Differential selection between the sexes and the evolution of recombination in diploid and haplodiploid species

A thesis
Submitted towards the partial fullfilment of
BS-MS dual degree programme
by

VINAYAK PATEL

MAY 2025

under the guidance of

DR. DENIS ROZE

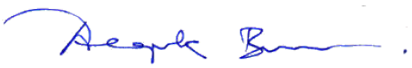
STATION BIOLOGIQUE DE ROSCOFF

from May 2024 to March 2025

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
PUNE

Certificate

This is to certify that this dissertation entitled "Differential selection between the sexes and the evolution of recombination in diploid and haplodiploid species " submitted towards the partial fulfillment of the BS-MS degree at the Indian Institute of Science Education and Research, Pune represents original research carried out by "Vinayak Patel" at "Station Biologique de Roscoff" under the supervision of "Denis Roze" from June 2024 to December 2024 and at "Indian Institute of Science Education and Research, Pune" under the co-supervision of "Deepak Barua" from January 2025 to March 2025.



Supervisor:
DENIS ROZE
DIRECTEUR DE RECHERCHE AU CNRS
STATION BIOLOGIQUE DE ROSCOFF

Co-Supervisor:
DEEPAK BARUA
ASSOCIATE PROFESSOR
IISER PUNE

Expert:
SUTIRTH DEY
PROFESSOR
IISER PUNE

DATE: 09/05/2025

Declaration

I, hereby declare that the matter embodied in the report titled "Differential selection between the sexes and the evolution of recombination in diploid and haplodiploid species" is the results of the investigations carried out by me at the "Station Biologique de Roscoff" from the period 21-06-2024 to 15-12-2024 under the supervision of Professor Denis Roze and at the "Indian Institute of Science Education and Research, Pune" from the period 7-01-2025 to 15-03-2025 under the co-supervision of Professor Deepak Barua and the same has not been submitted elsewhere for any other degree.

Student:
VINAYAK PATEL
20201178
BS-MS
IISER PUNE

DATE: 09/05/2025

Acknowledgements

I would especially like to thank my supervisor Professor Denis Roze for guidance and support throughout the duration of my thesis, which very often exceeded whatever expectations one might have from their supervisor. The amount of time he dedicated to guiding and teaching me has been invaluable. I would also like to thank my co-supervisor Professor Deepak Barua for his support and guidance, not only during this thesis but also earlier in my career. I would like to thank Professor Sutirth Dey for his helpful feedback during the mid-year evaluation. I would like to thank other members of the scientific unit UMR7144 at Roscoff, and the computational resources provided by ABIMS. I would further acknowledge the financial suppport provided by CNRS/Sorbonne University and KVPY.

I would like to acknowledge the support of my friends and family, and thank them for their consistent presence for me in the face of any difficulties.

Contributions

Contributor name	Contributor role
Denis Roze	Conceptualization Ideas
Denis Roze	Methodology
Denis Roze, Vinayak Patel	Software
Denis Roze, Vinayak Patel	Validation
Denis Roze, Vinayak Patel	Formal analysis
Denis Roze, Vinayak Patel	Investigation
Denis Roze	Resources
Denis Roze, Vinayak Patel	Data Curation
Vinayak Patel	Writing - original draft preparation
Denis Roze, Vinayak Patel	Writing - review and editing
Vinayak Patel	Visualization
Denis Roze, Deepak Barua	Supervision
Denis Roze	Project administration
Denis Roze	Funding acquisition

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy[1]

[1] https://journals.biologists.com/jcs/pages/author-contributions

Abstract

The exchange of genetic material between homologous chromosomes during meiosis (i.e., genetic recombination) is a fundamental process that generates genetic diversity and influences evolutionary dynamics. Important variation in recombination rates can be observed at different phylogenetic scales with several broad patterns. In particular, eusocial hymenopterans (which are almost invariably haplodiploid) display some of the highest recombination rates among all eukaryotes. Among diploid species, females often have a higher recombination rate than males. The explanations for these phenomena are still unclear. We explore to what extent differences in selection between males and females may explain these patterns. To test these hypotheses, we use analytical modelling and individual-based, multilocus simulations in which fitness variation is caused by recurrent deleterious mutations occurring along a linear chromosome and in which the genetic map length of the chromosome (number of crossovers at meiosis) can evolve. We examined the effects of stronger selection against and stronger negative epistasis between deleterious mutations in males, which could possibly be caused by sexual selection among males, on the evolution of recombination rates. Our results show that negative epistasis among deleterious mutations does result in higher recombination rates in haplodiploids than in diploids. Negative epistasis restricted to males also generates a sexual dimorphism in recombination rates (aka heterochiasmy) in diploids consistent with the empirical pattern of higher recombination in females than males. In haplodiploids, higher recombination rates are favoured in meioses that produce parthenogenetic eggs (that will develop into males) than in meioses that produce fertilised eggs (that will develop into females). These results show that negative epistasis and sex-based differences in selection may explain the elevated recombination rates of eusocial hymenopterans and the patterns of heterochiasmy in diploid species.

Contents

Chapter 1

Introduction

Genetic recombination, the process by which genetic material is exchanged between homologous chromosomes during meiosis, is a fundamental process which plays a key role in evolution by generating genetic diversity. Sexually reproducing eukaryotes almost universally utilise meiotic recombination to produce gametes with reduced chromosomal complement (Cavalier-Smith (2002)). The meiotic division consists of one round of chromosomal duplication followed by two rounds of segregation. Before the first round of segregation, the two homologous chromosomes from each parent come together to form synapses. They then exchange genetic material between themselves through a process known as crossing over. While the formation of at least one crossover, referred to as obligate crossover, is necessary for proper segregation of chromosomes during meiosis in most species (Bishop and Zickler (2004); Carlton *et al.* (2005); Deshong *et al.* (2014)), the empirically observed rates of crossover often exceed this. Further, meiosis in certain organisms is achiasmatic in one of the sexes Scott Hawley and Theurkauf (1993); Newnham *et al.* (2010); Papaioannou *et al.* (2021), suggesting that recombination is not strictly necessary for chromosome segregation during the first phase of meiosis. Additionally, the rates of genetic recombination vary substantially across different taxa, species, populations and even among different individuals of the same species (Coop and Przeworski (2007); Lenormand and Dutheil (2005); Wilfert *et al.* (2007); Stapley *et al.* (2017); Ritz *et al.* (2017); Samuk *et al.* (2020); Johnston (2024)). These variations indicate that recombination may offer adaptive benefits that extend beyond ensuring the proper segregation of chromosomes.

Meiotic recombination is a critical source of genetic variation in populations. During recombination, the homologous chromosomes exchange genetic information, breaking the association between the linked genes and generating new genetic associations in the gametes. It allows sexually reproducing organisms to produce an incredibly large variety of gametes. While recombination leads to the creation of novel genetic variants, it may also lead to the breakdown of existing favourable genetic variants. Therefore, recombination may have contrasting effects on fitness, and its overall effect depends on factors such as linkage disequilibrium, selection regime, epistasis, migration and mating system (assortative or random) (Agrawal (2006); Otto and Lenormand (2002)).

Recombination can either hinder or increase the efficiency of natural selection depending on the type of linkage disequilibrium present in the population. Linkage disequilibrium (LD) describes the non-random association of alleles at different loci within a population. For a pair of deleterious alleles, linkage disequilibrium is positive if the

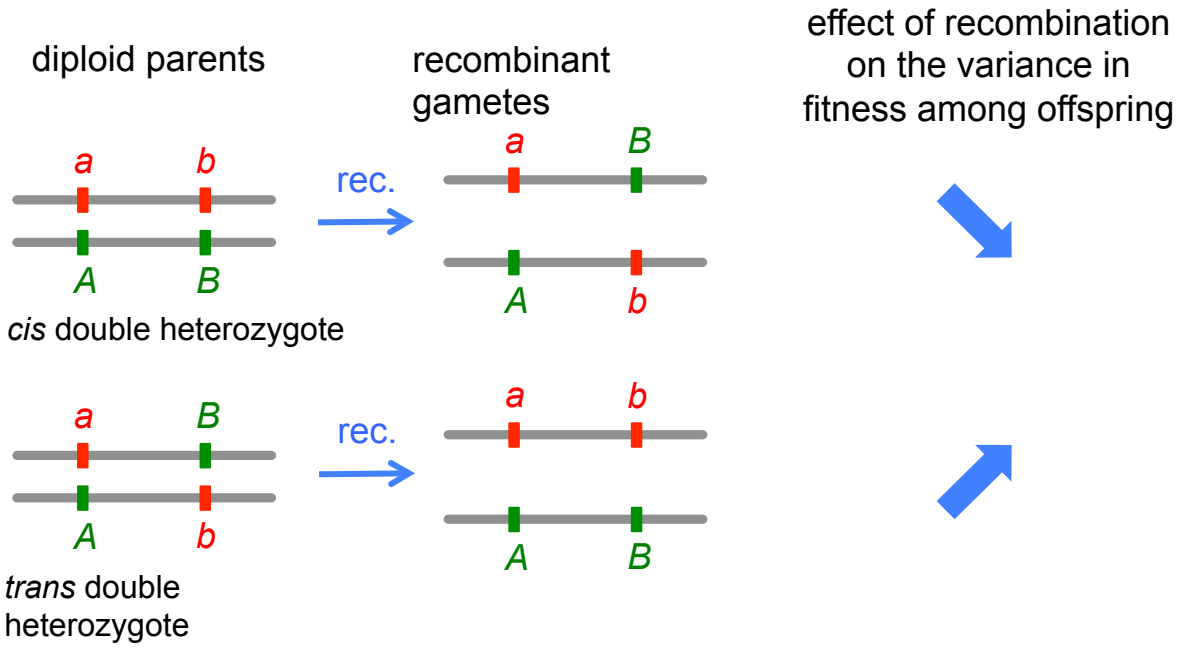


Figure 1.1: **Effects of recombination on the variation of fitness in offspring.**The schematic shows, using a simple two-locus case as an example, how recombination can either decrease or increase the variance in fitness in offspring depending on the kind of linkage disequilibrium in the population. Here a and b are the deleterious alleles.

deleterious alleles tend to be more frequently associated with each other than what would be expected if they were distributed randomly among individuals in the population. Similarly, linkage disequilibrium is negative if the deleterious alleles tend to be associated with beneficial alleles at the other locus (or loci).

In a simple two-locus case, recombination only has an effect in double heterozygotes. In trans double heterozygotes, it increases the variance in fitness among offspring. This occurs because recombination places the deleterious alleles and the beneficial alleles on two different chromosomes during meiosis, resulting in them being carried on different gametes (Figure 1.1). Similarly, it reduces the variance in fitness among offspring in cis double heterozygotes by putting the beneficial alleles with deleterious alleles together on the same chromosomes (and, therefore, on the same gametes). Thus, the net effect of recombination is to increase the variance in fitness if there is an excess of trans relative to cis double heterozygotes (that is, if there is negative linkage disequilibrium), and vice versa. Therefore, when there is negative disequilibrium in the population, recombination increases the genetic variance in fitness among the offspring (by producing extreme genotypes from intermediate ones) and, therefore, increases the efficiency of natural selection, as the efficiency of natural selection is directly proportional to the genetic variance in fitness. Similarly, when the population is in positive linkage disequilibrium, recombination reduces the variability in fitness in the offspring by shuffling away the linked beneficial (or deleterious alleles) in the parents and thus producing intermediate genotypes, contributing to a reduced efficiency of selection.

The different theories on the evolutionary advantage of recombination thus correspond to different possible causes of negative linkage disequilibrium within populations. The two main mechanisms that have been identified are 1) negative epistasis and 2) selec-

tion and genetic drift in finite populations. First, when epistasis is defined as a deviation from multiplicative fitness effects between loci, epistasis between selected loci tends to generate linkage disequilibrium (while multiplicative selection does not generate linkage disequilibrium). If selection is the only evolutionary force affecting genetic associations, the linkage disequilibrium generated by selection will be of the same sign as epistasis (see Agrawal (2006) for a comprehensive review). In particular, negative epistasis (meaning that deleterious alleles have stronger negative effects when combined, or beneficial alleles have weaker effects when combined, as compared to multiplicative fitness effects) generates negative linkage disequilibrium. Therefore, under negative epistasis, recombination increases the variance in fitness among offspring, which has a beneficial effect as it increases the efficiency of selection. However, one can show that in the presence of epistasis, recombination tends to decrease the mean fitness of offspring as it breaks down combinations of alleles that have been favoured by epistatic selection. Therefore, high recombination rates can be favoured only when epistasis is negative and sufficiently weak, as when epistasis is strong and negative, the costs of reduced mean fitness of offspring outweigh the benefits of increased genetic variance among them (Barton (1995)).

Another factor creating linkage disequilibrium in finite populations is genetic drift. Drift can create both negative and positive linkage disequilibrium. When it creates positive linkage disequilibrium, the improved selection (due to increased genetic variance) reduces the frequency of deleterious alleles (or increases the frequency of beneficial alleles), and therefore the positive linkage disequilibrium gets eliminated. When drift generates negative linkage disequilibrium, the response to selection is hindered due to reduced genetic variance, so the negative linkage disequilibrium tends to persist longer. Therefore, in finite populations experiencing genetic drift, disequilibrium is negative more often than positive, generating a selection pressure for increased recombination rates (Hill and Robertson (1966); Felsenstein (1974); Barton and Otto (2005); Keightley and Otto (2006); Roze (2021)).

1.1 Variation in Reccombination Rates and the Special Case of Eusocial Hymenopterans

Recombination rates show a huge amount of variation between organisms, populations and individuals, and they also vary greatly depending on the chromosomal region within the same individual. Genome-wide recombination rates are exceptionally high in social insects of the order Hymenoptera (Wilfert *et al.* (2007)). Hymenopterans are an order of insects comprising sawflies, ants, bees, and wasps. The vast majority of hymenopterans have a haplodiploid sex-determination system, where the males are haploid and produced from unfertilised ovules, while the females are diploid and the product of fertilisation. Therefore, meiosis and recombination happen only in females, as only the females are diploid. There is a large variation in life history, behaviour, and demography in the 15,000 species of hymenoterans. Hymenopterans tend to have very high recombination rates: in a review of eukaryotic recombination rates (Wilfert *et al.* (2007)), the average recombination rate in eight species of hymenopterans was reported to be 7.5 cM/Mb, compared to the average recombination rate of 2.2 cM/Mb for 15 insect species outside of Hymenoptera and 1.36 cM/Mb for 7 species of vertebrates. Further, there is also a huge range of variation in recombination rates within the hymenopterans, ranging from

25.1 cM/Mb in *Apis florea* to 1.0 cM/Mb in the solitary bee *Megachile rotundata* (Jones *et al.* (2019)). The members of the genus Apis (comprising honeybees) have the highest recombination rates among the hymenopterans: *Apis dorsata* (25.1 cM/Mb), *Apis mellifera* (21.6 cM/Mb), *Apis florea* (20.8 cM/Mb) and *Apis cerana* (17.4 cM/Mb) (Rueppell *et al.* (2016); DeLory *et al.* (2024)). Other eusocial bees have relatively lesser, but still high, recombination rates: *Bombus terrestris* (8.9 cM/Mb) and *Frieseomelitta varia* (12.5 cM/Mb) (Kawakami *et al.* (2019); DeLory *et al.* (2024)). The eusocial wasp *Vespula vulgaris* has a comparable recombination rate of 9.7 cM/Mb (DeLory *et al.* (2024)). The eusocial ants, too, have high recombination rates ranging from 14.0 cM/Mb in *Pogonomyrmex rugosus* to 5.5 cM/Mb in *Solenopsis invicta* (DeLory *et al.* (2024)). In contrast, the solitary bees *Megachile rotundata* (1.0 cM/Mb) and *Eucera pruinosa* (2.1 cM/Mb) have very low recombination rates (Jones *et al.* (2019); DeLory *et al.* (2024)). Five species of solitary parasitoid wasps were reported to have an average recombination rate of 4.74 cM/Mb, ranging from 1.5 cM/Mb in *Nasonia vitripennis* to 6.9 cM/Mb in *Cotesia typhae* (DeLory *et al.* (2024)). Therefore, eusocial hymenopterans have much higher recombination rates than non-social hymenopterans, and these high recombination rates have evolved multiple times independently among eusocial hymenopterans DeLory *et al.* (2024).

However, eusocial termites, which are not hymenopterans and are diploid, do not exhibit high recombination rates, indicating that both haplodiploidy and eusociality might be needed for the evolution of high recombination rates. The average recombination rates in eusocial termites have been estimated to be 0.34 - 0.43 cM/Mb in *Macrotermes bellicosus* and 0.31 - 0.38 cM/Mb in *Cryptotermes secundus* (Everitt *et al.* (2024)), which are much lower than the average recombination rates seen in eusocial hymenopterans.

Multiple hypotheses have been suggested to explain the unusually high recombination rates seen in social hymenopterans. The red queen hypothesis suggests that the high recombination rates evolved due to the evolutionary arms race between social insects and parasites or pathogens. Social insects are at a higher risk of disease transmission due to their crowded living conditions, high rate of social interactions, and the high genetic similarity between the members in the colony (Sherman *et al.* (1988)). Therefore, this hypothesis suggests that increased recombination rates confer resistance to disease propagation within the colony by increasing the genetic diversity within the colony (Kidner and Moritz (2015)). Increased genetic diversity within a colony could also have many other benefits; for example, it has also been shown to result in increased colony productivity and foraging efficiency (Oldroyd *et al.* (1993)). While there is some evidence that increased genetic diversity provides resistance to parasites and disease transmission, there is no clear evidence for parasites being responsible for the increased recombination rates (see DeLory *et al.* (2024) for a review of the evidence).

It has also been proposed that increased genetic diversity created by high recombination rates is needed to evolve elaborate caste specialisation. There are large behavioural and phenotypic differences between the queen and workers and within the workers in social insects. A higher rate of recombination reduces the selective interference (Hill-Robertson effect) on the genes affecting the different queen and worker phenotypes, respectively (Kent and Zayed (2013)). This allows the beneficial alleles affecting queen and worker phenotypes to be all on the same haplotype, favouring the evolution of an increased recombination rate. Further, reduced selective interference and increased gene duplication

events (a possible byproduct of increased recombination) of the worker-related genes can, in turn, lead to faster evolution of new worker phenotypes(DeLory *et al.* (2024)). The evidence for this hypothesis is inconsistent and mixed (Kent *et al.* (2012); DeLory *et al.* (2024); Everitt *et al.* (2024)).

DeLory *et al.* (2024) recently proposed that the increased recombination in eusocial hymenopterans might be due to the high longevity of their queens. The increased lifespan of the queen in social hymenopterans as compared to solitary hymenopterans results in an increased fitness cost of mutations accumulated due to oxidative stress and proliferation of transposable elements (DeLory *et al.* (2024)). Suppose the fitness costs of such accumulated mutations are only paid later in life (such as once the accumulated mutations cross a threshold). In that case, social hymenopterans, where the queens tend to live much longer than the solitary insects, will face a greater selection pressure against such mutations and transposable elements. An increased recombination rate can lead to improved selection against TEs and to an improved response to oxidative stress. The mechanisms that mitigate the harmful effects of oxidative stress and TE accumulation include ectopic recombination and DNA repair via homologous recombination, both of which show mechanistic overlap with meiotic recombination. Therefore, direct selection on mechanisms which overlap between DNA repair and meiotic recombination can lead to a selection for increased recombination rates. Further, indirect selection for increased recombination could result from selective interference between longevity-related alleles. This hypothesis is quite recent, and there is no strong or consistent evidence in its support(DeLory *et al.* (2024)).

Another explanation for the increased recombination rates in social insects is the reduced effective population size in social insects (DeLory *et al.* (2024)). Due to the strong reproductive skew in social insects (queens producing the majority of offspring), the effective population size in social insects is much smaller than in non-social insects with the same overall population size. Genetic drift, which increases negative linkage disequilibrium by the Hill-Robertson effect, tends to be higher in smaller populations. Therefore, selection for increased recombination is greater in social insects, which tend to have smaller effective population sizes. The evidence supporting this hypothesis, like that of the other hypotheses, is insufficient DeLory *et al.* (2024).

Another hypothesis for the increased recombination rates in eusocial insects suggests that the increased genetic diversity due to recombination could reduce kin conflict within the colony by reducing the variance of relatedness among the individuals (Sherman (1979)). Finally, it has been suggested that sex-specific recombination in haplodiploids and some other diploid insects (like many Lepidopterans) might be responsible for their high recombination rates, as the recombining sex might have evolved a high recombination rate to compensate for the lack of recombination in the other sex (Wilfert *et al.* (2007)).

Except for the last hypothesis, all of these hypotheses try to explain increased recombination rates as a consequence of eusociality. However, these hypotheses fail to explain the absence of high recombination rates in eusocial termites. Therefore, it was suggested by Everitt *et al.* (2024) that the increased recombination rates in eusocial Hymenopternas are due to strong sexual selection among males in eusocial hymenopterans. As recombination leads to a more effective purging of deleterious alleles by reducing selective interference, the evolutionary advantage of recombination could be greater in species with highly increased competition among the males, where the selection for reduced mu-

tations might be very strong . Since eusocial hymenopterans have highly male-biased operational sex ratios, they are likely to have high sexual selection in males. In contrast to this, eusocial termites tend to have fairly balanced sex ratios, and therefore, they are likely to have lesser sexual selection in males. Moreover, since selection is more effective in haploid males, heightened recombination rates could offer even greater advantages in eusocial hymenopterans, which are haplodiploid.

1.2 Heterochiasmy: Why do females recombine more than males?

Another interesting but relatively understudied aspect of recombination is the sexual dimorphism in recombination rates, also known as heterochiasmy. Heterochiasmy is widespread among taxa (Sardell and Kirkpatrick (2020)) and has important consequences for sex chromosome evolution, adaptation, population differentiation and speciation. Heterochiasmy is even seen among hermaphroditic organisms (Franch *et al.* (2006); Giraut *et al.* (2011); Jones *et al.* (2013); Theodosiou *et al.* (2016)). On a broad scale, females tend to have higher genome-wide recombination rates than males in most of the studied organisms Sardell and Kirkpatrick (2020). This is true for most eutherians ((Brandvain and Coop (2012)), including humans (Bhérer *et al.* (2017))), outcrossing angiosperms and teleost fish (Brandvain and Coop (2012)). The opposite trend is seen in marsupials, some grasshoppers and newts (Brandvain and Coop (2012)). Some plants and birds do not show much dimorphism in recombination rates Brandvain and Coop (2012). Besides the overall rates of recombination, the locations of the crossovers also differ between males and females. Males tend to have higher recombination towards the ends of chromosomes (i.e., telomeres), while in females, crossovers tend to be more uniformly distributed or elevated near the centromeres. This pattern was seen in 33 of the 51 species studied in a recent meta-analysis by Sardell and Kirkpatrick (2020) and was widespread in eutherian mammals, teleosts, frogs, plants and some birds, while the opposite pattern was seen in marsupials, domestic pigs and tomatoes (Sardell and Kirkpatrick (2020)). Besides such broad-scale differences, there are also sex differences in recombination at a finer scale. Sex-specific recombination hotspots have been observed in many mammals and maize(Sardell and Kirkpatrick (2020)). In mice, a larger proportion of recombination hotspots are binding sites for PRDM9 in males than in females (Brick *et al.* (2018)). In humans, recombination is elevated in promoter regions of genes in females while it is suppressed in promoters of males (Bhérer *et al.* (2017)). Female maize plants tend to have crossovers at transcription start sites, while male maize plants tend to have them around 400 bp upstream of that (Kianian *et al.* (2018)). In this thesis, we focus on the differential genome-wide recombination rates between the two sexes.

Various explanations for heterochiasmy have been proposed, including both mechanistic and adaptive hypotheses. The mechanistic explanations include sex differences in crossover interference, length of the synaptonemal complex, the temperature of gonads, genomic imprinting, epigenetic modifications and effects of ageing (reviewed in Sardell and Kirkpatrick (2020)). However, the existence of sex-specific genetic variation for recombination rates, along with the huge variation in sexual dimorphism in recombination rates and landscapes across taxa, indicates that heterochiasmy is not just an artefact of mechanical constraints but might be of adaptive significance.

Probably the very first proposed explanation for heterochiasmy was the suggestion by Haldane (Haldane (1922)) and Huxley (Huxley (1928)) that the heterogametic sex has suppressed recombination as a pleiotropic consequence of selection aimed at preventing recombination between heteromorphic sex chromosomes and, thereby, protecting the linkage of sex-determining regions. This hypothesis does explain achiasmy very well, i.e., when recombination is completely absent in one sex, then that sex is always the heterogametic sex. However, it does not explain heterochiasmy (differences in recombination rates between males and females when both sexes do recombine), as it fails to explain the high recombination rates observed in heterogametic females, and neither does it explain heterochiasmy in hermaphroditic species (Franch *et al.* (2006); Giraut *et al.* (2011); Theodosiou *et al.* (2016)).

There is not much prior mathematical literature on heterochiasmy. In his model, Lenormand (2003) found that a difference between the sexes in selection on autosomal loci in the diploid phase does not result in the evolution of sex differences in recombination. According to his model, heterochiasmy could evolve (1) if there is a difference in epistasis and/or directional selection between the gametes produced by males and females during the haploid phase; (2) if the difference between cis-epistasis (epistasis between genes present on the same chromosome) and trans-epistasis (epistasis between genes present on homologous chromosomes) differs between the two sexes during the diploid phase; and (3) if there is a difference in directional selection and/or epistasis between combinations of genes inherited maternally and paternally (which might be due to factors like imprinting). All three of these conditions are rather unlikely as haploid phases in most animals are short, and only a small fraction of genes (0.1 per cent in mammals Sardell and Kirkpatrick (2020)) tend to show parental imprinting. Based on these results Lenormand and Dutheil (2005), proposed that differences in selection during the haploid phase would result in a reduced recombination rate in the sex with the more intense selection, as there is a short-term disadvantage to recombination in the presence of epistasis, and a greater cost of recombination will be paid by the sex which faces more intense selection during the gametic phase. There is some evidence for this hypothesis in the case of plants (Lenormand and Dutheil (2005)), but it seems to be not very applicable in the case of animals where the haploid phases tend to be short without much gene expression (Joseph and Kirkpatrick (2004)).

Trivers (1988) proposed that the sex under strong sexual selection should evolve to have lower recombination rates to preserve the favourable gene combinations that are associated with enhanced male reproductive fitness. However Mank (2009), in a meta-analysis of ten eutherian mammals, found sexual dimorphism to be positively correlated with male-biased heterochiasmy (i.e., more sexually dimorphic mammals exhibited a higher recombination rate in males than in females).To explain these results,Mank (2009) proposed that fitness landscapes fluctuate with time more strongly in males than in females due to factors like male-biased dispersal and fluctuating female choice, which therefore result in an increased recombination rate in males to produce sons more likely to be able to adapt to a changed environment. However, it is not clear why this should increase the recombination rate in males more than in females, and neither is there any strong evidence for this hypothesis as of now. The meta-analysis by Mank (2009) only considered ten species of mammals, data for four of which (sheep, pig, cat, dog) was obtained from feral populations, which may limit the applicability of the results to natural populations.

This project used individual-based multilocus simulations to explore the effect of differences in selection between sexes on the evolution of recombination in both diploid and haplodiploid species. These models allow us to explore the hypothesis that strong sexual selection among males leads to increased recombination rates in haplodiploid species (assuming that sexual selection leads to stronger selection in males than in females). We further used analytical modelling to better understand our results. The project specifically looks at the following questions: 1) How does the recombination rate differ between diploid and haplodiploid organisms in the absence of epistasis? 2) How does epistasis affect recombination rates in haplodiploid and diploid species? 3) Can sex differences in selection lead to a sexual dimorphism in recombination rates?

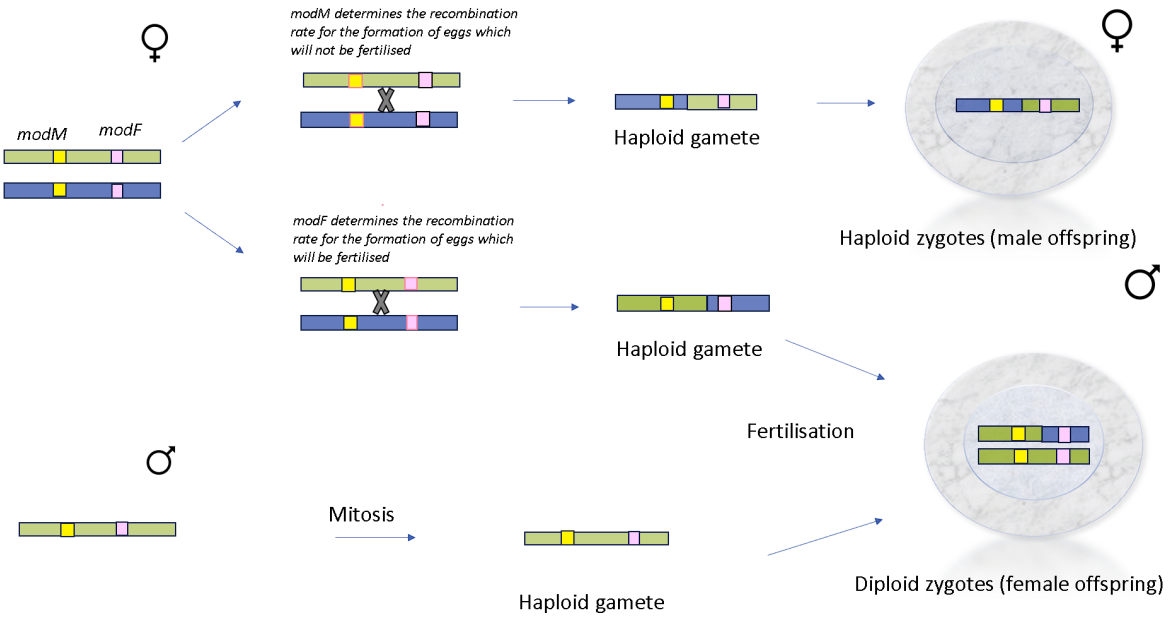
The main results obtained are: 1) Recombination rate does not differ significantly between diploid and haplodiploid organisms in the absence of epistasis, 2) Negative epistasis increases recombination rates in both haplodiploid and diploid species, but increases it to a much larger degree in haplodiploid species, 3) Sex differences in epistasis, but not fitness effects of mutations can lead to a pronounced sexual dimorphism in recombination rates.

Chapter 2

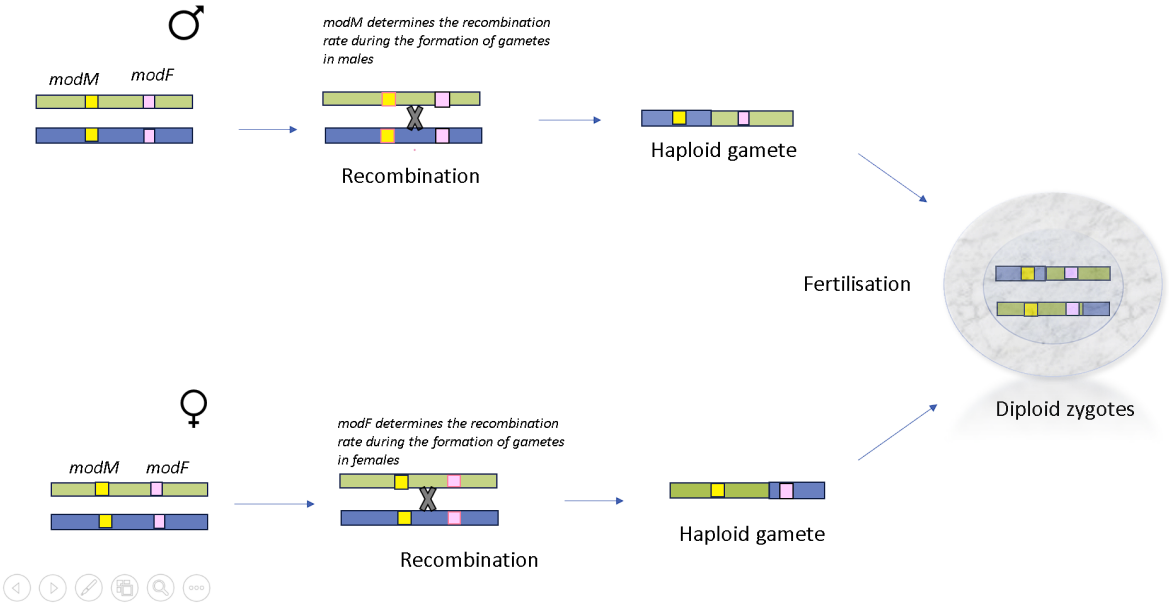
Materials and Methods

2.0.1 Simulations

We used multilocus individual-based simulations to study the evolution of recombination rates in different genetic systems. Our simulations consisted of a population of N individuals (N=10000), each having two or one copy of a linear chromosome (in diploid and haploid individuals, respectively). In the diploid case, the population consists of diploid males and females, where the sex of each offspring was randomly assigned. In each generation, males and females produce a very large (effectively infinite) number of gametes (by meiosis), which fuse at random to generate new diploid offspring. Among those, N offspring are sampled at random to form the next generation (Wright-Fisher model). The haplodiploid model is similar, except that males are haploid: they are produced by unfused female gametes, and they produce their gametes by mitosis. At every generation, a number of new deleterious mutations are introduced randomly on each chromosome (drawn from a Poisson distribution with parameter U_f for females and U_m for males), and their location along the chromosome is chosen independently by drawing from a uniform distribution between 0 and 1. Therefore, the number of sites at which mutations can occur is effectively infinite. Each mutation was assumed to have an equal effect on the fitness of the individual, i.e. all the mutations were equivalent.



(a) **The schematic shows the haplodiploid life cycle, specifically the Haplodiploid Two Modifier case.**



(b) **The schematic shows the diploid life cycle, specifically the Diploid Two Modifier case.**

Figure 2.1: Schematic for the Simulations.

Fitness

In the absence of epistasis, the fitness of the diploid individuals (W_m for males and W_f for females) is calculated as:

$$W_f = (1 - s_f h_f)^{n_{he}}(1 - s_f)^{n_{ho}} \tag{2.1}$$

$$W_m = (1 - s_m h_m)^{n_{he}}(1 - s_m)^{n_{ho}} \tag{2.2}$$

Where n_{he} and n_{ho} are the number of heterozygous and homozygous deleterious mutations, respectively, s_f and s_m are the selection coefficients in females and males, respectively, and h_f and h_m are the dominance coefficients in females and males respectively.

The fitnesses W_f and W_m for females and males, respectively, in the haplodiploid case are:

$$W_f = (1 - s_f h_f)^{n_{he}}(1 - s_f)^{n_{ho}} \tag{2.3}$$

$$W_m = (1 - s_m)^n \tag{2.4}$$

Here, n refers to the number of mutations in the haploid males. Therefore, the fitness effect of mutations in haploid males was considered equivalent to that of mutations in the diploid homozygous state.

It can be assumed that due to sexual selection, selection against deleterious mutations may be stronger in males than in females ($s_m > s_f$), given mutations also affect the reproductive success of males in sexual selection.

The effect of negative epistasis among deleterious mutations was studied by assuming two forms of negative epistasis: truncation selection and pairwise negative epistasis (Figure 2.2). In truncation selection, it was assumed that only males with fitness above a specified percentile threshold could reproduce. If the specified percentile threshold is denoted by P, only males with fitness values above or equal to the $P - th$ percentile of the male fitness distribution are allowed to reproduce. For example, a threshold of 0.5 for males corresponds to a situation where only males with fitness greater than the median fitness are allowed to reproduce. Therefore, the fitness equation in this case for a diploid male individual i is:

$$W_m(i) = (1 - s_m h_m)^{n_{he}}(1 - s_m)^{n_{ho}} \quad if \;\; W(i) \geq W_{X_m}^{-1}(P) \tag{2.5}$$

Here, $W_{X_m}$is the cumulative fitness distribution of males in the population.

$$W_m(i) = 0 \quad if \;\; W(i) < W_{X_m}^{-1}(P) \tag{2.6}$$

The equation for the fitness of diploid females remains unchanged.

Truncation selection among haploid males is defined similarly.

The case of epistasis (truncation selection or pairwise epistasis) restricted to males was considered because sexual selection could introduce a form of negative epistasis acting only in males by ensuring that only the best males could reproduce. Truncation selection, theoretically, is an extreme (strong) case of negative epistasis.

The second form of negative epistasis considered was pairwise negative epistasis, in which the fitness was modified by an epistatic factor for each pairwise interaction

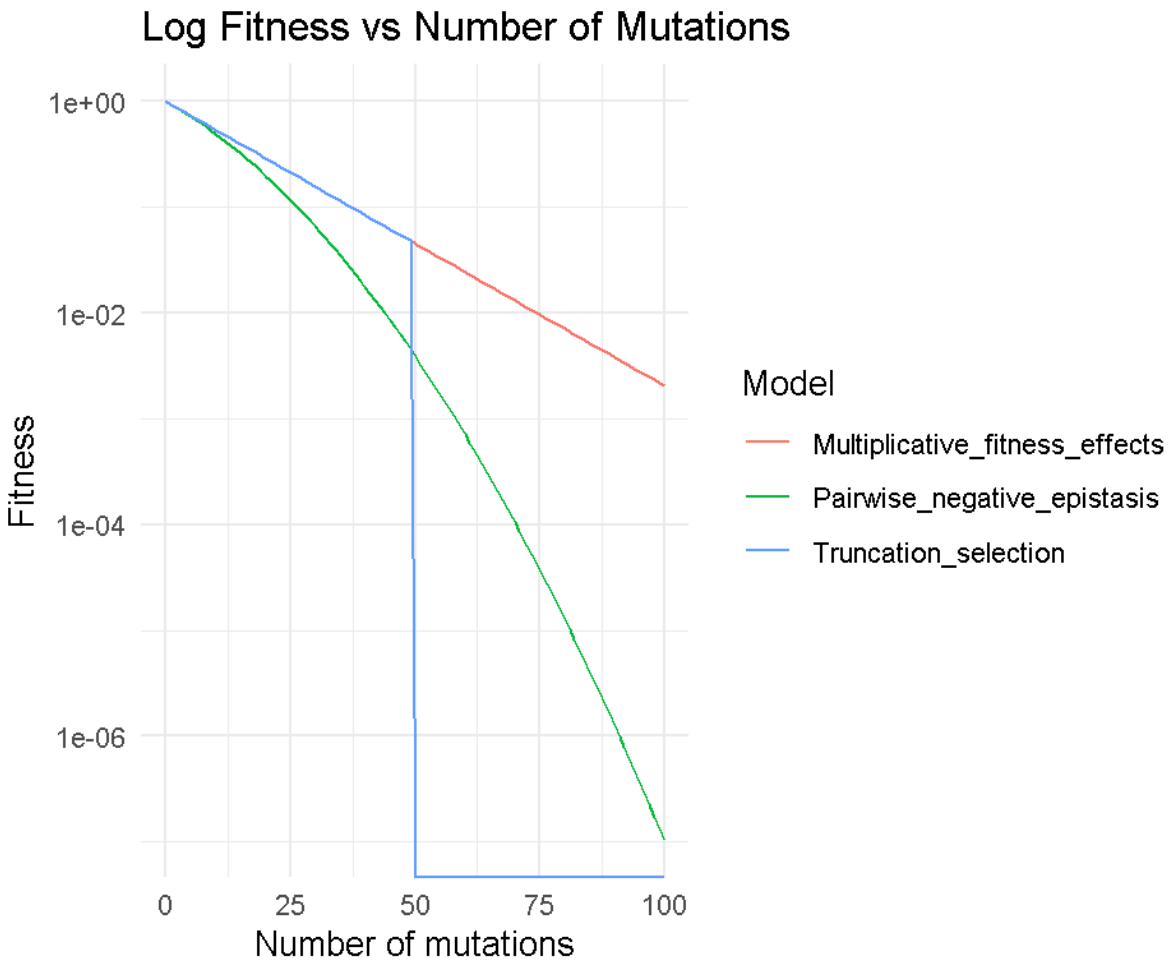


Figure 2.2: **The graph shows how the reproductive fitness of individuals varies with the number of deleterious mutations in their genome in our pairwise negative epistasis (green) and truncation selection (blue) regimes, compared to when mutations have multiplicative fitness effects or no epistasis (red).**

between deleterious alleles of two different loci. The corresponding fitness equations for pairwise negative epistasis in the diploid case are:

$$W_m = (1 - s_m h_m)^{n_{he}} (1 - s_m)^{n_{ho}} (1 - e)^{\binom{n_{he}+2n_{ho}}{2} - n_{ho}} \tag{2.7}$$

If epistasis is only in males, W_f is:

$$W_f = (1 - s_f h_f)^{n_{he}} (1 - s_f)^{n_{ho}} \tag{2.8}$$

and if epistasis is in both males and females, W_f:

$$W_f = (1 - s_f h_f)^{n_{he}} (1 - s_f)^{n_{ho}} (1 - e)^{\binom{n_{he}+2n_{ho}}{2} - n_{ho}} \tag{2.9}$$

Here, the sign of e is positive and e represents the magnitude of negative epistasis.

The corresponding fitness equations for the haplodiploid case with pairwise negative epistasis are:

$$W_m = (1 - s_m)^n (1 - e)^{\binom{n}{2}} \tag{2.10}$$

If epistasis is restricted to males, W_f is:

$$W_f = (1 - s_f h_f)^{n_{he}} (1 - s_f)^{n_{ho}} \tag{2.11}$$

and if epistasis is experienced by both males and females, W_f is

$$W_f = (1 - s_f h_f)^{n_{he}} (1 - s_f)^{n_{ho}} (1 - e)^{\binom{n_{he}+2n_{ho}}{2} - n_{ho}} \tag{2.12}$$

The chromosomes of a diploid parent undergo recombination during meiosis. The number of crossovers was drawn from a Poisson distribution with parameter R, corresponding to the chromosome genetic map length (in Morgans), and the position of each crossover was drawn independently from a uniform distribution along the chromosome, i.e. a state of no crossover interference was assumed.

For both diploid and haplodiploid cases, we considered one-modifier and two-modifier models. In the one-modifier models, the map length R is controlled by a single modifier locus located at the mid-point of the chromosome, affecting map length in males and females (in the diploid case) or in females only (in the haplodiploid case). In diploid individuals, R is the average of the map length coded for by the two alleles present at the modifier locus (one on each chromosome).

In two-modifier models, each individual has two modifier loci *modF* and *modM* (situated at one-third and two-thirds the length of the chromosome). In the diploid case, *modF* is expressed only in females and affects the map length only in females, and *modM* is expressed in males and affects the map length only in males. Both *modF* and *modM* can evolve independently of each other. In the haplodiploid case, *modM* affects R in meioses giving rise to parthenogenetic ovules (that will develop into males), and *modF* affects R in meioses that will give rise to fertilized ovules (that will develop into females). Both *modM* and *modF* are only expressed in females in the haplodiploid case, as only females are diploid and can undergo meiosis. While it is currently unknown whether recombination may differ between these two types of meioses in haplodiploid species, we considered it as an interesting theoretical possibility.

The model introduced a small direct cost of recombination by multiplying the fitness of each individual by a term $exp(-cR)$, where c represents a fitness cost per crossover and R is the map length expressed in the individual at meiosis. This was done to stabilize R during the simulation (ensure that it reaches an equilibrium). However, the results did not change qualitatively when the cost was removed.

The value of the chromosome map length was set at 1 during the beginning of the simulations for the first 20000 generations. After that, the modifier locus was allowed to mutate at a rate of m per generation, with mutations of small effect happening 95 per cent of the time, while the rest being mutations of large effect. To simulate mutations of small effect, the value of the modifier locus was multiplied by a value drawn from a Gaussian distribution with mean 1 and variance σ_m. To simulate mutations of large effect, a number drawn from a uniform distribution between -1 and 1 was added to the value of the modifier locus. If the resultant value was negative, it was set to zero.

The simulations were run for 5×10^6 generations. The average map lengths, the average fitness and the average number of mutations for the entire population were measured for the two sexes separately. These quantities were measured at every $200th$

Table 2.1: Parameters of the model and their default values

Variable	Description	Default Value (Other Values Considered)
s_f	Selection coefficient of deleterious mutations in females	$0.06(0.02, 0.04, 0.08, 0.1)$
s_m	Selection coefficient of deleterious mutations in males	0.06 (0.02, 0.04, 0.08, 0.1)
h_f	Dominance coefficient of deleterious mutations in females	0.2 (0.1, 0.5)
h_m	Dominance coefficient of deleterious mutations in males	0.2 (0.1, 0.5)
U_m	Deleterious mutation rate per generation per chromosome in males	0.2 (0.5)
U_f	Deleterious mutation rate per generation per chromosome in females	0.2 (0.1)
e	pairwise epistasis factor between deleterious mutations	0, 0.0001, 0.001,0.002, 0.005, 0.01, 0.02, 0.05, 0.1
c	e^{-c} is the direct fitness cost of recombination per unit map length	0.01 (0, 0.001)
N	Number of individuals	10000
m	Rate of mutation at the modifier locus per generation	0.0001
σ_m	determines the variance in magnitude of mutations of small effect (see text) at modifier loci	0.2
$NbGen$	Number of generations during which the simulation is run	5000000

generation during the first $20000th$ generations and at every $500th$ generation after that. The results reported everywhere in this document are the time averages of these quantities (excluding the initial burn-in period) unless otherwise specified. The average recombination rate was quantified as the average number of cross-overs happening in a chromosome.

The following table lists all the variables along with the values considered in the simulation :

Unless otherwise specified, the results shown throughout this document are for the default values of these parameters.

All the simulations were written in C++, while the code for the analysis of data was written in R (R Core Team (2023)). The simulation model is an extension of the one developed and used in Roze (2021) for diploid hermaphroditic organisms.

2.1 Analytical Model

To model the evolution of the rate of recombination, we used a three-locus model comprising one recombination modifier locus i and two loci j and k under direct selection. We used the method of modelling multilocus evolution proposed by Kirkpatrick *et al.* (2002) to model how the rate of recombination $r_{j,k}$ between the two loci under selection varies with the parameters of the model. Our basic notation also mostly follows Kirkpatrick *et al.* (2002). The three loci were assumed to be located on the chromosome physically in the order i, j, k. The recombination modifier gene is not under any direct selection but faces indirect selection due to its linkage with the loci j and k. The population was considered infinite (i.e., no genetic drift), and mating was random. There were no evolutionary forces besides selection, and forces like migration were not considered. Each locus was assumed to have two alleles (let us call them $\{I, i\}$, $\{J, j\}$, $\{K, k\}$ respectively). At meiosis, the rate of recombination between the loci j and k was $r_{jk}+\delta r$, $r_{jk}+\frac{\delta r}{2}$ and r_{jk} for the individuals with the genotype II, Ii and ii respectively. The recombination rate between the loci i and j was r_{ij} for all the three genotypes. The genotype of an individual at any locus i was denoted by the indicator variable X_i, whose value was either 0 or 1, depending on the allele. In our case, we consider the alleles I, J and K to take the values 1 for their respective indicator variables and vice versa. The alleles J and K reduce fitness in females by a factor $1 - s_f h_f$ when heterozygous and $1 - s_f$ when homozygous (and by $1-s_m h_m$ and $1-s_m$ respectively in males). The fitness is reduced by an additive epistatic factor of e for each pairwise epistatic interaction. We ignored the higher-order epistatic terms. For diploid individuals, the genotype of an individual at locus i was denoted by $X_{i,\emptyset}$ for the paternal haplotype,i.e. for the chromosome inherited from the male parent and $X_{\emptyset,i}$ for the maternal haplotype. Therefore, $X_{i,\emptyset}$ indicates the allele inherited from the father at locus i by an individual and vice versa. This positional way of referring to the alleles inherited from the male and female parent is used throughout the rest of the document.

We define new centred variables $\zeta_{i,\emptyset}$ by subtracting the reference frequencies (in our case, the allele frequencies) from the indicator variables as:

$$\zeta_{i,\emptyset} = X_{i,\emptyset} - p_{i,\emptyset} \tag{2.13}$$

and so on for the other two loci. Here, $p_{i,\emptyset}$ is the frequency of the allele I on chromosomes inherited from the male parent, which is not necessarily equal to the frequency of the allele I on chromosomes inherited from the female parent $p_{\emptyset,i}$, depending on the assumptions of the model.

The product of these centred variables for any set of loci U on the paternal chromosome and a set of loci V on the maternal chromosome of an individual is defined as:

$$\zeta_{U,V} = \Pi_{i \epsilon U} \zeta_{i,\emptyset} \Pi_{j \epsilon V} \zeta_{\emptyset,j} \tag{2.14}$$

Here, the product is taken over all loci i from the set U and all loci j from the set V. Please note that $\zeta_{U,V}$ is a random variable.

The expected value of $\zeta_{U,V}$ over all the individuals is a moment or a measure of genetic association and is analogous to the linkage disequilibrium, except that here, the loci can be on different chromosomes.

$$D_{U,V} = E_X[\zeta_{U,V}] \tag{2.15}$$

Here, E_X represents the average taken over all the individuals. The details of the model are specified in the appendix.

Chapter 3

Results

The results are first presented for simulations without epistasis. Here, we will explore the effect of differences in selection between males and females (s_m, s_f).

3.0.1 Without Negative Epistasis

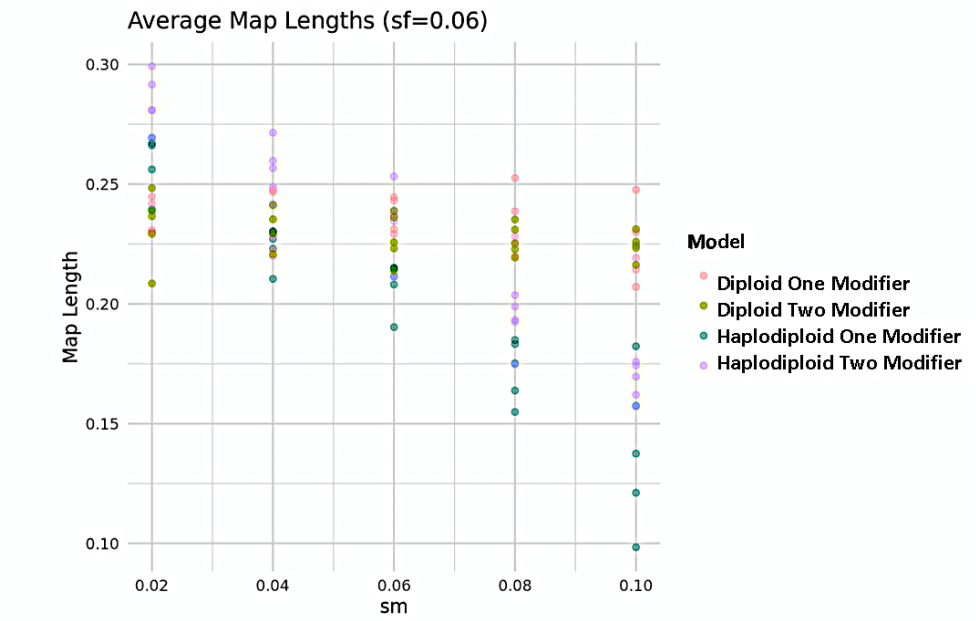


Figure 3.1: **The graph compares the average map lengths for the four cases without epistasis. Each dot represents a specific iteration of the respective model. The values used for the different parameters are the default values given in the Table 1.**

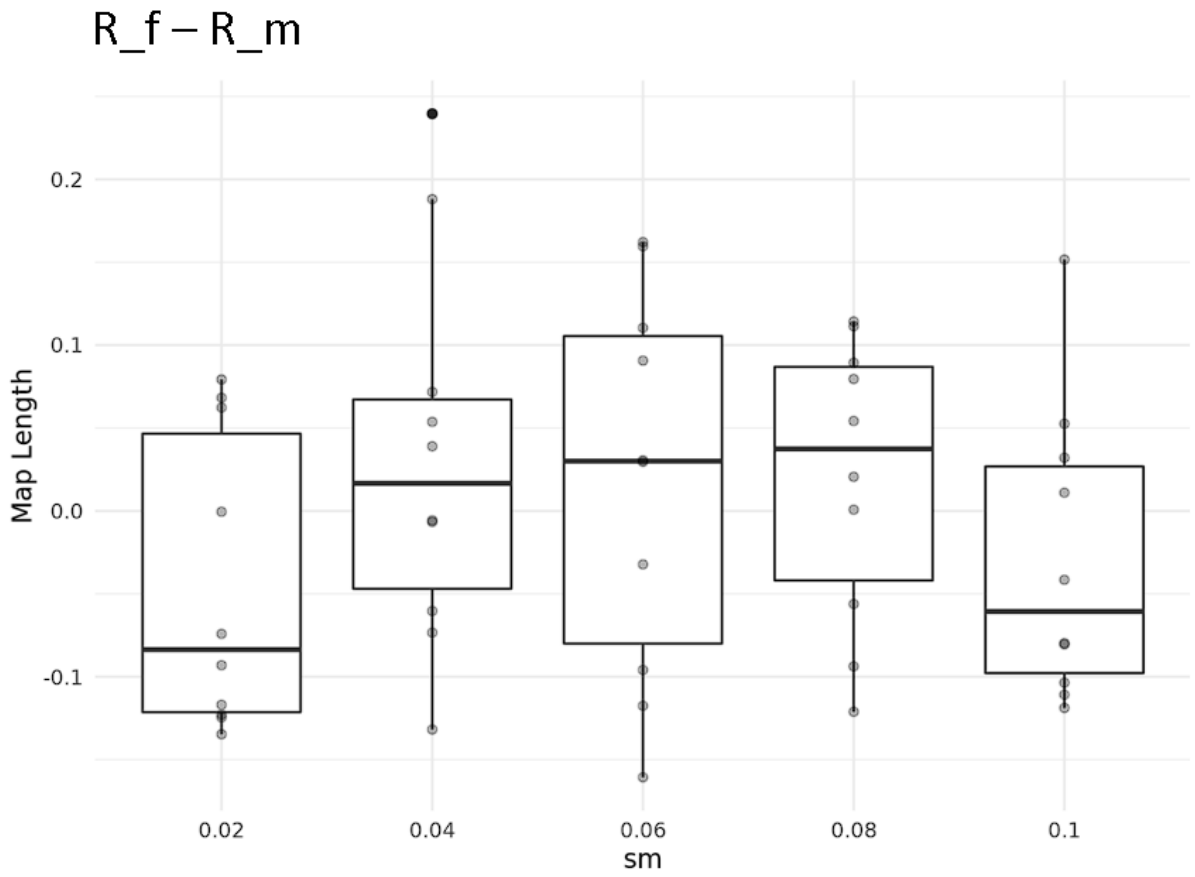
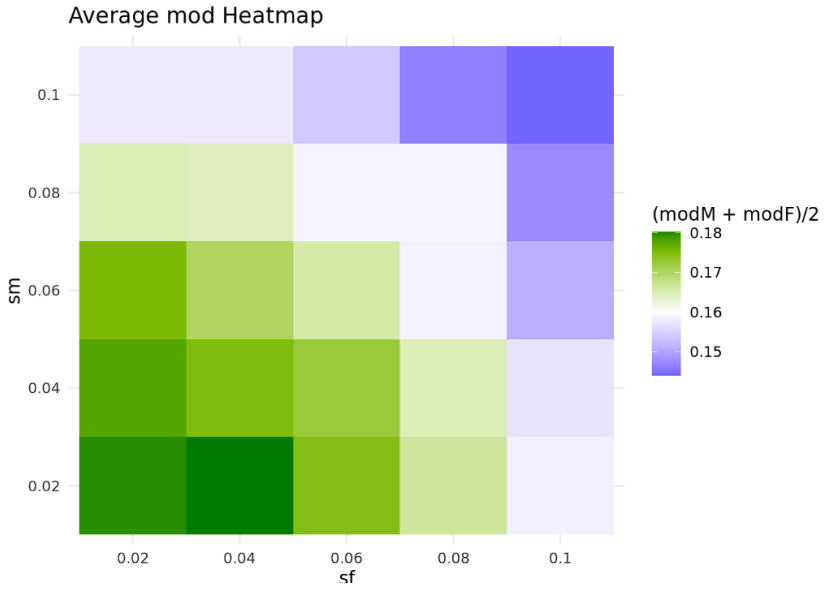
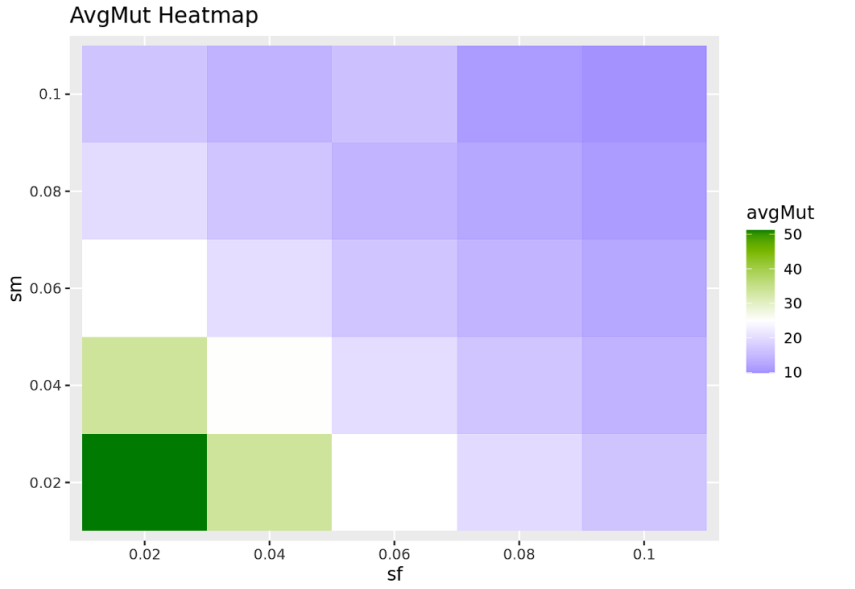


Figure 3.2: **The graph shows the difference in map lengths between the two sexes for the diploid two-locus case. Each dot represents a specific iteration of the respective model. The total number of iterations for each case is six. The values used for the different parameters are the default values given in the Table 1.**



(a)



(b)

Figure 3.3: **The graphs compare the a) average map lengths and b) the average number of mutations per chromosome for the diploid two modifier case (without epistasis) against the selection strengths.**Each square represents the average value of 25 iterations.

In the absence of epistasis, selection for recombination stems from the Hill-Robertson effect. Roze (2021) showed that selection for recombination mostly depends on U (mutation rate), N_e (effective population size) and R (chromosome map length), while s has less effect provided it is sufficiently small. As s increases, selection for recombination decreases (because deleterious alleles are rapidly eliminated from the population, and the negative linkage disequilibrium does not persist for long).

In the diploid one modifier case, the results show that when either s_m or s_f increases, the recombination rate decreases (Figure 3.3b and Figure 3.3a) in agreement with the results of Roze (2021)). In the diploid two-modifier model, the results are quite similar; the average R is similar in males and females in most cases and similar to the value obtained in the one-modifier case. Different fitness effects of mutations did not result in significant differences between the map lengths in males and females (Figure 3.2). However, different rates of mutations in the two sexes resulted in a slight sexual dimorphism in average map lengths, and this effect interacted with the differential effects of mutations in the two sexes (explained in appendix A.2).

In the haplodiploid one modifier case, the average map lengths were similar to the diploid cases. The results show that when $s_m = s_f$, R evolves to slightly lower values in the haplodiploid case compared to the diploid case, probably due to a greater elimination of deleterious mutations among haploid males due to the stronger effect of selection. This difference becomes more important as s_m increases (Figure 3.1). The average map lengths in the haplodiploid two-modifier case were similar to those in the one-modifier case.

3.0.2 With Negative Epistasis

Negative epistasis is known to favour recombination by generating negative linkage disequilibrium. We explored the effect of negative epistasis 1) in both the sexes and 2) only in males.

Negative Epistasis in Both Sexes

The average map lengths in both haplodiploids and diploids increased as the epistasis factor became more negative, before saturating and finally decreasing at a very high strength of negative epistasis (Figure 3.4a). The average map lengths in haplodiploids increased to a much larger degree than in diploids. The results of the one-modifier case were similar to the two-modifier case for both diploids and haplodiploids.

Negative Epistasis Only in Males

Across all models, the results were mostly similar for the two kinds of epistasis: truncation selection and pairwise negative epistasis. The differences were that: in the presence of truncation selection, the average map length became virtually zero when the threshold became too strong, an effect which can be seen for all four cases, while with pairwise negative epistasis, the map length saturated, continued to increase, or decreased slightly depending on the specific case. The reason for the behaviour seen with strong truncation selection is that under such strong selection, the mutational load at the mutation-selection balance became so low that there was hardly any benefit from

recombination. The results for truncation selection are shown and discussed in appendix (A.3) since they are qualitatively similar to the results for pairwise negative epistasis.

When there is negative epistasis only in males, average map lengths are lower in the diploid one-modifier and two-modifier cases compared to when there is epistasis in both the sexes (Figure 3.5b). In the diploid one-modifier case, they are even lesser than in the case of no epistasis. The average map lengths in the diploid two-modifier case are larger than those without epistasis. However, they are still lower than the average map length of haplodiploids (both one-modifier and two-modifier cases) in the presence of negative epistasis only in males (Figures 3.5b and 3.5a). In the haplodiploid one-modifier case, the average map lengths are lesser than when there is epistasis in both the sexes, while the opposite is true for the two-modifier case. Overall, the difference between diploids and haplodiploids is stronger when epistasis is only present in males, as compared to when it is present in both the sexes.

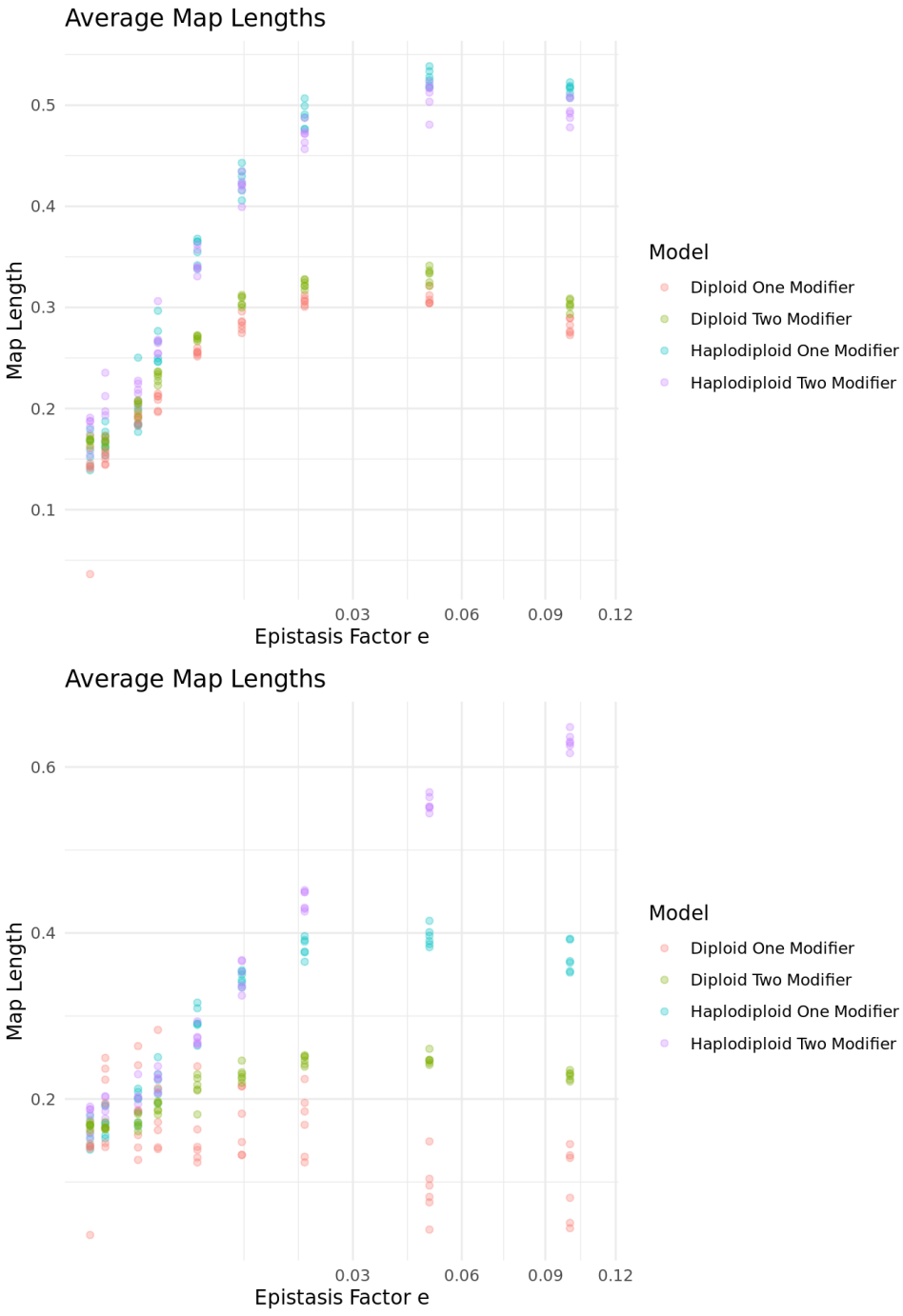
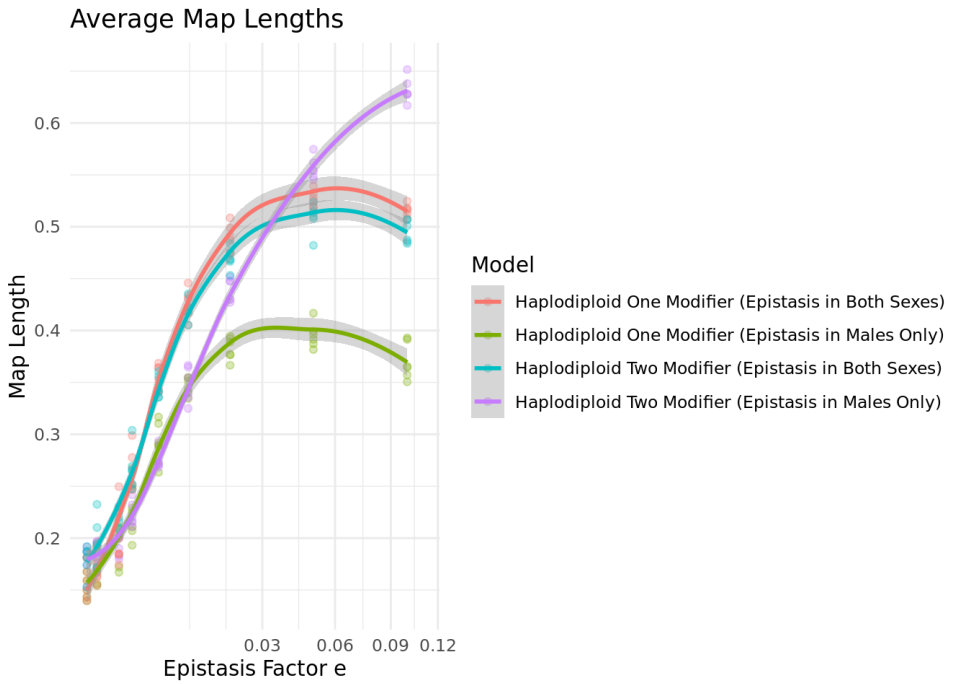
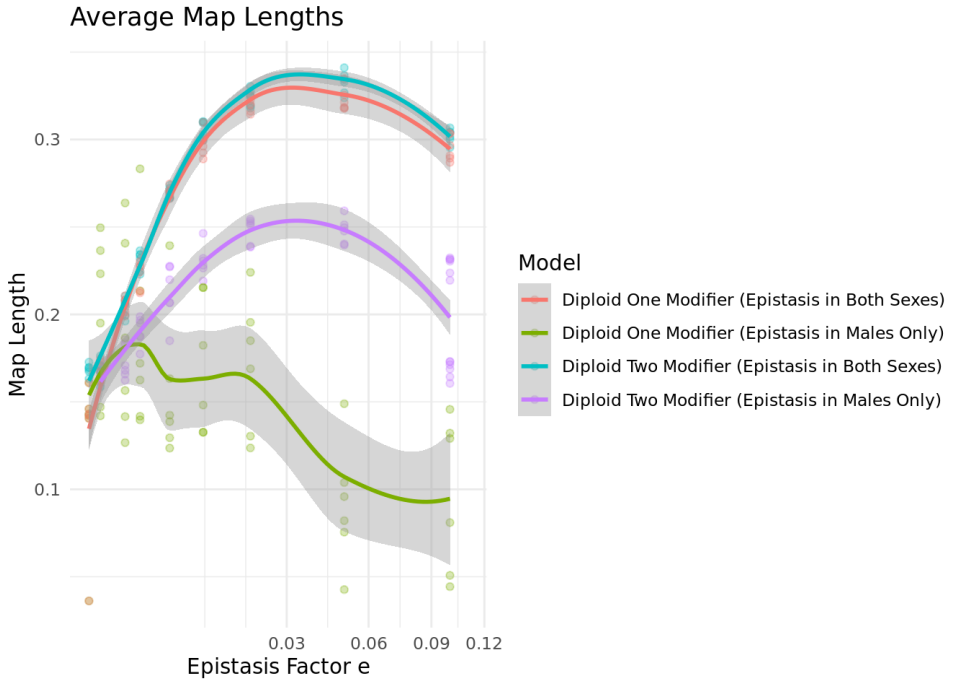


Figure 3.4: **The graphs compare the average map lengths for the four cases when there was a) pairwise negative epistasis in both sexes and b) in males only respectively.** The x-axis is square-root transformed, and the leftmost values correspond to e=0, i.e. the case of no epistasis.



(a)



(b)

Figure 3.5: **The graphs compare the average map lengths for the two a) haplodiploid and b) diploid cases with negative epistasis only in males and in both sexes, respectively.** The x-axis is square-root transformed, and the leftmost values correspond to e=0, i.e. the case of no epistasis.

When average map lengths can evolve independently in males and females (diploid two-locus case), they evolve to higher values in females than males (and the average is higher than in the one-locus model,Figure 3.6). The difference increased as the pairwise negative epistasis term became more negative.

When average map lengths can evolve independently in the two types of meioses (in the haplodiploid two-locus model), they evolve to higher values in meioses, leading to male offspring (Figure 3.7). The difference increased as the pairwise negative epistasis term became more negative.

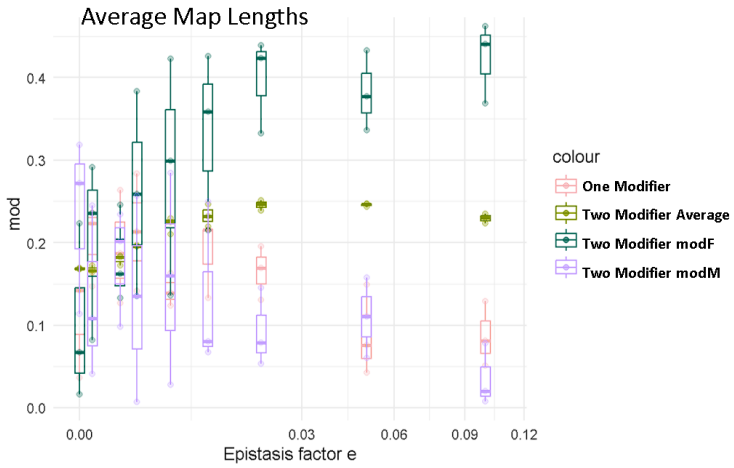


Figure 3.6: **The graph compares the average and sex-specific map lengths for the two diploid cases when there is pairwise negative epistasis in males only. modM and modF refer to map lengths in males and females, respectively.** The x-axis is square-root transformed, and the leftmost values correspond to e=0, i.e. the case of no epistasis.

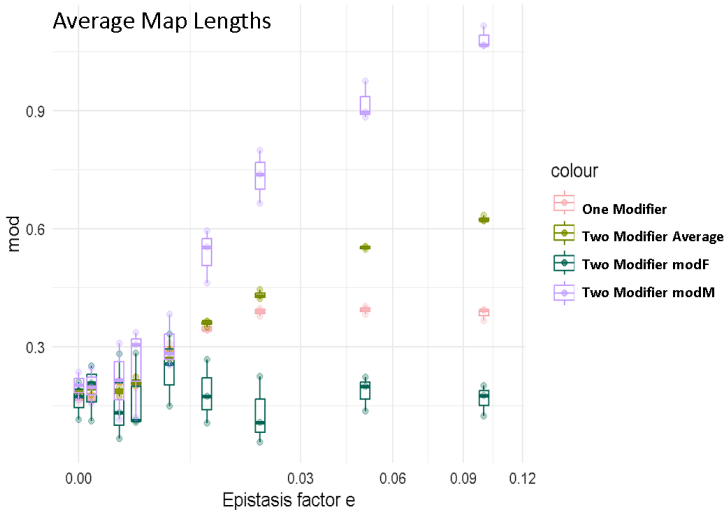


Figure 3.7: **The graph compares the average map lengths for the two haplodiploid cases when there is pairwise negative epistasis in males only. modM and modF refer to recombination rates (more specifically, average map lengths) during the production of parthenogenetic and to-be-fertilised ovules, respectively.**

3.0.3 Analytical Model

For the diploid one locus case, the rate of increase Δp_i in the frequency of the allele increasing recombination is:

$$\Delta p_i = -\frac{-\delta_r e p_i q_i p_j q_j p_k q_k}{(2r_{jk}r_{ijk})}[(a_j a_k)(\frac{1}{r_{ij}} + \frac{1}{r_{ik}} - 1) + e] \tag{3.1}$$

This equation was obtained by Barton (1995) using a slightly different formulation. Here,e is the pairwise epistasis factor (both in males and females); a_j and a_k are the selection coefficients for alleles at the loci j and k respectively, and both are approximately equal to $-sh$, r_{ik} and r_{ij} are the recombination rates between loci i and k, and i and j respectively, while r_{ijk} is $r_{ij}(1-r_{jk})+r_{jk}(1-r_{ij})$; p_j and p_k are the frequencies of the deleterious alleles at j and k loci respectively. This term is always negative when epistasis is positive, positive when epistasis is negative and sufficiently weak, and it becomes negative when epistasis is strongly negative. This equation shows that weakly negative epistasis generates a selection for increased recombination. At mutation-selection balance, $p_j q_j$ and $p_k q_k$ are approximately equal to u/a_j and u/a_k, respectively, where u is the mutation rate per locus.

The major result of our analytical model for the haplodiploid one locus case is that the rate of increase Δp_i in the frequency of the allele increasing recombination is:

$$\Delta p_i = \frac{1}{3}[\frac{-\delta r p_i q_i}{4r_{ijk}r_{jk}}[(2e^f + e^m)p_j q_j p_k q_k + 2\Delta_j \Delta_k]][A + T] \tag{3.2}$$

$$\Delta_j = \frac{a_j^m - a_j^f}{3} p_j q_j \tag{3.3}$$

$$\Delta_k = \frac{a_k^m - a_k^f}{3} p_k q_k \tag{3.4}$$

$$A = 2a_j^f {a_k}^f + 2e^f + a_j^m a_k^m + e^m + a_j^m a_k^f + a_j^f a_k^m \tag{3.5}$$

$$T = (2a_j^f + a_j^m)(2a_k^f + a_k^m)(\frac{1}{2r_{ik}} + \frac{1}{2r_{ij}} - 1) \tag{3.6}$$

$$a_j^f = a_k^f = -sh \tag{3.7}$$

$$a_j^m = a_k^m = -s \tag{3.8}$$

$$C = \frac{\delta r p_i q_i}{2} \tag{3.9}$$

When $a_j^m = a_k^m = e^m = 0$

$$\Delta p_i = \frac{1}{3}\Big[\frac{-\delta r p_i q_i}{r_{ijk} r_{jk}}(e^f + a_j^f a_k^f) p_j q_j p_k q_k\Big]\Big[e^f + (a_j^f a_k^f)(\frac{1}{r_{ik}} + \frac{1}{r_{ij}} - 1)\Big] \tag{3.10}$$

Here,e^f and e^m correspond to pairwise epistasis in males and females, respectively; a^fs and a^ms are the selection coefficients in males and females respectively, and r_{ik} is the recombination rate between allele i and k; p_j and p_k are the frequencies of the deleterious alleles at j and k loci respectively.

From this equation, it can be seen that Δp_i is positive only when the pairwise epistasis terms are negative. Further, Δp_i can become zero or negative when epistasis is too negative (when the second term in the product becomes negative). Compared to the diploid one locus case, we have an interesting additional factor of $-2\Delta_j\Delta_k$ in the numerator. This term is always negative and, therefore, tends to oppose the selection for increased recombination rates. This term corresponds to the differences in selection coefficients between the haploid males and the diploid females and becomes stronger as the selection difference between the sexes increases. A biological interpretation of this term and its role is given in the Discussion section.

For the haplodiploid two modifier case, if recombination rate during the meiosis which produces female offspring is fixed, but recombination rate during the meiosis which produces male offspring can evolve, the rate of change Δp_i in the frequency of the allele increasing recombination (during the production of male offspring) is:

$$\begin{aligned}\Delta p_i =& \frac{1}{3}\Bigg[\frac{D_j((a_j^f + a_j^m)R_{jm} + a_j^f R_{jf})}{1 - R_{jf} - R_{jm}} + \frac{D_k((a_k^f + a_k^m)R_m + a_k^f(R_f - C_f))}{1 - R_f - R_m + C_f} + \\ & (a_j^f a_k^m + a_j^m a_k^f + a_{jk}^m)D_{ijk}^{fm} + a_{jk}^f D_{ijk}^{ff}\frac{2}{1 + r_{ijk}^m}\Bigg]\end{aligned}$$

$$R_{jk}^f = \frac{1 - r_{jk}^f}{2} \tag{3.11}$$

$$R_{jk}^m = \frac{1 - r_{jk}^m}{2} \tag{3.12}$$

$$D_{jk}^{ff} = \frac{1 - R_{jk}^m + R_{jk}^f}{1 - R_{jk}^m - R_{jk}^f}(e^f p_j q_j p_k q_k + \Delta_j\Delta_k(1 - 2r_{jk}^f)) + \frac{R_{jk}^f}{1 - R_{jk}^m - R_{jk}^f}(e^m p_j q_j p_k q_k) \tag{3.13}$$

$$D_{jk}^{fm} = \frac{R_{jk}^m}{1 - R_{jk}^m}(D_{jk}^{ff} + e^m p_j q_j p_k q_k) + \frac{e^f p_j q_j p_k q_k + \Delta_j\Delta_k(1 - 2r_{jk}^m)}{1 - R_{jk}^m} \tag{3.14}$$

$$D_{ijk}^{ff} = \frac{-\delta_r p_i q_i(1 + r_{ijk}^m)}{4(r_{ijk}^m + r_{ijk}^f)}(D_{jk}^{fm} + D_{jk}^{ff} + e^m p_j q_j p_k q_k + 4\Delta_j\Delta_k) \tag{3.15}$$

$$D_{ijk}^{fm} = \frac{-\delta_r p_i q_i(1 - r_{ijk}^m)}{4(r_{ijk}^m + r_{ijk}^f)}(D_{jk}^{fm} + D_{jk}^{ff} + e^m p_j q_j p_k q_k + 4\Delta_j\Delta_k) \tag{3.16}$$

$$R_{jf} = \frac{1 - r^f_{ij}}{2} \tag{3.17}$$

$$R_{jm} = \frac{1 - r^m_{ij}}{2} \tag{3.18}$$

$$R_m = \frac{1 - r^m_{ik}}{2} \tag{3.19}$$

$$R_f = \frac{1 - r^f_{ik}}{2} \tag{3.20}$$

$$D_j = (a^m_k + a^f_k)D^{fm}_{ijk} + a^f_k D^{ff}_{ijk} \tag{3.21}$$

$$D_k = (a^m_j + a^f_j)D^{fm}_{ijk} + a^f_j D^{ff}_{ijk} \tag{3.22}$$

$$C_f = \frac{\delta_r(1 - 2p_i)(1 - 2r^f_{ij})}{4} \tag{3.23}$$

In all variables, the superscript (f or m) refers to the meiosis which produces female or male offspring respectively.

The corresponding equations for the case when recombination happens at a fixed rate during the meiosis which produces male offspring but its rate can evolve during the meiosis which produces female offspring are:

$$\Delta p_i = \frac{1}{3}\left[\frac{D_j((a^f_j + a^m_j)R_{jm} + a^f_j R_{jf})}{1 - R_{jf} - R_{jm}} + \frac{D_k((a^f_k + a^m_k)(R_m - C_m) + a^f_k R_f)}{1 - R_f - R_m + C_m} + (a^f_j a^m_k + a^m_j a^f_k + a^m_{jk} + a^f_{jk})D^{fm}_{ijk} + a^f_{jk} D^{ff}_{ijk}\right]$$

$$D^{ff}_{ijk} = \frac{-\delta_r p_i q_i (1 - r^f_{ijk})}{4(r^m_{ijk} + r^f_{ijk})}(D^{fm}_{jk} + D^{ff}_{jk} + e^m p_j q_j p_k q_k + 4\Delta_j \Delta_k) \tag{3.24}$$

$$D^{fm}_{ijk} = \frac{-\delta_r p_i q_i (1 + r^f_{ijk})}{4(r^m_{ijk} + r^f_{ijk})}(D^{fm}_{jk} + D^{ff}_{jk} + e^m p_j q_j p_k q_k + 4\Delta_j \Delta_k) \tag{3.25}$$

$$C_m = \frac{\delta_r(1 - 2p_i)(1 - 2r^m_{ij})}{4} \tag{3.26}$$

The equations for the other variables are same between the above two cases. In all variables, the superscript (f or m) refers to the meiosis which produces female or male offspring respectively.

On solving these equations numerically for the equilibrium value of r_{jk}, it was found that the equilibrium recombination rate for meiosis which produces male offspring was equal to or higher than the equilibrium recombination rate for meiosis which produces female offspring (3.8). This is consistent with our simulation results.

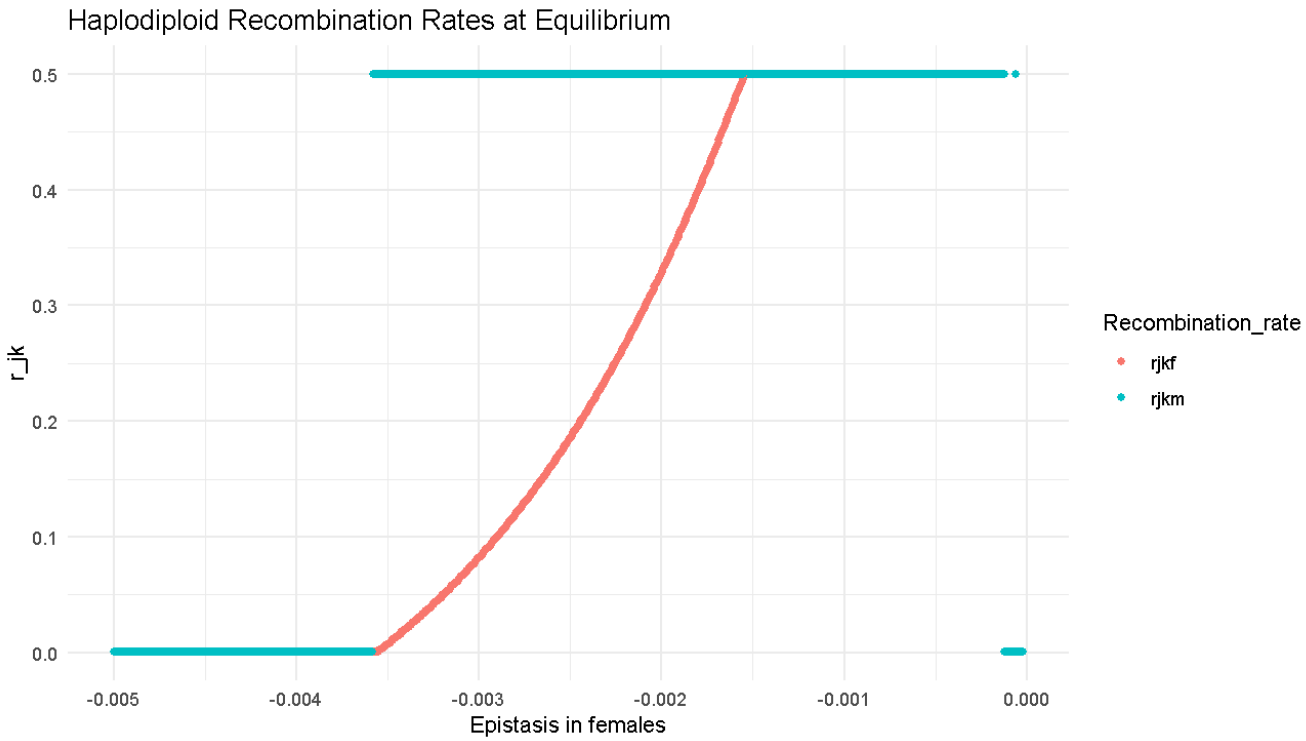


Figure 3.8: **The graph compares the recombination rate for the two kinds of meiosis in haplodiploids across a range of values of epistasis in females e_f. $rjkm$ and $rjkf$ refer to equilibrium recombination rates during the meiosis which results in the production of parthenogenetic and to-be-fertilised ovules, respectively.** The results shown are for $s_f = -0.03$,$s_m = -0.03$, and $em = -0.0001$. Qualitatively similar results were obtained for other values of s_m, s_f, and e_m.

Chapter 4

Discussion

Previous theoretical work (Barton and Charlesworth (1998); Agrawal (2006)) has shown that the evolution of increased recombination is favoured in the presence of negative linkage disequilibrium which may be generated either by negative epistasis or by selection and drift (Hill-Robertson interference), among other factors. Since in our model considered there is a direct selection on recombination due to an intrinsic cost of recombination, the equilibrium values of recombination rates are achieved when the direct selection and indirect selection (due to Hill-Robertson interference or negative epistasis) on the recombination modifier locus cancel each other out. In all cases considered the average map lengths and number of mutations was found to be inversely correlated with the selection strengths (or the strength of the fitness effects of mutations) (Figures 3.3b and 3.3a). This was because at greater selection strengths, the equilibrium mutation load in the population was reduced, leading to a reduced benefit of eliminating mutations by recombination. To recall, the main questions our work aimed to address were:1) How does recombination rate differ between diploid and haplodiploid organisms in the absence of epistasis? 2) How does epistasis affect recombination rates in haplodiploid and diploid species? 3) Can sex differences in selection lead to a sexual dimorphism in recombination rates?

There is not much difference between recombination rates in haplodiploids and diploids when there is no negative epistasis. Further, the results were similar irrespective of whether one or two loci determined the rate of recombination. This indicates that in the absence of epistasis, just haplodiploidy is not enough to result in a large increase in recombination rate. Haplodiploids had smaller map lengths than diploids when selection (especially in males) was stronger and vice versa. While the absence of recombination in males may tend to increase recombination in females (as was seen by us in our simulations for the case of sex-limited recombination in diploids, see Appendix A.4), this tendency seems to be balanced by the increased efficiency of selection in the haploid males reducing the mutation load, which in turn leads to a reduced benefit of increased recombination rates and hence tends to reduce recombination. Since selection is stronger in haploid males than diploid females, changes in fitness effects of mutations in males has a much larger effect on map lengths as compared to the effect of changes in fitness effects of mutations in females (Figure 4.1) In our equation 3.2 for the change in modifier frequency in the haplodiploid case, it can be seen that the difference in selection strengths between the two sexes tends to reduce selection for recombination in haplodiploids. This is because the different selection strengths result in a difference in mutation loads between the two sexes, and therefore, male gametes tend to carry fewer mutations than female gametes. This

effect tends to generate a slight positive linkage disequilibrium, resulting in the reduction in the overall magnitude of negative linkage disequilibrium in the population. This reduction in negative linkage disequilibrium reduces selection for increased recombination. Therefore, our results show that increased selection strength in haploid males and sex differences in selection strengths results in reduced recombination rates in haplodiploids compared to diploids, when selection is moderate or strong.

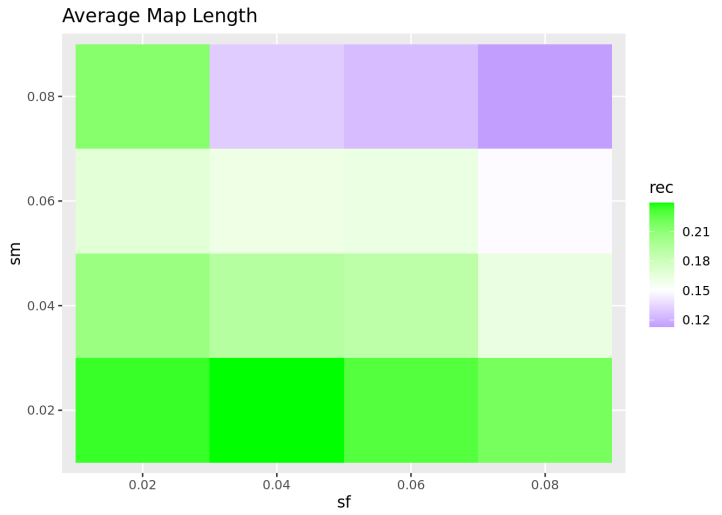


Figure 4.1: **The graph shows the average map length in the haplodiploid one modifier case against the selection strengths in males and females.** As can be seen from the graph, selection coefficients in males have a much stronger and clearer effect on map length than the selection coefficients in females. Each square represents the average of 25 iterations.

Different fitness effects of deleterious mutations in males and females did not by itself lead to any significant difference in male and female recombination rates. Therefore, if sexual selection acts by increasing the fitness effects of deleterious mutations in males (assuming that male success in sexual selection depends on their overall fitness), the simulation results, in agreement with the previous analytical results of Lenormand (2003), suggest that it would not result in a significant difference in recombination rates between the sexes. However, when the mutation rates differed between the two sexes, it resulted in a slight sexual dimorphism in recombination rates even in the absence of sex differences in fitness effects of mutations, the reasons for which are given in the appendix (A.2).

The average recombination rate in both haplodiploids and diploids (for the two locus case) increased considerably in the presence of pairwise negative epistasis, which is consistent with the established theory. However, interestingly, haplodiploid recombination rates evolved to a much greater degree than the diploid recombination rates. This indicates that just the presence of pairwise negative epistasis is enough to lead to an elevated recombination rate in haplodiploids compared to diploids, all else being equal. Similar results are seen when there is pairwise negative epistasis or truncation selection only in males, except that the recombination rates in diploid one locus case do not increase to high values in this case. Therefore, the difference between diploids and haplodiploids is greater when the epistasis is male-specific.

4.1 Sexual Dimorphism in Recombination Rates in Diploids

In the presence of male-specific negative epistasis (both truncation selection and negative pairwise epistasis), females were observed to evolve a much greater recombination rate than males. Our result therefore predicts that recombination rates should be lower in the sex under stronger epistatic selection if they can vary independently between the sexes, and are therefore consistent with the findings of a recent meta-analysis (Sardell and Kirkpatrick (2020)), which showed that crossovers in males are more concentrated near the telomeres, while those in females are more concentrated near the centromeres, which might result in an effectively higher recombination rate in females. However, our results are at odds with the theoretical model by (Lenormand (2003)) which predicted that any differences in selection between males and females in the diploid stage should not lead to sex dimorphism in recombination rates.

4.2 Differences in Recombination Rate Between Meioses Producing Male and Female Offspring Respectively in Haplodiploids

If the female could have different recombination rates depending on whether she is producing male (i.e., gametes which will not be fertilised) or female (i.e., gametes which are going to be fertilised) offspring, our results (both analytical and computational) show that the recombination rate will be much higher during the production of the first type of gametes. The likely explanation for this is the fact that the large reproductive skew in male reproductive fitness makes it necessary to have a large recombination rate so that there is at least some likelihood of producing a high fitness male. The analytical model indicates that in this case there are two benefits of recombination: in male offspring, as well as in the female offspring of the next generation (since the genotype stays almost unchanged between these two generations). Our result is of practical significance because empirical estimates of recombination rates in haplodiploids are almost exclusively inferred from the male offspring (since it is easier to estimate recombination rate in the female parent when the offspring is haploid). If the females can actually control the recombination rate depending on the sex of the future offspring, our results suggest that such estimates of recombination rates in haplodiploids are likely to be overestimates of the actual rates due to them being inferred only from the male offspring.

Therefore, the high recombination rates seen in haplodiploid social insects can be explained if there is sex-specific negative epistasis in these organisms. This could be due to increased sexual selection on males in these organisms, as is suggested by studies showing high levels of polyandry and male-biased sex ratio in many social Hymenopterans (Crozier and Fjerdingstad (2001); Everitt *et al.* (2024)). Male-biased sex ratio which often results in increased sexual selection in males have been observed in many eusocial hymenopterans (Baer (2005)), especially members of the genus Apis (which also have one of the highest recombination rates among all metazoans). A recent study of recombination rates in the diploid social termites (Everitt *et al.* (2024)) found much lower recombination rates than in eusocial Hymenopterans, indicating that eusociality by itself might not be responsible for evolution of high recombination rates. It is notable that eusocial termites also tend to

have roughly even sex ratios (Everitt *et al.* (2024)).

On the basis of the results it could be hypothesised that there should be a positive correlation between the intensity of sexual selection and the rate of recombination. It would be interesting to study the relation between recombination densities and proxies of sexual selection (such as operational sex ratio or sexual dimorphism) through phylogenetically controlled meta-analyses. Studies on hymenopterans that study both recombination rates and sexual selection will also be interesting to unravel the link between the two, and will test the hypothesis that sexual selection is the factor responsible for high recombination rates in eusocial hymenopterans. It would also be useful to conduct laboratory experiments which modulate the intensity of sexual selection to study the relationship between sexual selection and recombination rates. Artificial selection in general tends to result in increased recombination rates (Wilfert *et al.* (2007)) and it will be interesting to see if sexual selection has a similar effect. Since recombination rates effect fundamental evolutionary phenomena like the rate of adaptive evolution, speciation and sex chromosome evolution, understanding the interplay between sexual selection and recombination rates could have many important implications in evolutionary biology. Given the huge diversity in social systems, life histories and recombination rates in hymenopterans, they could serve as useful model systems to study the evolution of recombination rates.

References

Agrawal AF (2006). Evolution of sex: why do organisms shuffle their genotypes? Current Biology 16(17), R696–R704.

Baer B (2005). Sexual selection in apis bees. Apidologie 36(2), 187–200.

Barton NH (1995). A general model for the evolution of recombination. Genetics Research 65(2), 123–144.

Barton NH, Charlesworth B (1998). Why sex and recombination? Science 281(5385), 1986–1990.

Barton NH, Otto SP (2005). Evolution of recombination due to random drift. Genetics 169(4), 2353–2370.

Bhérer C, Campbell C, Auton A (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nature Communications 8.

Bishop DK, Zickler D (2004). Early decision: meiotic crossover interference prior to stable strand exchange and synapsis. Cell 117(1), 9–15.

Brandvain Y, Coop G (2012). Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics 190(2), 709–723.

Brick K, Thibault-Sennett S, Smagulova F, Lam KWG, Pu Y, Pratto F, Camerini-Otero RD, Petukhova GV (2018). Extensive sex differences at the initiation of genetic recombination. Nature 561(7723), 338–342.

Carlton PM, Farruggio AP, Dernburg AF (2005). A link between meiotic prophase progression and crossover control. PLoS Genetics preprint, e12.

Cavalier-Smith T (2002). Origins of the machinery of recombination and sex. Heredity 88, 125–141.

Coop G, Przeworski M (2007). An evolutionary view of human recombination. Nature Reviews Genetics 8(1), 23–34.

Crozier RH, Fjerdingstad EJ (2001). Polyandry in social hymenoptera—disunity in diversity? In Annales Zoologici Fennici, 267–285. JSTOR.

DeLory TJ, Romiguier J, Rueppell O, Kapheim KM (2024). Recombination rate variation in social insects: an adaptive perspective. Annual Review of Genetics 58.

Deshong AJ, Ye A, Lamelza P, Bhalla N (2014). A quality control mechanism coordinates meiotic prophase events to promote crossover assurance. PLoS Genetics 10, e1004291.

Everitt T, Rönneburg T, Elsner D, Olsson A, Liu Y, Larva T, Korb J, Webster MT (2024). Unexpectedly low recombination rates and presence of hotspots in termite genomes. bioRxiv 2024–03.

Felsenstein J (1974). The evolutionary advantage of recombination. Genetics 78(2), 737–756.

Franch R, Louro B, Tsalavouta M, Chatziplis D, Tsigenopoulos CS, Sarropoulou E, Antonello J, Magoulas A, Mylonas CC, Babbucci M, Patarnello T, Power DM, Kotoulas G, Bargelloni L (2006). A genetic linkage map of the hermaphrodite teleost fish sparus aurata l. Genetics 174, 851–861.

Giraut L, Falque M, Drouaud J, Pereira L, Martin O, Mézard C (2011). Genome-wide crossover distribution in arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. PLoS Genetics 7, e1002354.

Haldane J (1922). Sex ration and unisexual sterility in animal hybrids. J Genet 12(2), 101–109.

Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. Genetics Research 8(3), 269–294.

Huxley JS (1928). Sexual difference of linkage in gammarus chevreuxi. Journal of Genetics 20, 145–156.

Johnston SE (2024). Understanding the genetic basis of variation in meiotic recombination: past, present, and future. Molecular Biology and Evolution 41(7), msae112.

Jones DB, Jerry DR, Khatkar MS, Raadsma HW, Zenger KR (2013). A high-density snp genetic linkage map for the silver-lipped pearl oyster, pinctada maxima: a valuable resource for gene localisation and marker-assisted selection. BMC Genomics 14, 810.

Jones JC, Wallberg A, Christmas MJ, Kapheim KM, Webster MT (2019). Extreme differences in recombination rate between the genomes of a solitary and a social bee. Molecular Biology and Evolution 36(10), 2277–2291.

Joseph SB, Kirkpatrick M (2004). Haploid selection in animals. Trends in ecology & evolution 19(11), 592–597.

Kawakami T, Wallberg A, Olsson A, Wintermantel D, de Miranda JR, Allsopp M, Rundlöf M, Webster MT (2019). Substantial heritable variation in recombination rate on multiple scales in honeybees and bumblebees. Genetics 212(4), 1101–1119.

Keightley PD, Otto SP (2006). Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443(7107), 89–92.

Kent CF, Minaei S, Harpur BA, Zayed A (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. Proceedings of the National Academy of Sciences 109(44), 18012–18017.

Kent CF, Zayed A (2013). Evolution of recombination and genome structure in eusocial insects. Communicative & Integrative Biology 6(2), 18012–7.

Kianian PM, Wang M, Simons K, Ghavami F, He Y, Dukowic-Schulze S, Sundararajan A, Sun Q, Pillardy J, Mudge J, *et al.* (2018). High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. Nature Communications 9(1), 2370.

Kidner J, Moritz RF (2015). Host-parasite evolution in male-haploid hosts: an individual based network model. Evolutionary Ecology 29, 93–105.

Kirkpatrick M, Johnson T, Barton N (2002). General models of multilocus evolution. Genetics 161(4), 1727–1750.

Lenormand T (2003). The evolution of sex dimorphism in recombination. Genetics 163(2), 811–822.

Lenormand T, Dutheil J (2005). Recombination difference between sexes: a role for haploid selection. PLoS biology 3(3), e63.

Mank JE (2009). The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. Genetics research 91(5), 355–363.

Newnham L, Jordan P, Rockmill B, Roeder GS, Hoffmann E (2010). The synaptonemal complex protein, zip1, promotes the segregation of nonexchange chromosomes at meiosis i. Proceedings of the National Academy of Sciences 107(2), 781–785.

Oldroyd BP, Rinderer TE, Buco SM, Beaman LD (1993). Genetic variance in honey bees for preferred foraging distance. Animal Behaviour 45(2), 323–332.

Otto SP, Lenormand T (2002). Resolving the paradox of sex and recombination. Nature Reviews Genetics 3(4), 252–261.

Papaioannou I, Dutreux F, Peltier FA, Maekawa H, Delhomme N, Bardhan A, Friedrich A, Schacherer J, Knop M (2021). Sex without crossing over in the yeastsaccharomycodes ludwigii .

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritz KR, Noor MA, Singh ND (2017). Variation in recombination rate: adaptive or not? Trends in Genetics 33(5), 364–374.

Roze D (2021). A simple expression for the strength of selection on recombination generated by interference among mutations. Proceedings of the National Academy of Sciences 118(19), e2022805118.

Rueppell O, Kuster R, Miller K, Fouks B, Rubio Correa S, Collazo J, Phaincharoen M, Tingek S, Koeniger N (2016). A new metazoan recombination rate record and consistently high recombination rates in the honey bee genus apis accompanied by frequent inversions but not translocations. Genome biology and evolution 8(12), 3653–3660.

Samuk K, Manzano-Winkler B, Ritz KR, Noor MA (2020). Natural selection shapes variation in genome-wide recombination rate in drosophila pseudoobscura. Current Biology 30(8), 1517–1528.

Sardell JM, Kirkpatrick M (2020). Sex differences in the recombination landscape. The American Naturalist 195(2), 361–379.

Scott Hawley R, Theurkauf WE (1993). Requiem for distributive segregation: achiasmate segregation in drosophila females. Trends in Genetics 9(9), 310–316.

Sherman PW (1979). Insect chromosome numbers and eusociality. The American Naturalist 113(6), 925–935.

Sherman PW, Seeley TD, Reeve HK (1988). Parasites, pathogens, and polyandry in social hymenoptera. The American Naturalist 131(4), 602–610.

Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences 372(1736), 20160455.

Theodosiou L, McMillan WO, Puebla O (2016). Recombination in the eggs and sperm in a simultaneously hermaphroditic vertebrate. Proceedings of the Royal Society B: Biological Sciences 283, 20161821.

Trivers R (1988). Sex differences in rates of recombination and sexual selection. The evolution of sex 270, 286.

Wilfert L, Gadau J, Schmid-Hempel P (2007). Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98(4), 189–197.

Appendix A

Appendix

A.1 Appendix 1.1: Analytical Model

The

Selection

The fitness of an individual with genotype X relative to the mean fitness in the population is given by (Kirkpatrick *et al.* (2002)):

$$W(X) = 1 + \sum_{A \subseteq G} a_A(\zeta_A - D_A) \tag{A.1}$$

Here G refers to the set of all possible loci under direct selection on the two chromosomes of an individual. In our case, it refers to the following set $\{JK, JK\}$. Note here J and K refer to the locus and not the allele. The selection coefficients a_A defined here tend to be complicated functions of fitness effects of mutations, epistasis as well as the allele frequencies of mutations.

$$\begin{aligned} W(X) = 1 &+ a_{j,\emptyset}(\zeta_{j,\emptyset} - D_{j,\emptyset}) + a_{k,\emptyset}(\zeta_{k,\emptyset} - D_{k,\emptyset}) + a_{\emptyset,j}(\zeta_{\emptyset,j} - D_{\emptyset,j}) \\ &+ a_{\emptyset,k}(\zeta_{\emptyset,k} - D_{\emptyset,k}) + a_{j,j}(\zeta_{j,j} - D_{j,j}) + a_{k,k}(\zeta_{k,k} - D_{k,k}) \\ &+ a_{j,k}(\zeta_{j,k} - D_{j,k}) + a_{k,j}(\zeta_{k,j} - D_{k,j}) + a_{jk,\emptyset}(\zeta_{jk,\emptyset} - D_{jk,\emptyset}) \\ &+ a_{\emptyset,jk}(\zeta_{\emptyset,jk} - D_{\emptyset,jk}) + a_{jk,k}(\zeta_{jk,k} - D_{jk,k}) + a_{j,jk}(\zeta_{j,jk} - D_{j,jk}) \\ &+ a_{jk,j}(\zeta_{jk,j} - D_{jk,j}) + a_{k,jk}(\zeta_{k,jk} - D_{k,jk}) + a_{jk,jk}(\zeta_{jk,jk} - D_{jk,jk}) \end{aligned} \tag{A.2}$$

To recall, here, $X_{i,\emptyset} = \zeta_{i,\emptyset} + p_{i,\emptyset}$ indicates the allele inherited from father at locus i by an individual and vice versa. Throughout the document, in any set U, V the first element refers to the set of loci inherited from the male parent and the second element refers to the set of loci inherited from the female parent.

To simplify these equations, we use a quasi-linkage equilibrium approximation, assuming that selection is weak enough relative to recombination such that the allele frequencies are changing slowly, while the associations are broken down by segregation and recombination quickly enough to be in a state of quasi-equilibrium. That is, the values of the genetic associations are close to the equilibrium values they would achieve if the allele frequencies were not changing Kirkpatrick *et al.* (2002). In the following we assume that the selection coefficients are weak of the order ϵ, while epistasis factors are

weaker and of the order ϵ^2. This implies that the following coefficients are of order ϵ :$a_{j,\emptyset}, a_{k,\emptyset}, a_{\emptyset,j}, a_{\emptyset,k}, a_{j,j}, a_{k,k}$, while the other coefficients are of the order ϵ^2 or higher.

Since in our model mating is random, $D_{U,V} = D_U D_V$ for any set of positions of loci U and V. The moments after selection were calculated using the formula (Kirkpatrick *et al.* (2002)):

$$D'_{U,V} = D_{U,V} + \sum_{A \subseteq G} a_A (D_{AU,V} - D_A D_{U,V}) \tag{A.3}$$

To simplify the moments with repeated indices, the following reduction formula (Kirkpatrick *et al.* (2002)) was used:

$$D_{jjU,V} = p_j q_j D_{U,V} + (1 - 2p_j) D_{jU,V} \tag{A.4}$$

Therefore, in the diploid one-modifier case, D_{jk} up to the leading order is given by:

$$\begin{aligned}
D'_{jk,\emptyset} &= E'[(X_{j,\emptyset} - p'_{j,\emptyset})(X_{k,\emptyset} - p'_{k,\emptyset})] \\
&= E'[(X_{j,\emptyset} - p_{j,\emptyset} - \delta_{p_j})(X_{k,\emptyset} - p_{k,\emptyset} - \delta_{p_k})] \\
&= E'[(X_{j,\emptyset} - p_{j,\emptyset})(X_{k,\emptyset} - p_{k,\emptyset})] - \delta_{p_j}\delta_{p_k} \\
&= E'[\frac{W}{\overline{W}}\zeta_{j,\emptyset}\zeta_{k,\emptyset}] - \delta_{p_j}\delta_{p_k} \\
&= E'[\frac{W}{\overline{W}}\zeta_{j,\emptyset}\zeta_{k,\emptyset}] - \delta_{p_j}\delta_{p_k} \\
D'_{jk,\emptyset} &= D_{jk,\emptyset} + a_{jk,\emptyset} D_{jjkk,\emptyset} - \delta_{p_j}\delta_{p_k} \\
&= D_{jk,\emptyset} + a_{jk,\emptyset} p_j q_j p_k q_k - \delta_{p_j}\delta_{p_k} \\
&= D_{jk,\emptyset} + a_{jk,\emptyset} p_j q_j p_k q_k - a_{j,\emptyset} a_{k,\emptyset} p_j q_j p_k q_k
\end{aligned} \tag{A.5}$$

Here, δ_{p_j} and δ_{p_k} are the changes in allele frequencies at the J and K loci due to selection respectively:

$$\begin{aligned}
\delta_{p_j} &= E'[\frac{W}{\overline{W}}(X_{j,\emptyset} - p_{j,\emptyset})] \\
&= E'[\frac{W}{\overline{W}}(\zeta_{j,\emptyset})] \\
&= a_{j,\emptyset} D_{jj,\emptyset} \\
&= a_{j,\emptyset} p_j q_j
\end{aligned} \tag{A.6}$$

Similarly,

$$\delta_{p_k} = a_{k,\emptyset} p_k q_k \tag{A.7}$$

If we define the fitnesses of the different genotypes in the following way:

$$\begin{aligned}
&W(X_{j,\emptyset} = 1, X_{\emptyset,j} = 0, X_{k,\emptyset} = 0, X_{\emptyset,k} = 0) = (1 - sh) \\
&W(X_{j,\emptyset} = 1, X_{\emptyset,j} = 1, X_{k,\emptyset} = 0, X_{\emptyset,k} = 0) = (1 - s) \\
&W(X_{j,\emptyset} = 1, X_{\emptyset,j} = 0, X_{k,\emptyset} = 1, X_{\emptyset,k} = 0) = (1 - sh)^2 + e \\
&W(X_{j,\emptyset} = 1, X_{\emptyset,j} = 1, X_{k,\emptyset} = 1, X_{\emptyset,k} = 0) = (1 - s)(1 - sh) + 2e \\
&W(X_{j,\emptyset} = 1, X_{\emptyset,j} = 1, X_{k,\emptyset} = 1, X_{\emptyset,k} = 1) = (1 - s)^2 + 4e
\end{aligned} \tag{A.8}$$

and so on, then the difference $a_{jk,\emptyset} - a_{j,\emptyset} a_{k,\emptyset}$ upto the leading order is equal to the epistasis factor e. The equation for $D'_{jk,\emptyset}$ now becomes:

$$\begin{aligned}
D'_{jk,\emptyset} &= D_{jk,\emptyset} + a_{jk,\emptyset} p_j q_j p_k q_k - a_{j,\emptyset} a_{k,\emptyset} p_j q_j p_k q_k \\
&= D_{jk,\emptyset} + e p_j q_j p_k q_k
\end{aligned} \tag{A.9}$$

Recombination

The moments after recombination depend on the recombination rate between the different loci. Therefore, in the diploid one-modifier case, D^r_{jk} up to the leading order is given by:

$$\begin{aligned} D^r_{jk,\emptyset} &= E'\left[(1-r_{jk})\frac{(\zeta_{jk,\emptyset}+\zeta_{\emptyset,jk})}{2}+r_{jk}\frac{(\zeta_{j,k}+\zeta_{k,j})}{2}\right] \\ &= \left[(1-r_{jk})\frac{(D'_{jk,\emptyset}+D'_{\emptyset,jk})}{2}+r_{jk}\frac{(D'_{j,k}+D'_{k,j})}{2}\right] \\ &= \left[(1-r_{jk})\frac{(D_{jk,\emptyset}+ep_jq_jp_kq_k+D_{\emptyset,jk}+ep_jq_jp_kq_k)}{2}\right. \\ &\left.+\,r_{jk}\frac{(D_{j,k}+ep_jq_jp_kq_k+D_{k,j}+ep_jq_jp_kq_k)}{2}\right] \\ &= \left[(1-r_{jk})\frac{(2D_{jk,\emptyset}+2ep_jq_jp_kq_k)}{2}+r_{jk}\frac{(2ep_jq_jp_kq_k)}{2}\right] \end{aligned} \tag{A.10}$$

When the genetic associations are in a state of quasi-linkage equilibrium, i.e, $D^r_{jk,\emptyset} \approx D_{jk,\emptyset}$, the equation becomes:

$$D^r_{jk,\emptyset} = \frac{ep_jq_jp_kq_k}{r_{jk}} \tag{A.11}$$

Similarly, the expression for $D_{ijk,\emptyset}$ is given by:

$$\begin{aligned} D^r_{ijk,\emptyset} = E'\Bigg[&(1-r_{ij})(1-r_{jk})\frac{(\zeta_{ijk,\emptyset}+\zeta_{\emptyset,ijk})}{2}+(r_{ij})(1-r_{jk})\frac{(\zeta_{i,jk}+\zeta_{jk,i})}{2}+ \\ &(1-r_{ij})(r_{jk})\frac{(\zeta_{ij,k}+\zeta_{k,ij})}{2}+(r_{ij})(r_{jk})\frac{(\zeta_{ik,j}+\zeta_{j,ik})}{2}\Bigg] \end{aligned} \tag{A.12}$$

Here, r_{jk} is given by:

$$r_{jk} = \overline{r_{jk}} + \delta_r\frac{(\zeta_{i,\emptyset}+\zeta_{\emptyset,i})}{2} \tag{A.13}$$

We assume that δ_r is very small and therefore we neglect all the terms which have the order δ_r^2 or higher. The values of $D_{ij,\emptyset}$ and $D_{ik,\emptyset}$ are calculated similarly. Finally, the equation for change in allele frequency of the modifier allele Δ_i is obtained as:

$$\Delta p_i = a_j\frac{(D_{ij,\emptyset}+D_{\emptyset,ij})}{2}+a_k\frac{(D_{ik,\emptyset}+D_{\emptyset,ik})}{2}+a_{jk}\frac{(D_{ijk,\emptyset}+D_{\emptyset,ijk})}{2} \tag{A.14}$$

Here, $a_j = a_{j,\emptyset} = a_{\emptyset,j}$ and so on for a_k and a_{jk}, assuming the fitness effects of mutations are the same in both the sexes.

A.2 Appendix 1.2: Heterochiasmy Due to Sex Differences in Mutation Rate

When the mutation rates differed between the two sexes, it resulted in a slight sexual dimorphism in recombination rates even in the absence of sex differences in fitness effects

of mutations. The sex with higher mutation rate ended up evolving a higher recombination rate, with this difference being larger when the differential fitness effects of deleterious mutations was lower in this sex than the other and vice versa. The reasons for this puzzling obervation is explained by: there are transient sweeps in the population of mutation-free regions tightly linked with the recombination modifier locus coding for low recombination. Due to the low recombination, such regions do not get broken up during crossingover and tend to persist and proliferate in the population until they accumulate mutations. The proliferation of such regions leads to the proliferation of low recombination coding allele by hitchhiking and results in a drop in the average recombination rate in the population. Such regions are more likely to emerge in the sex with lesser mutation rate, and the selective advantage of such mutation free regions are greater when the fitness effects of such deleterious mutations are stronger. This hypothesis was supported by the observation that such differences in recombination rate disappeared when 1) the recombination rate was constrained to not fall below a minimum value, and 2) when the mutation rate was very high in both the sexes.

A.3 Appendix 1.3: Results for Truncation Selection

At very high levels of truncation selection, the amount of recombination becomes almost zero. This is simply because at such high levels of selection the mutation load becomes negligible (less than one mutation per chromosome) so the benefits of recombination disappear. The relevant figures for the map lengths when there is truncation selection in males are:

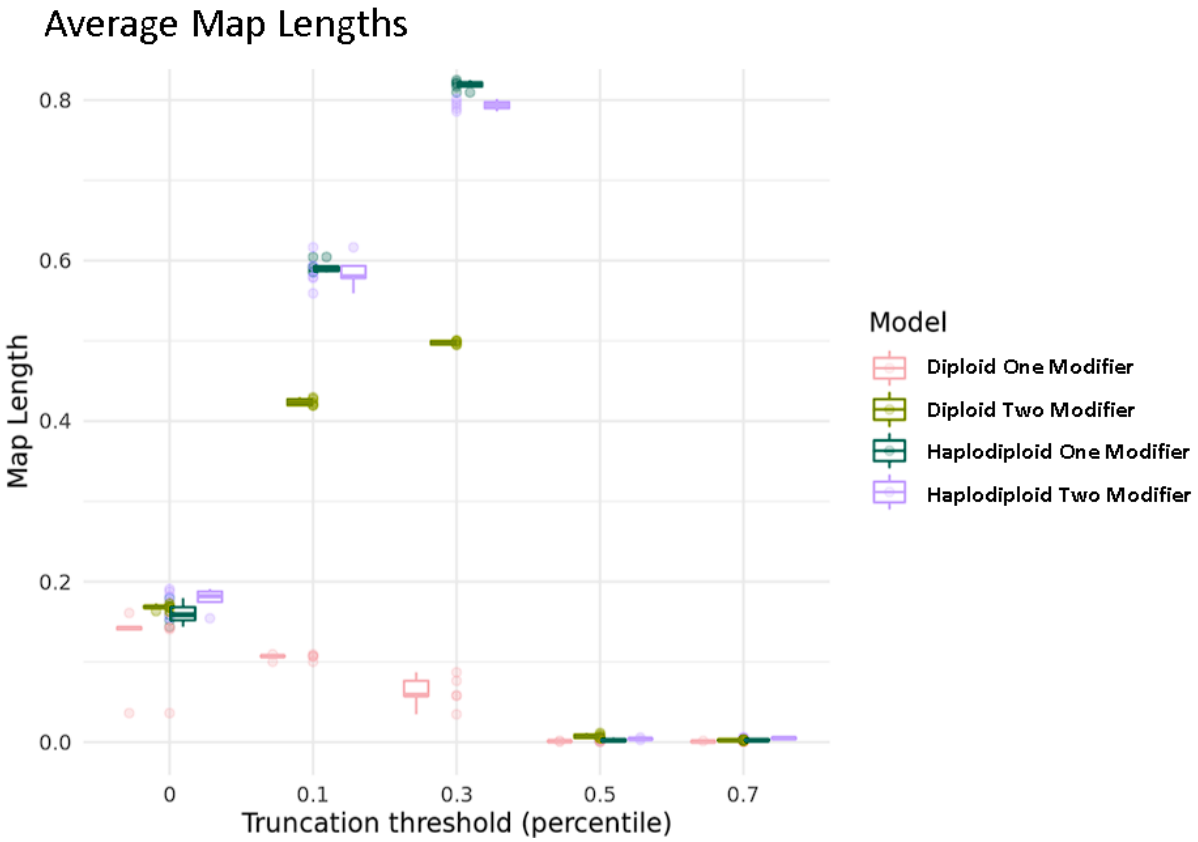


Figure A.1: **The graph compares the average map lengths for the four cases with truncation selection in males.**

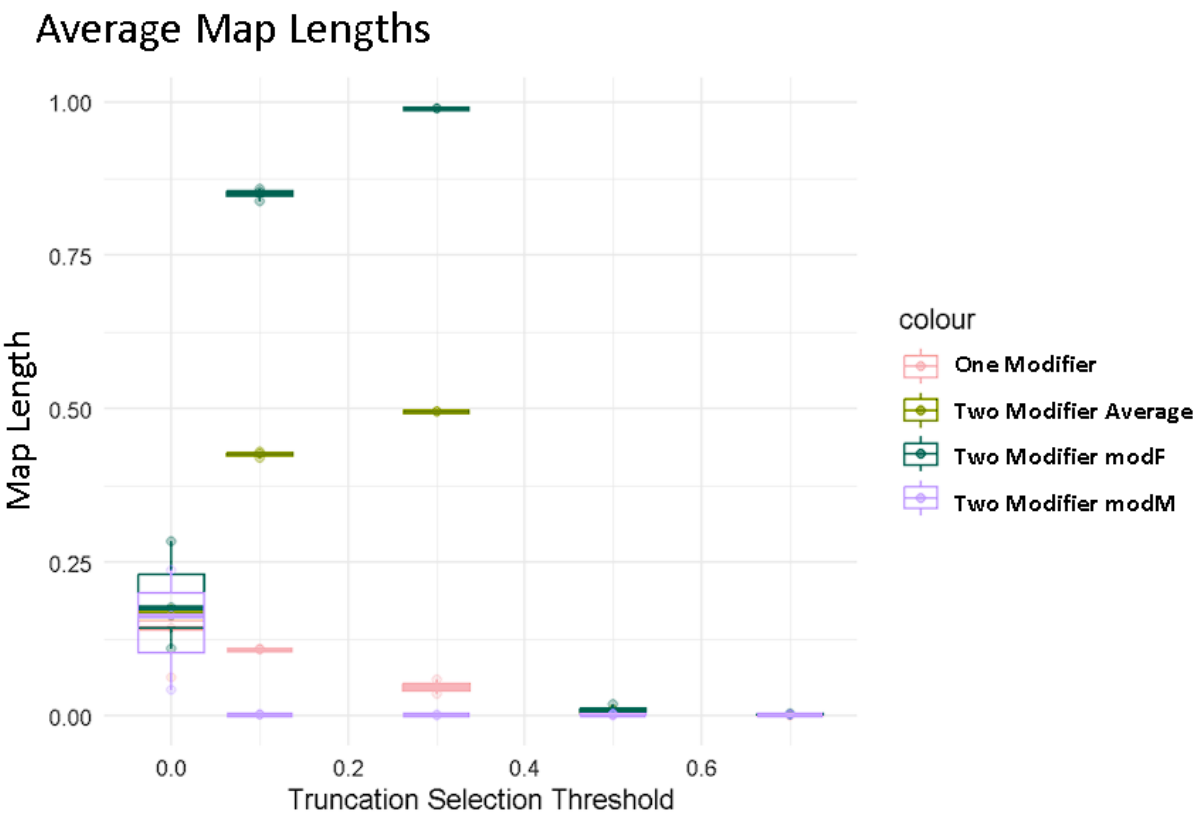


Figure A.2: **The graph compares the average and sex-specific map lengths for the two diploid cases with truncation selection in males.**

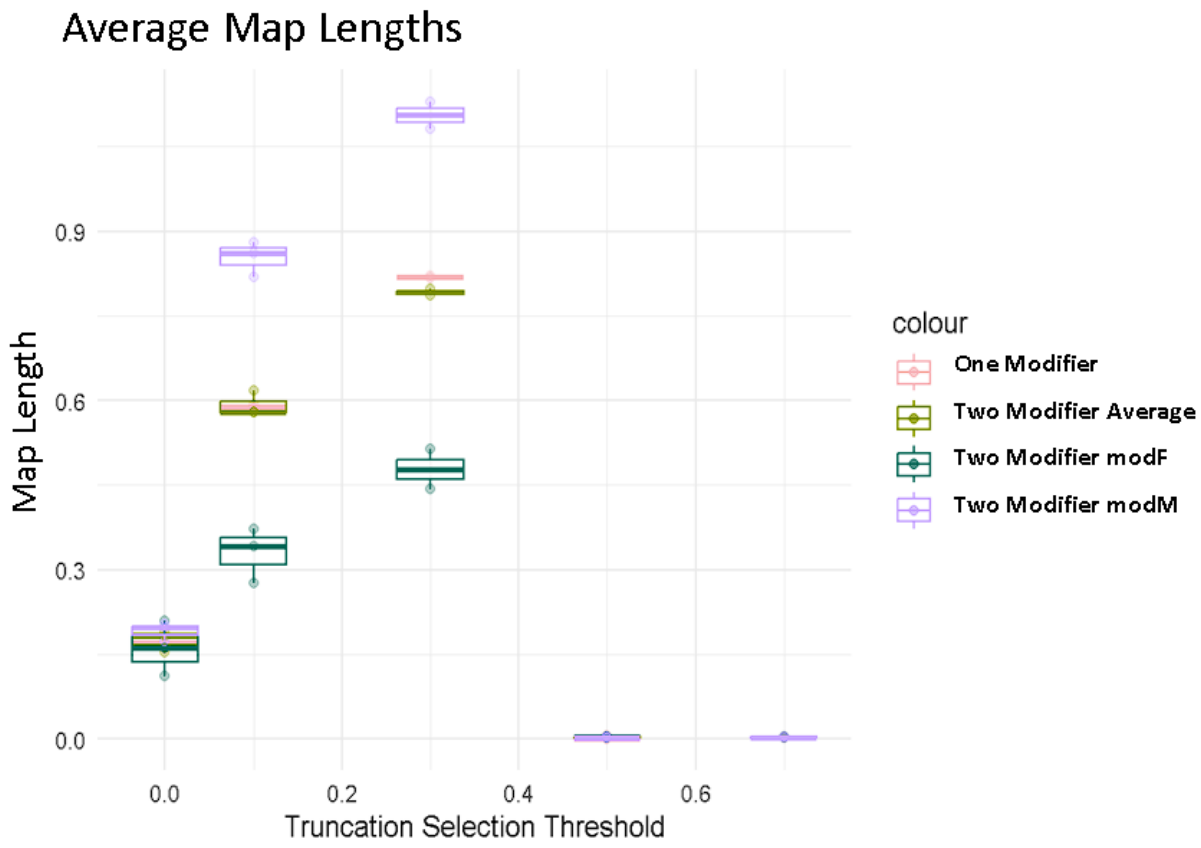


Figure A.3: **The graph compares the average map lengths for the two haplodiploid cases with truncation selection in males. modM and modF refer to recombination rates (more specifically, average map lengths) during the production of male and female gametes, respectively.**

A.4 Appendix 1.4: Sex-Limited Recombination in Diploids

When recombination in diploids was restricted to one sex, the other sex evolved high recombination rates (in our case, the two sexes are males and females respectively).

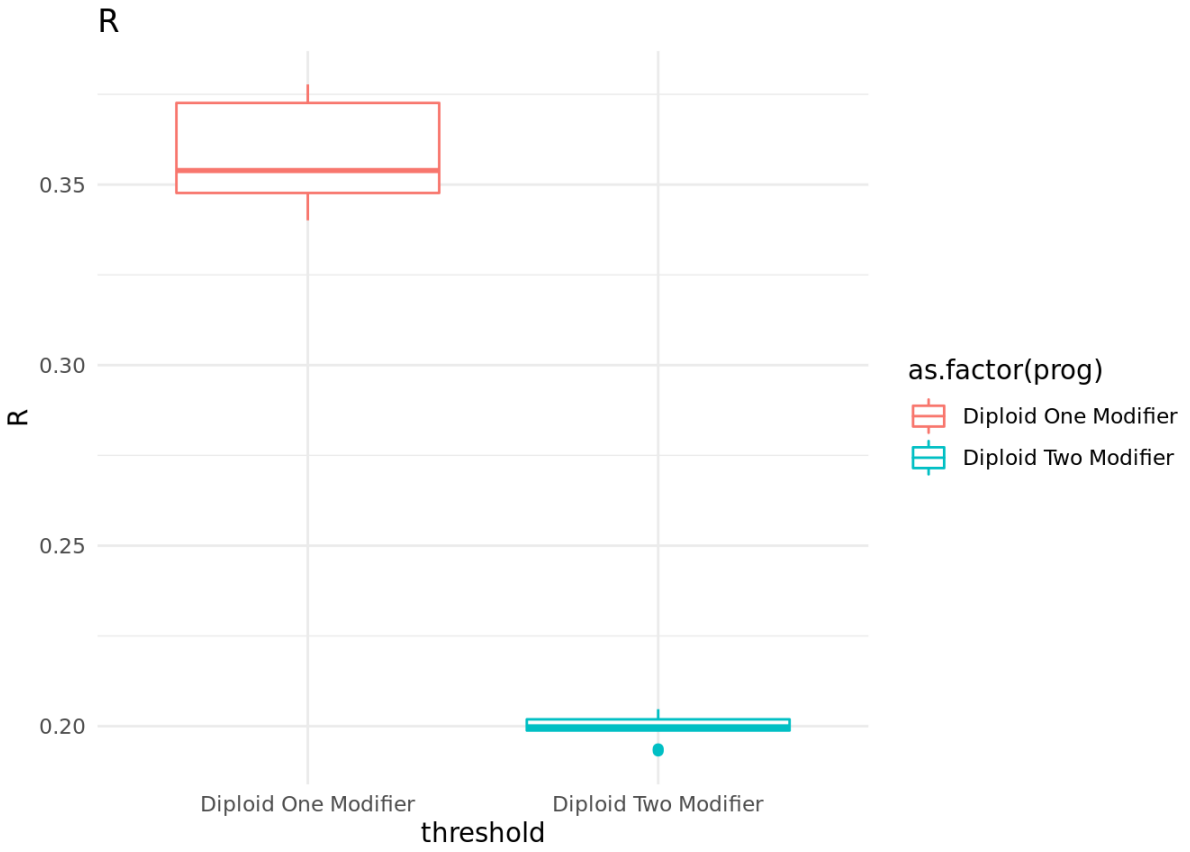


Figure A.4: **The graph compares the average map lengths for the diploid two modifier case with the case when recombination is limited to one sex.** As can be clearly seen, absence of recombination in one sex in diploids results in an increased recombination rate in the recombining sex.