol **284**: 1611–1624.

Hakes L., Lovell S. C., Oliver S. G., Robertson D. L., 2007 Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proc Natl Acad Sci **104**: 7999–8004.

Haldane J., 1937 The effect of variation of fitness. Am Nat **71**: 337–349.

Haldane J. B. S., 1957 The cost of natural selection. J Genet **55**: 511–524.

Haller B. C., Messer P. W., 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Mol Biol Evol **36**: 632–637.

Harrison P. W., Wright A. E., Zimmer F., Dean R., Montgomery S. H., Pointer M. A., Mank J. E., 2015 Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci **112**: 4393–4398.

Hayashi T. I., Vose M., Gavrilets S., 2007 Genetic differentiation by sexual conflict. Evolution **61**: 516–529.

Haygood R., 2004 Sexual conflict and protein polymorphism. Evolution **58**: 1414–1423.

Hirsh D., Vanderslice R., 1976 Temperature-sensitive developmental mutants of *Caenorhabditis elegans*. Dev Biol **49**: 220–235.

Hirsh D., Oppenheim D., Klass M., 1976 Development of the reproductive system of *Caenorhabditis elegans*. Dev Biol **49**: 200–219.

Hodgkin J., Horvitz H. R., Brenner S., 1979 Nondisjunction Mutants of the Nematode *Caenorhabditis Elegans*. Genetics **91**: 67–94.

Holland B., Rice W. R., 1998 Perspective: chase-away sexual selection: antagonistic seduction versus resistance. Evolution **52**: 1–7.

Hollis B., Houle D., Kawecki T. J., 2015 Evolution of reduced post-copulatory molecular interactions in Drosophila populations lacking sperm competition. J Evol Biol **29**: 77-85.

Hothorn T., Bretz F., Westfall P., 2008 Simultaneous inference in general parametric models. Biom J **50**: 346–363.

Howe K. L., Bolt B. J., Shafie M., Kersey P., Berriman M., 2017 WormBase ParaSite - a comprehensive resource for helminth genomics. Mol Biochem Parasit **215**: 2–10.

Hsu C.-C., Hou M.-F., Hong J.-R., Wu J.-L., Her G. M., 2009 Inducible male infertility by targeted cell ablation in zebrafish testis. Mar Biotechnol **12**: 466–478.

Huerta-Sanchez E., Durrett R., Bustamante C. D., 2008 Population genetics of polymorphism and divergence under fluctuating selection. Genetics **178**: 325–337.

Hunt V. L., Tsai I. J., Coghlan A., Reid A. J., Holroyd N., Foth B. J., Tracey A., Cotton J. A., Stanley E. J., Beasley H., Bennett H. M., Brooks K., Harsha B., Kajitani R., Kulkarni A., Harbecke D., Nagayasu E., Nichol S., Ogura Y., Quail M. A., Randle N., Xia D., Brattig N. W., Soblik H., Ribeiro D. M., Sanchez-Flores A., Hayashi T., Itoh T., Denver D. R., Grant W., Stoltzfus J. D., Lok J. B., Murayama H., Wastling J., Streit A., Kikuchi T., Viney M., Berriman M., 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet **48**: 299–307.

Immler S., Otto S. P., 2018 The evolutionary consequences of selection at the haploid gametic stage. Am Nat **192**: 241–249.

Innan H., Kondrashov F., 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet **11**: 97–108.

Innocenti P., Morrow E. H., 2010 The sexually antagonistic genes of *Drosophila melanogaster*. PLoS Biol **8**: e1000335–10.

Jaenike J., 2003 Sex chromosome meiotic drive. Annu Rev Ecol Syst **32**: 25–49.

Jakobsson M., Edge M. D., Rosenberg N. A., 2013 The relationship between fst and the frequency of the most frequent allele. Genetics **193**: 515–528.

Jones A. G., 2009 On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. Evolution **63**: 1673–1684.

Joseph S., Kirkpatrick M., 2004 Haploid selection in animals. Trends Ecol Evol **19**: 592–597.

Justine J. L., 2002 Male and female gametes and fertilisation. In:Lee D (Ed.), *The Biology of Nematodes*,

Kamei N., 2003 The species-specific egg receptor for sea urchin sperm adhesion is EBR1 ,a novel ADAMTS protein. Genes Dev. **17**: 2502–2507.

Kanke M., Nishimura K., Kanemaki M., Kakimoto T., Takahashi T. S., Nakagawa T., Masukata H., 2011 Auxin-inducible protein depletion system in fission yeast. BMC Cell Biol. **12**: 8.

Karr T. L., Pitnick S., 1999 Sperm competition: defining the rules of engagement. Curr Biol **9**: R787–90.

Kasimatis K. R., Phillips P. C., 2018 Rapid gene family evolution of a nematode sperm protein despite sequence hyper-conservation. G3 **8**: 353–362.

Kasimatis K. R., Moerdyk-Schauwecker M. J., Phillips P. C., 2018a Auxin-mediated sterility induction system for longevity and mating studies in *Caenorhabditis elegans*. G3 **8**: 2655–2662.

Kasimatis K. R., Moerdyk-Schauwecker M. J., Timmermeyer N., Phillips P. C., 2018b Proteomic and evolutionary analyses of sperm activation identify uncharacterized genes in *Caenorhabditis* nematodes. BMC Genomics **19**: 593.

Kasimatis K. R., Nelson T. C., Phillips P. C., 2017 Genomic signatures of sexual conflict. J Hered **108**: 780–790.

Käll L., Storey J. D., Noble W. S., 2008 Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. Bioinformatics **24**: i42–8.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A., 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**: 1647–1649.

Kempe K., Gils M., 2011 Pollination control technologies for hybrid breeding. Mol Breeding **27**: 417–437.

Kent W. J., Sugnet C. W., Furey T. S., Roskin K. M., Pringle T. H., Zahler A. M., Haussler D., 2002 The human genome browser at UCSC. Genome Res. **12**: 996–1006.

Kenyon C., 1988 The nematode *Caenorhabditis elegans*. Science **240**: 1448–1453.

Kenyon C., Chang J., Gensch E., Rudner A., Tabtiang R., 1993 A *C. elegans* mutant that lives twice as long as wild type. Nature **366**: 461–464.

Kidwell J. F., Clegg M. T., Stewart F. M., Prout T., 1977 Regions of stable equilibria for models of differential selection in the two sexes under random mating. Genetics **85**: 171–183.

King K. L., Stewart M., Roberts T. M., Seavy M., 1992 Structure and macromolecular assembly of two isoforms of the major sperm protein (MSP) from the amoeboid sperm of the nematode, *Ascaris suum*. J. Cell. Sci. **101 (Pt 4)**: 847–857.

Kiontke K., 2005 The phylogenetic relationships of *Caenorhabditis* and other rhabditids. WormBook: 1–11.

Kiontke K. C., Felix M.-A., Ailion M., Rockman M. V., Braendle C., Penigault J.-B., Fitch D. H., 2011 A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. BMC Evol Biol **11**: 339.

Kirkpatrick M., 2010 How and why chromosome inversions evolve. PLoS Biol **8**: e1000501–5.

Kirkpatrick M., 2017 The evolution of genome structure by natural and sexual selection. J Hered **108**: 3–11.

Kirkpatrick M., Guerrero R. F., 2014 Signatures of sex-antagonistic selection on recombining sex chromosomes. Genetics **197**: 531–541.

Klass M. R., Hirsh D., 1981 Sperm isolation and biochemical analysis of the major sperm protein from *Caenorhabditis elegans*. Dev Biol **84**: 299–312.

Kokko H., Jennions M. D., 2014 The Relationship between Sexual Selection and Sexual Conflict. Cold Spring Harbor Perspectives in Biology **6**: a017517–a017517.

Kosakovsky Pond S. L., Murrell B., Fourment M., Frost S. D. W., Delport W., Scheffler K., 2011 A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol **28**: 3033–3043.

Kosinski M., McDonald K., Schwartz J., Yamamoto I., Greenstein D., 2005 C. elegans sperm bud vesicles to deliver a meiotic maturation signal to distant oocytes. Development **132**: 3357–3369.

Kozielska M., Weissing F. J., Beukeboom L. W., Pen I., 2010 Segregation distortion and the evolution of sex-determining mechanisms. Heredity **104**: 100–112.

Kuijper B., Pen I., Weissing F. J., 2012 A guide to sexual selection theory. Annu Rev Ecol Evol Syst **43**: 287–311.

Kulkarni M., Shakes D. C., Guevel K., Smith H. E., 2012 SPE-44 implements sperm cell fate. PLoS Genetics **8**: e1002678.

L'Hernault S. W., 2006 Spermatogenesis. In: *WormBook*,, pp. 1–14.

LaMunyon C. W., Ward S., 2002 Evolution of larger sperm in response to experimentally increased sperm competition in *Caenorhabditis elegans*. Proc R Soc Lond B: 1125–1128.

Lee M. O., Bornelöv S., Andersson L., Lamont S. J., Chen J., Womack J. E., 2016 Duplication of chicken defensin7 gene generated by gene conversion and homologous recombination. Proc Natl Acad Sci **113**: 13815–13820.

Lee R. Y. N., Howe K. L., Harris T. W., Arnaboldi V., Cain S., Chan J., Chen W. J., Davis P., Gao S., Grove C., Kishore R., Muller H.-M., Nakamura C., Nuin P., Paulini M., Raciti D., Rodgers F., Russell M., Schindelman G., Tuli M. A., Van Auken K., Wang Q., Williams G., Wright A., Yook K., Berriman M., Kersey P., Schedl T., Stein L., Sternberg P. W., 2017 WormBase 2017: molting into a new stage. Nucleic Acids Research **46**: 1–6.

Lehtonen J., Jennions M. D., Kokko H., 2012 The many costs of sex. Trends Ecol Evol **27**: 172–178.

Lehtonen J., Parker G. A., Schärer L., 2016 Why anisogamy drives ancestral sex roles. Evolution **70**: 1129–1135.

Lenormand T., 2002 Gene flow and the limits to natural selection. Trends Ecol Evol **17**: 183–189.

Levitan D. R., 2006 Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science **312**: 267–269.

Lewontin R. C., 1974 *The genetic basis of evolutionary change*. Columbia University Press, New York.

Lewontin R. C., 1991 Twenty-five years ago in Genetics: electrophoresis in the development of evolutionary genetics: milestone or millstone? Genetics **128**: 657–662.

Lucanic M., Plummer W. T., Chen E., Harke J., Foulger A. C., Onken B., Coleman-Hulbert A. L., Dumas K. J., Guo S., Johnson E., Bhaumik D., Xue J., Crist A. B., Presley M. P., Harinath G., Sedore C. A., Chamoli M., Kamat S., Chen M. K., Angeli S., Chang C., Willis J. H., Edgar D., Royal M. A., Chao E. A., Patel S., Garrett T., Ibanez-Ventoso C., Hope J., Kish J. L., Guo M., Lithgow G. J., Driscoll M., Phillips P. C., 2017 Impact of genetic background and experimental reproducibility on identifying chemical compounds with robust longevity effects. Nature Communications **8**: 14256.

Lucotte E. A., Laurent R., Heyer E., Ségurel L., Toupance B., 2016 Detection of allelic frequency differences between the sexes in humans: a signature of sexually antagonistic selection. Genome Biol Evol **8**: 1489–1500.

Ma X., Zhu Y., Li C., Xue P., Zhao Y., Chen S., Yang F., Miao L., 2014 Characterisation of *Caenorhabditis elegans* sperm transcriptome and proteome. BMC Genomics **15**: 1–13.

Malik H. S., Henikoff S., 2003 Phylogenomics of the nucleosome. Nat Struct Biol **10**: 882–891.

Manier M. K., Palumbi S. R., 2008 Intraspecific divergence in sperm morphology of the green sea urchin, *Strongylocentrotus droebachiensis*: implications for selection in broadcast spawners. BMC Evol Biol **8**: 283–14.

Mank J. E., 2009 The W, X, Y and Z of sex-chromosome dosage compensation. Trends in Genetics **25**: 226–233.

Mank J. E., 2017a The transcriptional architecture of phenotypic dimorphism. Nature **1**: 1–7.

Mank J. E., 2017b Population genetics of sexual conflict in the genomic era. Nat Rev Genet **7**: 1–10.

Mank J. E., Hosken D. J., Wedell N., 2014 conflict on the sex chromosomes: cause, effect, and complexity. Cold Spring Harbor Perspectives in Biology **6**: a017715–a017715.

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

Markov G. V., Baskaran P., Sommer R. J., 2014 The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. J Mol Evol **80**: 18–36.

McDonough C. E., Whittington E., Pitnick S., Dorus S., 2016 Proteomics of reproductive systems: Towards a molecular understanding of postmating, prezygotic reproductive barriers. J. Proteomics **135**: 26–37.

McLaughlin R. N. Jr, Malik H. S., 2017 Genetic conflicts: the usual suspects and beyond. J Exp Biol **220**: 6–17.

Merrihew G. E., Davis C., Ewing B., Williams G., Kall L., Frewen B. E., Noble W. S., Green P., Thomas J. H., MacCoss M. J., 2008 Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations. Genome Res **18**: 1660–1669.

Merritt C., Rasoloson D., Ko D., Seydoux G., 2008 3′ UTRs are the primary regulators of gene expression in the *C. elegans* germline. Curr Biol **18**: 1476–1482.

Mezulis S., Yates C. M., Wass M. N., Sternberg M. J. E., Kelley L. A., 2015 The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocols **10**: 845–858.

Miller M. A., 2006 Sperm and oocyte isolation methods for biochemical and proteomic analysis. Methods Mol Biol **351**: 193–201.

Miller M. A., Nguyen V. Q., Lee M. H., Kosinski M., Schedl T., Caprioli R. M., Greenstein D., 2001 A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. Science **291**: 2144–2147.

Mills M., Yang N., Weinberger R., Vander Woude D. L., Beggs A. H., Easteal S., North K., 2001 Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: implications for the evolution of functional redundancy. Hum Mol Genet **10**: 1335–1346.

Mitchell D. H., Stiles J. W., Santelli J., Sanadi D. R., 1979 Synchronous growth and aging of *Caenorhabditis elegans* in the presence of fluorodeoxyuridine. J Gerontol **34**: 28–36.

Mordhorst B. R., Wilson M. L., Conant G. C., 2015 Some assembly required: evolutionary and systems perspectives on the mammalian reproductive system. Cell Tissue Res **363**: 267–278.

Morrow E. H., 1999 How the sperm lost its tail: the evolution of aflagellate sperm. Biological Reviews **79**: 795–814.

Murakami S., Johnson T. E., 1996 A genetic pathway conferring life extension and resistance to UV stress in *Caenorhabditis elegans*. Genetics **143**: 1207–1218.

Naito Y., Hino K., Bono H., Ui-Tei K., 2015 CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics **31**: 1120–1123.

Nam K., Munch K., Hobolth A., Dutheil J. Y., Veeramah K. R., Woerner A. E., Hammer M. F., Great Ape Genome Diversity Project, Mailund T., Schierup M. H., 2015 Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc Natl Acad Sci. **112**: 6413–6418.

Natsume T., Kiyomitsu T., Saga Y., Kanemaki M. T., 2016 Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. Cell Reports **15**: 210–218.

Nelson G. A., Ward S., 1980 Vesicle fusion, pseudopod extension and amoeboid motility are induced in nematode spermatids by the lonophore monensin. Cell **19**: 457–464.

Nelson G. A., Roberts T. M., Ward S., 1982 Caenorhabditis elegans spermatozoan locomotion: amoeboid movement with almost no actin. J Cell Biol **92**: 121–131.

Nielsen R., 2005 Molecular signatures of natural selection. Annu Rev Genet **39**: 197–218.

Nielsen R., Williamson S., Kim Y., Hubisz M. J., Clark A. G., Bustamante C., 2005 Genomic scans for selective sweeps using SNP data. Genome Res. **15**: 1566–1575.

Nishimura H., L'Hernault S. W., 2010 Spermatogenesis-defective (*spe*) mutants of the nematode Caenorhabditis elegansprovide clues to solve the puzzle of male germline functions during reproduction. Dev Dyn **110**: 1502–1514.

Nishimura K., Fukagawa T., Takisawa H., Kakimoto T., Kanemaki M., 2009 An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Meth **6**: 917–922.

Ohno S., 1970 *Evolution Through Gene Duplication*. Springer-Verlag, New York.

Paix A., Folkmann A., Rasoloson D., Seydoux G., 2015 High efficiency, homology-directed genome editing in caenorhabditis elegans using CRISPR-Cas9 ribonucleoprotein complexes. Genetics **201**: 47–54.

Palopoli M. F., Peden C., Woo C., Akiha K., Ary M., Cruze L., Anderson J. L., Phillips P. C., 2015 Natural and experimental evolution of sexual conflict within *Caenorhabditis* nematodes. BMC Evol Biol **15**: 1–13.

Park H.-E. H., Jung Y., Lee S.-J. V., 2017 Survival assays using *Caenorhabditis elegans*. Mol Cells **40**: 90–99.

Parker G., 1979 *Sexual selection and sexual conflict* (MS Blum and NA Blum, Eds.). Academic Press, New York.

Parker G. A., 2006 Sexual conflict over mating and fertilization: an overview. Phil Trans R Soc B **361**: 235–259.

Parker G. A., 2014 The sexual cascade and the rise of pre-ejaculatory (Darwinian) sexual selection, sex roles, and sexual conflict. Cold Spring Harbor Perspectives in Biology **6**: a017509–a017509.

Parker G. A., Partridge L., 1998 Sexual conflict and speciation. Phil Trans R Soc B **353**: 261–274.

Patten M. M., Haig D., 2009 Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. Evolution **63**: 2888–2895.

Patten M. M., Haig D., Úbeda F., 2010 Fitness variation due to sexual antagonism and linkage disequilibrium. Evolution **64**: 3638–3642.

Pehrson J. R., Fuji R. N., 1998 Evolutionary conservation of histone macroH2A subtypes and domains. Nucleic Acids Research **26**: 2837–2842.

Pennell T. M., Morrow E. H., 2013 Two sexes, one genome: the evolutionary dynamics of intralocus sexual conflict. Ecol Evol **3**: 1819–1834.

Pennell T. M., de Haas F. J. H., Morrow E. H., Van Doorn G. S., 2016 Contrasting effects of intralocus sexual conflict on sexually antagonistic coevolution. Proc Natl Acad Sci **113**: E978–E986.

Perry G. H., Dominy N. J., Claw K. G., Lee A. S., Fiegler H., Redon R., Werner J., Villanea F. A., Mountain J. L., Misra R., Carter N. P., Lee C., Stone A. C., 2007 Diet and the evolution of human amylase gene copy number variation. Nat Genet **39**: 1256–1260.

Perry J. C., Rowe L., 2015 The evolution of sexually antagonistic phenotypes. Cold Spring Harbor Perspectives in Biology **7**: a017558–18.

Perry J. C., Rowe L., 2018 Sexual conflict in its ecological setting. Philos Trans R Soc Lond B Sci. **373**: 20170418.

Petersen T. N., Brunak S., Heijne von G., Nielsen H., 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth **8**: 785–786.

Pischedda A., Rice W. R., 2012 Partitioning sexual selection into its mating success and fertilization success components. Proc Natl Acad Sci **109**: 2049–2053.

Pitnick S., Brown W. D., Miller G. T., 2001a Evolution of female remating behaviour following experimental removal of sexual selection. Proc Royal Soc B **268**: 557–563.

Pitnick S., Miller G. T., Reagan J., Holland B., 2001b Males' evolutionary responses to experimental removal of sexual selection. Proc Royal Soc B **268**: 1071–1089.

Pizzari T., Snook R. R., 2003 Perspective: sexual conflict and sexual selection: chasing away paradigm shifts. Evolution **57**: 1223–1236.

Poiani A., 2006 Complexity of seminal fluid: a review. Behav Ecol Sociobiol **60**: 289–310.

Pond S. L. K., Frost S. D. W., Muse S. V., 2005 HyPhy: hypothesis testing using phylogenies. Bioinformatics **21**: 676–679.

Pujolar J. M., Pogson G. H., 2011 Positive Darwinian selection in gamete recognition proteins of Strongylocentrotus sea urchins. Mol Ecol **20**: 4968–4982.

Qin D., Xia Y., Whitesides G. M., 2010 Soft lithography for micro- and nanoscale patterning. Nat Protocols **5**: 491–502.

R Core Team, 2015 R: A language and environment for statistical computing.

Reinke V., 2003 Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. **131**: 311–323.

Reuter M., Camus M. F., Hill M. S., Ruzicka F., Fowler K., 2017 Evolving plastic responses to external and genetic environments. Trends in Genetics **33**: 169–170.

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

Rice W. R., 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution **41**: 911–914.

Rice W. R., 1996 Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature **381**: 232–234.

Rice W. R., 1998 *Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproduction isolation* (DJ Howard and SH Berlocher, Eds.). Oxford University Press, New York.

Rice W. R., Chippindale A. K., 2001 Intersexual ontogenetic conflict. J. Evol. Biol. **14**: 685–693.

Rice W. R., Holland B., 1997 The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. Behav Ecol Sociobiol **41**: 1–10.

Richardson C. D., Ray G. J., DeWitt M. A., Curie G. L., Corn J. E., 2016 Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nature Biotechnology 2016 34:3 **34**: 339–344.

Roberts T. M., Ward S., 1982 Membrane flow during nematode spermiogenesis. J Cell Biol. **92**: 113–120.

Roesti M., Kueng B., Moser D., Berner D., 2015 The genomics of ecological vicariance in threespine stickleback fish. Nat Comm **6**: 8767.

Rödelsperger C., Streit A., Sommer R. J., 2013 Structure, function and evolution of the nematode genome. eLS **2**: 1.

Rudel D., Riebesell M., Sommer R. J., 2005 Gonadogenesis in *Pristionchus pacificus* and organ evolution: development, adult morphology and cell–cell interactions in the hermaphrodite gonad. Dev Biol **277**: 200–221.

Ryan M. J., Fox J. H., Wilczynski W., Rand A. S., 1990 Sexual selection for sensory exploitation in the frog *Physalaemus pustulosus*. Nature **343**: 66–67.

Sabeti P. C., Reich D. E., Higgins J. M., Levine H. Z. P., Richter D. J., Schaffner S. F., Gabriel S. B., Platko J. V., Patterson N. J., McDonald G. J., Ackerman H. C., Campbell S. J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E. S., 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419**: 832–837.

Sabeti P. C., Varilly P., Fry B., Lohmueller J., Hostetter E., *et al*., 2007 Genome-wide detection and characterization of positive selection in human populations. Nature **449**: 913–918.

Sackton T. B., Lazzaro B. P., Schlenke T. A., Evans J. D., Hultmark D., Clark A. G., 2007 Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet **39**: 1461–1468.

Schluter D., Price T. D., Rowe L., 1991 Conflicting selection pressures and life history trade-offs. Proc Royal Soc B **246**: 11–17.

Schrider D. R., Hahn M. W., 2010 Gene copy-number polymorphism in nature. Proc Royal Soc B **277**: 3213–3221.

Schrider D. R., Kern A. D., 2016 S/HIC: robust identification of soft and hard sweeps using machine learning. PLoS Genetics **12**: e1005928–31.

Schüpbach T., Wieschaus E., 1991 Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. Genetics **129**: 1119–1136.

Scienski K., Fay J. C., Conant G. C., 2015 Patterns of gene conversion in duplicated yeast histones suggest strong selection on a coadapted macromolecular complex. Genome Biol Evol **7**: 3249–3258.

Sharp P. M., Li W.-H., 1987 Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25**: 58–64.

Shi C., Murphy C. T., 2014 Mating induces shrinking and death in *Caenorhabditis* mothers. Science **343**: 536–540.

Shuster S. M., Wade M. J., 2003 *Mating Systems and Strategies*. Princeton University Press, New Jersey.

Simpson S. J., Sword G. A., Lo N., 2011 Polyphenism in insects. Curr Biol **21**: R738–49.

Singh A., Punzalan D., 2018 The strength of sex-specific selection in the wild. Evolution **72**: 2818–2824.

Sirot L. K., Wong A., Chapman T., Wolfner M. F., 2015 Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. Cold Spring Harbor Perspectives in Biology **7**: a017533.

Slatkin M., 1987 Gene flow and the geographic structure of natural populations. Science **236**: 787–792.

Smith H. E., Ward S., 1998 Identification of protein-protein interactions of the major sperm protein (MSP) of *Caenorhabditis elegans*. J Mol Biol **279**: 605–619.

Smith J. M., Haigh J., 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23–35.

Smith J. M., Hoekstra R., 1980 Polymorphism in a varied environment: how robust are the models? Genetics Research **35**: 45–57.

Stearns S. C., 1989 Trade-offs in life-history evolution. Functional Ecology **3**: 259.

Stroustrup N., Ulmschneider B. E., Nash Z. M., López-Moyado I. F., Apfeld J., Fontana W., 2013 The *Caenorhabditis elegans* lifespan machine. Nat Meth **10**: 665–670.

Sunday J. M., Hart M. W., 2013 Sea star populations diverge by positive selection at a sperm-egg compatibility locus. Ecol Evol **3**: 640–654.

Suzuki Y., Nijhout H. F., 2006 Evolution of a polyphenism by genetic accommodation. Science **311**: 650–652.

Swanson W. J., Vacquier V. D., 1997 The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci **94**: 6724–6729.

Swanson W. J., Vacquier V. D., 2002 Reproductive protein evolution. Annu Rev Ecol Syst **33**: 161–179.

Swanson W. J., Yang Z., Wolfner M. F., Aquadro C. F., 2001 Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci **98**: 2509–2514.

Tan X., Calderon-Villalobos L. I. A., Sharon M., Zheng C., Robinson C. V., Estelle M., Zheng N., 2007 Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature **446**: 640–645.

Tan Y., Bishoff S. T., Riley M. A., 1993 Ubiquitin revisisted. Mol Phylo Evol **2**: 351–360.

Tanphaichitr N., Kongmanas K., Kruevaisayawan H., Saewu A., Sugeng C., Fernandes J., Souda P., Angel J., Faull K., Aitken R., Whitelegge J., Hardy D., Berger T., Baker M., 2015 Remodeling of the plasma membrane in preparation for sperm-egg recognition: roles of acrosomal proteins. Asian J Androl **17**: 574–9.

Tarín J. J., Cano A. (Eds.), 2012 *Fertilization in Protozoa and Metazoan Animals*. Springer, Berlin.

Tarr D. E. K., Scott A. L., 2005 MSP domain proteins. Trends Parasitol. **21**: 224–231.

Teotonio H., Carvalho S., Manoel D., Roque M., Chelo I. M., 2012 Evolution of Outcrossing in Experimental Populations of *Caenorhabditis elegans*. PLoS ONE **7**: e35811–13.

Theriot J. A., 1996 Worm sperm and advances in cell locomotion. Cell **84**: 1–4.

Therneau T. M., Grambsch P. M., Pankratz V. S., 2012 Penalized survival models and frailty. journal of computational and graphical statistics **12**: 156–175.

Thompson J. D., Higgins D. G., Gibson T. J., 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**: 4673–4680.

Tomaiuolo M., Levitan D. R., 2010 Modeling how reproductive ecology can drive protein diversification and result in linkage disequilibrium between sperm and egg proteins. Am Nat **176**: 14–25.

Trivers R. L., 1972 *Sexual selection and parental investment*. Aldine Publishing Company, Chicago.

Trost M., Blattner A. C., Lehner C. F., 2016 Regulated protein depletion by the auxin-inducible degradation system in *Drosophila melanogaster*. Fly **10**: 35–46.

Vacquier V. D., Swanson W. J., 2011 Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harbor Perspectives in Biology **3**: a002931–a002931.

van Doorn G. S., Kirkpatrick M., 2007 Turnover of sex chromosomes induced by sexual conflict. Nature **449**: 909–912.

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

Vielle A., Callemeyn-Torre N., Gimond C., Poullet N., Gray J. C., Cutter A. D., Braendle C., 2016 Convergent evolution of sperm gigantism and the developmental origins of sperm size variability in Caenorhabditis nematodes. Evolution **70**: 2485–2503.

Viguerie N., Montastier E., Maoret J.-J., Roussel B., Combes M., Valle C., Villa-Vialaneix N., Iacovoni J. S., Martinez J. A., Holst C., Astrup A., Vidal H., Clément K., Hager J., Saris W. H. M., Langin D., 2012 Determinants of human adipose tissue gene expression: impact of diet, sex, metabolic status, and cis genetic regulation. PLoS Genetics **8**: e1002959–14.

Vitti J. J., Grossman S. R., Sabeti P. C., 2013 Detecting natural selection in genomic data. Annu Rev Genet **47**: 97–120.

Volpe A. M., Horowitz H., Grafer C. M., Jackson S. M., Berg C. A., 2001 *Drosophila* rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. Genetics **159**: 1117–1134.

Wagstaff B. J., 2005 Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. Genetics **171**: 1083–1101.

Walters J. R., Harrison R. G., 2010 Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. Mol Biol Evol **27**: 2000–2013.

Ward S., Miwa J., 1978 Characterization of temperature-sensitive, fertilization-defective mutants of the nematode *Caenorhabditis Elegans*. Genetics **88**: 285–303.

Ward S., Hogan E., Nelson G. A., 1983 The initiation of spermiogenesis in the nematode *Caenorhabditis elegans*. Dev Biol **98**: 70–79.

Washington N. L., Ward S., 2006 FER-1 regulates Ca2+ -mediated membrane fusion during *C. elegans* spermatogenesis. J. Cell. Sci. **119**: 2552–2562.

Watson P. J., Arnqvist G., Stallmann R. R., 1998 Sexual conflict and the energetic costs of mating and mate choice in water striders. Am Nat **151**: 46–58.

Watterson G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor Popul Biol **7**: 256–276.

Weatherhead P. J., Robertson R. J., 1979 Offspring quality and the polygyny threshold: "the sexy son hypothesis". Am Nat.

Weir, B. S., Cockerham C. C., 1984. Estimating F-statistics for the analysis of population structure. Evolution **38:**1358–1370.

Wigby S., Chapman T., 2005 Sex Peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol **15**: 316–321.

Wilburn D. B., Swanson W. J., 2016 From molecules to mating: rapid evolution and biochemical studies of reproductive proteins. J Proteomics **135**: 12–25.

Wilkinson G. S., Breden F., Mank J. E., Ritchie M. G., Higginson A. D., Radwan J., Jaquiery J., Salzburger W., Arriero E., Barribeau S. M., Phillips P. C., Renn S. C. P., Rowe L., 2015 The locus of sexual selection: moving sexual selection studies into the post-genomics era. J Evol Biol **28**: 739–755.

Williams G. C., 1975 *Sex and evolution.* (RM May, Ed.). Princeton University Press, Princeton, NJ.

Wolfe N. W., Clark N. L., 2015 ERC analysis: web-based inference of gene function via evolutionary rate covariation. Bioinformatics **31**: 3835–3837.

Wolfner M. F., 2002 The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. Heredity **88**: 85–93.

Woodruff G. C., Phillips P. C., Kanzaki N., 2017 Dramatic evolution of body length due to post-embryonic changes in cell size in a newly discovered close relative of *C. elegans*. bioRxiv: 1–34.

Wright A. E., Darolti I., Bloch N. I., Oostra V., Ben Sandkam, Buechel S. D., Kolm N., Breden F., Vicoso B., Mank J. E., 2017 Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nat Comm **8**: 1–10.

Wright A. E., Dean R., Zimmer F., Mank J. E., 2016 How to make a sex chromosome. Nat Comm **7**: 1–8.

Wright A. E., Fumagalli M., Cooney C. R., Bloch N. I., Vieira F. G., Buechel S. D., Kolm N., Mank J. E., 2018 Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. Evol Letters **215**: 403–10.

Wright P. E., Dyson H. J., 2015 Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol **16**: 18–29.

Wright S., 1931 Evolution in Mendelian Populations. Genetics **16**: 97–159.

Wright S., 1951 The genetical structure of populations. Ann Eugen **15**: 323–354.

Xu H., Xiao T., Chen C.-H., Li W., Meyer C. A., Wu Q., Wu D., Cong L., Zhang F., Liu J. S., Brown M., Liu X. S., 2015 Sequence determinants of improved CRISPR sgRNA design. Genome Res **25**: 1147–1157.

Xu X. Z. S., Sternberg P. W., 2003 A *C. elegans* sperm TRP protein required for sperm-egg interactions during fertilization. Cell **114**: 285–297.

Xue Y., Sun D., Daly A., Yang F., Zhou X., Zhao M., Huang N., Zerjal T., Lee C., Carter N. P., Hurles M. E., Tyler-Smith C., 2008 Adaptive evolution of UGT2B17 copy-number variation. Am J Human Genet **83**: 337–346.

Yang X., Schadt E. E., Wang S., Wang H., Arnold A. P., Ingram-Drake L., Drake T. A., Lusis A. J., 2006 Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res **16**: 995–1004.

Yang Z., 1998 Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol **15**: 568–573.

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

Yeaman S., Whitlock M. C., 2011 The genetic architecture of adaptation under migration-selection balance. Evolution **65**: 1897–1911.

Zahavi A., 1975 Mate selection—A selection for a handicap. J Theor Biol **53**: 205–214.

Zhang B., Chambers M. C., Tabb D. L., 2007 Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. J Proteome Res **6**: 3549–3557.

Zhang L., Ward J. D., Cheng Z., Dernburg A., 2015 The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in *C. elegans*. Development **142**: 4374–4384.