ABSTRACT

Title of Dissertation: GENOMIC AND REPRODUCTIVE CONSEQUENCES OF SELF-FERTILITY IN *CAENORHABDITIS* NEMATODES

Da Yin, Doctor of Philosophy, 2019

Dissertation directed by: Professor Eric S. Haag, Department of Biology

The evolution of a new reproductive strategy is expected to be reflected in an organism's genome and impact mating-related traits. Several species of *Caenorhabditis* nematodes have evolved the ability to self-fertilize from their outcrossing ancestors. Comparisons of species with different reproductive strategies may therefore reveal consequences of transition to self-fertilization. We compared chromosome-scale genome assemblies for the outcrossing nematode *Caenorhabditis nigoni* and its recently self-fertile sister species, *C. briggsae*. *C. nigoni* genome resembles that of outcrossing relatives but encodes 31% more protein-coding genes than *C. briggsae*. *C. nigoni* genes lacking *C. briggsae* orthologs were disproportionately small and male-biased in expression, including the *male secreted short (mss)* gene family that encodes sperm surface glycoproteins conserved only in outcrossing species. Sperm of *mss*-null males of an outcrossing species failed to compete with those of wild-type males, despite having normal fertility in non-competitive situations. Restoration of *mss* to *C. briggsae* males was sufficient to enhance sperm competitiveness. These results reveal the pervasive influence of sex on genome content that can be used to identify sperm competition factors. Further I found the fitness of *mss*+ genotype was influenced by

mating system and population subdivision. Specifically, *mss*+ is sufficient to increase male frequency and depress population growth in genetically homogenous androdioecious populations. Using experimental evolution, I demonstrated that when *mss*+ and *mss*-null (i.e. wild-type) genotypes compete, *mss*+ is positively selected in both mixed-mating and strictly outcrossing situations, though more strongly in the latter. I suggest that the lack of inbreeding depression and the strong subdivision thought to characterize natural *Caenorhabditis* populations impose selection on sex ratio that makes loss of *mss* adaptive in self-fertile species.

GENOMIC AND REPRODUCTIVE CONSEQUENCES OF SELF-FERTILITY IN *CAENORHABDITIS* NEMATODES

by

Da Yin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Eric S. Haag, Chair
Professor Najib M. El-Sayed
Professor Carlos A. Machado
Associate Professor Alexandra E. Bely
Associate Professor Antony M Jose

© Copyright by
Da Yin
2019

Acknowledgements

First and foremost, I would like to thank my advisor, Eric Haag. He welcomed me to Maryland and gave me the opportunity to explore science in his lab. His encouragement and guidance have helped me become a better researcher and thinker. I am constantly inspired by his optimism and enthusiasm. I cannot express enough of my gratitude for his support through all the years.

I would like to thank my committee members, Najib El-Sayed, Carlos Machado, Alexandra Bely, Antony Jose, for their valuable insight and helpful discussions throughout my graduate career.

I would like to acknowledge my past and current wonderful lab-mates, with whom I shared my happiness and struggles with. Qinwen Liu, Cristel Thomas, Gavin Woodruff, Rebecca Felde, Shuang Hu, Laren Ryan, John Ficklin, Ivy Obonyo, Yair Lupovitch and all other people I have worked with. I really appreciated the time we had together in the lab and the many discussion we had.

I would like to acknowledge all my collaborators. The one I have worked most closely with is Erich Schwarz. A big part of the Chapter one was made possible by him. The high-quality genome and transcriptome assembly, genome annotation, and phylogenomic searches were contributed by him. I am inspired by and thankful for his scientific rigorousness, expertise in genomics and all the helpful discussions. I would also like to acknowledge Cristel Thomas, who started the work on comparing genomes and discovered the *mss* gene family; Asher Cutter, Caitlin Schartner, Edward Ralston and Barbara Meyer for providing RNA-seq and genomic data; Zhongying Zhao, for

sharing reagents and helpful discussions; Katherine Pinter, Xiaojing Yuan and Iqbal Hamza for helping me with bombardment and worm-sorting; Lee Mendelowitz and Mihai Pop for helping me with genomic data visualization; Steve Mount, Thomas Kocher and Patrick Phillips for experimental suggestions; and people from the Baltimore Worm Club and the UMD Evo-devo and Evo-gen research groups for valuable discussions.

Table of Contents

List of Tables

Chapter 1

List of Figures

Chapter 3

Appendix

Introduction

Evolution of self-fertility and its genomic consequences

Sex between individuals is nearly ubiquitous in eukaryotic life (Schurko et al., 2009; Leonard et al., 2019). However, in multicellular organisms the costs of sex and scarcity of mates sometimes favor the evolution of uniparental reproduction through asexual parthenogenesis or self-fertilization (Bell, 1982). The evolutionary transition from outcrossing to selfing is most commonly found in flowering plants, nematodes, and crustaceans (Weeks, 2012). Selfing provides reproductive assurance when population densities are low, and facilitates colonization of new ecological niches (Baker, 1955; Zierold et al., 2007). Selfing also nearly eliminates the two-fold cost of males and can boost population growth because reproduction is egg-limited (Gibson et al., 2017; Smith, 1978).

Despite the above advantages, selfing can have consequences for both sexual traits and genome content. At the population genetics level, selfing leads to progressive homozygosis over generations and may result in reduced fitness because heterozygosity is lost, and deleterious recessive mutations are unmasked, known as inbreeding depression. Consequently, successful selfing species presumably found a way to continuously purge these deleterious mutations (Crnokrak and Barrett, 2002; Lande et al., 1994). Indeed, unlike their outcrossing congeners, selfing *Caenorhabditis* have healthy, naturally homozygous genotypes and are actually susceptible to outbreeding depression (Dolgin et al., 2007a; Gimond et al., 2013). Further, selfing reduces the effective population size for neutral alleles of autosomal genes (Pollak,

1987). As a consequence, the level of linkage disequilibrium across the genome of selfing species increases (Charlesworth and Wright, 2001; Cutter et al., 2008). In addition, severe population bottlenecks can result as a single self-fertile individual invades a new ecological niche. The above dynamics leads to self-fertilizing species having much lower genetic diversity than diecious species (Artieri et al., 2008; Charlesworth and Wright, 2001; Graustein et al., 2002) and alter the way natural selection acts on mutations. Low standing genetic variation could increase long-term extinction risks for selfers. Since in Rhabditidae, multi-taxon clades of selfing lineages are rare (Denver et al., 2011; Kiontke et al., 2011), it is plausible that extinction tends to happen before speciation.

Studies that compare two closely related species with alternative mating systems can provide insight into the genomic consequences of selfing. One study was carried out in the outcrossing Brassicaceae species *Arabidopsis lyrata,* and the self-fertilizing congener *A. thaliana.* The study reported non-coding DNA and transpososable elements (TEs) to be a major component responsible for the much larger *A. lyrata* genome. Further, the process of DNA loss was indicated to be an ongoing process in the Arabidopsis system by segregation analysis (Hu et al., 2011). Comparing selfing *Capsella rubella* to outcrossing *Capsella grandiflora*, Slotte et al. (2013) found a shift in expression of genes associated with flowering evolution and reduction of purifying selection in the selfing *C. rubella*. In *Caenorhabditis* nematodes, Fierst et al. (2015) found all functional genomic categories are lost in similar proportions in several selfing species; whereas Kanzaki et al. (2018) found transposable elements being the most significant contributor to the large genome size of outcrossing *C. inopinata*

compared to the small genome of *C. elegans*. However, as the taxa compared were highly diverged from each other in these studies, it was unclear how directly or how rapidly a change in mode of reproduction might cause shrinkage in genome size.

Newly discovered species and sequencing of new genomes in *Caenorhabditis* provides an exciting opportunity to look into the consequences of mating system transitions. In the *Elegans* group of *Caenorhabditis*, androdioecy arose independently at least three times from gonochoristic ancestors **(Figure 1)**. *C. elegans, C. briggsae* and *C. tropicalis* have an androdioecious mating system that is characterized by the coexistence of hermaphrodites and males (Chasnov and Chow, 2002; Ellis and Lin, 2014; Haag, 2005; Thomas et al., 2012a). This inference of convergent evolution of androdioecy in *Caenorhabditis* is based on parsimony (the evolution of hermaphroditism from the gonochoristic ancestors requires only three steps), as well as on growing evidence that the transition from selfing to outcrossing by hybridization or simple loss of hermaphrodite spermatogenesis is unlikely to produce a competitive outcrossing lineage (Garcia et al., 2007; Thomas et al., 2012a; Ting et al., 2014). The self-fertilizing *C. elegans* and its close relative *C. briggsae* diverged about 18 MYA and the origin of their sexual mode is estimated to be less than 4 million years ago (Cutter et al., 2008).

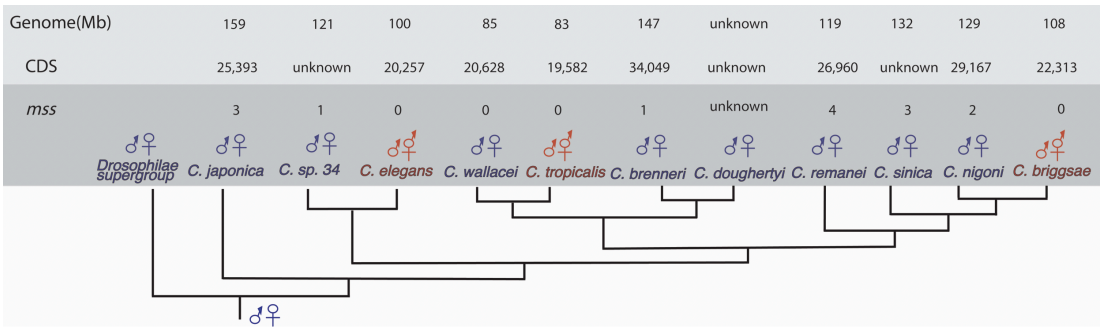

Figure 1. The phylogenetic relationship of *Caenorhabditis* and the evolution of self-fertility Phylogeny of Elegans supergroup *Caenorhabditis* adapted from (Kanzaki et al., 2018; Kiontke et al., 2011; Yin et al., 2018), with outcrossing species producing XX females indicated in blue, and self-fertile lineages with XX hermaphrodites indicated in red.

The genomes of androdioecious and gonochoristic species are dramatically different in size. The selfing *C. elegans, C. briggsae* and *C. tropicalis* have the smallest genome assembly sizes, near or below 100Mb **(Figure 1)**. In contrast, genome sizes of gonochoristic species are 30% larger than that of the androdioecious species (~120Mb or above). Barriere et al., reported high levels of residual heterozygosity exist in the large genomes (Barrière et al., 2009). But after heterozygous alleles are considered and pruned, their genomes are still substantially larger than the selfing species. The only exception is the *C. tropicalis* and *C. wallacei* sister pair, where the selfing *tropicalis* has almost the same (and very small) genome size as its outcrossing counterpart. There are two possible explanations for this. First, the *C. tropicalis* represents a reverted selfing lineage. Second, because the emergence of androdioecy is very recent, it is possible that the common outcrossing ancestor already had a small genome, and there has either not been enough time for the genome of the *C. tropicalis* to shrink, or it is already as streamlined as can be. Similar pattern of small genome size in selfing animal

species is also observed in *Eulimnadia texana*, a clam shrimp that has the smallest crustacean genome size (Baldwin-Brown et al., 2017).

In addition to genome size, RNAseq data indicates the de novo transcriptome assembly sizes of selfing species are only about two-thirds the size of those of the outcrossing species. The transcriptome assemblies of outcrossing species have more contigs, and a greater total assembly size (Thomas et al., 2012a). Taken together, it appears that selfing *Caenorhabditis* have smaller genomes, both in terms of base pairs and expressed genes, than their gonochoristic relatives. Similar pattern of parallel gene loss in selfing lineages was also observed in *Pristionchus* nematodes (Rödelsperger et al., 2018). The mechanisms for the genome size reduction in selfing species and the structural basis of the size difference (e.g. indel size, gene content, distribution on the chromosomes, etc.) remain generally unclear.

A recently discovered phenomenon of transmission genetics in *C. elegans* brought insight into the mechanism for the smaller genome size in selfing species (Wang et al., 2010). During meiosis in *C. elegans* males, the smaller version of the autosome tends to segregate with the X chromosome, whereas the larger version of the autosome tends to segregate away from the X chromosome. The Haag Lab refers to this as Indel Segregation Distortion (ISD). As a result, hermaphrodites inherit the smaller chromosome while males inherit the larger chromosome. As the outcrossing rate in selfing species is low (Barrière and Félix, 2005), the larger variants from the males are expected to be more rapidly lost through genetic drift. Computer simulations show that this phenomenon results in genome shrinkage in the selfing species (Wang et al., 2010).

Role of males in the androdioecious species

In *C. elegans, C. briggsae* and *C. tropicalis,* a very small fraction of males exists. Males are produced from either spontaneous nondisjunction of the X chromosome during meiosis, or through outcrossing. The frequency of males varies in different species and populations (Anderson et al., 2010; Denver et al., 2003; Sivasundar and Hey, 2005; Teotónio et al., 2006). For the selfing *C. elegans*, the common laboratory strain N2 Bristol has a male frequency ≤ 0.002 (Chasnov and Chow, 2002; Teotónio et al., 2006), close to the rate at which they are produced by nondisjunction (Cutter and Payseur, 2003; Hodgkin, 1983; Hodgkin et al., 1979). Another selfing species, *C. briggsae*, on the other hand, has higher frequency of males in the population, presumably due to the higher rate of outcrossing. The frequency of males may be strongly influenced by the male mating ability and hermaphrodite receptivity, which varies among species and populations (Hodgkin and Doniach, 1997). In order to successfully produce outcrossed progeny, a male must first be able to locate a hermaphrodite, initiate a complex suite of copulating behaviors culminating in the transfer of sperm (Emmons, 2006; Garcia et al., 2007). After insemination, male sperm must compete with hermaphrodite's self-sperm for fertilization. Variation among species and strains in any of these steps, either pre-mating or post-mating, could underlie the realized frequency of males (Wegewitz et al., 2008).

The low frequency of males in *C. elegans, C. briggsae* and *C. tropicalis* selfing lineages may often be adaptive. Under benign conditions, hermaphrodites enjoy a population growth rate advantage that allows invasion of male/female populations by eliminating the cost of males (Katju et al., 2008; Stewart and Phillips, 2002). In

addition, selfing provides reproductive assurance where mates or pollinators are limited, facilitating dispersal and colonization in new ecological niches (Baker, 1955; Theologidis et al., 2014; Zierold et al., 2007). Finally, the process of mating with males can physically and physiologically damage the hermaphrodites (Palopoli et al., 2015; Ting et al., 2014; Woodruff et al., 2014). This appears in part to be due to genetic conflict where males hijack the hermaphrodite's reproductive system for maximum reproductive output at the cost of hermaphrodite's life span (Shen et al., 2012; Shi and Murphy, 2014).

Experimental evolution studies reveal that the frequency of both males and outcrossing become elevated in androdioecious populations with increased mutation rate and/or facing drastic environmental changes, such as coevolving parasites (Anderson et al., 2010; Morran et al., 2009; Slowinski et al., 2016). A similar benefit for outcrossing in facilitating adaptation has also been observed in hermaphroditic snails (Noël et al., 2017).

Several models have been proposed for the evolution of androdioecy and the maintenance of males (Hedgecock, 1976; Otto et al., 1993; Stewart and Phillips, 2002). In theory, selfing rate, male fertilization success (α), inbreeding depression, the rate of nondisjunction at the X chromosome, relative viability difference between males and hermaphrodites all can impact the frequency of males. However since inbreeding depression is low in long-selfing species (Dolgin et al., 2007a; Gimond et al., 2013; Johnson and Hutchinson, 1993), X nondisjunction is rare, and male and hermaphrodite viability are equal in the early, reproductive phase of life, a simplified model for the

change of male frequency overtime would be determined largely by male fertilization success and initial male frequency in the population.

The simultaneous benefits of selfing and the need for at least occasional outcrossing may create an intermediate optimum sex ratio in facultative selfers like *C. elegans*. The XX-XO sex chromosome system would further imply that such modulation of ratio could be implemented at the level of mating success.

Degradation of mating related traits

The efficiency of mating in self-fertile *C. elegans* and *C. briggsae* is lower than outcrossing species, such as *C. remanei* (Anderson et al., 2010; Chasnov et al., 2007a; Garcia et al., 2007; Hodgkin, 1983; Kleemann and Basolo, 2007; Noble et al., 2015; Wegewitz et al., 2008; Yin et al., 2018). This is presumably because selfing species do not rely on mating for reproduction and traits related to mating are under relaxed selection. The poor mating efficiency of the male/hermaphrodite species starts at the precopulatory stage. Using chemotaxis assays, Chow et al. (2007) reported that *C. remanei* females readily secrete potent sex pheromones that attract conspecific males as well as males from C. *elegans* and *C. briggsae* from distance. In contrast, hermaphrodites secrete much weaker attractants (Leighton et al., 2014). Mate-finding behavior is also compromised. Lipton et al. (2004) reported that in the self-fertilizing *C. elegans* and *C. briggsae,* males have higher rate of leaving the bacteria lawn and wandering around in search of a mating partner while hermaphrodites are reluctant to leave food and look for mate (Lipton et al., 2004). In contrast, in gonochoristic *C. remanei* and *C. japonica*, males and females are both shown to leave the food source

and search for a partner at equal frequency. In addition, males from the selfing species do not discriminate between sexes and can exhibit same-sex and autocopulatory mating (Garcia et al., 2007; Noble et al., 2015). Of all the mating substeps, spicule insertion behavior is the most difficult for males to perform on the hermaphrodites. Garcia et al. (2007) showed *C. remanei* males penetrated their conspecific females instantaneously but did not penetrate *C. briggsae* hermaphrodites very efficiently. Similarly, *C. briggsae* males showed the same pattern of easily inserting their spicules into *C. remanei* females yet having trouble penetrating their own hermaphrodites.

In addition, Kleeman and Basolo (2007) demonstrated that hermaphrodites' mating decisions could be influenced by self-sperm status. When a male copulates with a hermaphrodite that has already developed sperm, the hermaphrodite tends to eject the male's sperm. Accordingly, the chance of hermaphrodite oocytes being fertilized by the male sperm becomes small. Furthermore, self-sperm-containing hermaphrodites also showed mating resistance by rapidly crawling away as males approach in an attempt to mate. The combined precopulatory and postcopulatory male sperm avoidance illustrates the low outcrossing efficiency in the self-sperm containing hermaphrodites.

Selfing species also invest minimally in their gametes. Seven hermaphrodites in the *Rhabditidae* produce much smaller sperm than those from males of their own species or from those of gonochoristic species (Baldi et al., 2011; LaMunyon and Ward, 1999). Accordingly, the hermaphroditic sperm crawl much slower compared to male sperm. In addition, sperm from males in selfing species are smaller than those from males in outcrossing species (Gimond et al., 2018; LaMunyon and Ward, 1998, 1999,

2002; Murray et al., 2011; Palopoli et al., 2015; Vielle et al., 2016). Larger sperm were shown to have a competitive advantage. It is plausible that the general degradation of mating related traits has its roots in the sweeping genomic changes that accompany it.

Present study

In the first chapeter, we compare the chromosome-scale genome assemblies for the outcrossing nematode *Caenorhabditis nigoni* and its closely related self-fertile sister species, *C. briggsae*. The landscape of structural variation will be characterized, including the insertion/deletion size, distribution, and mutational mechanisms inferred. In addition, we will characterize the genes present in *C. nigoni* but absent in *C. briggsae,* and vice versa, and infer their functional roles in reproductive biology. In the second chapter, we will characterize a *male secreted short (mss)* gene family that is conserved only in outcrossing species. The sequence features and phylogenomic conservation details will be examined; the expression pattern and cellular localization will be characterized; the function will be assessed through CRISPR knockout and micro-particle bombardment transformation. In the third chapter, we will use experimental evolution and mathematical modeling to assess how population structure and mating system influence the fitness of *mss* gene. We will also test the hypothesis that loss of *mss* in selfing species is adaptive through changing sex ratio and impact population growth.

Chapter 1 Genomic consequences of self-fertilization

Summary

The shift in reproductive system from outcrossing to selfing evolves rapidly in some eukaryotic lineages. The transition is expected to have pronounced consequences for genome size, mating related traits and the nature and intensity of sexual selection. Previous studies indicate that selfing species have smaller genomes and several thousand fewer protein-coding genes than their outcrossing ancestors (Fierst et al., 2015; Thomas et al., 2012b). However, as the taxa compared were highly diverged from each other, it was unclear how directly or how rapidly a change in mode of reproduction might cause shrinkage in genome size. In addition, the size, location, and gene content of specific deletions remain obscure for any animal system. Further whether the genomic reduction is simply due to the different sexual modes remains to be seen.

The *Caenorhabditis* genus of nematodes has a unique mix of closely related outcrossing and selfing species. The self-fertile C. *briggsae* (population has mostly hermaphrodites and a very small number of males) has a close male-female relative C. *nigoni*, yet C. *briggsae*'s genome is 20% smaller. These are the most closely related *Caenorhabditis* species known with alternative sexual modes (having diverged roughly one million years ago). In this chapter, our aim is to characterize the process of genome shrinkage and infer mutational mechanisms where possible.

First, we generated a high-quality, chromosome-scale *de novo* genome assembly of the male-female species *C. nigoni* to enable its comparison with the genome of its close self-fertile relative *C. briggsae*. Strikingly, and despite their partial interfertility, *C. briggsae* has 23% fewer protein-coding genes than *C. nigoni.* As *C. nigoni's* genome size is typical of related male-female *Caenorhabditis* species, we infer that the genome and proteome of *C. briggsae* has recently and dramatically contracted.

C. nigoni's six chromosomes are 6.6-16.6% larger than their *C. briggsae* homologs. While the two genomes are essentially colinear, they differ in many small species-specific segments. Most of the lineage specific DNA are located on the chromosomal arms where structural rearrangements and repeats are more frequent, and gene densities are lower.

The *C. nigoni* genes missing in *C. briggsae* encode proteins that are shorter and more male-biased in their expression than would be expected by chance, suggesting that small protein-coding genes related to male sexual traits are particularly susceptible to loss. The genome of the more widely studied self-fertile *C. elegans* also exhibits this shrunken state. Absent from *C. elegans* are hundreds of genes conserved between male-female *Caenorhabditis* species. Given the low frequency of males in self-fertilizing species, our observations point to sexual selection as a driving force in maintaining much of the genome's content. Overall, these findings demonstrate a strikingly large role for the mode of reproduction in determining the size and content of animal genomes.

Materials and Methods

Strains. The inbred wild-type *C. nigoni* strains JU1421 and JU1422 were obtained from Marie-Anne Félix (Ecole Normale Supérieure) and from the *Caenorhabditis* Genetics Center (CGC). Because each was derived by 25 rounds of inbreeding from the wild isolate strain JU1325(Kozlowska et al., 2012), we expected that they would be largely homozygous, but with some residual heterozygosity wild-type *C. briggsae* strain AF16 (Golden et al., 1983; Stein et al., 2003) and *C. remanei* strains EM464 (Baird et al., 1994; Sudhaus, 1974) and SB146 (Baird, 1999; Sudhaus, 1974) were also obtained from the CGC. *C. briggsae unc-119(nm67)* (Liu et al., 2012) and *she-1(v35)* (Guo et al., 2009) mutants were derived from AF16. All species were cultured on standard NGM plates, supplemented with additional agar (to 2.2%) to discourage burrowing. The *C. briggsae* strains RW20025 and JU936 were gifts from Zhongying Zhao (Hong Kong Baptist University) and Marie-Anne Félix, respectively.

Genomic DNA purification. For Illumina sequencing of *C. nigoni* JU1422, genomic DNA was prepared by methods optimized for short-insert paired-end reads (Quail et al., 2008). However, PacBio sequencing of *C. nigoni* JU1422 required genomic DNA of very high molecular weight that was free of contamination by both RNA and sequencing inhibitors (such as EDTA). To avoid fragmenting genomic DNA, wide-bore pipette tips were used throughout, generated by cutting off the ends of normal pipette tips with a sterile razor blade. Worms were grown on NGM plates, washed off with M9 buffer, and rinsed 2-3 times with microcentrifugation (1200 rpm

for 1 minute) at 4°C. The worm pellet was washed once in disruption buffer (200 mM NaCl; 50 mM EDTA; 100 mM Tris, pH 8.5) without SDS and again microfuged for 1 minute at 4°C. The pellet was then resuspended in 5 volumes of disruption buffer with a final concentration of 0.5% (w/v) added SDS, refrigerated in a -80°C freezer to soften tissues, and thawed to room temperature. Proteinase K (20 mg/ml stock solution) was added to a final concentration of 100-200 µl/ml, and the mix was incubated for ~5 hours at 68°C until dissolved, with periodic mixing by gentle inversion. Co-purified RNA was digested before phenol-chloroform extraction by adding 2 µl RNase A stock (10 mg/ml) for each 50 µl of sample, and incubating for 30 min. at 37°C. One volume of a 1:1 mixture of chloroform and buffer-saturated phenol (pH ~8) was added to the sample, and the sample's microfuge tubes were gently inverted until phases were mixed. Tubes were then microfuged at 13,000 RPM at room temperature for 2 minutes. The upper aqueous layer with its genomic DNA was similarly extracted two more times, followed by a final extraction with 24:1 chloroform/isoamyl alcohol. In a fresh tube, 0.1 volumes of 3 M sodium acetate (pH 5.2) and 1 volume of 100% isopropanol were added and mixed by gentle inversion. After incubation at -20°C overnight, DNA was pelleted by microfuging at 13,000 RPM for 10-15 minutes at 4°C. The DNA pellet was washed with 70% ethanol, inverting the sealed tube a few times, and reseating the DNA pellet with a spin at 13,000 RPM for 5 minutes at 4°C. Remaining supernatant was removed by pipetting, and the pellet allowed to air-dry at room temperature, taking care not to overdry it. The genomic DNA pellet was then resuspended in 50 µl EB buffer from Qiagen; to avoid shearing, the high molecular weight DNA was not resuspended by pipetting, but rather was allowed to slowly dissolve overnight at 4°C.

Genomic DNA sequencing. Genomic DNA from JU1422 was subjected to Blue Pippin selection, and sequenced to ~96x coverage (given an initial genome size estimate of 130 Mb) with Pacific Biosciences (PacBio) 20-kb libraries at the Cold Spring Harbor Laboratory genome facility. These data totaled 12,538,321,717 nt in 1,480,550 reads, with a mean length of 8,469 nt (maximum, 47,453 nt; minimum, 50 nt). In addition, we also sequenced JU1422 genomic DNA to ~100x coverage with Illumina paired-end libraries at the UC Berkeley genome facility.

Genome assembly. We generated an initial genome assembly from our PacBio data by using PBcR-MHAP 8.3rc2 (Berlin et al., 2015) (*https://sourceforge.net/projects/wgs-assembler/files/wgs-assembler/wgs-8.3*) to self-correct the longest 40x coverage of reads, and then to assemble the longest 25x coverage of reads. This yielded an initial assembly of 143 Mb and 485 contigs, with a contig N50 of 2.2 Mb. Individual PacBio reads can be used to improve existing genome assemblies by linking existing sequences. To do this with our decontaminated *C. nigoni* genome, we used PBJelly2 from PBSuite 15.8.24 (English et al., 2012) (*http://sourceforge.net/projects/pb-jelly/files*) and our error-corrected 40x coverage of reads. The accuracy of PacBio-based genome assemblies can be driven to final consensus accuracies in excess of 99.999% (QV of >50) by detailed comparison to the raw pulse and base-call information of their original PacBio read files. To do this with our PBJelly2-scaffolded *C. nigoni* genome, we used Quiver (Chin et al., 2013) from SMRTanalysis 2.3.0. Genome assemblies of diploid, heterozygous organisms can contain allelic sequences that can be computationally resolved into a single quasi-haploid consensus assembly. To resolve such minor alleles in the Quiver- and Pilon-polished *C. nigoni* genome assembly, we

used HaploMerger2 20151106 (Huang et al., 2017) (*http://mosas.sysu.edu.cn/genome/download_softwares.php*) by protocols given in the software documentation. We further scaffolded our *C. nigoni* genome assembly with *C. nigoni* non-genome-guided cDNA (assembled from RNA-seq data) and *C. briggsae* peptide sequences. To determine the completeness and homozygosity of our successive genome assembly versions, we used CEGMA 2.4 (Parra et al., 2009).

RNA-seq data. We used biologically triplicated RNA-seq data from male and female adult *C. nigoni*. These data for *C. nigoni* were generated as single-end 100-nt reads from the strain JU1421, which like JU1422 also is an inbred derivative of JU1325 (Kozlowska et al., 2012). We also used RNA-seq data for mixed-sex whole-animal *C. nigoni* that had been previously generated by the modENCODE consortium (Gerstein et al., 2010); We assembled all of our RNA-seq data into cDNA with Trinity 2.2.0 (Grabherr et al., 2011) (*https://github.com/trinityrnaseq/trinityrnaseq/releases*), in both genome-guided and ab initio form. The genome-guided cDNA assembly was subsequently used to guide gene parameter construction and gene prediction; to avoid circularity, the non-genome-guided cDNA assembly was subsequently used to enable cDNA scaffolding of the genome assembly. For the *C. nigoni* RNA-seq data, we generated expression values in transcripts per million (TPM) and estimated mapped read counts per gene with Salmon 0.7.0 (Patro et al., 2017). Equivalent analyses were performed for male-specific and female/hermaphrodite-specific RNA-seq data of *C. briggsae*, *C. remanei*, *C. brenneri*, and *C. elegans*, previously published by Thomas et al. (Thomas et al., 2012b).

Chromosome tiling. To determine a subset of the *C. nigoni* genome assembly that aligned syntenically (tiled) onto the *C. briggsae* genome, we used *show-tiling* from MUMmer 3.23 with the arguments "*-l 1 -g -1 -i 80.0 -v 1.0 -V 0*". We used the resulting tiling data with custom Perl scripts, along with visual inspection of MUMmer alignments to *C. briggsae*, to generate a chromosomally-aligned and pseudochromosomal version of the *C. nigoni* genome. Within each pseudochromosome, we linked successive contigs or scaffolds with 1000-nt blocks of N residues. This yielded a final assembly with 6 major pseudochromosomal superscaffolds that totaled 118 Mb in length, and a residuum of 150 unaligned contigs that totaled 11.7 Mb.

Gene predictions, annotations and orthologies. We used AUGUSTUS 3.2.2 (Stanke et al., 2008) (*http://bioinf.uni-greifswald.de/augustus/binaries/augustus-3.2.2.tar.gz*) to predict protein-coding genes. We determined motifs and traits for protein products of our predicted gene set as follows. We predicted signal sequences and transmembrane sequences with Phobius 1.01 (Käll et al., 2004), coiled-coils with NCoils (Lupas, 1996), and low-complexity domains with PSEG (Wootton, 1994). We predicted protein motifs from two databases: the PFAM database with hmmscan in HMMER 3.1b2 (Eddy, 2009; Finn et al., 2016). We predicted Gene Ontology (GO) terms (Gene Ontology Consortium, 2015) for gene functions with command-line Blast2GO v1.3.3 (Götz et al., 2008). Orthologies between genes were computed with Orthofinder 0.7.1 (*https://github.com/davidemms/OrthoFinder/releases/download/0.7.1/OrthoFinder-0.7.1.tar.gz*) (Emms and Kelly, 2015), aided by BLAST+ 2.2.31, mcl 11335, and SciPy 0.9.0. For analyses of motifs, GO terms, and orthology, we used

only the longest-predicted isoform for each gene in a genome as a representative of that gene. For comparisons of ortholog family size and protein length, We looked at the membership size for each OrthoFinder group (excluding 1-1 orthologs) from both *C. nigoni* and *C. briggsae* (**Figure 3a**), and calculated protein length distributions from the longest isoforms predicted by AUGUSTUS (**Figure 3b**). *Gene ontology (GO) enrichment analysis. C. nigoni* genes were classified by OrthoFinder as either having or lacking *C. briggsae* homologs. To find GO terms enriched in the *C. nigoni* gene set lacking *C. briggsae* homologs, we used the R package topGO (Alexa et al., 2006), and calculated the p-values of GO terms with the default method 'weight01'.

Pfam domain enrichment analysis. We ran custom R scripts on our Pfam domain analyses of *C. nigoni* (Supplementary Dataset S1) and *C. briggsae* (Supplementary Dataset S3) to count the genes associated with each Pfam domain in each species, and then to use Fisher's exact test to determine which Pfam domains in *C. nigoni* were significantly overrepresented by comparison with their abundance in *C. briggsae*. We likewise ran custom R scripts on our Pfam domain analyses of *C. elegans*, *C. remanei*, and *C. briggsae* to count the genes associated with each Pfam domain for each species, and to use Fisher's test in pairwise comparisons to find the Pfam domains overrepresented in outcrossing species as compared to the selfing species.

Aligning genomes and measuring indel sizes, frequencies, inversions, repeats, structural features and gene densities. We used *nucmer* from MUMmer 3.23 with the arguments "*-mincluster 35 -masgap1000 -minmatch 10*" to align the six *C. nigoni* pseudochromosomes to the *C. briggsae* chromosomes, which were themselves

generated from a high-resolution recombination map (Ross et al., 2011). We used *dnadiff* from MUMmer 3.23 to generate coordinates for the differences between the two genomes using default parameters. *nucmer* and *dnadiff* generated alignment statistics, gaps, duplications, inversions, etc. We calculated the frequency of indels of different sizes and used the computing language R to generate bar graphs. In addition, we calculated the probability of sequence deletions using a 200-kb non-overlapping sliding window. The probability equals the size of deletions divided by the 200-kb window length. We calculated repeat frequencies based on our analyses of the *C. nigoni* and *C. briggsae* genomes with RepeatModeler and RepeatMasker; inversion frequencies from *dnadiff*; and gene density (using, for each gene, the longest isoforms) from our AUGUSTUS predictions. We plotted these features circularly (**Figure 2**), using the same 200-kb window, with Circos-0.69-3 (*http://circos.ca/software/download/circos*) (*100*). The genome composition for CDS, intron, and intergenic sequences were extracted from the GFF3 file of our AUGUSTUS gene predictions. We used bedtools v2.26.0 (*http://bedtools.readthedocs.io*) (Quinlan and Hall, 2010) to intersect the coordinates of genomic features to the coordinates of differences between the genomes, and diagrammed the overlap.

Determining and comparing summed exon and intron lengths of orthologous genes. We selected strictly orthologous protein-coding genes for analysis that had the following properties. First, they were identified in *C. nigoni* as being either autosomal or X-chromosomal; we ignored genes whose chromosome in *C. nigoni* was unassigned. Second, they were predicted to have a single, strict ortholog by our OrthoFinder analysis in each of the following proteomes: *C. nigoni*; *C. briggsae* (the official gene

prediction set, from WormBase release WS254); *C. briggsae* (our prediction set, "briggsae-alt"); *C. remanei* (the official gene prediction set by Fierst et al., from WS254); *C. remanei*, our prediction set ("remanei_alt"); and *C. elegans* (official gene prediction set from WS254). Imposing this orthology requirement allowed us to make direct comparisons of summed exon and intron size for these genes for any pair of species and predictions. Using our own gene predictions for different species, and comparing these (for *C. briggsae* and *C. remanei*) to earlier gene prediction sets by other investigators, allowed to control for differences that might be due to different gene prediction methods. These criteria yielded 6,404 sets of orthologous autosomal genes and 1,394 sets of orthologous X-chromosomal genes for exon/intron size comparisons. For each gene in each set, we identified the isoform that encoded the largest protein product, and used that isoform for all exon/intron sums and comparisons. To enforce uniform predictions of exons, we extracted the coordinates of protein-coding sequences (CDS), which were available for all six gene sets and which did not depend on further predictions of non-coding exons (which were available for only some gene sets, and which were likely to be less reliable). For the largest isoform of each gene, we used custom Perl scripts to extract its CDS coordinates from its genome annotation file (in GFF3 format), to infer the coordinates of intervening introns from these CDS coordinates, and to sum up the sizes of both exons and introns. For mass comparisons of exon and intron sets we used the *gsl_fit_linear* and *gsl_stats_correlation* functions of the Math::GSL module in Perl to infer slopes, y-intersects, and Pearson r^2 correlations for sizes of exons and introns. GFF3 annotation

files for *C. briggsae*, *C. remanei*, and *C. elegans* were downloaded from WormBase release WS254.

Sex-biased expression analysis. We defined sex-biased expression, or the lack of it, as follows. Having determined gene expression levels by RNA-seq using biological triplicates of males and females, we defined a gene's expression as male-biased if it was ≥2-fold greater in males than in females (Male.v.Fem.logFC ≥ 1) with a false discovery rate (FDR; corrected for multiple hypothesis testing) of ≤ 0.001. We similarly defined a gene's expression as female-biased if it was ≥2-fold greater in females than in males (Male.v.Fem.logFC ≤ -1) with an FDR ≤ 0.001. All other genes with intermediate expression ratios (-1 < Male.v.Fem.logFC < 1) and for which these intermediate ratios were determined with high significance (FDR ≤ 0.001) were classified as unbiased. These criteria exclude genes whose male-female expression ratios were determined with weaker statistical significance, or for which edgeR assigned no significance at all. We chose these criteria (≥2-fold difference in expression, FDR ≤ 0.001) because they are likely to come closest to accuracy for small numbers of biological replicates (Schurch et al., 2016). To detect over- or under-representation of sex-biased expression in conserved versus non-conserved *C. nigoni* genes, we used our OrthoFinder groups to distinguish *C. nigoni* genes with or without *C. briggsae* homologs, compared the frequencies of sex-biased expression within both groups, and determined their significance with Fisher's exact test.

Results

Of ~50 known *Caenorhabditis* species, the most closely related pair with different sexual modes are the outcrossing *C. nigoni* and the selfing *C. briggsae* (Cutter; Félix et al., 2014; Woodruff et al., 2010). They remain partially interfertile, yet have numerous genetic and reproductive incompatibilities (Bi et al., 2015; Bundus et al., 2015; Li et al., 2016; Ting et al., 2014; Woodruff et al., 2010; Yan et al., 2012). To compare their genomes, we assembled the *C. nigoni* genome from 20-kb Pacific Biosciences (PacBio) and Illumina short-read libraries (**Table 1**). The final *C. nigoni* chromosome-scale genome assembly totaled 129 Mb with an N50 contig length of 3.3 Mb; it was estimated as 99.6% complete (Parra et al., 2009). The genome was 19% larger than *C. briggsae*'s (108 Mb), but similar in size to genomes of the more distantly related outcrossing species *C. remanei, C. sinica, C. brenneri,* and *C. japonica,* which range from 131-135 Mb (**Figure 1**) (Fierst et al., 2015). Therefore, larger genome sizes were probably the ancestral condition, and genomic shrinkage occurred in the *C. briggsae* lineage after it diverged from *C. nigoni*. Over 90% (118 Mb) of the assembly can be aligned to the chromosomes of *C. briggsae* without large translocations or inversions, despite megabase-sized contigs (**Figure 1**). Thus, the two genomes are essentially colinear, but differ in many small species-specific segments. *C. nigoni*'s six chromosomes are 6.6-16.6% larger than their *C. briggsae* homologs (**Table 2**).

Species:	***C. nigoni***	***C. briggsae***	***C. remanei***	***C. brenneri***	***C. elegans***
Total nt:	129,488,540	108,384,165	118,549,266	147,122,675	100,286,401
Scaffolds:	155	367	1,591	1,859	7
Contigs:	213	6,724	48,142	8,900	7
ACGT nt:	129,435,387	105,416,539	112,435,984	133,694,091	100,286,401
N-res. nt:	53,153	2,967,626	6,113,282	13,428,584	0
% non-N:	99.96	97.26	94.84	90.87	100.00
% GC:	37.75	37.35	37.89	38.58	35.44
% repetitive:	27.31	26.04	16.99	n/d	21.95
Scaffold N50 nt:	20,390,332	17,485,439	1,522,088	760,442	17,493,829
Scaffold N90 nt:	15,535,478	14,578,851	98,048	82,237	13,783,801
Scaf. max. nt:	23,648,458	21,540,570	18,579,143	4,147,112	20,924,180
Scaf. min. nt:	1,315	1,378	864	538	13,794
Contig N50 nt:	3,254,670	41,490	14,669	37,413	17,493,829
Contig N90 nt:	562,841	9,261	2,027	8,044	13,783,801
Contig max. nt:	9,436,569	516,571	207,691	448,409	20,924,180
Contig min. nt:	1,315	1	1	1	13,794
Protein-coding genes (this study):	29,167	22,313	26,960	34,049	[none]
Protein-coding genes (WB):	[none]	21,814	26,226	[none]	20,257
CEGMA genes	247	247	238	247	244
CEGMA % completeness	99.60	99.60	95.97	99.60	98.39
CEGMA Average	1.19	1.13	1.19	1.24	1.12

Table 1. Genome statistics for *C. nigoni* and other *Caenorhabditis* in this study. For *C. nigoni* and *C. briggsae*, repetitive DNA analyses were from this study; for *C. remanei* and *C. elegans*, they were taken from WormBase release WS254 . "% repetitive" was computed with respect to the number of non-scaffolding (non-N) residues, not the total assembly size. Protein-coding genes were from WormBase WS254 (in the cases of *C. briggsae* and *C. remane*i, from both sources). CEGMA scores were taken from complete, full-length matches to 248 single-copy eukaryotic

genes (Bundus et al., 2015): they denote the number of genes detected, the equivalent completeness score, and the average number of hits for these 248 genes in the genome.

C. nigoni	**chrI**	**chrII**	**chrIII**	**chrIV**	**chrV**	**chrX**
Size (Mb)	16.7	19.2	15.5	20.4	22.3	23.6
Species-specific DNA	5.2 (31%)	6.7 (35%)	5.1 (33%)	8.1 (40%)	7.3 (33%)	8.3 (35%)
Species-specific DNA (> 1kb)	2.1 (13%)	2.8 (15%)	2.0 (13%)	3.7 (18%)	3.2 (15%)	3.8 (16%)
C. briggsae	**chrI**	**chrII**	**chrIII**	**chrIV**	**chrV**	**chrX**
Size (Mb)	15.5	16.6	14.6	17.5	19.5	21.5
Species-specific DNA	3.7 (24%)	4.2 (25%)	3.7 (25%)	5.1 (29%)	4.4 (23%)	5.6 (26%)
Species-specific DNA (> 1kb)	0.67 (4.3%)	0.80 (4.8%)	0.68 (4.7%)	1.1 (6.3%)	0.80 (4.1%)	1.2 (5.4%)
***C. nigoni* size difference**	8.3%	15.7%	6.6%	16.6%	14.3%	9.8%

Table 2. Chromosome sizes and species-specific sequence content. Sizes, total unalignable (species-specific) sequences, and the subset of species-specific sequences larger than 1 kb for a given chromosome are shown for *C. nigoni* (top) and *C. briggsae* (bottom).

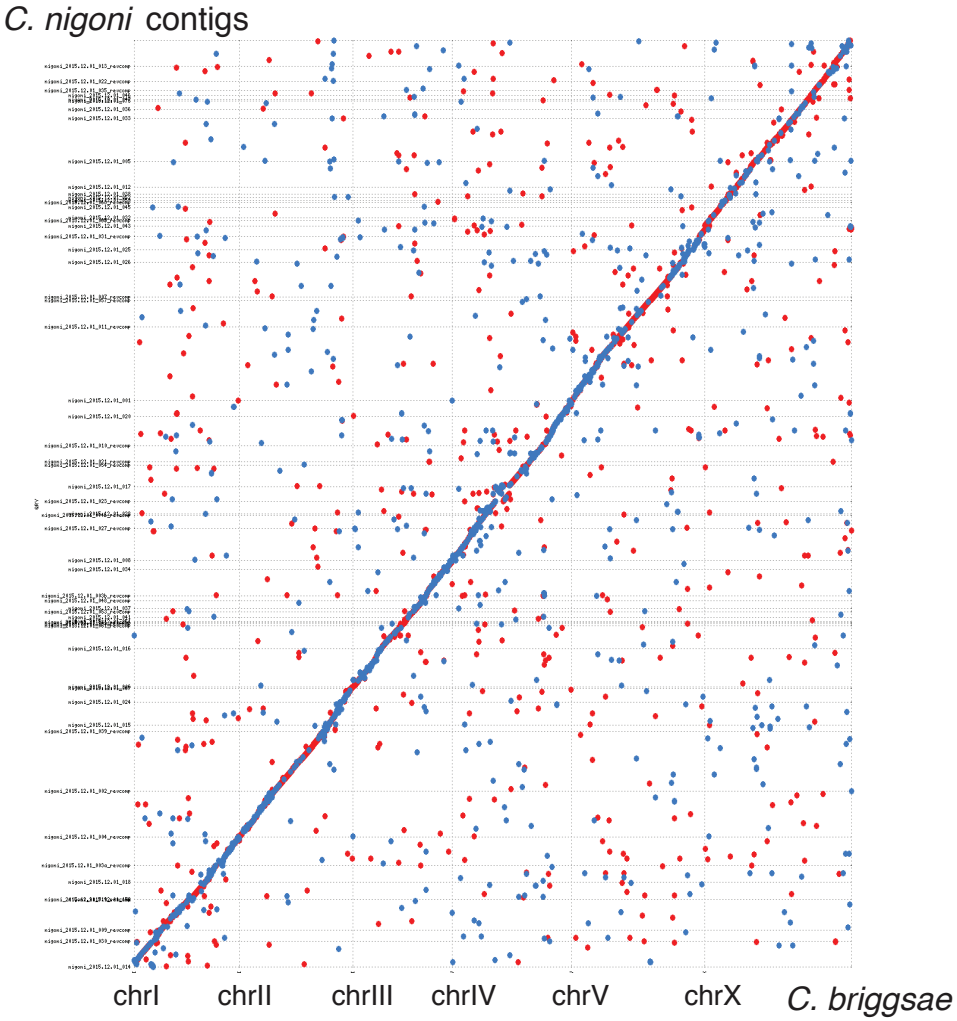

Figure 1. Tiling of *C. nigoni* contigs against *C. briggsae* chromosomes. MUMMER 3.23 was used to align *C. nigoni* contigs (after their full genome assembly, but before their being fused *in silico* into pseudochromosomes) against six chromosomes from the CB4 genome assembly of *C. briggsae*. Maximal unique matches (MUMs) are depicted as dots: red indicates forward matches; blue indicates reverse matches. Scaffolds from the CB4 assembly that were unassigned to chromosomes were omitted from the alignment.

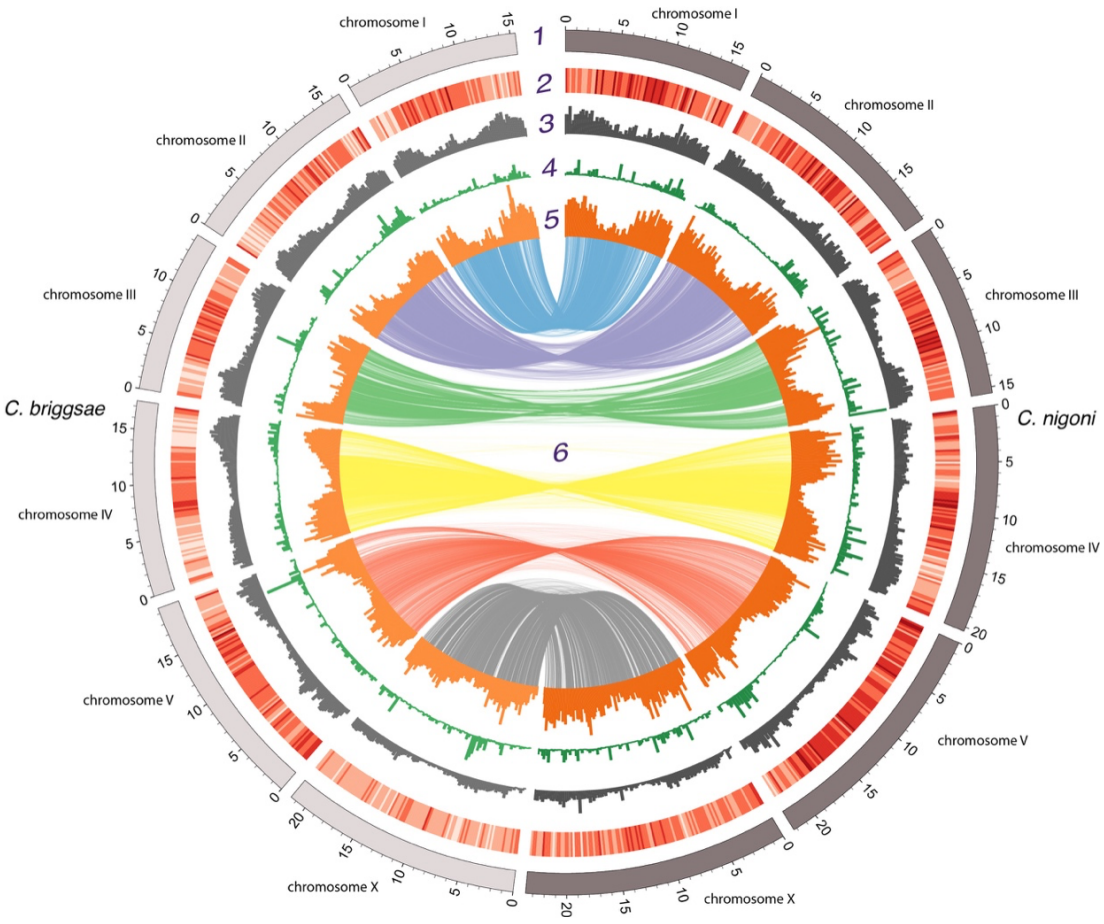

Figure 2. Comparison of the *C. nigoni* and *C. briggsae* genome assemblies. a, Circular visualization of the chromosomal alignments and genomic features over 200-kb chromosomal intervals. Tracks from outside to inside **1,** Positions (in Mb) of the six chromosomes of *C. nigoni* and *C. briggsae,* **2,** gene density heat map (darker shade means higher density), **3,** repeat frequency, **4,** inversion frequencies, **5,** percentage of sequence lacking homology in the other assembly (representing either deletions or species-specific gains), **6,** DNA sequence synteny.

We used whole-genome alignment to identify species-specific genomic segments. In *C. nigoni*, 47.7 Mb (36.9%) did not align with *C. briggsae*, and *C. briggsae* had 27.7 Mb (25.6% of 108.4 Mb) that did not align with *C. nigoni*. This 20.0 Mb difference accounted for 95% of the difference in genome sizes. Non-alignable genomic regions were concentrated on the distal arms of all six holocentric chromosomes, where small inversions and repetitive sequences were abundant and gene densities were low (**Figure 2**). These regions were mostly small (median ~500 bp; **Figure 3a**), but larger (1-65 kb) insertions or deletions accounted for 17 Mb (81%) of the genome size difference (**Figure 3b**). In both assemblies, non-alignable sequences were most common in intergenic regions and introns (**Figure 4a, b**). *C. nigoni* harbored 5.4 Mb more species-specific protein-coding sequences than *C. briggsae*, consistent with a net loss of genes in *C. briggsae* (see below). For orthologous genes in both species, exon lengths were highly correlated (**Figure 3c, 4c**). In contrast, ortholog intron content was weakly correlated and was significantly larger in *C. briggsae.* Because both genomes had similar repetitive DNA fractions (*C. nigoni* 27% versus *C. briggsae* 25%), disproportionate loss of repetitive sequences (seen in plants) did not contribute to different genome sizes (Fierst et al., 2015; Hu et al., 2011; Wright et al., 2008).

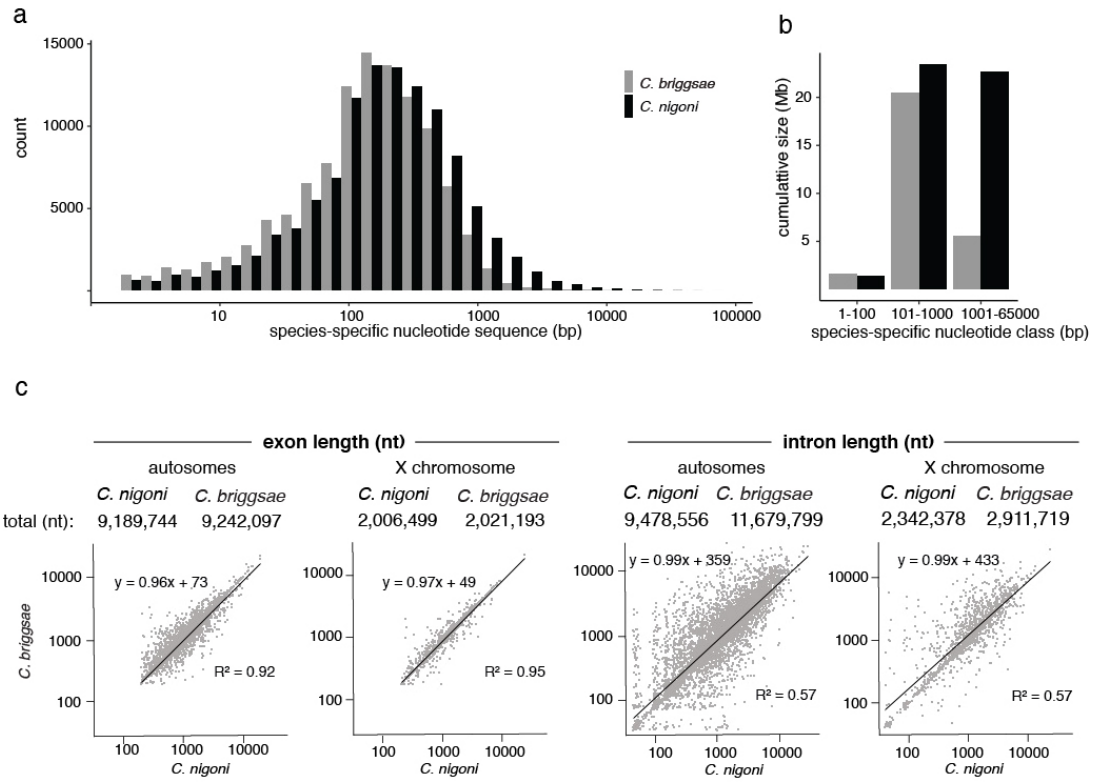

Figure 3. Size distribution of insertion-deletion variants. **a,** Size distribution of species-specific sequences in the *C. briggsae*-*C. nigoni* whole-genome alignment. Black, sequences present in *C. nigoni* alone; grey, sequences present in *C. briggsae* alone. **b,** Contribution of different species-specific sequence types to genome size. **c,** Regression analysis of total exon and intron lengths for 6,404 1:1 *C. briggsae*-*C. nigoni* orthologs on autosomes and 1,394 orthologs on the X chromosome. Interspecies differences were insignificant for either exon set (Wilcoxon rank-sum test with Bonferroni correction; $p = 0.378$ for autosomes; $p = 0.668$ for X), but introns on both autosomes ($p = 1.53 \cdot 10^{-10}$) and the X chromosome ($p = 1.2 \cdot 10^{-5}$) were significantly larger in *C. briggsae* (Wilcoxon rank-sum test with Bonferroni correction).

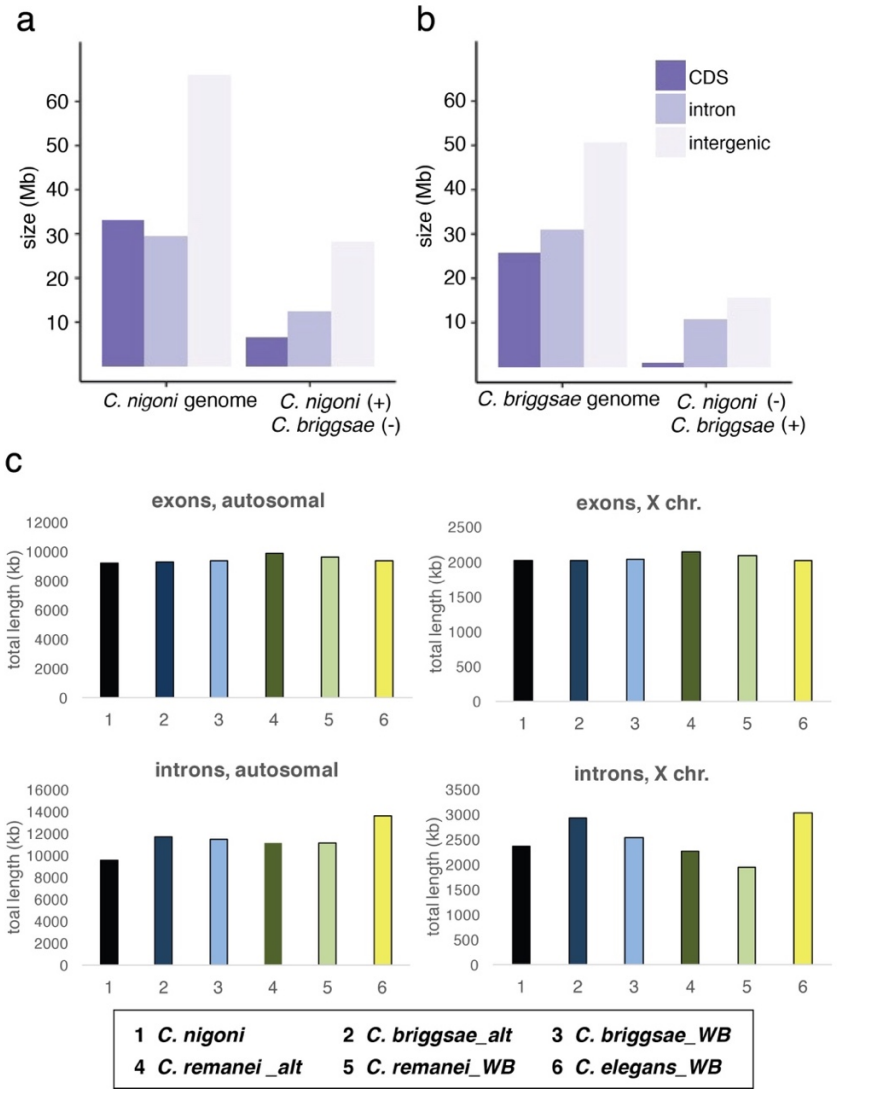

Figure 4. Coding, intronic, and intergenic proportions of the *C. nigoni* and *C. briggsae* genomes. Values are presented for the *C. nigoni* (**a**) and *C. briggsae* (**b**) genomes overall and for their species-specific sequences. **c**, summary of total exon and intron lengths extracted from predicted protein-coding genes with strict orthology (1:1 between all taxa and gene predictions) for *C. nigoni, C. briggsae, C. remanei,* and *C. elegans*. Gene sets with dark bars (including those with the "alt" suffix; 1, 2, 4) were predicted in this study by consistent methods. Gene sets with pastel bars (3, 5, 6) were taken from WormBase (*www.wormbase.org,* release WS254). X-linked *C. briggsae* (alt) orthologs have significantly more intron content than do those of *C. nigoni* or *C. remanei* (alt). For autosomal introns, *C. briggsae* (alt) and *C. remanei* (alt) orthologs have significantly greater intron content than *C. nigoni* than do their *C. nigoni* counterparts. Intron comparisons used a Wilcoxon rank-sum test with Bonferroni correction, $p < 0.05$.

We predicted 29,167 protein-coding genes for *C. nigoni*, with 88.9% (25,929) being expressed in adults (≥0.1 transcripts per million [TPM]). By equivalent methods, we predicted 22,313 genes in *C. briggsae*, 23.5% less than *C. nigoni*. The published gene annotations for *C. briggsae (Coghlan et al., 2008)* were even fewer (21,814 genes).

This 6,854-gene difference could have several causes, including gene family contraction and loss of sequence classes in *C. briggsae,* as well as *C. nigoni*-biased gain of novel sequences. We compared genes of *C. briggsae* and *C. nigoni* to genes of the outgroups *C. remanei*, *C. brenneri,* and *C. elegans*. In *C. nigoni*, 24,341 genes (83.5%) were orthologous to 21,124 *C. briggsae* genes, reflecting larger multigene families in *C. nigoni* versus *C. briggsae* (**Figure 5a**) (Emms and Kelly, 2015). Another 2,949 *C. nigoni* genes without *C. briggsae* orthologs (10.1%) represent losses in *C. briggsae* based on homologs in *Caenorhabditis* outgroups (**Figure 6)**. Finally, 1,877 *C. nigoni* genes (6.4%) lacked homologs entirely and were classed as orphans. These genes could be exceptionally divergent, recently arisen in *C. nigoni*, or arisen shortly before the *C. nigoni*-*C. briggsae* split but then lost in *C. briggsae*. Overall, gene loss in *C. briggsae* appears to be the primary driver of the gene number difference.

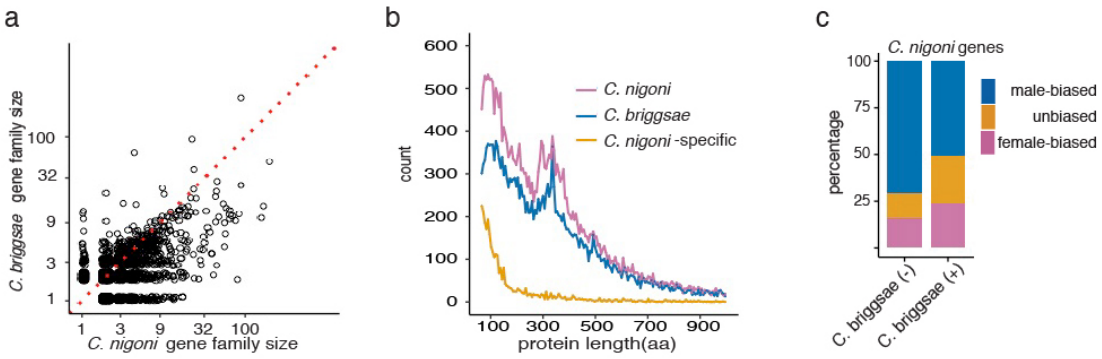

Figure 5. Impact of selfing on the *C. briggsae* proteome. a, Scatter plot of sizes of OrthoFinder gene families, excluding one-to-one orthologs. Of 2,367 families with unequal numbers of *C. nigoni* and *C. briggsae* genes, the majority (1,624) were larger in *C. nigoni* than in *C. briggsae* (Wilcoxon signed rank test, $p < 2.2 \cdot 10^{-16}$). Dotted line indicates equal family sizes. **b,** Length distributions of *C. nigoni* and *C. briggsae* proteins and of *C. nigoni* proteins that lack *C. briggsae* homologs. **c,** For genes with sex-biased expression, male bias was seen for 50.9% of 6,804 genes with *C. briggsae* homologs ("*C. briggsae (+)*"), but significantly overrepresented (70.9%) among 605 genes lacking *C. briggsae* homologs ("*C. briggsae (-)*"; Fisher's exact test, $p < 0.0001$; **Table 5**)

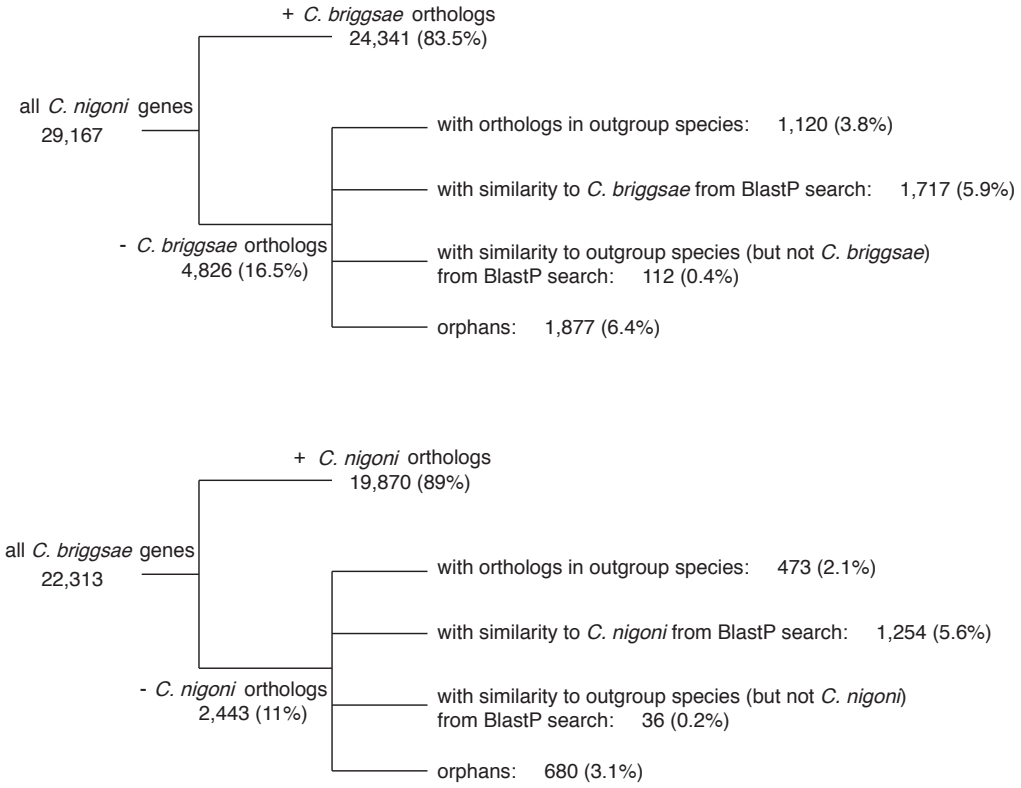

Figure 6. Gene homology categorization. A detailed classification of gene homologies is shown. For *C. nigoni*: 4,826 genes that lacked *C. briggsae* orthologs (16.5%) may represent both recent losses in *C. briggsae* and recent gains in *C. nigoni*. Within this group, 1,120 genes (3.8% of the total) had OrthoFinder orthologs in outgroup species, and therefore are very likely to represent losses in the *C. briggsae* lineage. Many of the remaining 3,706 genes specific to *C. nigoni* may also represent cases of *C. briggsae* gene loss, including 1,717 (5.9%) that exhibited some similarity to non-orthologous *C. briggsae* genes in BlastP searches ($E \leq 10^{-5}$) (*115*) and 112 (0.4%) that lacked *C. briggsae* BlastP similarity but had BlastP similarity to other *Caenorhabditis*. Finally, 1,877 *C. nigoni* genes (6.4%) lacking *C. briggsae* homologs had neither homologs nor similarities to other species, and were classed as orphans. Conversely, for *C. briggsae*: 2,443 genes (16.5%) lacked *C. nigoni* orthologs; 473 genes (2.1%) had other *Caenorhabditis* orthologs; 1,254 genes (5.6%) had some similarity to *C. nigoni*; 36 genes (0.2%) had similarity to other *Caenorhabditis*; and 680 genes (3.1%) were orphans.

To reveal the recent evolutionary dynamics of the *C. briggsae* gene set, we subjected it to comparative analyses similar to those described above for *C. nigoni* (**Figure 6**). A higher proportion of *C. briggsae* genes (using our equivalently predicted gene set as a baseline) were conserved in its sibling species: 19,870 genes (89.0%) had *C. nigoni* homologs; 473 genes (2.1%) had other *Caenorhabditis* homologs; 1,254 genes (5.6%) had some similarity to *C. nigoni*; 36 genes (0.2%) had similarity to other *Caenorhabditis*; and 680 *C. briggsae* genes (3.1%) were orphans. The low number and proportion of orphan genes is especially striking, but all categories skewed toward greater homology.

These comparisons demonstrate that loss of ancestral genomic content in *C. briggsae* does not fully explain the genome shrinkage. Some loss of ancestral DNA in *C. nigoni* has occurred despite the constant male-female sexual mode, and ongoing birth of novel sequences in both lineages also contributes to genomic divergence. However, the ratio of losses to gains has increased substantially in *C. briggsae*, resulting in net shrinkage of its genome. In addition, coding sequences have frequently been lost in *C. briggsae*, reducing its gene number by nearly one quarter.

To characterize genes lost in *C. briggsae*, we first compared Pfam protein domains encoded by *C. nigoni* versus *C. briggsae*. We found 26 Pfam domains that were overrepresented in *C. nigoni* **(Figure 7)**; of these, seven were consistently overrepresented in outcrossing *C. nigoni*, *C. remanei*, and *C. brenneri* relative to the selfing species *C. briggsae* and *C. elegans*. Three of these domains (F-box, FBA_2/F-box associated, and BTB) are predicted to mediate protein-protein interactions. Male-female *Caenorhabditis* had 272-1,074 genes in these families, while hermaphroditic

Caenorhabditis had only 101-258 genes per family. Two other domains (Peptidase_A17 and DNA_pol_B_2) are associated with repetitive DNA. The final two overrepresented domains were Asp_protease_2 (possibly associated with retroelements) and DUF3557 (a nematode-specific domain, currently of unknown function). One overrepresented domain specific to *C. nigoni* was zf.RING2_finger; the RING domain gene *spe-42* is important for sperm-egg interactions in *C. elegans* (Wilson et al., 2011).

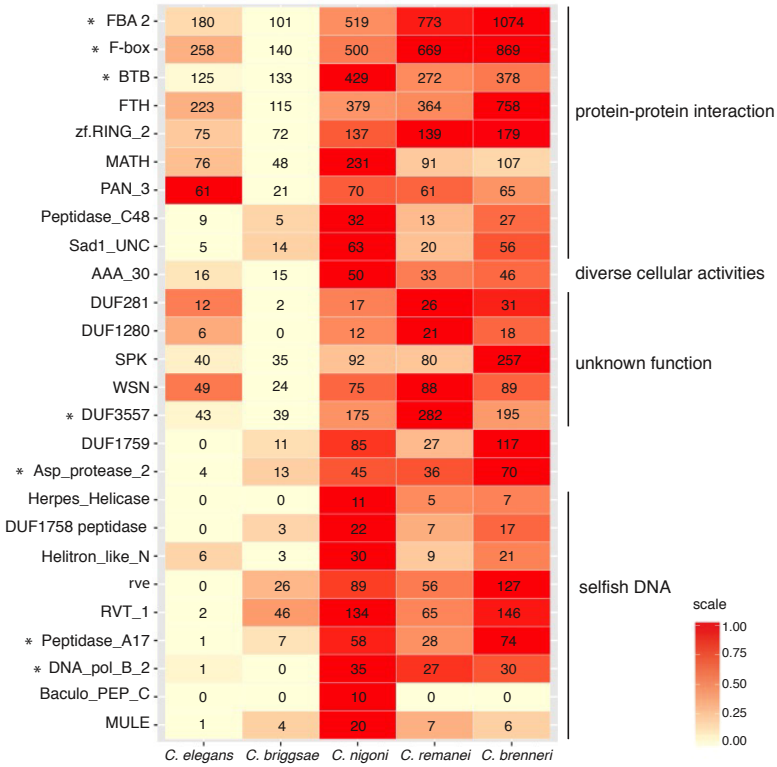

Figure 7. Enriched Pfam domains in *C. nigoni* relative to *C. briggsae.* A heatmap of all 26 Pfam domains encoded by significantly more genes in *C. nigoni* than in *C. briggsae* ($P < 0.01$, Fisher's exact test) is shown, with broad functional categories indicated to the right. By an equivalent criterion (pairwise Fisher's exact tests with $p < 0.01$, followed by Holm-Bonferroni correction), seven of these Pfam domains are encoded by significantly more genes in all three of the outcrossing species (*C. nigoni*, *C. remanei*, and *C. brenneri*) than in both of the selfing species (*C. briggsae* and *C. elegans*); these seven domains are marked with asterisks. For each domain, the numbers of genes encoding it in a given species are indicated in the boxes.

***C. nigoni* prots.:**	**66-99 aa**	**100-199 aa**	**200-399 aa**	**400-599 aa**	**600+ aa**	**Total**
***C. briggsae* [-]**	1,455	1,905	934	250	282	5,626
***C. briggsae* [+]**	1,436	5,031	9,264	4,511	4,099	23,541
Total	2,891	6,936	10,198	4,761	4,381	29,167
Bias	[-]	[-]	[+]	[+]	[+]	n/a
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	n/a

***C. briggsae* prots.:**	**66-99 aa**	**100-199 aa**	**200-399 aa**	**400-599 aa**	**600+ aa**	**Total**
***C. nigoni* [-]**	826	1,043	418	92	64	2,443
***C. nigoni* [+]**	1,105	4,290	7,294	3,764	3,417	19,870
Total	1,931	5,333	7,712	3,856	3,481	22,313
Bias	[-]	[-]	[+]	[+]	[+]	n/a
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	n/a

Table 3. Relationship of protein size to sibling-species conservation for *C. nigoni* and *C. briggsae* proteins. For all *C. nigoni* genes, the longest isoforms were divided into different size classes and categorized as either lacking or having *C. briggsae* homologs ("*C. briggsae* [-]" and "*C. briggsae* [+]"); the same was also done for *C. briggsae* genes with respect to *C. nigoni* homologs. Equivalent methods were used to predict both gene sets, and homologs were determined by OrthoFinder. In the two smallest size classes (66-99 and 100-199 residues), *C. nigoni* proteins without *C. briggsae* homologs are significantly overrepresented with respect to the total proteome (Fisher's two-tailed exact test, confidence level 0.99); conversely, in the median and two largest size classes (200-399, 400-599 and 600+ residues), *C. nigoni* proteins with *C. briggsae* homologs are significantly overrepresented. The same pattern is observed for *C. briggsae* proteins.

Species	**Total genes**	**XO-biased**	**XX-biased**	**XX sex**
C. nigoni	29,167	3,895	1,707	female
C. briggsae (alt.)	22,313	4,589	2,215	hermaphrodite
C. briggsae	21,814	4,334	2,100	hermaphrodite
C. remanei (alt.)	26,960	5,442	3,898	female
C. remanei	26,226	5,334	3,724	female
C. brenneri	34,049	6,689	4,486	female
C. elegans	20,257	5,166	3,448	hermaphrodite

Table 4. Statistics for sex-biased genes in *Caenorhabditis* genomes. Protein-coding genes, their observed sex-biased RNA-seq expression, and their XX sex are listed for the genomes of five *Caenorhabditis*. Sources of genomic and RNA-seq data for the non-*nigoni* species, genomic sequences are listed in Methods.

***C. nigoni* genes:**	Any/no homology	+ *C. briggsae* [+]	+ *C. briggsae* [-]
FDR ≤ 0.001	7,409 (100%)	6,804 (100%)	605 (100%)
+ male-biased	3,895 (52.6%)	3,466 (50.9%)	429 (70.9%)
+ unbiased	1,807 (24.4%)	1,728 (25.4%)	79 (13.6%)
+ female-biased	1,707 (23.0%)	1,610 (23.7%)	97 (15.0%)

***C. briggsae* genes:**	Any/no homology	+ *C. nigoni* [+]	+ *C. nigoni* [-]
FDR ≤ 0.001	7,696 (100%)	7,308 (100%)	388 (100%)
+ male-biased	4,589 (59.6%)	4,349 (59.5%)	240 (61.9%)
+ unbiased	892 (11.6%)	871 (11.9%)	21 (5.4%)
+ hermaph.-biased	2,215 (28.8%)	2,088 (28.6%)	127 (32.7%)

Table 5. Proportions of *C. nigoni* and *C. briggsae* genes with male-biased, unbiased, and female-biased expression. For each species, we identified genes that exhibited highly significant expression data (FDR ≤ 0.001) for differential RNA-seq expression in comparisons between males (XO chromosomes) and either females or hermaphrodites (XX chromosomes, for *C. nigoni* and *C. briggsae*). Note that this filters out roughly two-thirds of the genes in each species, but selects for the one-third of genes whose expression could be sharply distinguished between male-biased, female/hermaphrodite-biased, or unbiased. We defined male-biased as those genes exhibiting ≥2-fold higher expression in males than in females/hermaphrodites, and female-/hermaphrodite-biased as those genes exhibiting ≥2-fold higher expression in females/hermaphrodites than in males; genes falling in between these two sets were classified as unbiased. We further stratified genes by whether they had homologs in their sibling species or not (e.g., "*C. briggsae* [+]" and "*C. briggsae* [-]").

Because *C. nigoni*-specific genes might encode fast-evolving proteins that lacked known domains, we compared other gene properties. Genes encoding medium to large proteins (≥200 residues) were similar in frequency in both species, but *C. nigoni* encodes disproportionately more small proteins (<200 residues) than *C. briggsae* **(Figure 5b; Table 3).** As seen in other *Caenorhabditis (Thomas et al., 2012b)*, genes with male-biased expression outnumbered female-biased genes (**Figure 5c; Table 4**). However, even against this background, *C. nigoni* genes without *C. briggsae* homologs were disproportionately male-biased in expression. Preferential

loss of small and fast-evolving proteins thus occurred in *C. briggsae* after the adoption of selfing.

Discussion:

Comparison of the *C. nigoni* and *C. briggsae* genomes revealed that *C. briggsae* experienced rapid contraction of chromosomes and loss of protein-coding genes. However, loss of ancestral genomic content in *C. briggsae* does not fully explain their genomic divergence; the ongoing birth of novel sequences in both species, along with loss of ancestral DNA in *C. nigoni*, is also important. Net shrinkage of the *C. briggsae* genome therefore resulted from a substantial increase in the ratio of losses to gains. These losses included many coding sequences, reducing the *C. briggsae* gene count by nearly one quarter.

Multiple observations implicate the evolution of selfing as the cause of genome shrinkage in *C. briggsae*. Reduced genome and transcriptome sizes are observed in all three selfing *Caenorhabditis* (Fierst et al., 2015; Thomas et al., 2012b). Continued interfertility of *C. briggsae* and *C. nigoni* (Woodruff et al., 2010) indicates that self-fertility and genome shrinkage evolved in quick succession. Genes with male-biased expression, such as the *mss* family, are disproportionately and consistently lost from selfing species (Thomas et al., 2012b). This suggests that genes with male reproductive functions that are either dispensable or maladaptive in the new sexual mode are purged from the genome. Finally, the net genome shrinkage we observed has been predicted to arise from a partially selfing mating system coupled with transmission distortion of autosomal deletion alleles (Le et al., 2017; Wang et al., 2010). Such distortion is driven by imbalanced chromatin during meiosis I of XO males, and causes preferential

inheritance of shorter alleles by hermaphrodite progeny and their increased fixation in the population.

Larger autosomal deletions, influenced most by the deletion segregation distortion mechanism, are primarily responsible for the smaller genome of *C. briggsae* (**Figure 2**). However, such deletions and net shrinkage were also found on the X chromosome (**Table 2**), which should be unaffected. Moreover, orthologous genes have larger introns in *C. briggsae* than *C. nigoni* (**Figure 2**), and introns comprised a greater fraction of the *C. briggsae* genome (**Figure 4**). X-chromosomal *C. briggsae* introns are also larger than those of the outgroup *C. remanei* (Fierst et al., 2015) (**Figure 4**), suggesting that introns of many genes expanded in *C. briggsae*. Thus, additional processes must also contribute to shrinkage of the *C. briggsae* genome. Spontaneous short (1-5 nt) mutations in *C. elegans* are biased towards insertions rather than deletions (Denver et al., 2004), though biases in formation of larger indels remain uncharacterized. Regardless, the relative rates of insertion and deletion mutations likely evolve too slowly to explain *C. briggsae*'s reduced genome size, given its recent divergence from *C. nigoni (Thomas et al., 2015)*. Gene loss can sometimes be adaptive (Cutter and Jovelin, 2015; Olson, 1999), and has been proposed as a factor promoting genome shrinkage in selfing *Caenorhabditis* (Fierst et al., 2015). Our results for the *C. nigoni*-*C. briggsae* pair provides evidence in support of this hypothesis.

Chapter 2 Loss of sperm competition proteins in self-fertilizers

Summary

Several studies have shown selfing *Caenorhabditis* species have reduced performances in the mating process. For example, mate finding, copulation, sperm size/invasiveness are all compromised (Chasnov et al., 2007a; Cutter, 2008; Garcia et al., 2007; Hodgkin, 1983; Noble et al., 2015; Palopoli et al., 2015; Stewart and Phillips, 2002; Ting et al., 2014). The genetic bases of these changes remain largely unknown and maybe associated with the sweeping genomic changes since the evolution of selfing. Thomas et al. reported a family of short, male-specific proteins named MSS (male secreted short) that are conserved in male-female *Caenorhabditis* species but lost from hermaphroditic ones. We found these proteins are glycosylated and localized to the surface of activated sperm. We reasoned that MSS proteins might be involved in reproductive functions that are dispensable after self-fertilization evolves. We used CRISPR/Cas9-mediated genome editing in the male-female *C. remanei* to precisely delete all four of its *mss* paralogs. Though mutant males have normal fertility in a homogenous population, loss of *mss* profoundly impairs the ability of their sperm to compete with wild-type males in siring progeny. Thus, MSS proteins are novel sperm competition factors. Remarkably, restoration of the MSS genes to males of the self-fertile *C. briggsae* is sufficient to confer a large advantage over wild-type *C. briggsae*

males, whose sperm naturally lack MSS proteins. Furthermore, *C. briggsae* populations harboring *mss* genes maintain higher male frequencies by more effectively blocking self-fertilization in hermaphrodites. That *C. briggsae* would lose a gene which confers a clear fitness advantage to males is counterintuitive. We present some preliminary evidence that *mss* genes were not simply lost due to relaxed sexual selection, but were actually driven out of the genome by positive selection to reduce male frequency, a situation predicted by W.D. Hamilton to occur when males primarily compete against close relatives in small isolated populations (Hamilton, 1967). We will further test the hypothesis that loss of *mss* confers adaptation in the next chapter.

Materials and Methods

Identification of mss and msrp genes in Caenorhabditis. Because of their limited sequence conservation, *mss* and *msrp* gene products were difficult to identify reliably with BlastP or psi-BLAST. However, we found that we could identify phylogenetically coherent sets of *mss* and *msrp* genes by iterative sequence analysis. We began with the first *mss* protein sequence that we had identified, Cre-MSS-1/FL81_17790-RA, and carried out psi-BLAST with it against *Caenorhabditis* proteomes that had been selected for proteins ≤200 residues in length. psi-BLAST was run with the arguments "*-evalue 1e-03 -seg no -num_iterations 20 -inclusion_ethresh 1e-03*". By constraining the psi-BLAST search to small proteins, we found that we could increase our ability to detect *mss* homologs while lowering the background rate of hits to large proteins. This highly specialized psi-BLAST search converged on a set of 13 *mss* and *msrp* homologs. We repeated psi-BLAST with each of these homologs;

their searches all converged, to yield a cumulative set of 43 *mss* and *msrp* homologs. Having acquired an extensive and carefully defined initial sequence set, we performed several rounds of profile-based sequence searches of *Caenorhabditis* and other nematode proteomes. In each of these rounds, we aligned homologs with MAFFT v7.305b (Katoh and Standley, 2013), used their alignment to construct an HMM with hmmbuild and hmmpress in HMMER 3.1b2, and scanned nematode proteomes with this HMM via hmmsearch from HMMER. MAFFT alignments were run in slow/sensitive mode (L-INS-i), with the arguments "*--maxiterate 1000 --localpair*"; hmmsearch was run with the argument "*--domE 0.5*", to allow weak hits to be examined. For each of these rounds, we selected new hits both for the statistical strength of their alignment to the HMM (e.g., E-values of $\leq 10^{-6}$) and for their general structural characteristics (marginal hits to small domains of large proteins were rejected). Although we did not use sex-biased gene expression data to select genes, we were able to use sex-biased gene expression data for *Caenorhabditis* as a guide to whether the iterative search was detecting credible new hits. We observed that one protein sequence from *C. remanei*, FL81_26171-RA, was identical to FL81_17791 in its first 147 residues, and lacked the last 10 residues of *mss* consensus sequence in its C-terminus; we therefore deleted FL81_26171-RA as a possible misprediction. Having identified a set of 68 nonredundant *mss* and *msrp* homologs in *C. nigoni*'s closer relatives (*C. briggsae*, *C. sinica*, *C. remanei*, *C. elegans*, and *C. nigoni* itself), we constructed both HMMs for *mss* and *msrp* families and a global 68-sequence HMM, which we used to search the proteomes of three outgroup species, *C. japonica*, *C. afra*, and *C.* sp. 34; genes which gave positive results for both the family-specific HMMs

and the global HMM were further examined as possible *mss*/*msrp* homologs. We concluded this search with 81 *mss* and *msrp* homologs from *Caenorhabditis*.

Protein sequence alignment and phylogenetic tree for mss and msrp. We used MAFFT to align the protein sequences of all *mss* and *msrp* genes. From the remaining *mss*/*msrp* alignment, we filtered weakly aligned residues having >50% gaps with trimAl v1.4.rev15 (*http://github.com/scapella/trimal*) (Capella-Gutiérrez et al., 2009), using the argument '*-gt 0.5*'. On the filtered alignment, we used FastTree 2.1.9 (*http://www.microbesonline.org/fasttree*) (Price et al., 2010) to compute maximum-likelihood trees with posterior probabilities on the tree nodes, with the arguments '*-pseudo -wag*'. We used JalView 2.9.0b2 (*http://www.jalview.org*) (Waterhouse et al., 2009) with ClustalW residue colors to draw the *mss* alignment, and FigTree v1.4.3 (*http://tree.bio.ed.ac.uk/software/figtree*) to draw the *mss*/*msrp* phylogeny.

Predicting exons for the pseudogene Cbr-mss-3-ps *in* C. briggsae. We used exonerate 2.2.0 and the protein sequence of MSS-3 from *C. nigoni* to identify actual or potential exons from both the *Cni-mss-3* genomic locus and the *Cbr-mss-3-ps* genomic locus. In both cases, we ran exonerate with the arguments "*-E -m protein2genome:bestfit --useaatla FALSE -Q protein -q [Cni-mss-3/Cnig_chr_III.g11661.t1 protein sequence] -T dna -t [genomic locus DNA sequence]*". To provide input sequences to exonerate, we used *extractseq* from EMBOSS to extract the predicted region (based on BlastN with the coding sequence of *Cni-mss-3*) and 500 nt of flanking DNA from the appropriate *C. nigoni* or *C. briggsae* genome scaffold. We also ran AUGUSTUS with *C. briggsae*-specific parameters on each region to confirm that there was an observable protein-coding gene in the *Cni-*

mss-3 region, but no observable protein-coding gene in the *Cbr-mss-3-ps* region (i.e., we checked to make sure that there were no local compensatory mutations that might restore coding capacity to *Cbr-mss-3-ps*).

Identifying the pseudogene Cbr-mss-3-ps *in wild isolates of C. briggsae*. Cutter and coworkers have identified and sequenced genomic DNA from wild isolates of *C. briggsae* worldwide (Thomas et al., 2015). Although there were no preexisting genome assemblies for these isolates, their raw genomic sequence data were available. We therefore downloaded genomic sequence reads (generally, though not universally, paired-end) for the following 12 wild isolates (NCBI SRA accession numbers given in parentheses): Hubei_VX0034 (SRR1793007); Kerala_JU1341 (SRR1792996 and SRR1793000); Kerala_JU1348 (SRR1793004); Nairobi_ED3101 (SRR1793002); Quebec_QR24 (SRR1793005); Quebec_QR25 (SRR1793006); Taiwan_NIC19 (SRR1793010); Taiwan_NIC20 (SRR1793012); Temperate_EG4181 (SRR1792978); Temperate_JU516 (SRR1792992); Tropical_JU1399 (SRR1792934); and Tropical_QX1410 (SRR1792974). In particular, we chose Tropical_QX1410 because it is phylogenetically almost identical to the reference wild-type strain AF16, and chose the Kerala isolates because they are currently the most divergent outgroups known in *C. briggsae* (*27, 106*). For each wild isolate's reads, we used ABySS (abyss-pe 2.0.2) to assemble a draft genome, with the arguments "*np=8 j=8 k=51*". From 11 of the 12 draft assemblies (including one for Kerala), we identified a full-length copy of the *Cbr-mss-3-ps* locus with BlastN, and extracted its first exon with *extractseq* from EMBOSS. We used MAFFT (with arguments "*--maxiterate 1000 --localpair*") to align the first

exons for all wild-isolate alleles of *Cbr-mss-3-ps* with *Cni-mss-3*, and to confirm that two mutations inactivating *Cbr-mss-3-ps* in AF16 are also present in all wild isolates.

Quantitative RT-PCR. For *Cre-mss-2* five replicate populations of staged animals were washed off NGM plates in M9 buffer, or individually picked in M9 buffer for male and female preparations. After three washes in RNAse-free water, samples were resuspended in 50 µl of RNAse-free water. 250 µl TRI-Reagent (Molecular Research Center) was added and the samples were frozen at -80˚C. The samples were thawed, pelleted and lysed using a plastic pestle. RNA was purified using manufacturer's instructions. Samples were treated with DNase I (New England Biolabs), phenol/chloroform extracted, isopropanol precipitated, and resuspended in RNase- free water. cDNA was synthesized using 1 µg of total RNA using Superscript III (Invitrogen) in 50 µl according to manufacturer's instructions. 2 µl cDNA were used as template with the Light Cycler 480 SYBR Green I kit (Roche) according to manufacturer's instructions. Primers spanning exon-exon junctions were designed so that the amplicons sizes were between 161 and 220 bp. Primers were CGT057 (5'-GGA TCT TCT GGG GCT TTC GG-3') and CGT058 (5'-GGA TTT CCG ACTCCA CCA TCT G-3'). Control reactions with no template and multiple non-sex-biased transcripts for each primer pair were performed, and all reactions for a gene were run simultaneously on a single 96-well plate on a Roche Light Cycler 480 machine using manufacturer's software. Data were analyzed using the program LinRegPCR (Ramakers et al., 2003; Ruijter et al., 2009). cDNA was synthesized as above, and PCR was carried out using PrimeSTAR Max DNA Polymerase using primers DY112 (5'-

CTG GTC CAT TCA CAG TCA CAG C-3') and DY113 (5'-ATG ATG GTG GTG CTG GAG GC-3') for detection of *mss-1*.

In situ hybridization. Gonads and intestines were dissected from adult *C. remanei,* fixed in a glutaraldehyde-based fixative, and incubated with digoxygenin-labelled antisense ssDNA probes as previously described (Hill et al., 2006; Jones et al., 1996). The templates for production of antisense and sense *Cr-mss-1* ssDNA probes via asymmetric PCR was amplified from JU1422 genomic DNA using the following oligonucleotide primers (IDT), where underlined sequence represents the phage T7 promoter. RLM001 (for sense probe PCR and IVT): 5'-TAA TAC GAC TCA CTA TAG GGA GAG CAT TGT TGG CCA CCG-3'. RLM002 (for antisense template PCR): 5'-GCA TTG TTG GCC ACC G-3'. RLM003 (for antisense probe PCR and IVT): 5'-TAA TAC GAC TCA CTA TAG GGA GAT CCT TCG GCT GGT GCT TCT GGT-3'. RLM004 (for sense template PCR): 5'-TCC TTC GGC TGG TGC TTC TGG T-3'.

Immunohistochemistry and microscopy. After adult *C. remanei* males and mated females were dissected, testis and activated sperm were isolated on ColorFrost Plus slides using a modification of standard protocols (Shakes et al., 2009). In brief, after dissection, a cover slip was placed over the sample and then the slide was placed in dry ice for 30 min to freeze and crack. Dissected gonads or sperm were fixed in 4% formaldehyde for 10 min, followed by a 30 min post-fix in 100% methanol at -20°C. After multiple washing steps (PBS containing 0.1% Tween) and blocking (PBS containing 0.1% Tween + 0.75% BSA), incubations with primary antibody were carried out at room temperature in a humid chamber overnight with a 1:100 dilution of

Anti-HA-Peroxidaxe-3F10 (Roche). Secondary antibody incubations were carried out for 2 hours with 1:100 Alexa Fluo 555 anti-rat IgG (Invitrogen). Hoechst (Sigma-Aldrich) at 100 µg/ml was used for staining DNA for 15 min. Slides were prepared with Vectashield mounting media. Images were acquired on Leica SP5 X confocal microscope and processed with Zen Lite software (Zeiss). For anti-HA immunoblots, worms were homogenized and boiled in Laemmli sample buffer under standard reducing conditions (Wright, 1989). The resulting proteins were fractionated via 15% SDS-PAGE gels and transferred to nitrocellulose. Filters were incubated sequentially with rat anti-HA monoclonal antibody (Santa Cruz), goat anti-rat horseradish peroxidase (HRP)-conjugated secondary (Thermo Fisher), and HRP luminescence detection reagents (Pierce) according to manufacturer's instructions.

Generation of mss knock-out in C. remanei using CRISPR. Cas9 protein containing an NLS (PNA Bio Inc) was reconstituted by dissolving in 40 µl of nuclease-free water, creating a stock at 1250 ng/µl and stored in small aliquots at −80°C. The sgRNAs were transcribed *in vitro* using the Ambion Megascript SP6 kit. We followed a published protocol for steps including annealing oligos, transcribing sgRNA and cleanup of sgRNA, except we used Taq polymerase for fill-in instead of T4 DNA polymerase (Gagnon et al., 2014). Multiple *in vitro* transcription reactions were set up in order to get very concentrated sgRNA. Prior to the microinjection, *in vitro* DNA cleavage assay was carried out to confirm the effectiveness of the gRNA and Cas9 protein for cutting the target DNA template. In the final injection mixture, the concentration of Cas9 protein was ~800 ng/µl while the concentration of each of gRNAs was ~670 ng/µl. The injection mixture was incubated at 37°C for 10 minutes

before being loaded into needles for injection. Young gravid females were injected, put individually on NGM plates, and allowed to lay embryos for 32 hr. After the F1 progeny become adults and mated, the mated females were divided into five worms/plate. After the F2s laid embryos, the F1 females on each plate were picked and genotyped using single-worm PCR to detect any edit. Positives were identified using primers flanking the *mss* genes or the HA epitope tag that generates amplicons of different sizes between edit and non-edit. Once the edit was confirmed by sequencing, the L4s from corresponding plate(s) were used to set up crossings to isolate homozygous mutants. *mss* knock-out mutants were independently generated in both SB146 and EM464 strains. The SB146-derived *mss* knockout strain, CP157, was rendered homozygous for its *nmDf1* deletion by multiple generations of backcrossing. However we were not able to render the EM464 knockout allele, *nmDf2,* homozygous, presumably due to a lethal or harmful allele linked to the *mss* locus. Therefore, only heterozygous EM464 mutants were generated. To compensate for the effects of inbreeding depression, interstrain progeny were generated by crossing heterozygous EM464 mutants and homozygous SB146 mutants.

HA knock-in. The hemaglutinin (HA) epitope tag was inserted using CRISPR/Cas9 through homologous recombination. The nine amino acid HA epitope was placed between *C. remanei* (EM464*)* MSS-1 residues 22 and 23, one residue downstream of the predicted mature N-terminus after signal peptide cleavage. We designed sgRNA that was complementary to sequence downstream of the signal peptide, so that even a processed peptide would retain the HA epitope at its N-terminus. Edited line CP158 (*Cre-mss-1(nmIs9)*) was identified by PCR (primer DY70 5'-ACG

ACG TTC CAG ACT ACG CC-3' and DY41 5'-TGA GTG TCT TTG GGT GCG TT-3') of offspring of injected mothers, after they had participated in pair-wise full-sib mating and laid abundant progeny (see detailed screening method above in the *mss* knock-out section). In the microinjection mixture, the concentration for Cas9 was 900 ng/µl; sgRNA 900 ng/µl; donor (repair) template oligonucleotide, 1000 nM. Repair template: DY68 donor template (for homologous recombination after Cas9 cutting) was a PAGE-purified Ultramer custom oligonucleotide (IDT). DY68 is 89 nt in total. 27 nt (lower case) indicates the HA epitope tag sequence. 31 nt (upper case) is homologous to *C. remanei mss-1* sequence on each side of the Cas9 cleavage site: 5'-CGC ATT GTT GGC CAC CGT CGC TCG TGG AGC Tta ccc ata cga cgt tcc aga cta cgc cGA CGG TGA TAA CGT AGA AGC AGG AGA TGC AC-3'. sgRNA EH75 (*C. remanei mss-1* gene-specific oligo used for generating sgRNA): 5'-att tag gtg aca cta taG ATG CAC AAC TAC CATCAG Ggt ttt aga gct aga aata gca-3'. Lower case DNA on the left is the SP6 promoter; upper case sequence is the gene-specific sequence for making sgRNA; lower case DNA on the right is homologous sequence are anneals with the Constant Oligo during the template fill-in reaction. Constant Oligo: 5'-AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATT TCT AGC TCT AAA AC-3'.

C. remanei MSS-1 immunoblots and deglycosylation assay. ~500 *C. remanei* (EM464) adults were washed off plates with M9 buffer, washed twice with more M9, washed once in lysis buffer (150 mM NaCl, 20 mM Tris, pH 8.0, 1 mM EDTA, 0.1% NP40) without protease inhibitors, and then washed twice (and left in) 100 µl lysis buffer with protease inhibitors (Complete EDTA-free, Roche). Worms were

homogenized by grinding with a small plastic pestle. 50 µl of the lysate was used as an untreated control. The other 50 µl was used for deglycosylation treatment with a protein glycosidase cocktail (New England Biolabs P6044S, Mix II, used with reducing Buffer 2). We varied the concentration of lysate by having either 5 µl or 3 µl in the final deglycosylation reactions. The mixture was incubated at 75°C for 10 mins, cooled to 15°C, and then the protein deglycosylation mix II was added for a final reaction volume of 50 µl. Digestions were incubated at room temperature for 30 min before incubation overnight (~13 hr) at 37°C. For both treated and untreated control, equal volumes of 2x (Laemmli + BME) solution were added before loading for SDS-polyacrylamide gels, with pre-stained PageRuler Plus (Thermo Scientific) or ExcelBand 3 color Extra Range (SMOBIO). Gel electrophoresis was conducted for 1.5 hours at 150 V, and transfer was conducted overnight at 4°C with 30 V. The membrane was blocked with a 5% BSA+PBST solution and washed with PBST. The primary antibody (rat monoclonal αHA 3F10; Roche) was diluted 300-fold and incubated for 90 mins at room temperature before washing. The secondary antibody (goat anti-rat HRP conjugate; Santa Cruz) was diluted 600-fold and incubated for 1 hour at room temperature. Antibody was detected with the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

Generation of stable transgenic C. briggsae strains expressing MSS *from C. nigoni*. *mss-1* and *mss-2* from *C. nigoni* were amplified by PrimeSTAR Max DNA Polymerase (TAKARA Bio Inc) using primers flanking the gene family (also including 1,500 nt 5' from the start codon of *mss-1*). *Cnig-mss-1/2* was inserted into a reporter construct pZZ0031 (gift from the laboratory of Zhongying Zhao, Hong Kong Baptist Univ.). The construct has the *Cbr-myo-2* promoter fused with GFP, and *unc-119*

rescuing sequence as selection marker for bombardment. *Cbr-unc-119* mutants were bombarded and stable transgenic lines CP161(*nmIs7*) and CP162(*nmIs8*) expressing strong GFP expression were isolated as described (Liu et al., 2012; Yan et al., 2012). The primer pair used for amplifying *mss-1* and *mss-2* was as follows. Forward (DY107): 5'-TAG ACT GGG CCC CGA ATT TCC CTG ACG AAT GCT CC-3'. Reverse (DY108): 5'-TAC ATT GGG CCC TTG CGG ACA GAG CCACAG AG-3'. The final insert size of *mss-1* and *mss-2* was 4.5 kb. PrimeSTAR Max DNA Polymerase (TAKARA Bio Inc) was used with a 15-second annealing at 55°C, and a 5-minute extension at 72°C in the PCR. Both primers used above have 6-nt overhangs and GGGCCC ApaI restriction sites added to the 5' end of their *C. nigoni* sequences. The plasmid pZZ0031 was digested with ApaI (New England Biolabs; NEB), followed by dephosphorylation using NEB shrimp alkaline phosphatase (rSAP). ApaI-digested primers were then ligated to the vector using T4 ligase. Transformation was performed on NEB 5-alpha competent *E. coli* and plasmids from colonies were isolated and sequenced to confirm the correct insertion of *mss-1* and *mss-2.*

Sperm competition assay in C. remanei. To create competitor males, heterozygous EM464 *mss(nmDf2)* knockout mutants were crossed to SB146 homozygous *mss(nmDf1)* knockout mutants to create interstrain F1 hybrid progeny. The progeny male genotypes were *nmDf1/+* and *nmDf1/nmDf2*. The males were 24 hr post-L4 virgins. In the MSS- defense scenario, *nmDf1/nmDf2* males were first allowed to mate with 24 hr post-L4 virgin adult females for 4 hours, after which the females were moved to another plate and allowed to mate with wild-type males for 4 hours. In the MSS- offense scenario, the order between the first male and second male were

switched. As a control, *nmDf1*/+ sibling males were subjected to the same defensive and offensive mating assays. Each mating was set up with one male and one female and allowed to mate for four hours on 3.5-cm NGM plates seeded with 0.5 cm *E. coli* to maximize the chance of contact and mating. Copulation with both males was closely observed under a stereomicroscope to confirm successful transfer of sperm. After mating, each male was genotyped by single-worm PCR. After a total of eight hours of mating, each female was picked into individual plates and allowed to lay embryos for 24 hours. Progeny laid during this period were scored by PCR to score paternity. 24-30 progeny from each plate were scored.

Competition of male sperm against male sperm in C. briggsae. To test CP161(*nmIs7*) MSS+ sperm's defense ability, 20 virgin young adult males [*Cni-mss-1, mss-2, cbr-myo*-2::GFP] were placed with 15 wild-type AF16 hermaphrodites for four hours, after which only hermaphrodites with visible embryos in their uteri were transferred to a new plate with 20 RW20025 (*Cbr-unc-119*, *stIs*20025[*Cbr-HIS-72*::mCherry]) control males for 4 hours. Individual hermaphrodites were then transferred to new plates, and allowed to lay embryos for 18 hr. Individual hermaphrodites were transferred to fresh plates and allowed to lay embryos for another 24-hr period. After two days the progeny were scored under a fluorescent microscope. Progeny sired by CP161(*nmIs7*) MSS+ had green fluorescent protein (GFP) expression in the pharynx, whereas those sired by control male expressed red fluorescent protein (RFP). Since both the GFP and RFP were dominant markers, non-fluorescent progeny were sired by hermaphrodite self sperm. For CP161(*nmIs7*) MSS+ sperm offense, the same experiment was carried out but the hermaphrodites were first mated to RW20025

males, then CP161(*nmIs7*) MSS+ males. Cases in which the second males' mating completely failed, as indicated by 100% cross progeny from the first males, were excluded.

Competition of male sperm vs. hermaphrodite self-sperm. To test the effect of MSS for male sperm precedence over hermaphrodite self-sperm, 20 virgin young adult males of each genotype were placed with 15 AF16 hermaphrodites for 4 hr. The control male genotypes were JU936 *[Ce-lip-1::GFP, Ce-myo-2::GFP],* RW20025 (*Cbr-unc-119*, *stIs*20025 [*Cbr-HIS-72*::mCherry]), and CP161(*nmIs7*) MSS+ experimental male (*Cbr-unc-119*, *Cni-mss-1, mss-2, Cbr-myo-2::*GFP). Individual hermaphrodites were transferred to fresh plates and allowed to lay embryos for a 24-hr period. Progeny were scored for GFP or mCherry after reaching adulthood. The ratio of progeny expressing fluorescent protein to dark progeny was calculated.

Experimental evolution assay for male percentages. 20 L4 males and 20 L4 hermaphrodites were set up for crossing for 24 hours, after which 10 males and 10 mated hermaphrodites (50% ratio of males) were transferred to a new plate as a first-generation population. Both CP161(*nmIs7*) MSS+ and wild-type AF16 strains had five replicates. All populations were grown on 6 cm NGM agar Petri plates seeded with *E. coli* strain OP50 and transferred by excising a 1cm square of agar to new plates every generation (~3 days/generation). Each excised agar had approximately 200-300 L1 stage larvae and a very small number of adults. When the majority of larvae reached adulthood, adult males and hermaphrodites were counted and ratios of males were calculated. The adults were allowed to lay embryos, that were allowed to develop into L1 before each transfer. The whole procedure was repeated for a total of 12 generations.

Male ratios were scored at the 4th, 8th and 12th generations. The populations were incubated at 25°C during this assay.

Results:

We hypothesized that genes with highly male-biased expression that are present in outcrossing species, but lost in selfing species, might function in sexual selection. Among such genes the *mss* (*male secreted short)* family stood out. One to four *mss* genes are present in the outcrossing species *C. nigoni, C. sinica, C. remanei, C. brenneri*, *C.* sp. 34, *C. japonica*, and *C. afra,* but none were found in the selfing *C. elegans, C. briggsae,* and *C. tropicalis*. The *mss* family encodes small proteins (median 111 residues) with N-terminal signal sequences, rapidly evolving central domains with several predicted O-glycosylation sites, and C-terminal glycosylphosphatidylinositol (GPI) anchor membrane attachment signals (**Figure 1**). Enzyme treatments confirmed that MSS proteins were heavily glycosylated (**Figure 2**).

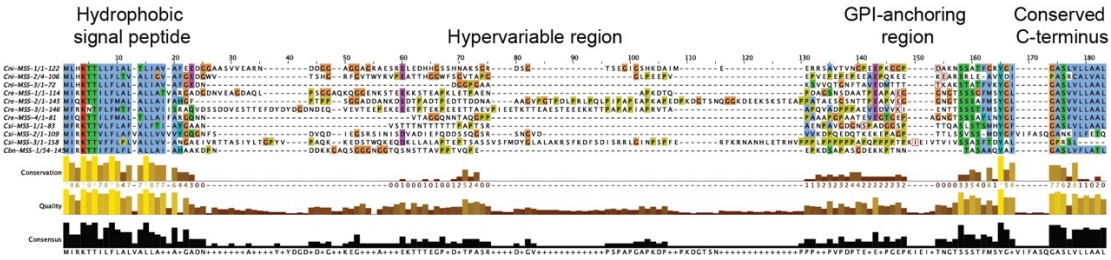

Figure 1. Alignment of predicted MSS homologs from outcrossing *C. nigoni, C. sinica, C. remanei,* and *C. brenneri*, with protein domains indicated above. Alignment of predicted MSS homologs (from top to bottom) of the outcrossing *C. nigoni* (three homologs), *C. sinica* (three homologs), *C. remanei* (four homologs), and *C. brenneri* (one homolog). Important sequence domains are indicated above.

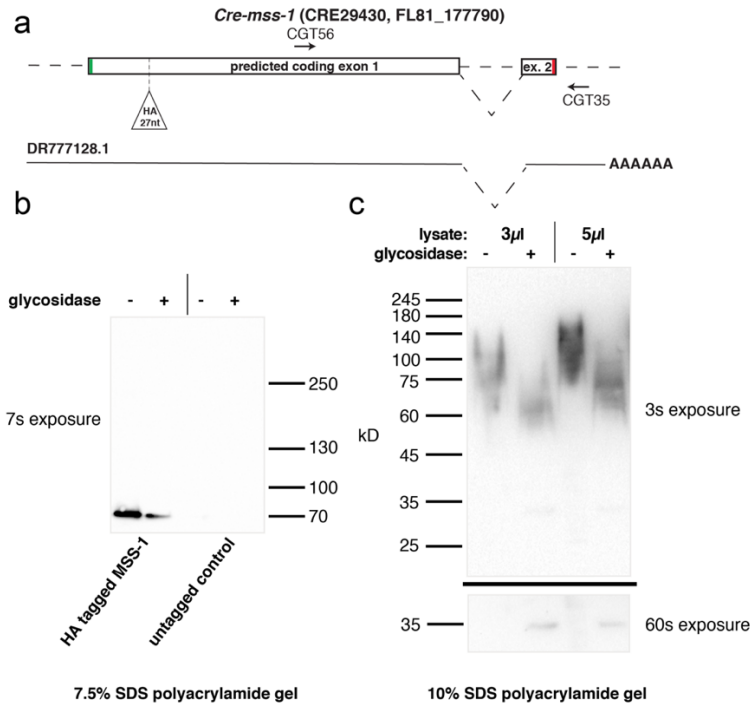

Figure 2. Cre-MSS-1 is post-translationally glycosylated. a, Gene model. Two independent predictions of protein-coding genes for the genome assemblies of *C. remanei* are shown. CRE29430 is from the Washington University Genome Center assembly of *C. remanei* strain PB4641 (available at *http://www.wormbase.org*). FL81_17790 is from the assembly of *C. remanei* strain PX356 by Fierst et al. (Fierst et al., 2015). Both of these are derivatives of the strain EM464, which was first described as "*C. vulgaris*" (*49*) and later found to be synonymous with *C. remanei* (*116*). These gene predictions are 100% identical. Predicted start (green) and stop (red) codons are indicated. The intron location and part of the 3' UTR was confirmed by RT-PCR and sequencing using the primers CGT56 and CGT35, as indicated. The EST DR777128 is derived from the strain SB146, and reveals the extent of the 5' and 3' untranslated regions, including the site of 3' cleavage and polyadenylation. It differs from the EM464-derived models in some codons in the hypervariable domain. The site of insertion of the HA epitope tag just downstream of the signal peptide by CRISPR-Cas9 editing is indicated. **b** and **c,** Western analysis of HA-tagged Cre-MSS-1. Proteins from 500 *C. remanei* homozygous for the HA-tagged MSS-1 allele as well as untagged negative controls were resolved by SDS-PAGE on a 7.5% gel (**b**) and a 10% gel (**c**), then detected with chemiluminescence with different exposures. The image merges the anti-HA fluorescent signal (black) with a reflected light image of pre-stained markers. On the 7.5% gel (**b**), MSS-1 runs near the dye front. On the 10% gel (**c**), deglycosylated (+) and untreated (-) HA-tagged MSS-1 show distinct mobilities. Either 5 µl or 3 µl of lysate was used in the final deglycosylation reactions. Upon longer exposure (**c**, bottom view), a tight band with an apparent mass of 34 kD appears in glycosidase-treated samples.

Although no *mss* genes were detected in selfing species, a larger family of *mss*-related protein (*msrp)* genes was discovered, within which *mss* forms a monophyletic clade (**Figure 3**). Notably, *msrp* genes are found both in outcrossing *Caenorhabditis* and in the hermaphroditic *C. elegans*, *C. briggsae*, and *C. tropicalis* (**Figure 3**). Like MSS, MSRP proteins are small, and are predicted to be secreted, O-glycosylated, and (often) GPI-anchored. Both *mss* and *msrp* genes show male-biased expression in *C. nigoni* and other species. In cases where their chromosomal loci can be identified, *mss* and *msrp* genes are autosomal; this linkage fits a general pattern in heterogametic male species of male-biased genes being autosomal rather than X-chromosomal (Cassone et al., 2017).

Because *mss* genes were also present in two *C. elegans* outgroups (*C. japonica* and *C. afra*, **Figure 3**), their absence from hermaphrodites most likely reflects independent gene losses rather than phylogenetic restriction to close relatives of *C. nigoni*. Examination of the *C. briggsae* genomic region syntenic to the *C. nigoni mss* locus revealed fragments of *mss-1* and *mss-2* coding sequences and a nearly complete *mss-3* pseudogene (**Figure 4**). Mutations that ablate *Cbr-mss-3-ps* function in the AF16 reference strain also occur in 11 wild isolates that span the known diversity of *C. briggsae* (**Figure 5**) (Thomas et al., 2015). Orthologs of all three *C. nigoni mss* genes were therefore present in the common ancestor of *C. nigoni* and *C. briggsae,* but lost in *C. briggsae* before its global diversification.

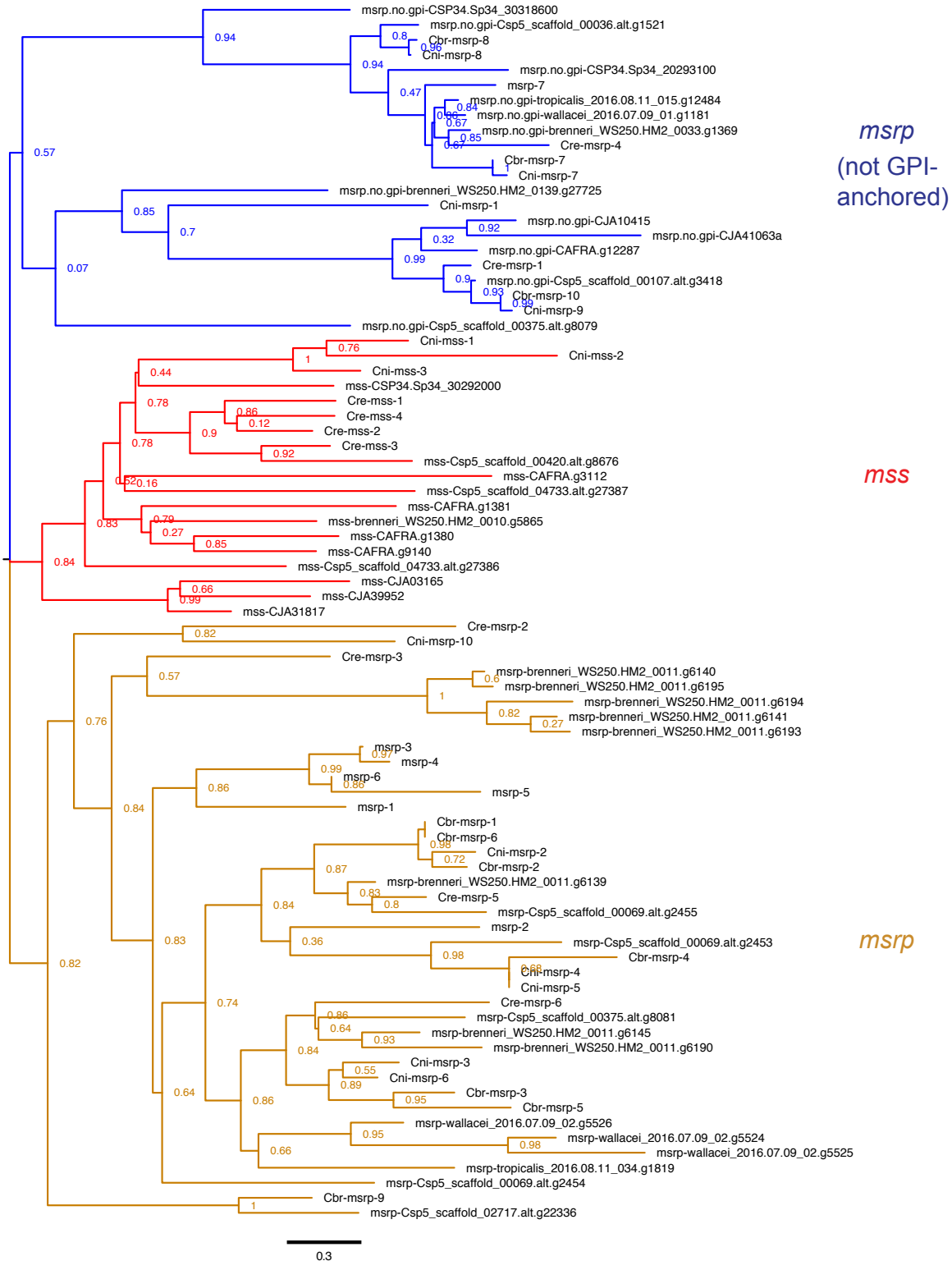
msrp.no.gpi-CSP34.Sp34_30318600
msrp.no.gpi-Csp5_scaffold_00036.alt.g1521
Cbr-msrp-8
Cni-msrp-8
msrp.no.gpi-CSP34.Sp34_20293100
msrp-7
msrp.no.gpi-tropicalis_2016.08.11_015.g12484
msrp.no.gpi-wallacei_2016.07.09_01.g1181
msrp.no.gpi-brenneri_WS250.HM2_0033.g1369
Cre-msrp-4
Cbr-msrp-7
Cni-msrp-7
msrp.no.gpi-brenneri_WS250.HM2_0139.g27725
Cni-msrp-1
msrp.no.gpi-CJA10415
msrp.no.gpi-CJA41063a
msrp.no.gpi-CAFRA.g12287
Cre-msrp-1
msrp.no.gpi-Csp5_scaffold_00107.alt.g3418
Cbr-msrp-10
Cni-msrp-9
msrp.no.gpi-Csp5_scaffold_00375.alt.g8079
Cni-mss-1
Cni-mss-2
Cni-mss-3
mss-CSP34.Sp34_30292000
Cre-mss-1
Cre-mss-4
Cre-mss-2
Cre-mss-3
mss-Csp5_scaffold_00420.alt.g8676
mss-CAFRA.g3112
mss-Csp5_scaffold_04733.alt.g27387
mss-CAFRA.g1381
mss-brenneri_WS250.HM2_0010.g5865
mss-CAFRA.g1380
mss-CAFRA.g9140
mss-Csp5_scaffold_04733.alt.g27386
mss-CJA03165
mss-CJA39952
mss-CJA31817
Cre-msrp-2
Cni-msrp-10
Cre-msrp-3
msrp-brenneri_WS250.HM2_0011.g6140
msrp-brenneri_WS250.HM2_0011.g6195
msrp-brenneri_WS250.HM2_0011.g6194
msrp-brenneri_WS250.HM2_0011.g6141
msrp-brenneri_WS250.HM2_0011.g6193
msrp-3
msrp-4
msrp-6
msrp-5
msrp-1
Cbr-msrp-1
Cbr-msrp-6
Cni-msrp-2
Cbr-msrp-2
msrp-brenneri_WS250.HM2_0011.g6139
Cre-msrp-5
msrp-Csp5_scaffold_00069.alt.g2455
msrp-2
msrp-Csp5_scaffold_00069.alt.g2453
Cbr-msrp-4
Cni-msrp-4
Cni-msrp-5
Cre-msrp-6
msrp-Csp5_scaffold_00375.alt.g8081
msrp-brenneri_WS250.HM2_0011.g6145
msrp-brenneri_WS250.HM2_0011.g6190
Cni-msrp-3
Cni-msrp-6
Cbr-msrp-3
Cbr-msrp-5
msrp-wallacei_2016.07.09_02.g5526
msrp-wallacei_2016.07.09_02.g5524
msrp-wallacei_2016.07.09_02.g5525
msrp-tropicalis_2016.08.11_034.g1819
msrp-Csp5_scaffold_00069.alt.g2454
Cbr-msrp-9
msrp-Csp5_scaffold_02717.alt.g22336
0.3
msrp
(not GPI-anchored)
mss
msrp

Figure 3. Phylogenetic relationships of the MSS family and MSS-related proteins (MSRPs). Using MSS homologs (**Figure 1**), iterative searches were made to identify other, more distantly related sequences in sequenced *Caenorhabditis* genomes (see Methods). Retrieved sequences were aligned, revealing conserved hydrophobic sequences at the N- and C-termini. Phylogenetic analysis of the 108 best-conserved positions revealed four distinct, major clades; three contained distinct subsets of *msrp* genes, while one encompassed all *mss* genes. Different clades varied with respect to the presence or absence of predicted GPI lipid anchor modification sites.

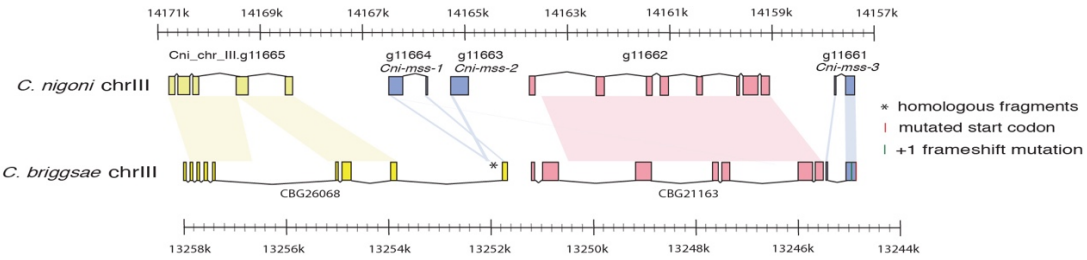

Figure 4. Comparison of *mss* gene regions in *C. nigoni* and *C. briggsae.* Pastel shapes connect homologous sequences. Except for *Cni-mss-3,* all genes are transcribed from left to right. Genes surrounding the three *C. nigoni mss* paralogs are conserved in *C. briggsae,* but only fragments and a pseudogene (*Cbr-mss-3-ps*) of the *mss* genes remain. The pseudogene has a lost start codon and a +1 frameshift. CBG26068 has a novel 3' exon derived from part of the *Cni-mss-1* second exon.

a

```
Cnig_chr_III.g11661.t1 (Cni-mss-3 ORF) vs. Cbr-mss-3-pseudo gDNA

  1 :  M  L  H  K  T  T  L  L  F  L  A  L  A  L  I  A  V  A  F  G  E :   21
       !!:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||!
        I  L  H  K  T  T  L  L  F  L  A  L  A  L  I  A  V  A  F  G  A
  1 : ATTCTCCACAAAACGACCTTGCTCTTTTTGGCTCTTGCACTGATCGCTGTAGCTTTTGGAGC :   60

 22 :    D  N  D  G  G  P  G  A  A <-><-> K  S  V  V  Q  T  G ---- N  :   38
       !:!!|||||||||||||| !!||||!.!|||       ! |||  ! !!! !!.!..!####|||
         N  N  D  G  G  A  G  G  A  Y  D  H  S  R  F  P  N  S #### N
 61 : AAATAATGATGGTGGAGCTGGGGGGGCTTACGACCATTCCCGTTTTCCAAATTCTTCCAAAT :  121

 39 :  F  T  A  V  E  D  M  T  T  T  T  K  A  K  S  T  A  T  F  V  K :   59
       ! !:!!.!! !!  !|||:!!|||  ! !! !!|||!.!||||||||||:!!||| !!||||||
        S  S  T  F  I  D  V  T  D  P  P  K  G  K  S  T  S  T  V  V  K
122 : TCTTCCACATTTATAGATGTGACAGATCCACCCAAGGGTAAAAGTACATCAACTGTTGTGAA :  184

 60 :   Y  G  I  {  }  >>>> Target Intron 1 >>>>  {G } A  S  L  V  L   :   69
       ||||||||||{|}            538 bp           {||}||||||||||!.!|||||
         Y  G  I  {  }++                        ++{G } A  S  L  A  L
185 : GTATGGAATT{G}gt.......................ag{GT}GCATCATTGGCCCTTC :  252

 70 : L  A  A  L  :   72
       ||||||||||
       L  A  A  L
253 : TGGCTGCTCTC :  263
```

b

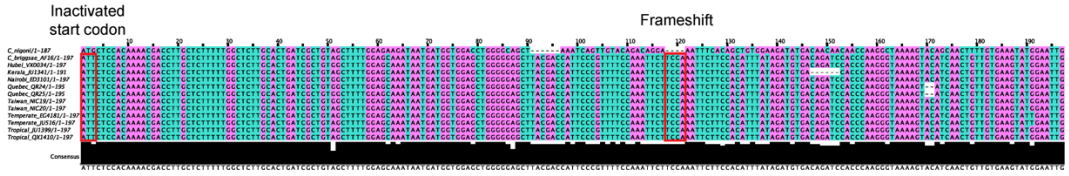

Figure 5. Alignment of the *Cbr-mss-3-ps* pseudogene across world diversity of *C. briggsae*. a, An alignment, generated by exonerate (Slater and Birney, 2005), of homologous codons in the predicted gene *Cni-mss-3* (top) versus the predicted pseudogene *Cbr-mss-3-ps* (middle; from the *C. briggsae* strain AF16 reference genome), with the genomic DNA sequence of the latter (bottom). A 10-nt repeat present two and half times in the frame-shifted region of *Cbr-mss-3-ps* is underlined. **b,** Alignment of *Cbr-mss-3-ps* alleles found in wild isolates of *C. briggsae* versus the AF16 reference allele, with start codon and frameshift mutations indicated. Homologous fragments were identified in genome assemblies produced independently for a collection of *C. briggsae* wild isolates representing all known global diversity (Thomas et al., 2015).

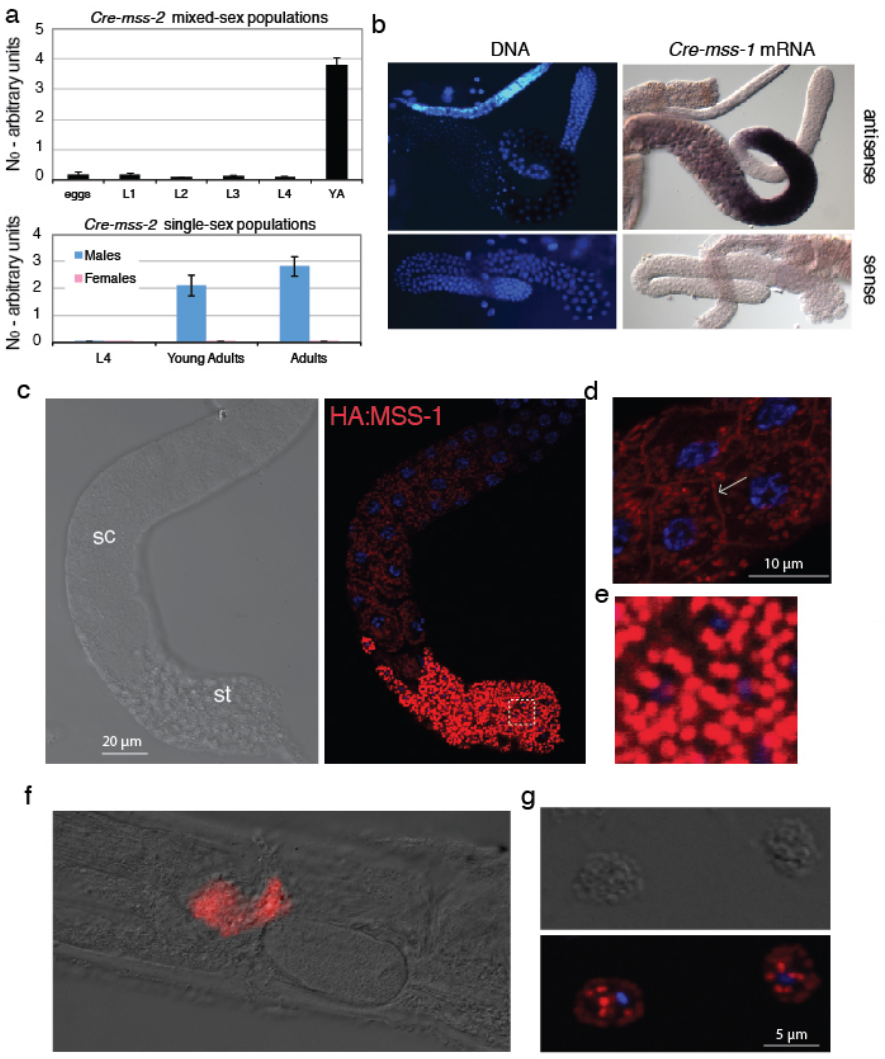

Figure 6. *C. remanei* MSS is a male-specific protein localized to the surface of activated sperm. a, qRT-PCR transcript quantification (for *Cre-mss-2*) of mixed-sex (top) versus larval and adult sex-specific populations (bottom), showing that *mss* expression is specific to adult males. Mean values are shown with standard error of the mean. Female data are 2-3 orders of magnitude below that for males. **b,** *Cre-mss-1* transcripts are detected in pachytene-stage primary spermatocytes. **c,** HA-tagged *Cre*-HA-MSS-1 is first detectable in spermatocytes (sc), and becomes enriched in spermatids (st). **d,** Some *Cre*-HA-MSS-1 is localized to the plasma membrane of spermatocytes, as indicated by the arrow. Blue fluorescence: Hoechst-stained DNA. **e,** Enlarged view of the boxed region in c, showing complete restriction to membranous organelles (MOs). **f,** *Cre*-HA-MSS-1 remains attached to sperm after activation and transfer to the female. **g,** Sperm cells dissected from a female show *Cre*-HA-MSS-1 in plasma membrane and fused MO remnants.

In the outcrossing species *C. remanei, mss* transcripts were expressed only in adult males (**Figure 6a**), with strongest expression in spermatocytes during mid-pachytene of meiosis I (**Figure 6b**). To determine subcellular localization of MSS peptides, we used CRISPR/Cas9 editing to tag the *Cre-mss-1* gene of *C. remanei* with the hemagglutinin (HA) epitope. *Crem*-MSS-1::HA expression was first detected in large vesicles and on the plasma membrane of spermatocytes, with intensity increasing and localization restricted to secretory vesicles in mature spermatids (**Figure 6c, d, e**). The secretory vesicles of nematode sperm, known as membranous organelles (MOs), fuse with the plasma membrane upon ejaculation and sperm activation (Ward et al., 1983).

MSS peptides might be processed by a signal peptidase to release a soluble fragment into the MO lumen, which could then be dumped into seminal fluid upon sperm activation. However, their transient plasma membrane localization in spermatocytes and predicted C-terminal GPI attachment signals (**Fig. 4**) suggested that MSS peptides might instead be attached to membranes. Consistent with this latter hypothesis, *Crem*-MSS::HA remained associated with activated sperm dissected from inseminated females (**Fig. 6f**). We observed staining of the plasma membrane and of MO-derived punctae (**Fig. 6g**), which may be fused vesicles that remain as cup-like invaginations (Ward et al., 1981). Persistence of MSS on the surface of sperm after activation suggested that MSS acts cell-autonomously, rather than through the seminal fluid.

Because the four *C. remanei mss* paralogs form a 7-kb tandem array (**Figure 7a**), we deleted the entire *mss* cluster via CRISPR/Cas9 editing. To generated homozygous *mss* null mutants, the edited *C. remanei* heterozygous strain was backcrossed many times. Presumably because of inbreeding depression, the resulting homozygous strain showed lower fecundity (**Figure 7b**). To avoid inbreeding depression associated with homozygosity of entire chromosomes, we generated the *mss* deletion in two different *C. remanei* strains and crossed them to create hybrid *mss*-null mutants. The resulting males showed no intrinsic fertility defects as judged by overall brood size (**Figure 7c**). However, when competing against heterozygous *mss*(null/+) males, *mss* mutants sired fewer progeny than non-mutants in both offense (mutant male second) and defense (mutant male first) scenarios (**Figure 8a, b**). The *mss* family is therefore not required for fertility itself, but for male sperm competitiveness in multiple mating situations. Sperm lacking MSS compete poorly even when the female reproductive tract is conditioned by wild-type sperm. Thus, MSS proteins probably do not function as a secreted signal, but instead act cell-autonomously.

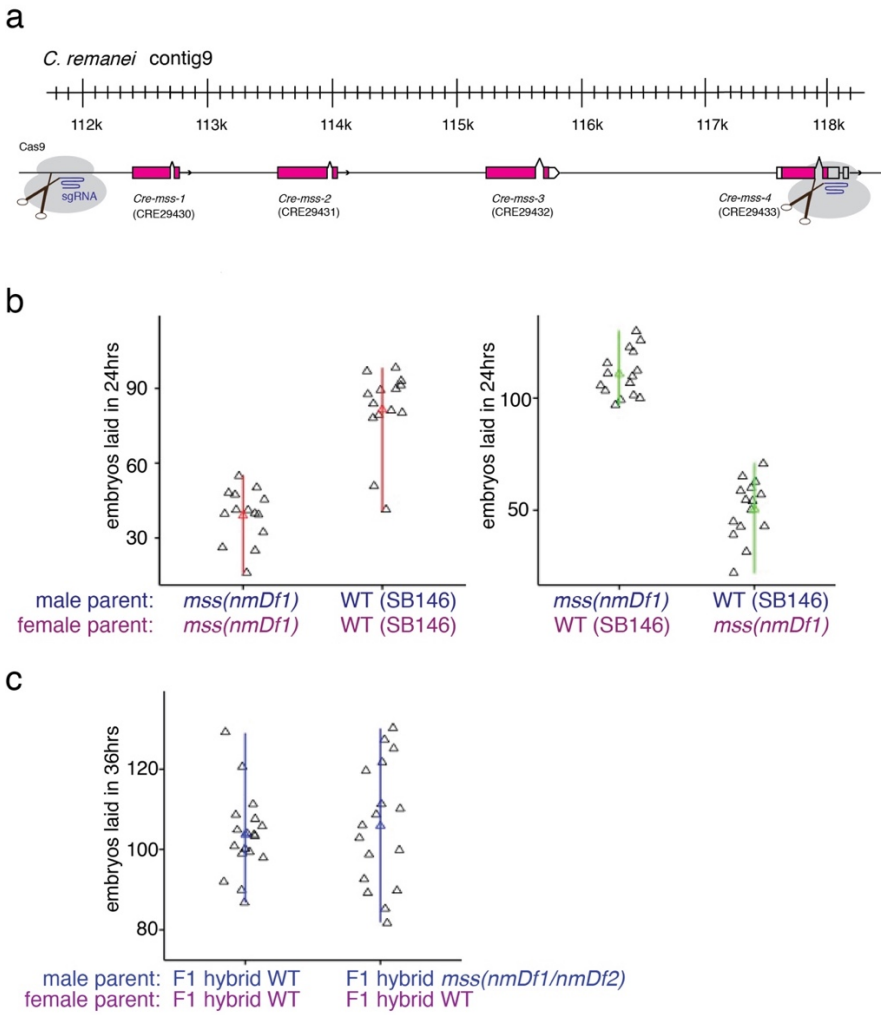

Figure 7. MSS does not affect intrinsic fertility when inbreeding depression is eliminated. a Strategy for CRISPR-mutagenic deletion of the entire tandem array of *mss* paralogs from the genome of *C. remanei*, which yielded the deletion alleles *Cre-mss*(*nmDf1*) and *Cre-mss*(*nmDf2*). **b** (left), *C. remanei* strain SB146 *mss(nmDf1)* parents produced fewer embryos in a 24 hr period than wild-type *C. remanei* SB146 parents. **b** (right), *C. remanei* SB146 *mss(nmDf1)* males fertilized wild-type females and produced more embryos than wild-type males and *C. remanei* SB146 *mss(nmDf1)* females were the parents ($P < 0.001$, two-sample Kolmogorov–Smirnov test). **c,** Hybrid (SB146/EM464) wild-type parents produced embryos in a 36 hr period that were very similar in number to hybrid *mss(nmDf1/nmDf2)* mutant males and wild-type females (statistically non-significant, two-sample Kolmogorov–Smirnov test).

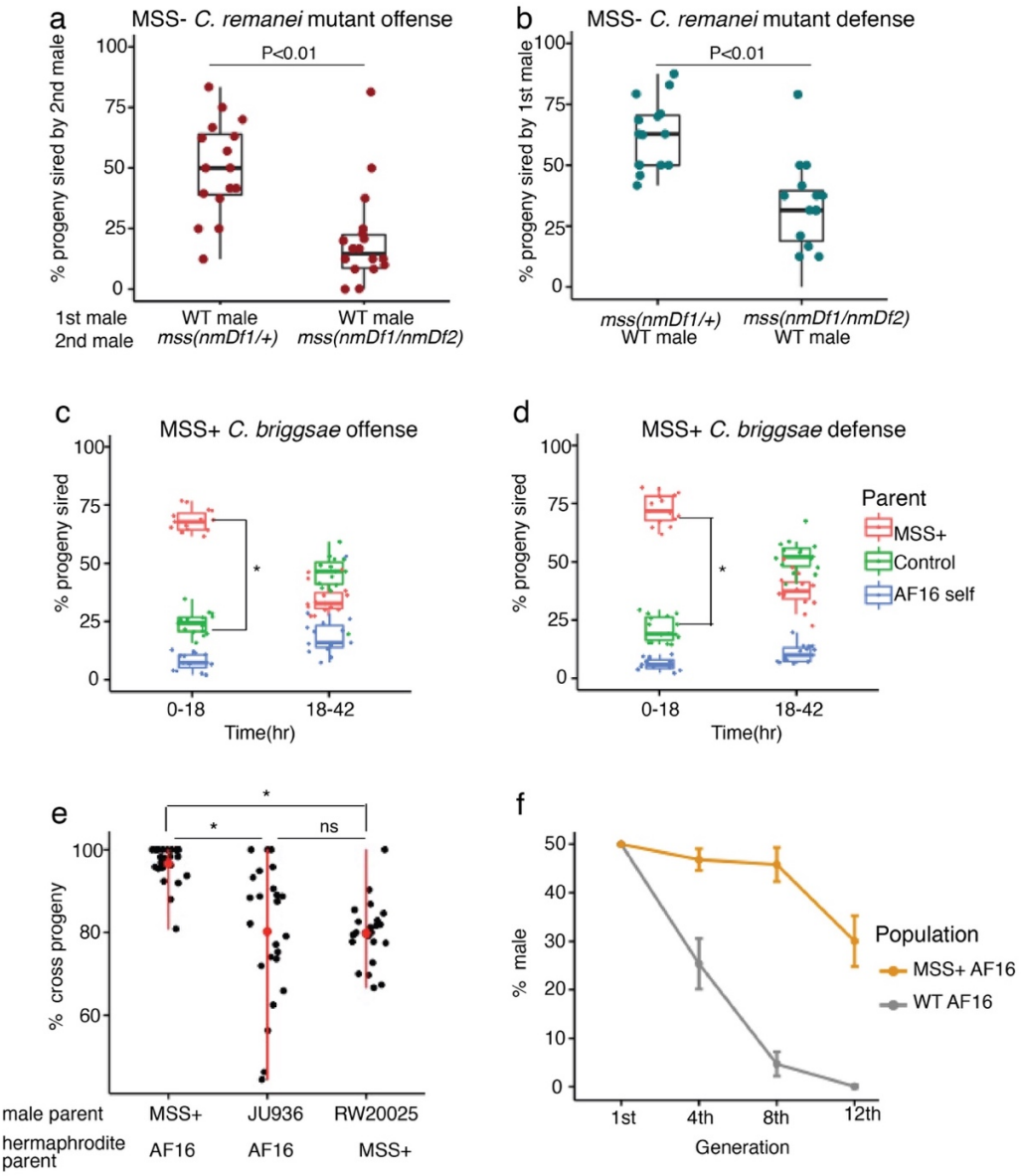

Figure 8. *mss* genes are necessary for sperm competitiveness in an outcrossing species, and sufficient to enhance it in a selfing species. a, When mated after a wild-type male ("offense"), *C. remanei mss(nmDf1/+)* males sire more than twice the progeny of *nmDf1/nmDf2 mss*-null mutants (N = 16 for both). **b,** When allowed to mate first ("defense"), heterozygous *C. remanei mss*(*nmDf1/+*) males have a slight advantage over wild-type males. *mss*-null mutants, in contrast, do not (N = 15 for both). Heterozygote success is assumed to be double the observed *nmDf1* frequency in their progeny. For both defense and offense, $p < 0.01$ (two-sample Kolmogorov–Smirnov test). **c, d,** Wild-type young *C. briggsae* hermaphrodites were mated sequentially (4 h each) with conspecific males carrying either an *C. nigoni mss(+)* transgene or control mCherry::histone reporter (RW0025). Progeny laid 0-18 h and 18-42 h after the second

mating were scored for green (MSS+), red (RW0025), or no (self) fluorescent markers. In both offense (c) and defense (d), MSS+ males sire several-fold more progeny than control males in the first laying window. *, $p < 0.001$. **e,** MSS+ *C. briggsae* males suppress selfing more effectively than control AF16 (wild-type) males. Strain JU936 is a second control strain bearing two transcriptional GFP reporters in the AF16 background. (*, $p < 0.001$; ns, not significant Kolmogorov–Smirnov test) **f,** Male frequency in MSS+ and wild-type AF16 *C. briggsae* populations in which male frequency was artificially elevated to 50% at the start of the experiment.

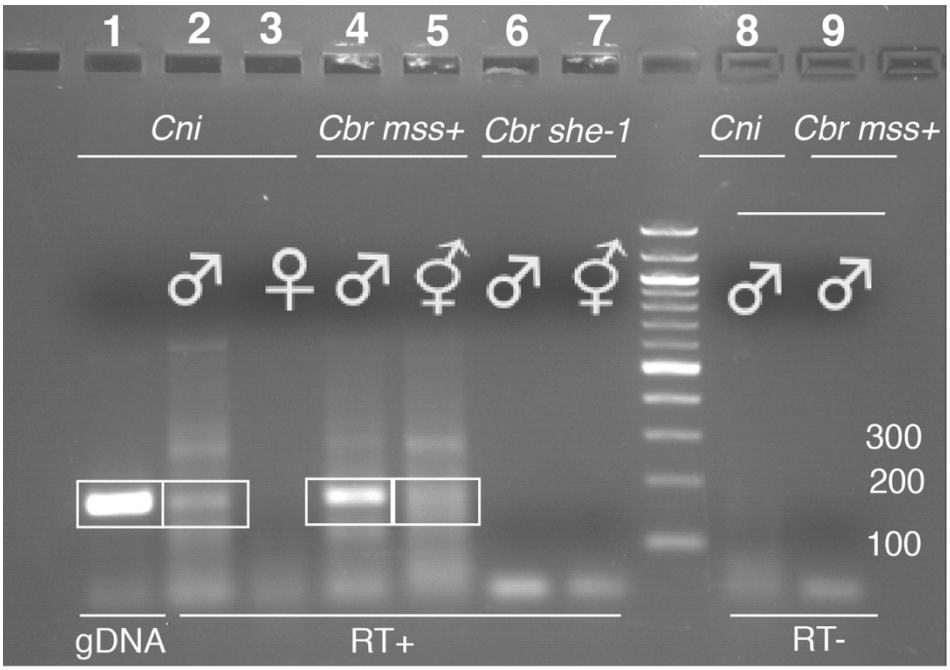

Figure 9. *Cni-mss-3* expression in transgenic *C. briggsae.* *Cni-mss-3* transcripts from transgenic *C. briggsae* males (strong expression) and hermaphrodites (weak expression) were detected by RT-PCR (lanes 4, 5). Positive controls were performed using *C. nigoni* male genomic DNA and cDNA (lanes 1, 2). Negative controls were performed using *C. nigoni* female (lane 3); *C. briggsae she-1(v35)* mutant females (lanes 6, 7) and RNA from *C. nigoni* and transgenic *C. briggsae* without reverse transcription (lanes 8, 9).

We then introduced *mss-1* and *mss-2* genes from *C. nigoni* into *C. briggsae* via a low-copy, germline-expressed MSS transgene; this transgene was strongly expressed in *C. briggsae* males, while also being detectable in hermaphrodites (**Figure 9**). Remarkably, sperm from transgenic *mss*(+) *C. briggsae* males outcompeted those of wild-type males (**Figure 8c, d**). After *mss*(+) sperm were exhausted, however, wild-type *mss*(null) sperm were still fertilization-competent (**Figure 8c, d**). In addition, *mss*(+) males were more consistently able to suppress use of a hermaphrodite mate's self-sperm (**Figure 8e**).

Because 50% of outcross progeny are male, but selfed progeny are almost exclusively hermaphrodites, we examined the effect of transgenic *mss* on long-term sex ratios in *C. briggsae* populations. We started both wild-type and *mss*(+) *C. briggsae* populations with a 1:1 male-to-hermaphrodite sex ratio and examined them over time. Wild-type *C. briggsae* showed a rapid decline of males, as previously seen in *C. elegans* (Chasnov and Chow, 2002; Stewart and Phillips, 2002). However, male frequency remained elevated in the *mss*(+) strain **(Figure 8f)**, only declining after 12 generations. The expression of MSS proteins was thus sufficient to shift population sex ratios towards parity.

Discussion:

Genes encoding small proteins with male-biased expression are disproportionately lost in *C. briggsae*, with *mss* providing an instance affecting reproduction. Unlike *comp-1,* which encodes a kinase required for male versus hermaphrodite sperm competition in *C. elegans (Hansen et al., 2015)*, and which is conserved regardless of mating system, we found *mss* orthologs only in outcrossing species. In interspecies matings, sperm from males of outcrossing species rapidly invade the ovaries and body cavities of selfing hermaphrodites, sterilizing or killing them (Ting et al., 2014). This cryptic toxicity of outcrossing sperm is likely due to ongoing sexual selection in outcrossing species. Given their pronounced role in sperm competition, MSS proteins may contribute to sperm invasiveness.

How MSS improves sperm competitiveness remains unclear, but mature MSS proteins are substantially glycosylated (**Figure 2**). Such posttranslational modification may impose little constraint on MSS proteins, explaining how they can have weak sequence conservation yet strong functional conservation. Another poorly conserved O-glycosylated protein, the mucin PLG-1, forms a copulatory plug found in all male-female *Caenorhabditis* species but lost in many wild isolates of *C. elegans* (Palopoli et al., 2015). Glycoproteins form the glycocalyx coat of mammalian sperm, and play important roles in fertility (Tecle and Gagneux, 2015). *Caenorhabditis* provides a useful model for interactions between the glycocalyx and female tissues, and how they affect sperm competition.

Independent loss of *mss* in the three known hermaphroditic *Caenorhabditis* species could reflect either relaxed sexual selection coupled with mutation and drift, or

adaptive convergence. Other changes in selfing species, such as loss of *plg-1* and of *plep-1,* which mediates reliable male discrimination between the vulva and excretory pore (Noble et al., 2015; Palopoli et al., 2015), are likely due to relaxed selection. However, restoring *mss to C. briggsae* enhances male fitness (**Figure 8c,d**), and mutations inactivating the *Cbr-mss-3-ps* pseudogene are not deletions that would be subject to loss via transmission ratio distortion (**Figure 5**). Could the loss of *mss* reflect adaptive convergence, permitting proto-hermaphrodites to adapt to a selfing lifestyle, or to resolve emergent sexual conflicts related to mating (Chasnov, 2010; Glémin and Ronfort, 2013; Sicard and Lenhard, 2011)? Selfing *Caenorhabditis* lack inbreeding depression (Dolgin et al., 2007b) and reproduce in spatially isolated habitats colonized by small numbers of founders (Kiontke et al., 2011). Reduced male mating success creates hermaphrodite-biased sex ratios (**Figure 8f**), which may be adaptive under these conditions (Chasnov, 2010; Hamilton, 1967; Lively and Lloyd, 1990). Thus, we hypothesize evolutionary transitions in reproductive mode may produce conditions for selection to rapidly eliminate formerly constrained reproductive genes. We will test some of the hypothesis in Chapter three.

Chapter 3: Mating system and population structure interact to determine the fitness of an evolutionarily labile sperm competition gene.

Summary

Several species of *Caenorhabditis* nematodes have evolved the androdioecious mating system, where hermaphrodites and rare males coexist. These species reproduce mostly through self-fertilization, but can also outcross. While selfing produces XX hermaphrodites, cross-fertilization produces 50% XO male progeny. Androdioecious *Caenorhabditis* have evolved reduced mating success, but the extent of this (and hence male frequency) varies in different populations and environments. The maintenance of males at intermediate frequencies is an interesting evolutionary problem. Here, we focus on the contribution of the *male secreted short (mss)* gene family to male mating success, sex ratio, and population growth. The *mss* family is essential for full sperm competitiveness in gonochoristic species, but has been lost in parallel in androdioecious species. Using a transgene to restore *mss* function to the androdioecious *C. briggsae,* we examined how mating system and population subdivision influence the fitness of the *mss*+ genotype. Consistent with theoretical expectations, *mss*+ is sufficient to increase male frequency and depress population growth in genetically homogenous androdioecious populations. When *mss*+ and *mss*-null (i.e. wild-type) genotypes compete, *mss*+ is positively selected in both mixed-mating and strictly outcrossing situations, though more strongly in the latter. Thus, sexual system greatly affects the

fitness of *mss*+. These results confirm that the evolution of self-fertility relaxed the sexual selection that historically favored *mss* retention. They also suggest that the lack of inbreeding depression and the strong subdivision thought to characterize natural *Caenorhabditis* populations impose selection on sex ratio that makes loss of *mss* adaptive.

Methods

C. briggsae strains. AF16, a *C. briggsae* wild isolate originally isolated in India, was obtained from the *Caenorhabditis* Genetics Center. CP161 (*nmIs7[Cni-mss-1(+) Cni-mss-2(+) Cbr-myo-2::GFP unc-119(+)]; Cbr-unc-119(nm67) III*) and CP162 (*nmIs8[Cni-mss-1(+) Cni-mss-2(+) Cbr-myo-2::GFP unc-119(+)]; Cbr-unc-119(nm67) III)* are independent *mss*+ chromosomally integrated transgenic strains generated through micro-particle bombardment as described previously (Yin et al., 2018). In addition to *mss* genes from *C. nigoni,* the construct has a dominant *Cbr-myo-2::*GFP reporter and *Cbr-unc-119(+)*, which serves as a marker for identification of successful transgenic lines. CP161(*nmIs7*) and CP162(*nmIs8*) have been confirmed for expression of *mss* by RT-PCR. Males have strong *mss* expression while hermaphrodites have weak but above background expression. CP164 was generated by crossing a *C. briggsae* strain *Cbr-she-1(v49)* (Guo et al., 2009) to CP161(*nmIs7*), followed by sibling matings so that the resulting strain is homozygous for both *mss* and *she-1* allele. *she-1* loss-of-function mutations cause a feminization of XX germline. CP164 therefore propagates itself as male/female species through outcrossing. Since worms carrying

she-1 alleles are inherently temperature sensitive, the strains were maintained in at 25° C.

F1 sex ratios and population sizes. For each strain (wild-type AF16, CP161 and CP162), fifty L4 males and fifty L4 hermaphrodites were set up for crossing for 24 hours, after which individual mated(as verified by mating plug) hermaphrodite was transferred to a 6 cm NGM agar Petri plates seeded with *E. coli* strain OP50(3cm diameter) and allowed to lay embryos for 40 hours before the hermaphrodite mother was killed. Then after 3 days of incubation at room temperature, the number of males and hermaphrodites (L4 and adult stages) was counted. To count the worms, M9 buffer with 0.05% Tween (to prevent worms sticking to pipette tips) was used to wash them off the plates and into a 2ml Eppendorf tube and resuspended in 200ul. Then the suspension was transferred onto a plain glass slide (AmScope) as thin streaks of liquid for counting. Every worm in the suspension was counted.

F2 sex ratios and population sizes. The strains were cultured the same way as described above except an individual mated P0 hermaphrodite was transferred to a 10 cm NGM agar Petri plate (OP50 bacterial spot diameter: 6cm) as the starting P0. The large NGM plates used here ensured the worms do not run out of food source after two generations. After 5 days at room temperature, the worms were washed into a 15ml Falcon tube. Then the worm suspension was diluted 10 times with M9 buffer and aliquoted to 96-well microtiter plates at approximately ten worms per microliter of M9 medium. The 96-well microtiter plate was sorted to quantify the number of animals using the large particle nematode sorter (COPAS Biosort, Union Biometrica). No specific size gating was set, therefore worms of all sizes and stages, including young

L1, L2 and embryos were included in the initial counts. Adult worms were distinguished from larva worms based on extinction (EXT) value (Smith et al., 2009). Only L4 and adult worms that represent F2s were used in the summary statistics boxplots and calculation of its statistical significance. A separate set of culture plates were used for sex ratio estimation. ~100 worms on each plate were washed off and transferred onto a plain glass slide for sexing.

Experimental evolution assay for GFP ratios. For each strain (genotype), wild-type AF16 and CP161, fifty L4 males and fifty L4 hermaphrodites were set up for crossing for 24 hours, after which 5 mated hermaphrodites of AF16 and 5 mated hermaphrodites of CP161 with the same number of males were transferred to a new plate and as a mixed population. Six replicates were set up. All populations were grown on 6 cm NGM agar Petri plates seeded with *E. coli* strain OP50. Each week (with ~2 generations/week), roughly 3% of a culture was transferred to a new plate by excising a 1 cm square of agar. The ratios of worms with and without GFP expression were counted every generation from 1st to 10th generation, then counted at 14th, 18th, and 25th generation. *Cbr-she-1(v49)* and CP164 experiments were set up in the same way except there were sixteen replicates. The population were maintained at 25° C.

Experimental evolution and determination of mss genotype frequency. For the mixed-mating experiment described in Figure 4a, we estimated the fractions of GFP+ animals that were homozygous and heterozygous for the *mss*+ transgene by isolating 6-7 individual GFP+ hermaphrodites from each of the 14 replicates at generation 25, and allowed them to lay progeny. For the obligate outcrossing *she-1* experiment in **Figure 4b**, the frequencies of homozygous and heterozygous *mss*+ were determined

by mating GFP+ virgin female worms at the 25th generation with *Cbr-she-1* (GFP-) males. Offspring were scored for GFP expression, and in all cases showed either all progeny were GFP+ (*mss*+/+) or a roughly equal mixture of GFP+ and GFP- progeny (for *mss*+/-). Chi-squared tests were performed on the deviation between observed and expected frequency based on Hardy-Weinberg equilibrium. One-tailed Sign-test were used to test whether observed increase in *mss* is significant compared to the initial frequency. All statistical tests were computed using R version 3.5.1.

Modeling interaction of male frequency, mating success, population size and mss+ allele frequency. We used a deterministic, discrete-generation model with no resource limitation, which should be accurate in early stages of population expansion. The expected male frequency in a given generation, m', is calculated from male frequency at the previous generation (m) and male fertilization success (α) as m' = mα/2. Hermaphrodite frequency equals (1 – m). In the first generation, the population is composed of a single mated hermaphrodite, whose fecundity is *F* offspring. The number of males in the second generation is $F(\alpha/4)$, while number of hermaphrodites is $F(1-(\alpha/4))$. The population size at generation three equals $F^2(1-(\alpha/4))$. With accurate counts for *F* and F^2 from second and third generation cultures, α can be estimated. The presence of the *mss*+ transgene is assumed to only impact α, through its effect on sperm competitiveness. For the mixed-genotype models (**Figure 1c** and **1d**), we assumed dominance of *mss*+ in *mss-/mss*+ males and no impact of *mss* genotype on hermaphrodites or females. For the gonochoristic model **(Figure 1d)**, we assume all eggs are cross-fertilized, so that a male genotype's contribution to the next generation

is the product of male frequency and a sperm competition factor, *C,* associated with that genotype.

Results

Modeling sex ratio, population size, and mss+ allele frequency.

Several models have been proposed for the evolution of androdioecy and the maintenance of males (Hedgecock, 1976; Otto et al., 1993; Stewart and Phillips, 2002). In theory, selfing rate, male fertilization success (α), inbreeding depression, the rate of nondisjunction at the X chromosome, relative viability difference between males and hermaphrodites all can impact the frequency of males. However since inbreeding depression is low in long-selfing species (Dolgin et al., 2007a; Gimond et al., 2013; Johnson and Hutchinson, 1993), X nondisjunction is rare, and male and hermaphrodite viability are equal in the early, reproductive phase of life, a simplified model for the change of male frequency overtime would be m' = αm/2, where m and m' are male frequency at current and subsequent generation, respectively, and α is the male fertilization success. α combines both male mate-finding (pre-mating) and sperm precedence (post-mating) success, and can be anywhere from 0 to 2. αm is the proportion of hermaphrodite eggs that are outcrossed. When α is zero, males cannot sire any progeny and the population is pure selfing. When α is maximal, m' = m, and males are maintained. If males additionally start as half the population, all progeny are cross-fathered and the population is a male-female equivalent. We can use this model to empirically determine the male fertilization success based on change of male frequency. Assuming a constant value of α at all male frequencies, any value of less

than 2 leads to eventual loss of males (**Figure 1A**). There is some indication that α itself may vary as a function of male frequency (Stewart and Phillips, 2002), but in common lab strains, populations enriched for males steadily lose them (Chasnov and Chow, 2002; Stewart and Phillips, 2002; Yin et al., 2018).

The simultaneous benefits of selfing and the need for at least occasional outcrossing may create an intermediate optimum sex ratio in facultative selfers like *C. elegans*. The XX-XO sex chromosome system would further imply that such modulation of ratio could be implemented at the level of mating success. Previously we identified a *Caenorhabditis* gene family, *mss* (for "male secreted short"), that can affect male frequency in the population. *mss* genes are conserved in the outcrossing species, but are consistently lost in the selfing androdioecious species (Yin et al., 2018). MSS proteins are heavily glycosylated surface factors that confer male sperm competitiveness. In the outcrossing *C. remanei,* mutant males lacking all *mss* gene paralogs are fully fertile, yet sire fewer progeny than wild-type *mss*+ males in competition. In the selfing *C. briggsae,* transgenic males in which *mss* has been restored sire more outcrossed progeny than *mss*- (i.e. wild-type) males, and are more effective at suppressing hermaphrodite self-sperm use. The strong advantage conferred by *mss* thus presents a mystery: Why would the males of mostly selfing species consistently lose an important weapon that evolved over millions of years of post-copulatory sexual selection? It is possible that the reduction of sexual selection leads to loss of *mss* purely through drift. Here, we consider an alternative, namely that *mss* loss was adaptive in the incipient selfing ancestors of *C. elegans* and *C. briggsae* because it resolved a sexual conflict over mating (Chasnov, 2010).

We have used standard plate cultures to simulate natural populations of the sort that may have existed as *C. briggsae* was evolving, and adapting to, selfing. We hypothesized that a *C. briggsae* population without *mss* (which we will refer to here as *mss*-) would grow faster than one with (*mss*+), owing to a lower value of α, lower male frequency, and increased egg output in the former. This is modeled in **Figure 1B**. In a mixed-genotype population, however, *mss*+ males would still have an advantage over *mss*- males, and the *mss*+ genotype would increase in frequency so long as males are present in appreciable numbers (**Figure 1C**). In the extreme case of obligate outcrossing, *mss*+ should eventually fix in the population (**Figure 1D**), consistent with its consistent presence in gonochoristic *Caenorhabditis*. In this study, we test these hypotheses by examining sex ratio, population growth, and fitness of the *mss*+ genotype in both selfing and obligately outcrossing *C. briggsae* populations. The results suggest that mating system and population structure may have interacted to promote the adaptive loss of the *mss* family in selfing *Caenorhabditis.*

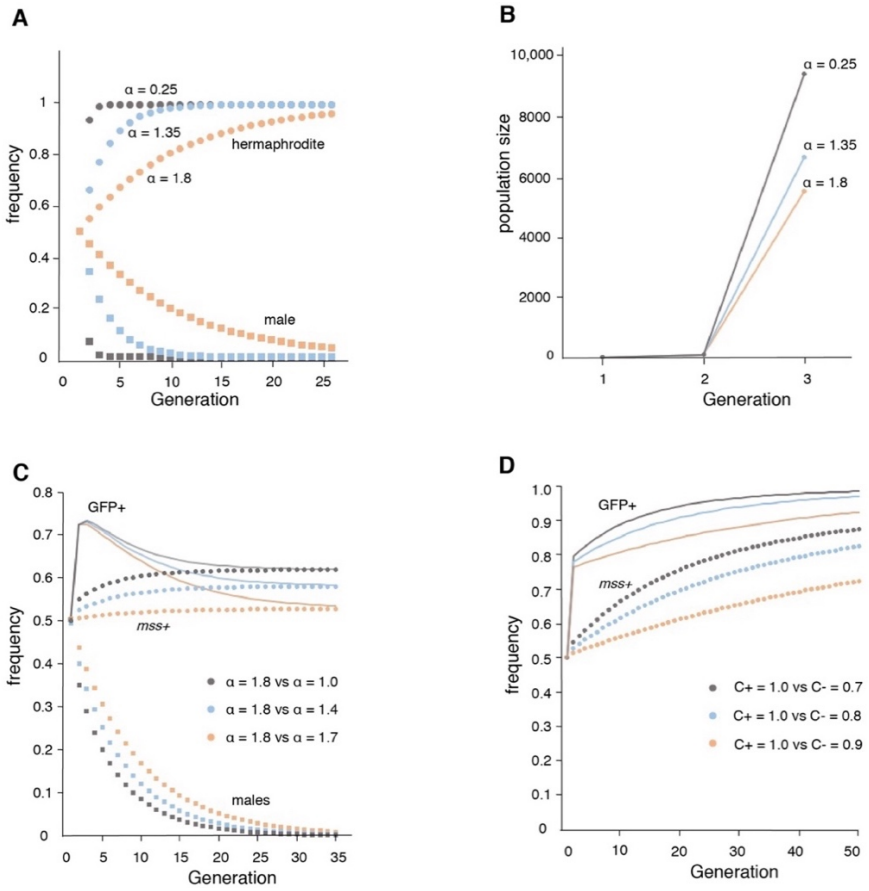

Figure 1. Modeling sex ratio, population size, and *mss*+ allele frequency. (A) Impact of varying fertilization success (α) on sex ratio over time in a genetically homogenous population. Males and hermaphrodites are both at 50% frequency in the first generation. At any value of $\alpha < 2$ male frequency declines and hermaphrodite frequency increases. (B) Expected impact of α on population growth. At generation one, the population is founded with one mated hermaphrodite. At generation two, population size is determined solely by the fecundity of this hermaphrodite (here we assume each produces 100 eggs). The advantage of low male fertilization success is seen in the substantially larger number of grand-offspring in generation three. (C) Competition between *mss*+ and *mss*- genotypes in a panmictic population, founded with equal proportions of selfing hermaphrodites and males, half of which are homozygous for each genotype. Over time the *mss*+ allele frequency (colored dots) increases, but slows as males are eliminated (because $\alpha < 2$). The expected sum of *mss*+/+ and *mss*+/- diploid genotype frequencies at Hardy-Weinberg equilibrium is also shown (GFP+, thin solid lines) for comparison with experimental data in Figure 4. (D) Competition between *mss*+ and *mss*- genotypes in a panmictic, obligately outcrossing (i.e. male-female) population. The sperm competitiveness factor (*C*) for *mss*+ is set at 1, while that for *mss*- males can vary from 0 (sterility) to 1.

mss+ increases male production and depresses growth

We previously observed that CP161 and CP162, two transgenic *mss*+ *C. briggsae* strains, were better at maintaining males as compared to the *mss*- AF16 wildtype strain (Yin et al., 2018). In this study, first we sought to track male frequency over each generation and directly assess how the changes in the sex ratio impact population growth. We hypothesized that when the number of individuals are the same between two populations, the population with fewer males (and thus more hermaphrodites) would grow faster.

We started by crossing individual AF16 wildtype, CP161, and CP162 hermaphrodites with males of corresponding genotype, and then allowed each hermaphrodite to lay embryos for 40 hours. The progeny were sexed when they became adults. We found the percentage of males in the AF16 wildtype (*mss*-) population was significantly lower than in the two *mss*+ populations in the first generation (**Figure 2a**). With a separate set of cultures, we allowed the F1s to produce F2s and sexed the population when F2 reached adulthood. Again we found significantly higher percentage of males in the CP161 and CP162 populations as compared to the AF16 wildtype (**Figure 2b**).

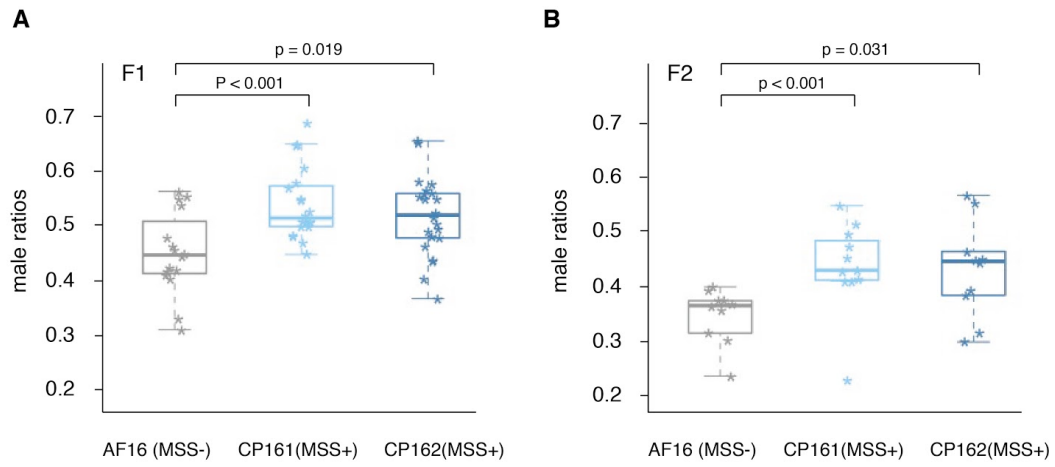

Figure 2. *C. briggsae* hermaphrodite mated with *mss*+ males produces higher ratio of males (A) In the F1 population (progeny produced by a single mated hermaphrodite), CP161(*mss*+)(n=20) and CP162(*mss*+)(n=22) both have a significantly higher ratio of males compared to AF16 (wild-type)(n=16). The median percentage is 0.448, 0.516 and 0.521 for each strain. (B) In the F2 generation, CP161 (*mss*+) (n=12) and CP162 (*mss*+) (n=10) both showed a significantly higher ratio of males compared to AF16 (wild-type) (n=12). The median percentage is 0.363, 0.430 and 0.446. P value for each comparison is indicated (Kolmogorov-Smirnov test)

Using these results, we can estimate the proportion of hermaphrodite eggs that are outcrossed (αm) for each of the genotypes. From P0 to F1, they are 89.2% (AF16), 100% (CP161), and 100% (CP162) respectively. From F1 to F2, they are 72.6% (AF16), 86% (CP161), and 89.2% (CP162). We can also calculate the male fertilization success α (0-2) for each of the genotypes. In the P0 to F1 generation, α is estimated to be 1.79 (AF16), 2 (CP161) and 2 (CP162). Because only hermaphrodites that bore copulatory plugs were picked for the P0 founders, the pre-mating component of α can be assumed to be 100%, and differences are likely wholly due to the post-mating component of α. From the F1 to F2 generations, where successful sperm transfer by the F1 males was not guaranteed, α was estimated to be 1.62 for AF16, 1.67 for CP161,

and 1.71 for CP162. Thus, male fertilization success is consistently higher in the *mss*+ populations than in the *mss*- AF16. Nevertheless, by the F2 generation some hermaphrodites were not outcrossed despite the presence of many males.

Next, we compared the growth of *mss*+ and *mss*- *C. briggsae* populations. The number of progeny laid by a hermaphrodite in 40 hours was not significantly different (**Figure 3a**). Thus, an intrinsic fertility difference between *mss*+ *C. briggsae* and *mss*- *C. briggsae* can be ruled out, validating a key assumption of our model. After the F1s produced F2s, we used an automated worm sorter to precisely count all progeny in these much larger populations. We found WT AF16 produced a significantly higher number of F2 (n=16 cultures, median=16,140) than CP161 (n=12, median=12,870) and CP162 (n=12, median=14,050) populations (**Figure 3b**). These results confirm the expected reduction of population growth incurred by high male mating success. However, when all counted progeny were included, the effect is even greater than predicted from the first generation sex ratio. We can model population growth as $p' = (p) \times (h) \times (f)$, where p' is the population size at generation N, p is the population size at generation N-1, h is the fraction of hermaphrodites, and f is the fecundity of each hermaphrodite. The median fractions of hermaphrodites at the first generation (0.552 for AF16 WT, 0.484 for CP161, and 0.479 for CP162) predict that in the second generation AF16 WT will have 1.14 times more progeny than CP161, and 1.15 times more than CP162. However, we observed WT AF16 (median=24,207) has 1.43 times more progeny than compared to CP161 (median=16,878) and 1.47 times more than CP162 (median=16,500) population.

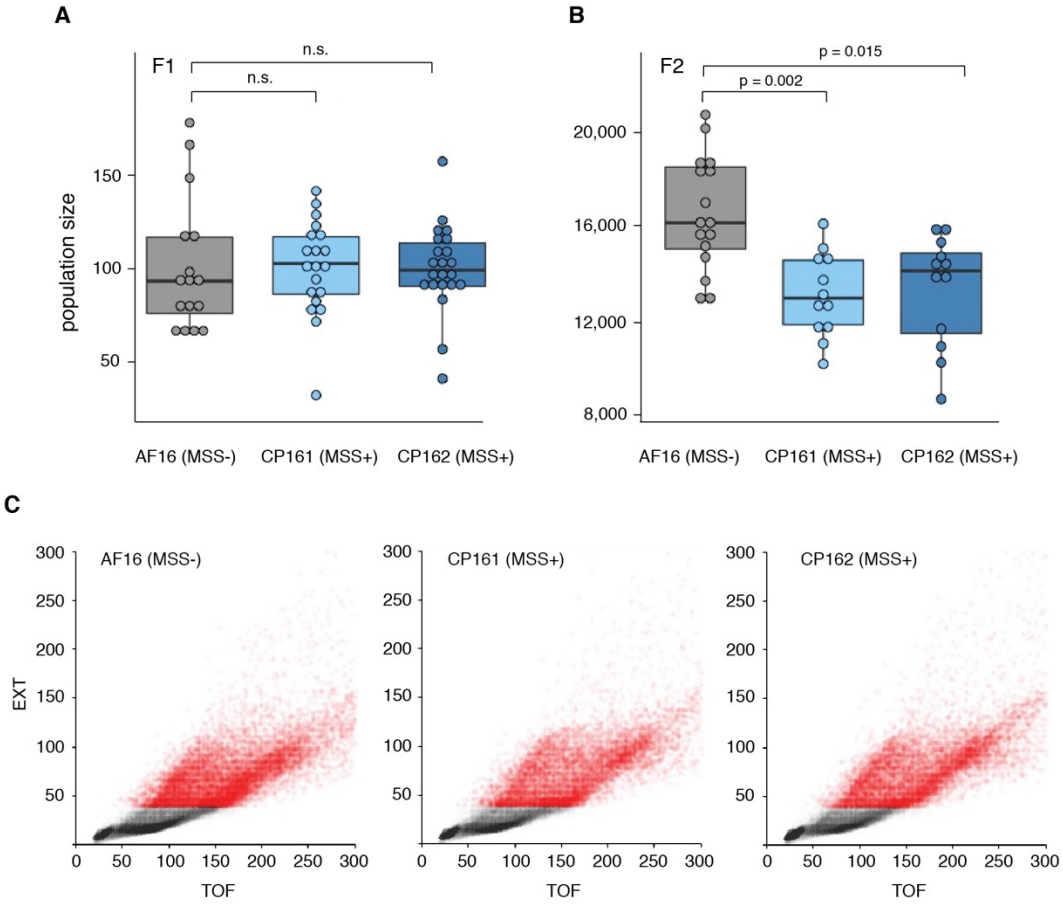

Figure 3. Populations with the *mss* transgene grow slower (A) The F1 progeny produced by a single mated hermaphrodite of strains AF16 (n=16), CP161 (n=20), and CP162 (n=21) are not significantly different. (B) The *mss*+ strains CP161 (n=12) and CP162 (n=12) both produced significantly smaller numbers of F2 progeny than the *mss*- AF16 (n=16). P value indicated for each comparison is by the Kolmogorov-Smirnov test. (C) Eight replicates with the number of worms closest to the medium value were used to make the scatterplots. Extinction (EXT) vs time-of-flight (TOF) data for individual worms were counted in the automated worm sorter. Data points in grey represent small, translucent embryos and larvae, many of which are F3 progeny. Data points in red represent animals developed enough to be F2 (or older) after 5 days of growth.

The above difference could indicate that the growth-suppressing effect of *mss*+ and/or the higher frequency of males associated with it exaggerates a baseline fecundity

difference. However, we noticed that at the 5th day of growth, the oldest progeny from the F2 populations have laid embryos (i.e. of the F3 generation); these were also counted by worm-sorter. Indeed, when the worm sorter data are gated to eliminate embryos, L1, and L2 larvae, the remaining number of worms (which must be F2 or residual F1) are substantially closer to expectation: AF16 has 1.25 times more progeny than CP161 (expected 1.14), and 1.15 times more progeny than CP162 (expected 1.15). We conclude that once F3 contamination is eliminated, the F2 growth of these experimental cultures can be predicted by F1 sex ratios.

Fate of mss+ allele in competitive context varies with mating system

In extant *C. briggsae*, traces of *mss* genes remain as pseudogenes (Yin et al., 2018), indicating they were lost recently. To model how the change to selfing may have impacted the fitness of the *mss+* genotype, we carried out two experimental evolution studies. One employed self-fertile stocks mixed with males, and the other a self-sterile, obligately outcrossing *she-1* (*spermless hermaphrodite)* loss-of-function mutation (Guo et al., 2009). *she-1* mutants reproduce effectively as a male/female strain, similar to *fog-2* mutant *C. elegans* (Schedl and Kimble, 1988). The frequency of the *mss+* transgene, which is marked with a dominant *myo-2::*GFP reporter, was tracked in mixed-genotype populations. Each began with equal numbers of mated AF16 (*mss*-/-, GFP-) and CP161 (*mss*+/+, GFP+) hermaphrodites.

In the mixed-mating populations, GFP+ frequency is expected to increase from the initial 0.50 due to the formation of heterozygotes by crosses between the parental strains, and potentially by selection for the *mss+* allele. However, only with positive selection will the *mss+* allele frequency be consistently elevated above the initial 0.50.

In our experiment, GFP+ frequency rose to a median of 63.5% (**Figure 4A**). Guided by an empirical estimate of heterozygote and homozygote fractions, we converted observed GFP+ frequencies at generation 25 to *mss*+ allele frequencies. These ranged from 0.42 to 0.67, had a mean of 0.57 (S.E.M. +/- 0.02), and for 11 of 14 replicates were greater than 0.50. This distribution is shifted significantly above the starting frequency of 0.50. We conclude that the *mss*+ is allele is under modest but detectable positive selection in the mixed-mating case. Surprisingly, male frequency remained relatively high (0.09) at generation 25.

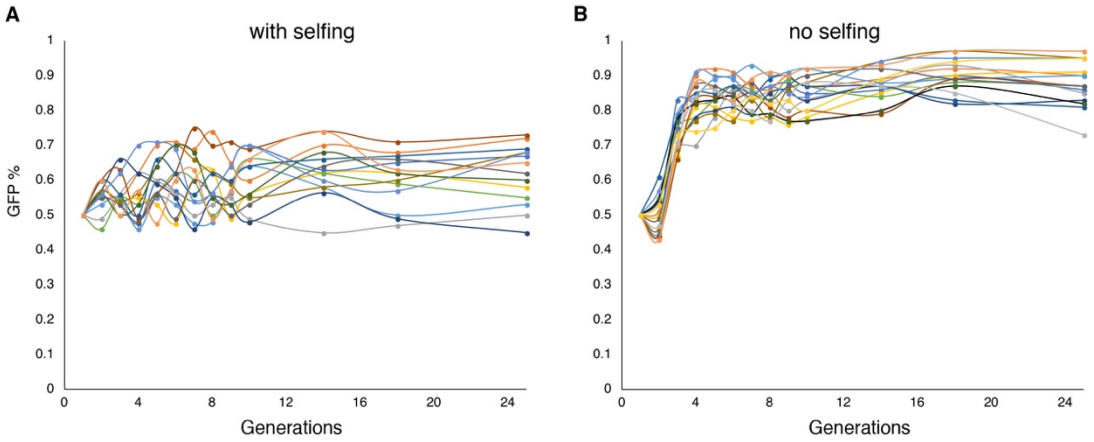

Figure 4. Relative fitness of *mss*+ and *mss*- genotypes in androdioecious and gonochoristic populations. (A) In a mixed population of *C. briggsae* WT AF16 (*mss*-, GFP-) and CP161 (*mss*+, GFP+), where selfing and outcrossing co-exist, GFP+ (sum of *mss*+/+ and *mss*+/- diploid genotype) frequency increases from 50% to a median of 63.5% over 25 generations. The estimated mean and median *mss* frequency at the 25th generation, 0.59 and 0.57, are significantly higher than 0.50 (P = 0.0287; one-tailed sign-test, n=14). (B) Competition in a mixed population of *C. briggsae she-1(v49); mss*-, GFP- and CP164 (*she-1(v49), mss*+, GFP+), where all XX animals lack self-sperm. Median GFP+ frequency increased to 0.72 at the 3rd generation, and was 0.87 at the 25th generation. The GFP+ frequency distribution at generation 25 is significantly higher than 0.75 (P = 0.0002594; one-tailed sign-test, n=16). Similarly, the mean *mss*+ allele frequency at the 25th generation is 0.67, significantly higher than the initial 0.50. (P = 7.629e-06; one-tailed sign-test, n=16)

Next we asked how the frequency of *mss*+ would change in a *C. briggsae* strain that can only outcross. To do this, we generated an *mss*+, GFP+ strain, CP164, that also carries the *she-1* mutation (Guo et al., 2009). When mixed populations were established with equal numbers of *mss*-; *she-1* (GFP-) and *mss*+; *she-1* (GFP+) worms, GFP+ frequency increased to a median 0.72 at the 3rd generation. This initial increase can be fully explained by the formation of the first heterozygotes. However, the frequency of GFP+ genotypes continued to rise, to a median value of 0.87 at the 25th generation (**Figure 4B**). This is significantly higher than the 0.75 expected from genetic drift alone. In addition, the mean *mss*+ allele frequency at the 25th generation is 0.67, significantly higher than the initial 0.50. These results indicate that *mss*+ enjoys higher fitness than *mss*- in *she-1* mutants, and that outcrossing selects more strongly for it than the mixed-mating case.

Using mating tests to distinguish homozygous and heterozygous individuals among 100 GFP+ (*mss*+) animals at the 25th generation, we estimated that the *mss*+/+, *mss*+/-, *mss*-/- frequencies were 0.47, 0.39, and 0.14, respectively. Therefore, the *mss*+ and *mss*- allele frequencies were 66.5% and 33.5%, respectively. Given these allele frequencies, the observed genotype frequencies are not significantly different from the Hardy-Weinberg equilibrium expectations (40, 46 and 14, respectively; $x^2 = 1.14$, $df = 1$; $P = 0.57$).

Discussion

The repeated evolution of self-fertility suggests that selfing provides some benefits to the species, perhaps through reproductive assurance at low densities and ease of invasion to new ecological niches. Provided they produce minimal sperm, selfers do not incur the two-fold cost of males, and thus enjoy a large boost in intrinsic growth rate (Bell, 1982; Smith, 1978). The advantage of selfing can been seen in the obligately outcrossing populations of *C. elegans fog-2* mutant, where independent gene conversion events that repair the loss-of-function mutation in the *fog-2* locus and revert the mode of reproduction from outcrossing back to selfing rapidly spread (Katju et al., 2008). Similarly outcrossing populations that start with 50% males can be invaded by *C. elegans* hermaphrodites (Chasnov and Chow, 2002; Stewart and Phillips, 2002; Teotónio et al., 2006).

While hermaphrodite *Caenorhabditis* lay more embryos after crossing than if limited to self-fertilization (Hodgkin and Doniach, 1997), total brood size is not likely to be the most significant contributor to the growth of the population. First, the reproductive value of late progeny is discounted as compared to the progeny laid earlier (Hoogendyk and Estabrook, 1984). Furthermore, *C. elegans* and *C. briggsae* hermaphrodites do not show the same attractiveness to males because hermaphrodites secrete limited sex pheromones (Chasnov and Chow, 2002; Chasnov et al., 2007b; Leighton et al., 2014; Simon and Sternberg, 2002). Hermaphrodites are also less receptive to males, (Garcia et al. 2007). When hermaphrodites still have self-sperm, they tend to move away when males attempt to mate. Even after insemination, hermaphrodites have been shown to eject the male sperm (Barker, 1994; Kleemann and

Basolo, 2007). The resistance of hermaphrodites to mating thus appears to have made outcrossing inefficient. These findings indicate outcrossing in the nature is minimized in androdioecious species by a suite of changes in both sexes, yet complete male sterility has not evolved.

Loss of mss boosts population growth

The parallel loss of *mss* in three distinct lineages is striking, and suggested to us that it may not be via drift alone. Hamilton (Hamilton, 1967) proposed the local mate competition (LMC) model, which posits that in isolated small populations where only mated females disperse, interdemic selection favors a female-biased sex ratio. In addition, the inbreeding associated with this mating system imposes additional selection to minimize male frequency (Herre, 1985). Though these models were produced with fig wasps and other gonochoristic organisms in mind, they are relevant to *Caenorhabditis*. The male-female ancestor of *C. nigoni* and *C. briggsae*, like other *Caenorhabditis,* likely colonized spatially isolated rotting fruits in small numbers (Cutter; Kiontke et al., 2011). These circumstances resemble the LMC scenario, but offer the extreme case of zero mating as a viable option to selfing hermaphrodites. Selfing thus provides reproductive assurance in the colonization of novel substrate patches at low density, and in the absence of inbreeding depression also maximizes growth rate. We set out to test the plausibility of the idea that loss of *mss* family was a response to LMC-like conditions that selected for minimal outcrossing.

To approximate the LMC scenario, we compared male frequencies and population growth rates in populations of pure *mss*- and *mss*+. In the first generation, *mss*+ and *mss*- *C. briggsae* hermaphrodites do not differ in their intrinsic fecundity or

viability (**Figure 3A**). Consistent with the superior competitive ability of *mss*+ sperm against hermaphrodite self-sperm (Yin et al., 2018), however, we found higher frequencies of males in *mss*+ populations (**Figure 2**), even in the first generation when all hermaphrodites (regardless of genotype) had been successfully mated. We can thus attribute a portion of the enhanced frequency of fertilization success (α) of *mss*+ males to post-mating sperm competition. By the third generation, *mss*+ populations had lower census sizes, consistent with the prediction (**Figure 3B**). If animals in this or subsequent generations form dispersive L3-dauer larvae, the loss of the *mss*+ allele could increase out-migration, and thus the probability of successful colonization of the next patch.

Effects of mating system on mss fitness

We were also interested in trying to understand how the *mss*- and *mss*+ alleles interact in large, genetically mixed populations. Our simple model (**Figure 1A)** predicted that in both mixed-mating populations with selfing hermaphrodites and males, and in obligately outcrossing populations, the *mss*+ transgene will increase in frequency. However, this is expected to continue to fixation in the outcrossing competition, while it is expected to be muted and dependent on the presence of males in the mixed mating situation. Our experimental populations generally support these conclusions, but produced some unexpected results.

In the mixed-mating case, we did not observe an initial sharp increase in GFP+ frequency as expected (**Figure 4A**), indicating that mating between *mss*- and *mss*+ animals in the first generations was less than the estimated values of α would predict. Nevertheless, after 25 generations the median *mss*+ frequency rose modestly but

significantly, in keeping with our predictions. Given the poor initial mating inferred in the first few generations, however, we were surprised that male frequency (*m*) remained as high as it did (median of 0.09) at generation 25. For example, using the empirically calculated value of α for the *mss*+ CP161 strain in the second generation (1.67) for all subsequent generations, *m* is predicted to be .007 in generation 25, an order of magnitude smaller than observed. One possible explanation may be that α (fertilization success) of males varies as a function of male frequency, as suggested by Stewart and Phillips (Stewart and Phillips, 2002). For example, as males become rarer, they may encounter virgin hermaphrodites (as opposed to previous mated hermaphrodites or other males) more reliably, allowing multiple matings. It is also possible that some assortative mating occurred with regard to *mss* genotype, though no obvious mechanism to mediate this is apparent.

In the male/female (*she-1*) mixed population, the GFP+ allele frequency increased to 87% by the 25th generation, and corresponds to an *mss*+ allele frequency of 0.67. Thus, while the *mss*+ allele experienced positive selection in both mixed mating and obligately outcrossing populations, the effect is substantially stronger in the latter. This provides a clear explanation for the retention of *mss*+ in obligately outcrossing species.

Conclusion and future directions

Caenorhabditis nematodes have a unique mix of closely related outcrossing and selfing species. Prior studies noted that self-fertilizing *Caenorhabditis* species have smaller genomes than male-female *Caenorhabditis* species (Fierst et al., 2015; Thomas et al., 2012b). However, as the taxa compared were highly diverged from each other, it was unclear how directly or how rapidly a change in mode of reproduction might cause shrinkage in genome size. In addition, self-fertilizing species were known to have less aggressive sperm and atrophied mating behavior, but a link between genome shrinkage and inefficient hermaphrodite-male outcrossing had not been established. In the three main chapters of my dissertation, we use comparative genomics, functional analysis and predictive modeling/experimental evolution to demonstrate a direct link and establish the underlying mechanism. This study demonstrates that the precise mode of sexual reproduction plays a critical role in shaping the size and content of the genome.

First, we compared the high-quality, chromosome-scale *de novo* genome assembly of the male-female species *C. nigoni* to the genome of its close self-fertile relative *C. briggsae*. Strikingly, and despite their partial interfertility, *C. briggsae* has 23% fewer protein-coding genes than *C. nigoni.* As *C. nigoni's* genome size is typical of related male-female *Caenorhabditis* species, we infer that the genome and proteome of *C. briggsae* has recently and dramatically contracted.

The *C. nigoni* genes missing in *C. briggsae* encode proteins that are shorter and more male-biased in their expression than would be expected by chance, suggesting that small protein-coding genes related to male sexual traits are particularly susceptible to loss. The genome of the more widely studied self-fertile *C. elegans* also exhibits this

shrunken state. Absent from *C. elegans* are hundreds of genes conserved between male-female *Caenorhabditis* species. Given the low frequency of males in self-fertilizing species, these observations point to sexual selection as a driving force in maintaining much of the genome's content. Overall, these findings demonstrate a strikingly large role for the mode of reproduction in determining the size and content of animal genomes.

Second, we characterized a family of short, male-specific proteins named MSS (male secreted short) that are conserved in male-female *Caenorhabditis* species but lost from hermaphroditic ones. These are glycoproteins localized to the surface of activated sperm. We reasoned that MSS proteins might be involved in reproductive functions that are dispensable after self-fertilization evolves. We used CRISPR/Cas9-mediated genome editing in the male-female *C. remanei* to precisely delete all four of its *mss* paralogs. Though mutant males have normal fertility in a homogenous population, loss of *mss* profoundly impairs the ability of their sperm to compete with wild-type males in siring progeny. Thus, MSS proteins are novel sperm competition factors.

Remarkably, restoration of the MSS genes to males of the self-fertile *C. briggsae* is sufficient to confer a large advantage over wild-type *C. briggsae* males, whose sperm naturally lack MSS proteins. Furthermore, *C. briggsae* populations harboring *mss* genes maintain higher male frequencies by more effectively blocking self-fertilization in hermaphrodites. That *C. briggsae* would lose a gene which confers a clear fitness advantage to males is counterintuitive.

Third, we hypothesized that *mss* genes were not simply lost due to relaxed sexual selection, but were actually driven out of the genome by positive selection to

reduce male frequency, a situation predicted by W.D. Hamilton to occur when males primarily compete against close relatives in small isolated populations (Hamilton [1967], *Science* 156, 477-488). In natural populations of androdioecious *Caenorhabditis* males are rare, and considerably lower than observed in the laboratory conditions (Barrière and Félix, 2005, 2007). Selection may often oppose outcrossing, both for the growth-retarding effects of males addressed above and because selfers experience outcrossing depression and other genetic signatures (Baird and Stonesifer, 2012; Dolgin et al., 2007a; Ross et al., 2011; Seidel et al., 2008). Nevertheless, males persist and are generally fertile world-wide. The value of outcrossing can be demonstrated, and appears to be based environment. Exposure to the dauer lifestage, environmental stresses such as heat shock, mutational load, and novel pathogens all can lead to increased outcrossing (Manoel et al., 2007, 2007; Morran et al., 2009).

Taken together, we propose the sex ratio of selfing species may represent a balance between selection to maintain the ability to outcross under stressful conditions and selection to maximize hermaphrodite selfing under more benign conditions. The *mss* gene family is repeatedly lost in selfing species, suggesting that it may not be due solely to the slow process neutral mutational decay from weakened sexual selection. The results support the alternative, that *mss* loss is adaptive because it promotes a more hermaphrodite-biased sex ratio without blocking outcrossing completely.

Sperm competition is a ubiquitous feature of sexual reproduction in nature. When females undergo multiple mating, it creates the opportunity for postcopulatory sexual selection favoring males with sperm that has a fertilization advantage. Functional traits of sperm, such as migration velocity, retention ability, ability to block

subsequent access have all been shown to affect sperm competitiveness (Gomendio and Roldan, 2004; Pizzari and Parker, 2009; Wigby and Chapman, 2004). In *Drosophila* and mammals, some genetic loci underlying reproductive success have been identified (Civetta and Finn, 2014; Fiumera et al., 2005; Sutton et al., 2008; Yeh et al., 2012). In some cases, specific seminal fluid components that affect fertilization success have been characterized (Avila et al., 2010; Mueller et al., 2008; Simmons and Fitzpatrick, 2012). However, examples of a specific protein whose function can be entirely tied to sperm competition are few (Hansen et al., 2015; Yin et al., 2018).

In this study, we showed evidence *mss* is both necessary and sufficient for sperm competition. In the future, it will be interesting to understand what aspect of sperm MSS changes that allow it to have a reproductive advantage over the sperm that doesn't have MSS. In addition, it is now known *mss* belongs to a large family of sperm proteins, *msrps*. Using sequence similarity searches, a total of 81 *mss* and *msrp* homologs from *Caenorhabditis* were found. *mss* and *msrp* share many sequence similarities in the conserved N-terminal and C-terminal. Many *msrp* are even more highly expressed than *mss* and are found in both selfing and outcrossing species. It remains to be tested whether *msrp* has similar functions as *mss*.

Appendix A: Generation and characterization of *clec-223* mutants in *C. elegans.*

Introduction

C-type lectins are calcium-dependent glycan-binding proteins characterized by their ability to recognize and bind carbohydrate, thus representing "receptors" for which glycoproteins, such as MSS and MSRP could interact with in the female's reproductive tract. Because restoration of *mss* in *C. briggsae* allows it to regain its sperm competitiveness, it is conceivable that if MSS is interacting with a receptor, the receptor is still present in *C. briggsae. C. elegans* has over 200 C-type lectin genes. Of these, only 12 are hermaphrodite- biased with at least two-fold higher expression. Only two, *clec-190* and *clec-223* are only expressed in hermaphrodites. Provocatively, *C. elegans clec*-223's expression is restricted to the spermatheca, thus representing a particularly attractive candidate. We generated a *clec*-223 likely loss-of-function mutant using CRISPR/Cas9, and phenotyped this mutant for reproductive trait changes. Traits examined thus far are fecundity and sperm localization relative to wild-type.

Methods

Generation of clec-223 mutants through CRISPR/Cas9. DY120 and DY121 were used for making a linear template for *in vitro* transcription (IVT). IVT was carried out using the SP6 Megascript kit (Invitrogen). Three reactions for each gRNA (total 60µl) were carried out, followed by ethanol precipitation. The precipitated gRNA was resuspended in 12 µl water. In the final injection mixture, 0.5µl DY120 gRNA, 0.5µl

DY121 gRNA, 0.5µl *dpy-10* gRNA, 0.5µl *dpy-10* repair template DNA (600 µm), and 4µl Cas9 (1250ng/ µl) were added. 12 hermaphroditic young adult N2 worms were injected. 4/12 injected mothers laid progeny that have roller phenotypes. The roller mothers were allowed to lay progeny before they were screened using single worm PCR. DY129 and DY134 primer pair was used in PCR to screen for an amplicon smaller than 250bp, indicating a deletion. Once a mutant was found through PCR, the mutation was confirmed by sequencing using primer DY131 or DY134. Then the mutant was rendered homozygous by backcrossing. Then to get rid of any background effect, the mutant was outcrossed to the wild-type N2 three times, to generate the final strain CP167 (*clec-223*).

N2 vs clec-223 brood size assay. N2 and CP167 (*clec-223*) strains were stage-synchronized by egg preparation protocol. Both were cultured at room temperature and allowed to develop into L4 stages. Then individual L4 worms were picked into individual plates. After 24 hours, the number of embryos were counted (1^{st}-24^{th} hour post L4). In a separate set of plates, L4 hermaphrodites were allowed to develop another 33 hours before they were isolated into individual plates. These hermaphrodites were allowed to lay embryos for 30 hours (34^{th}-64^{th} hour post L4). Their embryos were allowed to develop into adult worms and then the adult worms were counted.

Sperm localization in outcrossed clec-223 hermaphrodites. Mitotracker CMXRos solid was dissolved in 50µl DMSO and 50µl M9 solution. The stock solution concentration was 1mM, then aliquoted to tubes with 20µl each. *fog-2* (CB4108) adult

worms were washed off the plate and into an Eppendorf tube, then washed with M9 for at least three times to get rid of bacteria. Worms were then resuspended in 800µl M9 solution. 20µl of Mitotracker CMXRos stock solution was added to the worm suspension. The final working concentration was 25 µM. The worms were incubated in the dark for 2 hours, transferred to an NGM plate seeded with *E. coli*, and then ~150 males were transferred again for overnight recovery. For sperm localization assays, 30 adult hermaphrodites (N2 or clec-223) were anesthetized with 0.1% tricaine and 0.01% tetramisole hydrochloride in M9 buffer for a period of 30 minutes. The adults were transferred to an NGM plate containing an ~1 cm circle of bacteria and ~60 Mitotracker-labelled males. Males were allowed to mate for 100 minutes. Outcrossed hermaphrodites were separated from males for a period of 1 hour before mounting on 2% agarose pads for microscopy. Imaging was performed using a Zeiss Axioskop 2 plus using DIC and Rhod channels, a 25X objective, PC computer, and ImageJ (Fiji) software. The reproductive tract from spermatheca to vulva was scanned and images were acquired.

Results

To screen for *clec-223* CRISPR/Cas9 edits, we isolated F1s that showed roller phenotypes. DY129 and DY134 primer pair was used in PCR to screen for an amplicon smaller than 250bp, indicating a deletion event. In one of the single worm PCR reactions, a double band was observed. A 250bp and a 350bp bands indicate there was a potential insertion. The 250bp and 350bp allele segregated through selfing and the homozygous 350bp allele was isolated and sequenced. The insertion was confirmed by sequencing using DY131 or DY134 primer. The insertion was upstream of the *clec-*

223 start codon (near the sgRNA DY120 cutting site). 70bp of the insertion appears to be random sequence that does not match anywhere in the genome, the other 30bp matches exactly to part of the coding sequence of *dpy-10*. This may not create a loss of function in the gene. However, we also found a small deletion and a frame shift mutation in the first exon off *clec-223* (near the sgRNA DY121 PAM sequence), as shown in **Figure 1.** This should create a premature stop codon and a strong loss of function.

Figure 1. Mutation in the *clec-223* gene. CRISPR/Cas9 generated a six nucleotides deletion and one nucleotide insertion in the first exon of the *clec-223* gene. *clec-223* has a predicted signal peptide sequence on the N terminal (shown in blue). The one nucleotide insertion is predicted to create a frame shift mutation in the amino acid, as shown in red. The translated protein sequences were also shown, with the frame shift mutation creating a premature stop codon.

We carried out two assays to phenotype the CP167 *clec-223* mutants. First we asked whether losing a *clec-223* would have any effect on the rate of egg laying. This might be expected if sperm, from which oocytes receive an MSP-based maturation signal (Miller et al., 2001). We measured the number of embryos laid in two periods of

time. We found in the initial 24 hours following L4 stage, *clec-223* laid significantly fewer embryos than N2 wild-type (**Figure 2)**. Then we measure the number of embryos laid in the 34th to 64th hour time period post L4 stage, and during this 30-hour time frame, there was not a significant difference in the number of embryos laid.

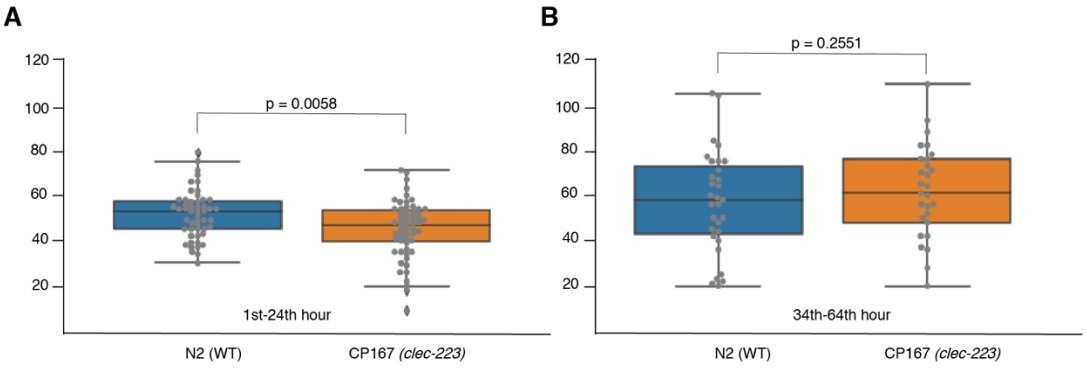

Figure 2. N2 wild-type hermaphrodites laid more embryos in the 24 hours following L4 stage. (A). For the 24 hours period following L4 stage, N2 wild-type hermaphrodites laid significantly more embryos (median = 53, n = 51) than CP167 *clec-223* mutants (median =47, n = 59), P = 0.0058. (B). For the 34th to 64th hour period following L4 stage, the number of embryos N2 laid (median = 58, n = 31) is not significantly different from those laid by CP167 *clec-223* (median = 61, n = 29). P = 0.2551. Statistical test performed by Mann Whiteney U test.

Since *clec-223* is expressed in the spermatheca and could bind the carbohydrates on the surface of sperm cells, we asked if we could detect any abnormal sperm localization in the *clec-223* mutant. We crossed mito-tracker labeled male to hermaphrodites and used flurorescence microscopy to visualize male sperm localization in the reproductive tract. We expected sperm to be localized inside spermatheca. In both N2 and *clec-223,* we saw this was mostly the case (**Figure 3**). The number of hermaphrodites with sperm localized outside of spermatheca was not significantly different between the mutant and the wild-type.

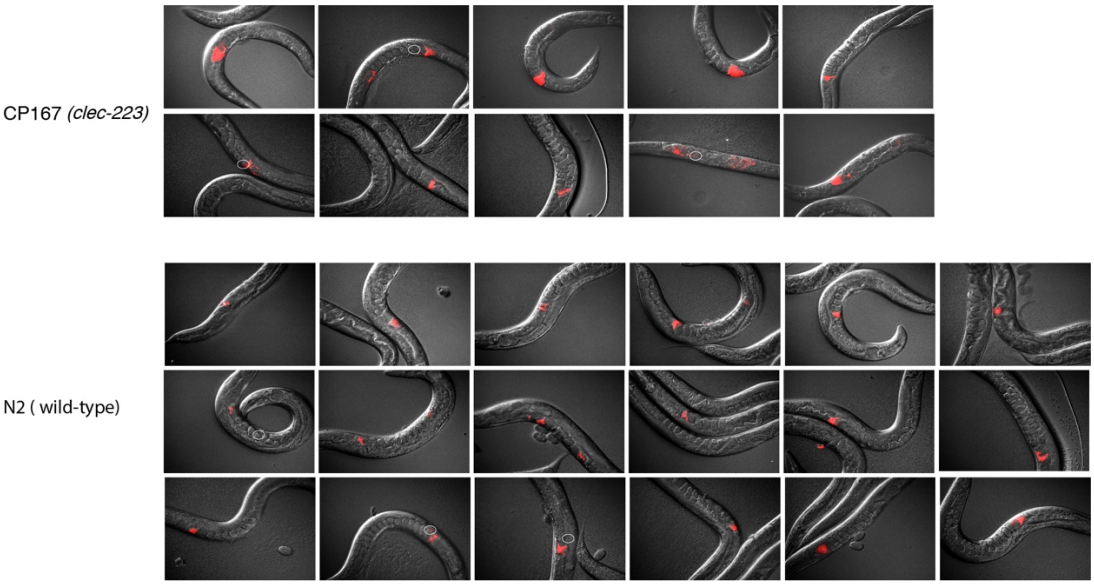

Figure 3. sperm localization in the reproductive tract of hermaphrodites. Upper panel shows the male sperm localization in the *clec-223* hermaphrodites. 7/10 hermaphrodites have male sperm localized entirely inside the spermatheca; 3/10 hermaphrodites have one or a few male sperm localized outside spermatheca (in the uterus). Lower panel shows male sperm localization in the N2 hermaphrodites. 15/18 hermaphrodites have male sperm localized entirely inside the spermatheca; 3/18 hermaphrodites have one or a few male sperm localized outside spermatheca (in the uterus). The occurrences of sperm localization outside of spermatheca is not significantly different between the two genotypes. Chi-square test, P = 0.41.

Conclusions

Our data suggest there was a small reduction in the number of embryos laid in the first 24 hours post L4 stage in the *clec*-223 mutant as compared to WT N2. However, when mitotracker- labeled male sperm was imaged in the reproductive tract of *clec*-223 mutant hermaphrodites, sperm localization was not significantly different from what we observed in a WT. In addition, because *clec-223* has no predicted transmembrane domain, the specific location and orientation of

the CLEC-223 in the hermaphrodite's spermatheca still remain unclear. An alternative approach to study the function of *clec*-223 is to generate a *clec*-223 knock-out (or *clec*-223/*clec*-190 double KO) in *C. briggsae,* and then crossed to males bearing the *mss* transgene. If MSS is indeed interacting with CLEC-223 to confer the sperm competitiveness, then we might see the *mss*+ transgene mutant loses its competitive advantage in *clec*-223 mutants.

References:

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinforma. Oxf. Engl. *22*, 1600–1607.

Anderson, J.L., Morran, L.T., and Phillips, P.C. (2010). Outcrossing and the maintenance of males within *C. elegans* populations. J. Hered. *101*, S62–S74.

Artieri, C.G., Haerty, W., Gupta, B.P., and Singh, R.S. (2008). Sexual selection and maintenance of sex: evidence from comparisons of rates of genomic accumulation of mutations and divergence of sex-related genes in sexual and hermaphroditic species of *Caenorhabditis*. Mol. Biol. Evol. *25*, 972–979.

Avila, F.W., Sirot, L.K., LaFlamme, B.A., Rubinstein, C.D., and Wolfner, M.F. (2010). Insect Seminal Fluid Proteins: Identification and Function. Annu. Rev. Entomol. *56*, 21–40.

Baird, S.E. (1999). Natural and experimental associations of *Caenorhabditis remanei* with Trachelipus rathkii and other terrestrial isopods. Nematology *1*, 471–475.

Baird, S.E., and Stonesifer, R. (2012). Reproductive isolation in *Caenorhabditis briggsae*. Worm *1*, 189–195.

Baird, S., Fitch, D., and Emmons, S. (1994). *Caenorhabditis Vulgaris Sp.N.* (Nematoda: Rhabditidae): a Necromenic Associate of Pill Bugs and Snails. Biol. Sci. Fac. Publ. 1–11.

Baker, H.G. (1955). Self-Compatibility and Establishment After "Long-Distance" Dispersal. Evolution *9*, 347–349.

Baldi, C., Viviano, J., and Ellis, R.E. (2011). A bias caused by ectopic development produces sexually dimorphic sperm in nematodes. Curr. Biol. CB *21*, 1416–1420.

Baldwin-Brown, J.G., Weeks, S.C., and Long, A.D. (2017). A New Standard for Crustacean Genomes: The Highly Contiguous, Annotated Genome Assembly of the Clam Shrimp Eulimnadia texana Reveals HOX Gene Order and Identifies the Sex Chromosome. Genome Biol. Evol. *10*, 143–156.

Barker, D.M. (1994). Copulatory plugs and paternity assurance in the nematode *Caenorhabditis elegans*. Anim. Behav. *48*, 147–156.

Barrière, A., and Félix, M.-A. (2005). High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr. Biol. CB *15*, 1176–1184.

Barrière, A., and Félix, M.-A. (2007). Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. Genetics *176*, 999–1011.

Barrière, A., Yang, S.-P., Pekarek, E., Thomas, C.G., Haag, E.S., and Ruvinsky, I. (2009). Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. Genome Res. *19*, 470–480.

Bell, G. (1982). The Masterpiece of Nature :: The Evolution and Genetics of Sexuality (CUP Archive).

Berlin, K., Koren, S., Chin, C.-S., Drake, J.P., Landolin, J.M., and Phillippy, A.M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat. Biotechnol. *33*, 623–630.

Bi, Y., Ren, X., Yan, C., Shao, J., Xie, D., and Zhao, Z. (2015). A Genome-wide hybrid incompatibility landscape between *Caenorhabditis briggsae* and *C. nigoni*. PLoS Genet. *11*, e1004993.

Bundus, J.D., Alaei, R., and Cutter, A.D. (2015). Gametic selection, developmental trajectories, and extrinsic heterogeneity in Haldane's rule. Evol. Int. J. Org. Evol. *69*, 2005–2017.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinforma. Oxf. Engl. *25*, 1972–1973.

Cassone, B.J., Kay, R.G.G., Daugherty, M.P., and White, B.J. (2017). Comparative Transcriptomics of Malaria Mosquito Testes: Function, Evolution, and Linkage. G3 Genes Genomes Genet. *7*, 1127–1136.

Charlesworth, D., and Wright, S.I. (2001). Breeding systems and genome evolution. Curr. Opin. Genet. Dev. *11*, 685–690.

Chasnov, J.R. (2010). The evolution from females to hermaphrodites results in a sexual conflict over mating in androdioecious nematode worms and clam shrimp. J. Evol. Biol. *23*, 539–556.

Chasnov, J.R., and Chow, K.L. (2002). Why are there males in the hermaphroditic species *Caenorhabditis elegans*? Genetics *160*, 983–994.

Chasnov, J.R., So, W.K., Chan, C.M., and Chow, K.L. (2007a). The species, sex, and stage specificity of a *Caenorhabditis* sex pheromone. Proc. Natl. Acad. Sci. U. S. A. *104*, 6730–6735.

Chasnov, J.R., So, W.K., Chan, C.M., and Chow, K.L. (2007b). The species, sex, and stage specificity of a *Caenorhabditis* sex pheromone. Proc. Natl. Acad. Sci. U. S. A. *104*, 6730–6735.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.

Civetta, A., and Finn, S. (2014). Do Candidate Genes Mediating Conspecific Sperm Precedence Affect Sperm Competitive Ability Within Species? A Test Case in *Drosophila*. G3 Genes Genomes Genet. *4*, 1701–1707.

Coghlan, A., Fiedler, T.J., McKay, S.J., Flicek, P., Harris, T.W., Blasiar, D., nGASP Consortium, and Stein, L.D. (2008). nGASP--the nematode genome annotation assessment project. BMC Bioinformatics *9*, 549.

Crnokrak, P., and Barrett, S.C.H. (2002). Perspective: purging the genetic load: a review of the experimental evidence. Evol. Int. J. Org. Evol. *56*, 2347–2358.

Cutter, A.D. (2008). Reproductive Evolution: Symptom of a Selfing Syndrome. Curr. Biol. *18*, R1056–R1058.

Cutter, A.D. *Caenorhabditis* evolution in the wild. BioEssays *37*, 983–995.

Cutter, A.D., and Jovelin, R. (2015). When natural selection gives gene function the cold shoulder. BioEssays News Rev. Mol. Cell. Dev. Biol. *37*, 1169–1173.

Cutter, A.D., and Payseur, B.A. (2003). Rates of deleterious mutation and the evolution of sex in *Caenorhabditis*. J. Evol. Biol. *16*, 812–822.

Cutter, A.D., Wasmuth, J.D., and Washington, N.L. (2008). Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. Genetics *178*, 2093–2104.

Denver, D.R., Morris, K., and Thomas, W.K. (2003). Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. Mol. Biol. Evol. *20*, 393–400.

Denver, D.R., Morris, K., Lynch, M., and Thomas, W.K. (2004). High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature *430*, 679–682.

Denver, D.R., Clark, K.A., and Raboin, M.J. (2011). Reproductive mode evolution in nematodes: Insights from molecular phylogenies and recently discovered species. Mol. Phylogenet. Evol. *61*, 584–592.

Dolgin, E.S., Charlesworth, B., Baird, S.E., and Cutter, A.D. (2007a). Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. Evol. Int. J. Org. Evol. *61*, 1339–1352.

Dolgin, E.S., Charlesworth, B., Baird, S.E., and Cutter, A.D. (2007b). Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. Evol. Int. J. Org. Evol. *61*, 1339–1352.

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform. Int. Conf. Genome Inform. *23*, 205–211.

Ellis, R.E., and Lin, S.-Y. (2014). The evolutionary origins and consequences of self-fertility in nematodes. F1000Prime Rep. *6*.

Emmons, S.W. (2006). Sexual behavior of the *Caenorhabditis elegans* male. Int. Rev. Neurobiol. *69*, 99–123.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., et al. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PloS One *7*, e47768.

Félix, M.-A., Braendle, C., and 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.

Fierst, J.L., Willis, J.H., Thomas, C.G., Wang, W., Reynolds, R.M., Ahearne, T.E., Cutter, A.D., and Phillips, P.C. (2015). Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. PLoS Genet. *11*, e1005323.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279-285.

Fiumera, A.C., Dumont, B.L., and Clark, A.G. (2005). Sperm Competitive Ability in *Drosophila melanogaster* Associated With Variation in Male Reproductive Proteins. Genetics *169*, 243–257.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., et al. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PloS One *9*, e98186.

Garcia, L.R., LeBoeuf, B., and Koo, P. (2007). Diversity in mating behavior of hermaphroditic and male-female *Caenorhabditis* nematodes. Genetics *175*, 1761–1771.

Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049-1056.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al. (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science *330*, 1775–1787.

Gibson, A.K., Delph, L.F., and Lively, C.M. (2017). The two-fold cost of sex: Experimental evidence from a natural system. Evol. Lett. *1*, 6–15.

Gimond, C., Jovelin, R., Han, S., Ferrari, C., Cutter, A.D., and Braendle, C. (2013). Outbreeding depression with low genetic variation in selfing *Caenorhabditis* nematodes. Evol. Int. J. Org. Evol. *67*, 3087–3101.

Gimond, C., Vielle, A., Soares, N.S., Zdraljevic, S., McGrath, P., Andersen, E., and Braendle, C. (2018). Evolution of sperm competition: Natural variation and genetic determinants of *Caenorhabditis elegans* sperm size. BioRxiv 501486.

Glémin, S., and Ronfort, J. (2013). Adaptation and maladaptation in selfing and outcrossing species: new mutations versus standing variation. Evol. Int. J. Org. Evol. *67*, 225–240.

Golden, J.W., Nelson, F.K., Riddle, D.L., and Fodor, A. (1983). Comparison of a New Wild-Type *Caenorhabditis briggsae* With Laboratory Strains of *C. briggsae* and *C. elegans*. Nematologica *29*, 203–216.

Gomendio, M., and Roldan, E.R.S. (2004). Implications of diversity in sperm size and function for sperm competition and fertility. Int. J. Dev. Biol. *52*, 439–447.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. *36*, 3420–3435.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Graustein, A., Gaspar, J.M., Walters, J.R., and Palopoli, M.F. (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *caenorhabditis*. Genetics *161*, 99–107.

Guo, Y., Lang, S., and Ellis, R.E. (2009). Independent Recruitment of F Box Genes to Regulate Hermaphrodite Development during Nematode Evolution. Curr. Biol. *19*, 1853–1860.

Haag, E.S. (2005). The evolution of nematode sex determination: *C. elegans* as a reference point for comparative biology. WormBook Online Rev. C Elegans Biol. 1–14.

Hamilton, W.D. (1967). Extraordinary Sex Ratios. Science *156*, 477–488.

Hansen, J.M., Chavez, D.R., and Stanfield, G.M. (2015). COMP-1 promotes competitive advantage of nematode sperm. ELife *4*.

Hedgecock, E.M. (1976). The Mating System of *Caenorhabditis elegans*: Evolutionary Equilibrium between Self- and Cross-Fertilization in a Facultative Hermaphrodite. Am. Nat. *110*, 1007–1012.

Herre, E.A. (1985). Sex ratio adjustment in fig wasps. Science *228*, 896–898.

Hill, R.C., de Carvalho, C.E., Salogiannis, J., Schlager, B., Pilgrim, D., and Haag, E.S. (2006). Genetic flexibility in the convergent evolution of hermaphroditism in *Caenorhabditis* nematodes. Dev. Cell *10*, 531–538.

Hodgkin, J. (1983). Male Phenotypes and Mating Efficiency in *CAENORHABDITIS ELEGANS*. Genetics *103*, 43–64.

Hodgkin, J., and Doniach, T. (1997). Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics *146*, 149–164.

Hodgkin, J., Horvitz, H.R., and Brenner, S. (1979). Nondisjunction Mutants of the Nematode *CAENORHABDITIS elegans*. Genetics *91*, 67–94.

Hoogendyk, C.G., and Estabrook, G.F. (1984). The consequences of earlier reproduction in declining populations. Math. Biosci. *71*, 217–235.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., et al. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. *43*, 476–481.

Huang, S., Kang, M., and Xu, A. (2017). HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinforma. Oxf. Engl. *33*, 2577–2579.

Johnson, T.E., and Hutchinson, E.W. (1993). Absence of strong heterosis for life span and other life history traits in *Caenorhabditis elegans*. Genetics *134*, 465–474.

Jones, A.R., Francis, R., and Schedl, T. (1996). GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. Dev. Biol. *180*, 165–183.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036.

Kanzaki, N., Tsai, I.J., Tanaka, R., Hunt, V.L., Liu, D., Tsuyama, K., Maeda, Y., Namai, S., Kumagai, R., Tracey, A., et al. (2018). Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. Nat. Commun. *9*, 3216.

Katju, V., LaBeau, E.M., Lipinski, K.J., and Bergthorsson, U. (2008). Sex Change by Gene Conversion in a *Caenorhabditis elegans* fog-2 Mutant. Genetics *180*, 669–672.

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

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

Kleemann, G.A., and Basolo, A.L. (2007). Facultative decrease in mating resistance in hermaphroditic *Caenorhabditis elegans* with self-sperm depletion. Anim. Behav. *74*, 1339–1347.

Kozlowska, J.L., Ahmad, A.R., Jahesh, E., and Cutter, A.D. (2012). Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis sp. 9*. Evol. Int. J. Org. Evol. *66*, 1180–1195.

LaMunyon, C.W., and Ward, S. (1998). Larger sperm outcompete smaller sperm in the nematode *Caenorhabditis elegans*. Proc. Biol. Sci. *265*, 1997–2002.

LaMunyon, C.W., and Ward, S. (1999). Evolution of sperm size in nematodes: sperm competition favours larger sperm. Proc. Biol. Sci. *266*, 263–267.

LaMunyon, C.W., and Ward, S. (2002). Evolution of larger sperm in response to experimentally increased sperm competition in *Caenorhabditis elegans*. Proc. Biol. Sci. *269*, 1125–1128.

Lande, R., Schemske, D.W., and Schultz, S.T. (1994). High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evol. Int. J. Org. Evol. *48*, 965–978.

Le, T.S., Yang, F.-J., Lo, Y.-H., Chang, T.C., Hsu, J.-C., Kao, C.-Y., and Wang, J. (2017). Non-Mendelian assortment of homologous autosomes of different sizes in males is the ancestral state in the *Caenorhabditis* lineage. Sci. Rep. *7*, 12819.

Leonard (2019). Transitions Between Sexual Systems: Understanding the Mechanisms of, and Pathways Between, Dioecy, Hermaphroditism and Other Sexual Systems (New York, NY: Springer).

Leighton, D.H.W., Choe, A., Wu, S.Y., and Sternberg, P.W. (2014). Communication between oocytes and somatic cells regulates volatile pheromone production in *Caenorhabditis* elegans. Proc. Natl. Acad. Sci. *111*, 17905–17910.

Li, R., Ren, X., Bi, Y., Ho, V.W.S., Hsieh, C.-L., Young, A., Zhang, Z., Lin, T., Zhao, Y., Miao, L., et al. (2016). Specific down-regulation of spermatogenesis genes targeted by 22G RNAs in hybrid sterile males associated with an X-Chromosome introgression. Genome Res.

Lipton, J., Kleemann, G., Ghosh, R., Lints, R., and Emmons, S.W. (2004). Mate searching in *Caenorhabditis elegans*: a genetic model for sex drive in a simple invertebrate. J. Neurosci. Off. J. Soc. Neurosci. *24*, 7427–7434.

Liu, Q., Stumpf, C., Thomas, C., Wickens, M., and Haag, E.S. (2012). Context-dependent function of a conserved translational regulatory module. Dev. Camb. Engl. *139*, 1509–1521.

Lively, C.M., and Lloyd, D.G. (1990). The Cost of Biparental Sex Under Individual Selection. Am. Nat. *135*, 489–500.

Lupas, A. (1996). Prediction and analysis of coiled-coil structures. Methods Enzymol. *266*, 513–525.

Manoel, D., Carvalho, S., Phillips, P.C., and Teotonio, H. (2007). Selection against males in *Caenorhabditis elegans* under two mutational treatments. Proc. R. Soc. Lond. B Biol. Sci. *274*, 417–424.

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

Morran, L.T., Parmenter, M.D., and Phillips, P.C. (2009). Mutation load and rapid adaptation favour outcrossing over self-fertilization. Nature *462*, 350.

Mueller, J.L., Linklater, J.R., Ram, K.R., Chapman, T., and Wolfner, M.F. (2008). Targeted Gene Deletion and Phenotypic Analysis of the *Drosophila melanogaster* Seminal Fluid Protease Inhibitor Acp62F. Genetics *178*, 1605–1614.

Murray, R.L., Kozlowska, J.L., and Cutter, A.D. (2011). Heritable determinants of male fertilization success in the nematode *Caenorhabditis elegans*. BMC Evol. Biol. *11*, 99.

Noble, L.M., Chang, A.S., McNelis, D., Kramer, M., Yen, M., Nicodemus, J.P., Riccardi, D.D., Ammerman, P., Phillips, M., Islam, T., et al. (2015). Natural Variation in plep-1 Causes Male-Male Copulatory Behavior in *C. elegans*. Curr. Biol. CB *25*, 2730–2737.

Noël, E., Jarne, P., Glémin, S., MacKenzie, A., Segard, A., Sarda, V., and David, P. (2017). Experimental Evidence for the Negative Effects of Self-Fertilization on the Adaptive Potential of Populations. Curr. Biol. CB *27*, 237–242.

Olson, M.V. (1999). When less is more: gene loss as an engine of evolutionary change. Am. J. Hum. Genet. *64*, 18–23.

Otto, S.P., Sassaman, C., and Feldman, M.W. (1993). Evolution of Sex Determination in the Conchostracan Shrimp Eulimnadia texana. Am. Nat. *141*, 329–337.

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

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. Nucleic Acids Res. *37*, 289–297.

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

Pizzari, T., and Parker, G.A. (2009). 6 - Sperm competition and sperm phenotype. In Sperm Biology, T.R. Birkhead, D.J. Hosken, and S. Pitnick, eds. (London: Academic Press), pp. 207–245.

Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics *117*, 353–360.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE *5*, e9490.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome center's improvements to the Illumina sequencing system. Nat. Methods *5*, 1005–1010.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. *26*, 841–842.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., and Moorman, A.F.M. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci. Lett. *339*, 62–66.

Rödelsperger, C., Röseler, W., Prabh, N., Yoshida, K., Weiler, C., Herrmann, M., and Sommer, R.J. (2018). Phylotranscriptomics of *Pristionchus* Nematodes Reveals Parallel Gene Loss in Six Hermaphroditic Lineages. Curr. Biol. *28*, 3123-3127.e5.

Ross, J.A., Koboldt, D.C., Staisch, J.E., Chamberlin, H.M., Gupta, B.P., Miller, R.D., Baird, S.E., and Haag, E.S. (2011). *Caenorhabditis briggsae* recombinant inbred line

genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. *7*, e1002174.

Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., van den Hoff, M.J.B., and Moorman, A.F.M. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. *37*, e45.

Schedl, T., and Kimble, J. (1988). fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. Genetics *119*, 43–61.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA N. Y. N *22*, 839–851.

Schurko, A.M., Neiman, M., and Logsdon, J.M. (2009). Signs of sex: what we know and how we know it. Trends Ecol. Evol. *24*, 208–217.

Seidel, H.S., Rockman, M.V., and Kruglyak, L. (2008). Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science *319*, 589–594.

Shakes, D.C., Wu, J.-C., Sadler, P.L., Laprade, K., Moore, L.L., Noritake, A., and Chu, D.S. (2009). Spermatogenesis-specific features of the meiotic program in *Caenorhabditis elegans*. PLoS Genet. *5*, e1000611.

Shen, Y., Wollam, J., Magner, D., Karalay, O., and Antebi, A. (2012). A steroid receptor-microRNA switch regulates life span in response to signals from the gonad. Science *338*, 1472–1476.

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

Sicard, A., and Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Ann. Bot. *107*, 1433–1443.

Simmons, L.W., and Fitzpatrick, J.L. (2012). Sperm wars and the evolution of male fertility. Reproduction *144*, 519–534.

Simon, J.M., and Sternberg, P.W. (2002). Evidence of a mate-finding cue in the hermaphrodite nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. *99*, 1598–1603.

Sivasundar, A., and Hey, J. (2005). Sampling from natural populations with RNAI reveals high outcrossing and population structure in *Caenorhabditis elegans*. Curr. Biol. CB *15*, 1598–1602.

Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

Slowinski, S.P., Morran, L.T., Parrish, R.C., Cui, E.R., Bhattacharya, A., Lively, C.M., and Phillips, P.C. (2016). Coevolutionary interactions with parasites constrain the spread of self-fertilization into outcrossing host populations. Evolution *70*, 2632–2639.

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

Smith, M.V., Boyd, W.A., Kissling, G.E., Rice, J.R., Snyder, D.W., Portier, C.J., and Freedman, J.H. (2009). A Discrete Time Model for the Analysis of Medium-Throughput *C. elegans* Growth Data. PLOS ONE *4*, e7018.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinforma. Oxf. Engl. *24*, 637–644.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. *1*, E45.

Stewart, A.D., and Phillips, P.C. (2002). Selection and maintenance of androdioecy in *Caenorhabditis elegans*. Genetics *160*, 975–982.

Sudhaus, W. (1974). Zur Systematik, Verbreitung, Okologie und Biologie neuer und wenig bekannter Rhabditiden (Nematoda) 2. Teil. Zool Jb Syst Bd *101*, 417–465.

Sutton, K.A., Jungnickel, M.K., and Florman, H.M. (2008). A polycystin-1 controls postcopulatory reproductive selection in mice. Proc. Natl. Acad. Sci. *105*, 8661–8666.

Tecle, E., and Gagneux, P. (2015). Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycocalyx. Mol. Reprod. Dev. *82*, 635–650.

Teotónio, H., Manoel, D., and Phillips, P.C. (2006). Genetic variation for outcrossing among *Caenorhabditis elegans* isolates. Evolution *60*, 1300–1305.

Theologidis, I., Chelo, I.M., Goy, C., and Teotónio, H. (2014). Reproductive assurance drives transitions to self-fertilization in experimental *Caenorhabditis* elegans. BMC Biol. *12*, 93.

Thomas, C.G., Woodruff, G.C., and Haag, E.S. (2012a). Causes and consequences of the evolution of reproductive mode in *Caenorhabditis* nematodes. Trends Genet. *28*, 213–220.

Thomas, C.G., Li, R., Smith, H.E., Woodruff, G.C., Oliver, B., and Haag, E.S. (2012b). Simplification and desexualization of gene expression in self-fertile nematodes. Curr. Biol. *22*, 2167–2172.

Thomas, C.G., Wang, W., Jovelin, R., Ghosh, R., Lomasko, T., Trinh, Q., Kruglyak, L., Stein, L.D., and Cutter, A.D. (2015). Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. Genome Res. *25*, 667–678.

Ting, J.J., Woodruff, G.C., Leung, G., Shin, N.-R., Cutter, A.D., and Haag, E.S. (2014). Intense sperm-mediated sexual conflict promotes reproductive isolation in *Caenorhabditis* nematodes. PLoS Biol. *12*, e1001915.

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

Wang, J., Chen, P.-J., Wang, G.J., and Keller, L. (2010). Chromosome Size Differences May Affect Meiosis and Genome Size. Science *329*, 293–293.

Ward, S., Argon, Y., and Nelson, G.A. (1981). Sperm morphogenesis in wild-type and fertilization-defective mutants of *Caenorhabditis elegans*. J. Cell Biol. *91*, 26–44.

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

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinforma. Oxf. Engl. *25*, 1189–1191.

Weeks, S.C. (2012). The Role of Androdioecy and Gynodioecy in Mediating Evolutionary Transitions Between Dioecy and Hermaphroditism in the Animalia. Evolution *66*, 3670–3686.

Wegewitz, V., Schulenburg, H., and Streit, A. (2008). Experimental insight into the proximate causes of male persistence variation among two strains of the androdioecious *Caenorhabditis elegans* (Nematoda). BMC Ecol. *8*, 12.

Wigby, S., and Chapman, T. (2004). Sperm competition. Curr. Biol. *14*, R100–R103.

Wilson, L.D., Sackett, J.M., Mieczkowski, B.D., Richie, A.L., Thoemke, K., Rumbley, J.N., and Kroft, T.L. (2011). Fertilization in *C. elegans* requires an intact C-terminal RING finger in sperm protein SPE-42. BMC Dev. Biol. *11*, 10.

Woodruff, G.C., Eke, O., Baird, S.E., Félix, M.-A., and Haag, E.S. (2010). Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of *Caenorhabditis* nematodes. Genetics *186*, 997–1012.

Woodruff, G.C., Knauss, C.M., Maugel, T.K., and Haag, E.S. (2014). Mating damages the cuticle of *C. elegans* hermaphrodites. PLoS One *9*, e104456.

Wootton, J.C. (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput. Chem. *18*, 269–285.

Wright, K. (1989). Antibodies a laboratory manual: By E Harlow and D Lane. pp 726. Cold Spring Harbor Laboratory. 1988. $50 ISBN 0-87969-314-2. Biochem. Educ. *17*, 220–220.

Wright, S.I., Nano, N., Foxe, J.P., and Dar, V.-U.N. (2008). Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. Genet. Res. *90*, 119–128.

Yan, C., Bi, Y., Yin, D., and Zhao, Z. (2012). A method for rapid and simultaneous mapping of genetic loci and introgression sizes in nematode species. PloS One *7*, e43770.

Yeh, S.-D., Do, T., Chan, C., Cordova, A., Carranza, F., Yamamoto, E.A., Abbassi, M., Gandasetiawan, K.A., Librado, P., Damia, E., et al. (2012). Functional evidence that a recently evolved *Drosophila* sperm-specific gene boosts sperm competition. Proc. Natl. Acad. Sci. *109*, 2043–2048.

Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., and Haag, E.S. (2018). Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science *359*, 55–61.

Zierold, T., Hanfling, B., and Gómez, A. (2007). Recent evolution of alternative reproductive modes in the “living fossil” *Triops cancriformis*. BMC Evol. Biol. *7*, 161.