The Evolutionary Interaction of Intraspecific Competition and Genetic Recombination

Sol Noah Ackerman

A Thesis
In the Department
of
Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Biology) at
Concordia University
Montreal, Quebec, Canada

December 2011

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Sol Ackerman

Entitled: The Evolutionary Interaction of Intraspecific Competition and Genetic Recombination

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Judith Kornblatt Chair

Dr. Sabin Lessard External Examiner

Dr. Lea Popovic External to Program

Dr. James Grant Examiner

Dr. William Zerges Examiner

Dr Donal Hickey Thesis Supervisor

Approved by Dr. Selvadurai Dayanandan
Chair of Department or Graduate Program Director

Dr. Brian Lewis
Dean of Faculty

ABSTRACT

The Evolutionary Interaction of Intraspecific Competition and Genetic Recombination

Sol Ackerman, Ph.D.
Concordia University, 2011

The reason for the prevalence of sexual reproduction among multicellular eukaryotes is a long-standing unanswered question in evolutionary biology. It is widely believed that sexual reproduction and the resulting genetic recombination provide a selective advantage by increasing a population's genotypic variance. However, recombination will only do so when the population is in negative linkage disequilibrium. It has been proposed that if alleles do not contribute multiplicatively to fitness, but instead display a negative epistatic fitness curve, then the population will be in such a state following selection, and sex will be advantageous. However, there is no *a priori* reason to believe that fitness values should generally be negatively epistatic, as opposed to positive or zero.

In this study, we explore the relationship between contest competition and the maintenance of sexual reproduction. In Chapter 2, we develop a two-locus bi-allelic haploid model to examine the relationship between phenotype, competitive selection, and realized fitness. We assume that competitive ability is directly proportional to the phenotypic value and that the outcome of pairwise competition is dependent on the ratio of competitive abilities of the competing individuals. The stronger competitor does not always win, but it wins more frequently than the weaker competitor does. Using this very simple model, we find that intraspecies competition can lead to frequency dependent changes in genotypic fitness. In addition,

competition can result in negative epistasis at the level of realized fitness. This leads to the build-up of negative linkage disequilibrium among the alleles that affect fitness following selection.

In Chapter 3, we explore the effects of competition when the selection is repeated over several generations. We find that the amount of negative linkage disequilibrium builds up continuously in the absence of recombination, and this hinders the progress of selection. This accumulation of negative linkage disequilibrium is alleviated when recombination is present. We show, using both a numerical modeling approach and through individual-based simulations, that sexual individuals will increase in frequency in a mixed population of sexual and asexual individuals. The selective advantage of recombination is strongest when there is a large difference in competitive ability between genotypes, and when the selected alleles are initially rare.

Finally in Chapter 4, we consider a different mapping of phenotype onto fitness to ensure that our findings are robust. In this case, the outcome of pairwise competition is not determined by the ratio of competitive abilities; instead the stronger competitor always wins. In this case, as expected, there is even stronger negative epistasis at the fitness level. As a result, we find that advantageous alleles go to fixation at a faster rate when recombining, and sexual individuals increase in frequency when competing against asexual individuals in a mixed population.

Our results indicate that competition can play a significant role in the maintenance of sexual reproduction, and that the advantage of sexual reproduction may not lie in fertility selection, but in viability selection. In a competitive situation, it is better to produce one winner rather than two losers.

ACKNOWLEDGEMENTS

To Dr. Hickey, whose wisdom goes far beyond his research. And to my family, my friends, and to Rachel, whose support and encouragement helped me through.

CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 4

CHAPTER 5

LIST OF SYMBOLS AND ABBREVIATIONS

p_1	The frequency of the *A* allele in the population
p_2	The frequency of the *B* allele in the population
q_1	The frequency of the *a* allele in the population
q_2	The frequency of the *b* allele in the population
x	The effect of allele *A* on the phenotype
y	The effect of allele *B* on the phenotype
$\in$	The deviation from multiplicative phenotypic effects on the *AB* genotype
V_{ab}	The phenotypic value of the of the *ab* genotype
$Freq(ab)$	The frequency of the *ab* genotype in the entire population
$Freq(s)$	The frequency of the sexual reproducers in the entire population
$Freq(ab, a)$	The frequency of asexual *ab* genotyped reproducers in the entire population
$Freq(ab, s)$	The frequency of sexual *ab* genotyped reproducers in the entire population
$Freq(ab)'$	The frequency of the *ab* genotype after selection
$Freq(ab)''$	The frequency of the *ab* genotype after recombination
w_{ab}	The fitness of the *ab* genotype as calculated by $\frac{Freq(ab)'}{Freq(ab)}$
ε	The value of fitness level epistasis as calculated by $\ln \frac{w_{ab}w_{AB}}{w_{Ab}w_{aB}}$
D	The value of linkage disequilibrium as calculated by $Freq(ab)Freq(AB) - Freq(Ab)Freq(aB)$

Z	The value of linkage disequilibrium as calculated by $\ln \frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)}$
c	The cull rate within the population

CHAPTER 1

Introduction and Literature Review

Genetic recombination is of fundamental importance in Biology and it has been extensively studied at all levels of biological organization, from the molecular through the cytological, individual, population and evolutionary levels. These studies have spanned the past hundred years and significant amounts of new information on recombination continue to be gathered. This research has deepened our understanding of both the mechanisms of recombination and its biological effects. Significant questions remain, however, regarding the evolutionary origin and the maintenance of genetic recombination.

1.1 Molecular mechanisms of recombination

The intricacies of genetic recombination at the molecular level are still being elucidated, but the general mechanism is well understood. Holliday (1964) proposed that initially, single strand breaks occur at adjacent locations from opposite homologous DNA molecules. The DNA then unravels to form single strands, the strands cross-over, and each strand anneals to the complimentary strand of the other molecule. The resulting X-shaped structure is now referred to as a Holliday junction (see Figure 1.1). The junction is resolved by the single stranded breakage and rejoining of complementary DNA strands. Depending on which two strands are broken, this may, or may not, result in recombinant DNA molecules. If the cleaved strands are also the strands that initially crossed over, then there will be no recombination of outside markers; but

there will be recombination if the other two strands are cleaved instead. In both cases, the expected products should contain heteroduplex DNA (i.e. where complementary strands from different parental molecules are base paired) in the regions where the crossover strands have annealed. Normal base pairing will occur in the heteroduplex region if the DNA strands are homozygous; but if this region is heterozygous, there will be a mispairing of bases. These mismatches are detectable by DNA repair mechanisms, and can result in gene conversion: where one allele of a pair is replaced by a copy of the other allele.

Meselson and Radding (1975) contributed a modification to this mechanism in an effort to explain why the heteroduplex DNA region is not always present on both molecules, which they called asymmetric heteroduplex DNA. They proposed that a recombination event could begin with a single strand break on only one of the two recombining DNA molecules. One of the ends of the single strand can then be displaced, and cross over to pair with a complementary sequence in another molecule; a process known as strand invasion. This causes a single strand break in the displaced strand, followed by ligation to the invading strand on one side of the break, and exonuclease activity on the other side of the break (see Figure 1.2). DNA polymerase action would then begin at the original single strand break, and continue to the other homologous molecule where the nucleotides were excised. The result is a Holliday junction with asymmetric heteroduplex DNA on only the one of the two interacting molecules. If the Holliday junction were to change locations by "zipping" up or down the DNA molecules, then both molecules would have heteroduplex regions.

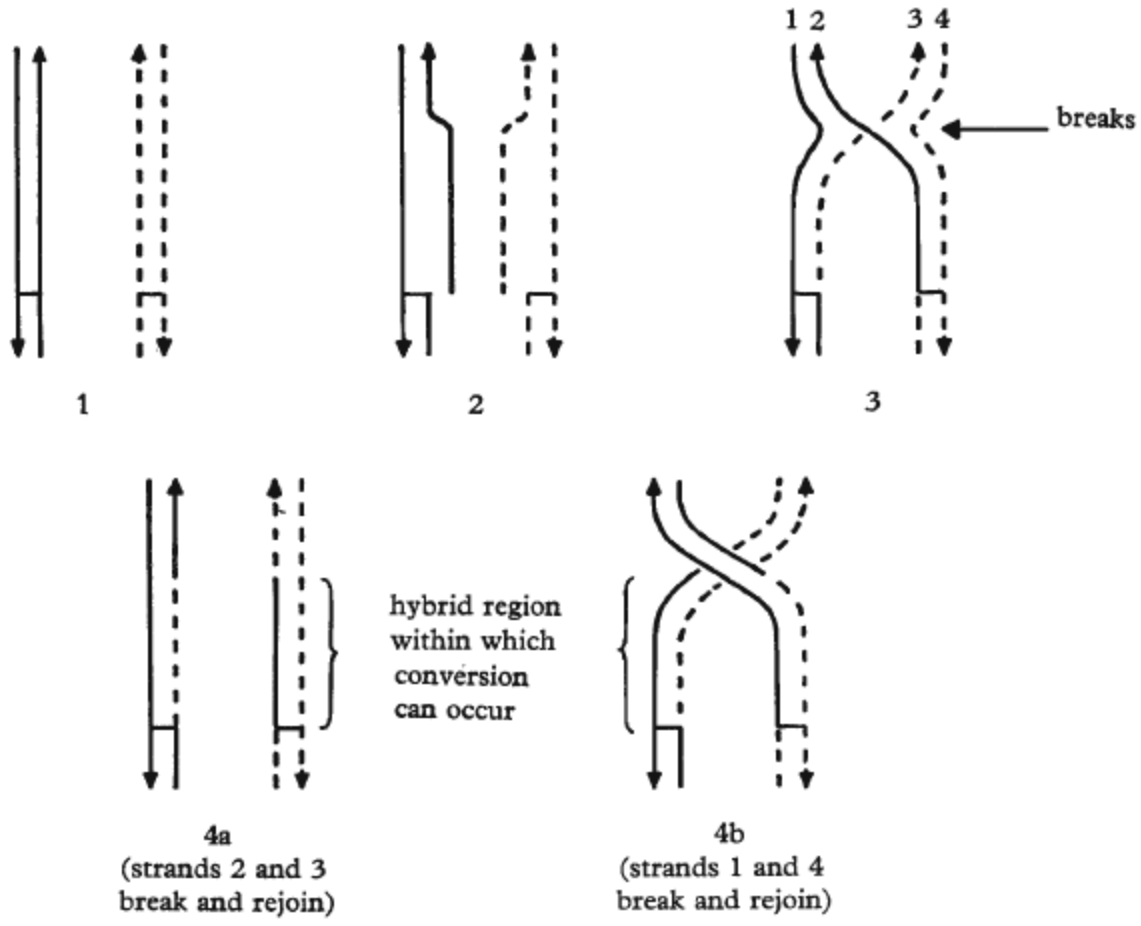


Figure 1.1: The Holliday model of genetic recombination (image taken from Holliday, 1964).

Vertical lines represent DNA strands, with arrows to depict their anti-parallel nature. One strand is depicted with solid lines, and the other with dashed lines as a way to differentiate between parental strands. Short horizontal lines show the positions single strand breaks. 1) Single stranded breaks are created at similar loci on DNA molecules from opposite homologous strands. 2) The DNA molecules then unravel to form single strands. 3) The single strands then cross-over, and anneal to the complimentary strand of the other DNA molecule. The resulting X-shaped structure is resolved by the single stranded breakage and rejoining of complementary DNA strands. Depending on which two strands are broken, 4a) this can result in no recombination, 4b) or this can result in recombinant DNA molecules.

Experimental evidence later showed that the initiating strand is likely also to undergo a genetic conversion. Thus, the double-strand-break-repair model (Szostak et. al, 1983) was proposed as a way of explaining this observation. In this model, recombination is initiated by an endonuclease that creates a double-strand-break on one of the DNA molecules. Exonucleases then cut away to form a gap, producing 3' single stranded termini in the process. One of the 3' ends then invades a homologous region of the other DNA molecule, displacing the paired strand and producing a "D loop" (see Figure 1.3). DNA polymerization occurs along the invading strand, which enlarges the D loop. The displaced strand then anneals to the other 3' single strand of the endonuclease digested molecule, and acts as a template for DNA repair. In this way, the initiating strand becomes the target of a gene conversion. Furthermore, this explanation allows for the existence of asymmetric heteroduplex DNA on both strands: in the region of strand invasion on the invaded molecule, and in the region where the displaced strand anneals to the initiating molecule. DNA repair is completed by ligating the polymerization chains to the 5' ends of the initiating strands, resulting in a double Holliday junction. Symmetric heteroduplex DNA can be created if the Holliday junctions zip away from the site of the double strand break. Each Holliday junction is resolved as before, resulting in either crossover or non-crossover products.

In is now known that recombination is often initiated by the meiosis-specific endonuclease Spo11 (Cao et. al., 1990) and its orthologs. Furthermore, there is some evidence to suggest that the endonuclease is location specific, and the initiating strand's sequence is often replaced by the homologous sequence during gene conversion (Nicolas et. al., 1989). Recent studies in humans and other species (Jeffreys et. al., 2001; Winkler et. al. 2005) have inferred that recombination rates are not uniform across the genome, but are clustered into regions

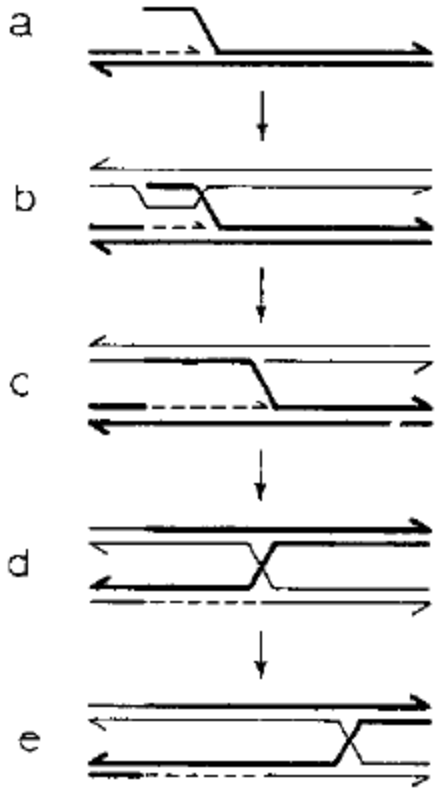


Figure 1.2: The Meselson-Radding model of recombination (image taken from Szostak et. al., 1983).

Horizontal lines represent DNA strands, with arrows to depict their anti-parallel nature. One strand is bold, and the other is normal as a way to differentiate between parental strands. DNA replication fragments are depicted with dashed lines. a) A single strand is nicked on one of the two involved DNA molecules. One of the ends of the single strand can then be displaced, and polymerase action then begins to repair the original single strand break. b) The displaced strand crosses over to pair with a complementary sequence on another DNA molecule. c) This causes a single strand break on the displaced strand, followed by ligation to the invading strand on one side of the break, and exonuclease activity on the other side of the break. d) Polymerase action continues to the other DNA strand where the nucleotides were excised, resulting in a Holliday junction with heteroduplex DNA only on the top molecule (asymmetric). e) If the Holliday junction zips away from the breakage site, there will be heteroduplex DNA on both molecules (symmetric).

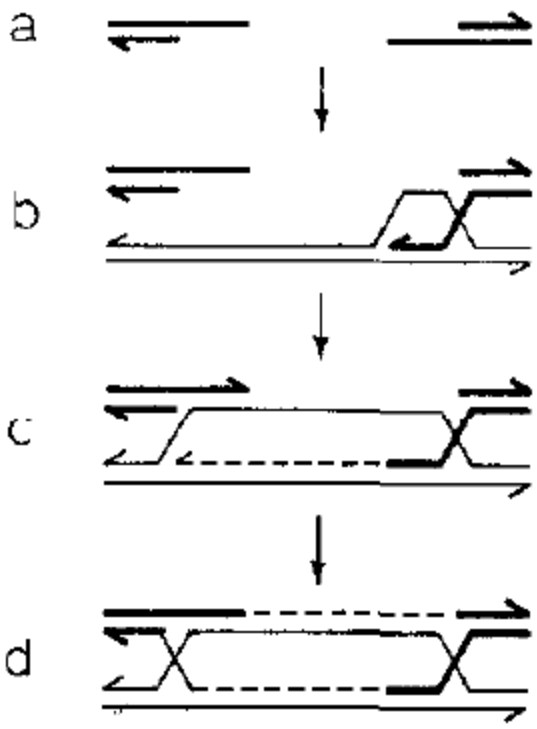


Figure 1.3: The double strand break model of recombination (image taken from Szostak et. al., 1983)

For a description, see Figure 1.2. a) Recombination is initiated with a double-strand-break on one of the DNA molecules. Exonucleases then cut away to form a gap, producing 3' single stranded termini in the process. b) One of the 3' ends invades a homologous region of another strand, displacing the paired strand and producing a D loop. c) DNA polymerization occurs along the invading strand, which enlarges the D loop. The displaced strand anneals to the other 3' single strand of the endonuclease digested molecule, and acts as a template for DNA repair. d) DNA repair is completed by ligating the polymerization chains to the 5' ends of the initiating strands, resulting in a double Holliday junction.

of under 2kb in length, known as recombination hotspots. These observations have led some to believe that recombination events are sequence specific, but due to biased gene conversion on the initiating strand, the hotspot initiator sequence is likely to be replaced by its "coldspot" homologue. Genome sequence analysis has found evidence of a possible sequence motif in humans (Myers et. al., 2005; Myers et. al., 2008), but direct evidence of a region from human sperm samples seems to contradict this hypothesis (Neumann and Jeffreys, 2006). While human and chimpanzee genomic sequences show a 99% sequence identity, their hotspot regions are not conserved (Winkler et. al., 2005). Furthermore, it has been shown that if there exists a recombination initiating sequence, then gene conversion will remove the hotspot sequences from the population very quickly (Boulton et. al., 1997; Pineda-Krch and Redfield 2005). This brings forward a paradox as to how recombination rates are maintained over an evolutionary time frame, given that recombination selects against itself. To make a long story short, the sequence specificity of a recombination site is still being hotly debated.

1.2 Recombination at the cytological level

Early on in the 20th century, a parallel was made between the process of chromosomal segregation in meiosis, and Mendel's theory of allelic segregation (Sutton, 1902). This spawned the chromosomal theory of inheritance, which is the theory that chromosomes are the carriers of genetic material, and are the basis of genetic inheritance. Morgan noticed that while some traits show inheritance patterns of Mendelian segregation, others show a pattern of being coupled. He concluded that some traits must be linked, with their genes located on chromosomes, and that chiasma formation allows genes on the same chromosome to be uncoupled (Morgan, 1911). For

his studies, Morgan calculated the fraction of meiotic products with non-parental allelic combinations at two loci, now referred to as the recombinant fraction.

Soon after, Sturtevant (1913) proposed that with genes located on linear chromosomes, the proportion of crossovers could be used as an index of the distance between any two genes. Sturtevant used the recombination frequency as a unit of genetic distance, which is the frequency that a single chromosomal crossover will take place between two genes during meiosis. He named this unit of measurement the Morgan in honour of his supervisor. In his experiment, Sturtevant showed that with three genes - *A*, *B* and *C* - located close to each other on the same chromosome, the recombinant fraction between *A* and *C* is equal to the recombinant fraction between *A* and *B*, plus the recombinant fraction between *B* and *C*. This demonstration of the linear arrangement of genes was the first genetic map. This proposal of a linear arrangement of genes met with some resistance. A genetic map length greater than 0.5 M (50 cM) is quite possible, whereas a recombinant fraction greater than 0.5 is never observed. The reason for this is that double crossovers on the same chromatid negate each other, reducing the recombinant fraction (Sturtevant et. al., 1919). In short, only an odd number of crossover events between allelic markers on a chromatid contribute to the recombinant fraction. Because of the disparity between the recombinant fraction and the inferred recombination frequency at larger distances, it is more common to refer to mapping distances in centiMorgans as opposed to Morgans.

Excluding crossovers between sister strands, there are three types of possible double crossovers. These are designated two strand, three strand, and four strand (Beadle and Emerson, 1935). Two strand double exchanges involve the same two strands at both crossover positions. Three strand double crossovers have three different strands involved in crossing-over, with one strand crossing over twice; and four strand doubles involve two strands at one crossover position,

and the two remaining strands at the other position. In diploids, these types are expected to occur in the ratio 1:2:1, which results in a recombinant fraction of 0.5 for outside markers (Beadle and Emerson, 1935).

As it turns out, the positions of chiasmata are not independent of each other. Haldane (1931) made the point that if the probability of crossing-over was unaffected by events elsewhere, then the number of crossovers on a chromosome should follow a Poisson distribution. They do not. His study showed that there are far fewer multiple crossovers than would be expected if their locations were randomly distributed, and that the variance in the number of crossovers is less than the mean. He concluded that the occurrence of a crossover reduces the likelihood of another one in its vicinity. This phenomenon is known as genetic interference.

For almost two decades, a genetic map was the only way to determine the relative position of a gene. Its physical location along a chromosome was still unknown. Creighton and McClintock (1931) were the first to show evidence of the physical position of a gene along a chromosome. By looking at the physical characteristics of a chromosome —such as its length after the centromere and the presence of a thumb-like extremity— they were able to show a correlation between the phenotype displayed, and a chromosomal feature. This allowed them to conclude that "pairing chromosomes exchange parts at the same time they exchange genes" (Creighton and McClintock, 1931). The process of constructing a physical map of the chromosome was refined by Painter (1933), when he noticed that each *Drosophila* chromosome has a characteristic pattern of chromatic lines or bands. By studying the changes in a chromosome's banding pattern, one can associate a unique band with a known genetic character. In this way, it is possible to give morphological positions to genetic loci. These techniques are

obsolete nowadays, as whole genome shotgun sequencing can create a physical map of the entire genome down to the base pair (Fleischmann et. al., 1995).

Recombination takes place during the early stages of meiosis, which is the sexually reproductive process of cell division that halves the number of chromosomes in cells. We will not go into great details about the meiotic process, but provide a basic description, with an emphasis on the stages pertinent to the recombinant process.

Prior to the onset of meiosis, the chromosomes have completed replication, but sister chromatids are still attached at the centromere. Meiosis consists of two cellular divisions, with no replication taking place between stages. The first and second meiotic divisions follow a similar sequence of events, but with slightly different results. The result of the first meiotic division is the production of two daughter cells, each with half the ploidy number of the parent cell. Homologous chromosomes are segregated into separate cells, but each chromosome still has two copies of itself, and they remain joined at the centromere. Each daughter cell divides again in the second meiotic division, but this time, sister chromatids are segregated into separate cells. The end result is the production four daughter cells whose chromosomes are single stranded, and whose ploidy number is half that of the parent cell.

Recombination occurs in the first stage of the first cellular division of meiosis, known as prophase I. Prophase I is broken down into 5 sub-stages which are listed chronologically as follows: leptotene, zygotene, pachytene, diplotene, and diakinesis (Hartl and Jones, 2005). During leptotene, chromosomes condense to the point where they become visible under a microscope. In zygotene, the chromosomes laterally pair up with their respective homologues along their entire length (a.k.a synapsis). Each pair of synapsed homologous chromosomes is referred to as a bivalent. In pachytene, chromatids continue to condense to the point where sister

chromatids become distinguishable. It is at this stage that crossing-over occurs. Their strands come into close proximity at a given point, then break and rejoin such that the strands switch tracks at the junction. Contrary to some textbook diagrams, the chromatids do not physically crisscross each other. The junction points are called chiasma (pluralized "chiasmata"). Although chiasmata occur during pachytene, they only become visible at the onset of diplotene,. This is because in diplotene, the synapsed chromosomes begin to separate, except for in the regions with a chiasma. All four chromatids are now visible in each bivalent, as well as their cross-connections. The final period of prophase I is diakinesis, where homologous chromosomes are seen to repel each other, while remaining attached at chiasmata. The first cellular division continues on with the chiasmata separating, and the homologous chromosomes segregating into separate daughter cells. But at this stage, the recombinant aspect of meiosis is complete.

There is a strong correspondence between a species' genetic map and its physical map, which can be explained through the details of meiosis. The genetic map has the same number of linkage groups as the physical map has homologous pairs of chromosomes. The meiotic explanation is that each bivalent makes up its own linkage group. Genes are both physically and genetically linked to each other along a chromosome, with different chromosomes segregating independently during meiosis. Within each homologous pair, loci that are close together are often inherited together, and those far apart might switch homologues. The linear arrangement of genes within a linkage group is the same as their linear arrangement on a chromosome. This can be explained as a recombination event along a chromosome dissociates all the genes to the other side of the break. Therefore, the recombination frequency can only monotonically increase along the length of a chromosome, meaning that the order of loci on both the physical and genetic map is identical. Roughly speaking, genetic distances correlate with physical distances. The meiotic

reason for this is that generally speaking, the greater the physical distance between loci, the greater the likelihood that a recombination event will take place between them.

As both physical mapping and genetic mapping progressed, it became clear that while a monotonic relationship exists among the number of base pairs and centiMorgans between loci, the relationship is not a constant one. In other words, physical distance is not equivalent to genetic distance. The recombination rate varies between species. For instance, the average physical distance represented by 1 cM is approximately 1 Mbp in humans, and 2 Mbp in mice (Jensen-Seaman et. al., 2004). The recombination rate can also differ between sexes in the same species. A case in point is *Drosophila melanogaster*, where the females undergo meiotic recombination, but the males do not (Hiraizumi, 1971). The recombination rate also varies between chromosomes of the same species. In humans, chromosomes 19 and Y both have a physical length of 59 Mbp, but while chromosome 19 has a genetic map length over 100 cM, Y has a map length of 1 cM (Doniskeller et. al., 1987). The recombination rate is not constant across the entire length of the chromosome. There is a significant deficiency of crossing-over near the centromere, and often also at the telomeres (Morton et. al., 1976). A small distance on the genetic map corresponds to a large distance on the chromosome within these regions. Finally, some evidence is suggesting that recombination rates are not even consistent at the sequence level. Inferred recombination rates from sequence data suggest that genetic map lengths increase sharply in short regions under 2 kilobases in length, called recombination hotspots, and very gradually everywhere else (Myers et. al., 2005). In other words, while there is a good correlation between genetic maps and physical maps, it is impossible to determine the genetic map directly from the physical map, and vice versa.

1.3 Recombination at the population level

Sexual reproduction has an effect on the genotypic make-up of a population. In this section, we will explore the effects of genetic shuffling on a population's genotype frequencies.

Consider the case of a haploid population with two loci of interest, each locus having two alleles. We label the alleles at the first locus *A* and *a*, and at the second locus *B* and *b*. Furthermore, we will assume that the four alleles have frequencies p_1, q_1, p_2, q_2, respectively, where $p+q=1$. It follows then that if these alleles were randomly assorted into genotypes, then the frequencies of the *AB*, *Ab*, *aB*, and *ab* genotype would be p_1p_2, p_1q_2, q_1p_2, and q_1q_2, respectively. While it might seem counterintuitive, it is unlikely that the alleles in the A gene are in random association with the alleles in the *B* gene. When genes *A* and *B* are not in random association with each other, they are said to be in linkage disequilibrium; and when they *are* in random association with each other, they are said to be in linkage equilibrium. Numerically, linkage disequilibrium can be expressed as the difference between the observed and expected frequency of a genotype (Robbins, 1918).

In a haploid population with genotype *AB* whose frequency is $Freq(AB)$, and where p_1 is the frequency of the *A* allele in the population, and p_2 is the frequency of the *B* allele, we can write the value of linkage disequilibrium as

$$D = Freq(AB) - p_1p_2 \tag{1.1}$$

Note that the population is in linkage equilibrium when $D=0$, and is in linkage disequilibrium otherwise. The value of *D* can also be expressed using only the genotype frequencies as variables, provided that each gene has only two alleles (see Appendix 2 for the derivation).

$$D = Freq(ab)Freq(AB) - Freq(Ab)Freq(aB) \tag{1.2}$$

A state of linkage equilibrium is eventually attained with random mating and recombination, but it is attained gradually. Bennett (1954) showed that a random-mating population that is initially in linkage disequilibrium with value *D* will decrease to value *D'* in the next generation as follows:

$$D' = (1 - r)D \tag{1.3}$$

where r is the recombinant fraction between loci. Therefore, the value of *D* decays by a factor of (1 - *r*) in every subsequent generation, eventually attaining linkage equilibrium (Bennett, 1954; see Appendix 3 for a more detailed analysis). Note that the lower the recombinant fraction, the longer it will take to attain linkage equilibrium.

The conclusion that a sexually recombinant population will eventually reach a state of linkage equilibrium is only true in a very restrictive scenario. More often than not, at least one of the implicit assumptions will be broken in a natural population. For example, directional selection can change the genotype frequencies, resulting in a loss of linkage equilibrium. Not all genotypes are equally well adapted to their environments: some are better at surviving to adulthood, and some are more adept at producing offspring. These genotypes will have a greater contribution to the next generation, and consequently, the alleles that make up such genotypes will increase in frequency as well. Unless alleles contribute multiplicatively to the fitness of the genotype -for example, if genotypes *ab*, *Ab*, *aB*, and *AB* have relative fitness values 1, 1 + *s*, 1 + *s*, $(1 + s)^2$, respectively- their frequencies will change to be in linkage disequilibrium (Felsenstein, 1965). When allelic contributions to fitness deviate from a multiplicative relationship, they are said to show epistatic effects (Cordell, 2002). It should be noted, however, that the word "epistasis" has a slightly different meaning at the fitness level than it does at the phenotypic level.

Epistasis is a measure of interaction between the genetic effects at different loci. In the simplest case, if the phenotypic effects of alleles at different loci are additive, we say that there is no epistasis. In this case, which comes from physiological genetics, epistasis is a measure of the deviation of the phenotypic values from additivity (Cordell, 2002). In population genetics, the emphasis is on the genetic effects on fitness (measured as viability and fertility) rather than the effects on a phenotypic value (such height or weight). In population genetics, epistasis is a measure of the deviation of the fitness values from multiplicativity (Cordell, 2002), because the genetic effects on fitness multiply over successive generations. In practice, the distinction between phenotype and fitness is usually overlooked and the phenotypic value is used as a convenient proxy for the fitness value. Generally speaking, provided that there is directional selection, a negative epistatic fitness curve (i.e. one that is less than multiplicative) will cause negative linkage disequilibrium, and a positive epistatic fitness curve will cause positive linkage disequilibrium (Felsenstein, 1965).

A second source of deviation from linkage disequilibrium is the fact that populations are finite, and furthermore, not always very large. A very large population generally follows the expected trajectory into linkage equilibrium, but a small population does not. In small populations (i.e. under 10 000), what often happens is that the genotype frequencies drift away from linkage equilibrium. In other words, the mean value of D will be zero, though the variance won't be (Hill and Robertson, 1968). This process of genetic drift can cause a genotype to be lost entirely from the population (Hill and Robertson, 1968; Ohta and Kimura, 1969). In fact, in the absence of mutation, all alleles will eventually either become fixed or lost (Kimura, 1968). The smaller the population size, the faster this happens (Kimura and Ohta, 1969a; Kimura and Ohta, 1969b).

Mutations can also alter the genotypic composition of a population. Provided that the population size is finite and the mutation rate is low enough that new mutations never occur at the same locus twice, then there will often be considerable linkage disequilibrium between loci (Ohta and Kimura, 1971). The extent of linkage disequilibrium is greater when the population size is small, and the recombination rate is low. It should be noted, however, that while each pair of loci will often show linkage disequilibrium, the mean value over all pairs of sites is expected to be zero (Ohta and Kimura, 1971).

Mating is not always random. When individuals generally prefer mates that are similar to themselves, it is called positive assortative mating; and it is called negative assortative mating when individuals prefer dissimilar mates. Generally speaking, the value of linkage disequilibrium increases with positive assortative mating, and decreases with negative assortative mating (Wilson, 1978). In both cases, introducing non-random mating into a population will produce linkage disequilibrium, and the population will remain in linkage disequilibrium upon reaching its steady-state (Wilson, 1978).

Populations are not always panmictic, but are sometimes segregated into smaller breeding subpopulations. When the population is very segregated and there is limited migration, random genetic drift prevails in each subpopulation. This can cause each subpopulation to show significant levels of linkage disequilibrium, as well as the population as a whole (Ohta, 1982). Moreover, a higher recombinant fraction between loci acts to reduce the variance in values of linkage disequilibrium, but will not eliminate it entirely (Ohta, 1982).

To summarize, a population will eventually reach a state of linkage equilibrium provided that mating is random; the population size is very large; there are insignificantly low levels of mutation; the population either shows no segregation, or migration rates are very high; and there

is either no selection, or non-epistatic selection. It is unlikely that all of these conditions will be met in a natural population.

It should be noted that the gradual attainment of a random assortment of alleles at two loci is a different dynamic than genetic segregation in diploids at a single locus, which attains randomness in a single generation. Consider a single gene where alleles *A* and *a*, have frequencies p_1 and q_1, respectively. Assuming that individuals are diploid, then an individual can have one of three possible genotypes: *AA*, *Aa*, or *aa*. Regardless of the initial genotype frequencies, the gametic frequency of *A* will be p_1, and the frequency of *a* will be q_1. With random mating, the frequencies the *AA*, *Aa*, and *aa* genotypes among zygotes will be p_1^2, $2p_1q_1$, q_1^2, respectively, in the next generation. These frequencies will not change among adults as each genotype has an equal likelihood of survival. Furthermore, the genotypes will keep these frequencies in all subsequent generations (Hardy, 1908; Weinberg, 1908). Therefore, both the allele and genotype frequencies will remain constant from generation to generation, provided that there are no other forces acting on the population. A population that is in this state is said to be in Hardy-Weinberg equilibrium.

1.4 The evolutionary origin of recombination

There are several theories to explain the evolutionary origin of sex and recombination. Some of these theories explain the origin of sex, others explain the origin of recombination, but none so far can explain the maintenance of either.

Perhaps the most cited theory for the evolutionary origin of recombination is that it originated as a DNA repair process (Bernstein et. al., 1981). As was stated in the section 1.1, the

recombinant site undergoes a genetic conversion. Therefore, the recombination process can also be thought of as a method of DNA repair. This theory proposes that first, haploid cells would fuse as a method of repairing their genetic material, with the diploid stage being initially transient. Later on, the diploid stage would come to dominate, for its redundancy provides protection against deleterious mutations. The masking effect in a diploid genome would make a large increase in genome size quite possible, after which a return to haploidy would be lethal, as it would expose these acquired deleterious mutations. Soon after this origin of recombination theory was published, Bernstein *et. al*. further proposed that sex and recombination are maintained due to its enhanced abilities of DNA repair (Bernstein et. al., 1985), but this theory seems less accepted. Recombination as a DNA repair process can still take place in an asexual parthenogenic species, and furthermore, it has been shown that a parthenogenic species would have a higher mean fitness if recombination's only advantage comes from DNA repair (Szathmáry and Kövér, 1991).

Perhaps the most cited theory of the origin of outbreeding is that it originated from selfish genetic elements (Hickey, 1982). The theory proposes that sex may have emerged from sections of DNA similar to a transposon or a conjugative plasmid. The requirements of such elements are that it be capable of transposition, self-replication, and that it induce a form of syngamy. Selfish DNA can only spread within an individual and its descendants when individuals are asexual, but can spread quite easily in a sexual population through horizontal gene transfer. Such elements could increase their frequencies even when their presence is detrimental to the host. When such elements are close to fixation, their fitness approaches that of the host (Hickey, 1982). Therefore, if selfish genetic elements are deleterious, they need not become fixed in the population (Rose, 1983). When such elements reach high frequencies, there will be selective pressures to reduce

their deleterious effects on the host. Consequently, this hypothesis can explain the origin of sex, but not its maintenance (Hickey and Rose, 1988).

It should be noted that these two origin theories (Bernstein et. al. 1981; Hickey, 1982) are not mutually exclusive, but somewhat complementary. Recombination could have first evolved as a method of DNA repair, and after this process became established, selfish genetic elements could have promoted syngamy (Hickey and Rose, 1988).

Kondrashov (1994) noted that an asexual population will have a lower mutational load if alternating between haploid and diploid phases, than if remaining permanently diploid. Some asexual species do in fact undergo such ploidy cycles, which acts as evidence for the hypothesis that sex may have arisen from such a cycle, immediately following the origin of syngamy.

A recent theory detailing the origin of meiosis has been put forward by Wilkins and Holliday (2009) where they proposed a step-wise process for meiosis to have evolved from mitosis. Since synapsis has been observed in the mitotic cells of some species, but terminating in either interphase or prophase, they argue that meiosis developed from a synapsis that extended into metaphase. Sister chromatids are still attached following a cellular division with synapsis, which acts to inhibit a subsequent replication phase. In this way, the evolution to meiosis could have occurred as a gradual process, emerging from a species with a ploidy cycle. The initial advantage of synapsis would be to prevent non-homologous recombination events, which have been observed in mitotic cells, and are often detrimental.

1.5 The maintenance of recombination by natural selection

The prevalence of sexual reproduction among multicellular eukaryotes is a long-standing unanswered question in evolutionary biology. Generally speaking, multicellular organisms can reproduce in one of two ways: sexually, or asexually. Many species are capable of both, many are exclusively sexual, but comparatively few are exclusively asexual. For example, among vertebrates there are known to be over 42 000 species (Wilson, 1992), yet less than 100 of these are exclusively asexual (Vrijenhoek et al., 1989). Furthermore, the majority of asexual species have recently descended from sexual species. The known exceptions are *bdelloid rotifers*, which have been asexual for over 40 million years (Welch et al., 2000); and *darwinuloid ostracods* have been asexual for maybe 200 million years (Martens et al., 2003).

Sex is not a necessity of life; this is clearly illustrated by the many instances of asexual reproduction that exist, especially among microbes. Yet, most biologists believe that sexual reproduction and the resulting genetic recombination do provide some selective advantage and, based on the broad patterns of distribution of sex among taxa, we can infer that the advantage of recombination is more important for large multicellular eukaryotes than it is for small single-celled prokaryotes: recombination and sexual reproduction are ubiquitous among derived eukaryotes (Hadany and Comeron, 2008), whereas bacteria are primarily asexual (Narra and Ochman, 2006).

Perhaps the first biologist to muse about the maintenance of sexual reproduction was August Weismann (1891), who wrote that the advantage of sex is that it provides variation for natural selection. Fisher (1930) elaborated on this theory, proposing that sex is advantageous because it can bring good genes together in a finite population. In the initial stages, descendants of a sexual individual might have difficulties increasing their numbers, but if this was to happen,

then the advantage of sexual reproduction would be proportional to the number of genes under natural selection. In making this argument, Fisher suggested that sex was an adaptation which favours the survival of groups instead of individuals. Muller (1932, 1964) then provided a counter benefit, in that genetic shuffling can also have the advantage of bringing bad mutations together in a single individual, allowing for negative mutations to be better purged from the population. He further noted that a deleterious mutation in an asexual individual is passed on to all of its descendants. In this way, deleterious mutations can build-up in an asexual lineage, and like a ratchet that tightens click-by-click, slowly kill off the lineage (Muller, 1964). This concept has since been referred to as Muller's ratchet (Felsentstein, 1974).

1.5.1 The maintenance of recombination by natural selection: defining the problem

Up until this point in time, the arguments given to explain the maintenance of sex had been verbal ones, not quantitative. Crow and Kimura (1965) then gave a mathematical context to the question of the maintenance of sex, utilizing a population genetics approach inspired by Muller. By comparing mean fitness values between sexual and asexual populations under directional selection, they showed that recombination greatly increases the fitness of sexuals by bringing good mutations from different lineages together in the same individual. The advantages of sexual reproduction are greatest when the mutation rate is high, when the population size is large, the mutant effects are small, and a double mutant has a higher fitness than either single mutant (a.k.a. directional selection). Furthermore, they demonstrated that two mutations that are deleterious on their own, but beneficial when together are initially selected against in a sexually

recombinant population. Therefore, it is unlikely that sex is maintained by this kind of allelic interaction.

John Maynard Smith then made some strong devil's advocate arguments. First (Maynard Smith, 1968), he countered Crow and Kimura's contention by showing that if mutations are allowed to reoccur with multiplicative allelic effects on fitness, then there would be no benefit to sexual reproduction. Later (Maynard Smith, 1978) he went on to illustrate the extent as to which sexual reproduction can reduce an individual's fertility. All things being equal, if a female were to defect to parthogenesis, she would double her genetic contribution to the next generation. This argument has since been dubbed the two-fold cost of sex, or more accurately, the two-fold cost of males (Hadany and Comeron, 2008). This immediate fitness gain of defecting to asexuality makes it clear that sexual reproduction is unlikely to be maintained through group selection (the idea that an allele can be selected for due to the benefit it gives to groups, regardless of its fitness effect on the individuals within that group), but must also provide a benefit to its offspring at the level of the individual. Maynard Smith made a similar argument for the maintenance of recombination rates, in that it too would require a short-term individual-based benefit (Maynard Smith, 1978). Therefore, an explanation that claims to provide an answer for the maintenance of sex must provide both short and long term benefits that allow sexual individuals to increase in frequency in a mixed population, and must also provide short and long term benefits that allow recombination rates to be maintained. For these reasons, my thesis only looks at selection at the individual level, as opposed to higher levels of selection (i.e. selection at the group, population, or species level).

1.5.2 The maintenance of recombination by natural selection: some solutions to the problem

Perhaps the most prominent ecological attempt at explaining the ubiquity of sex is the Red Queen Hypothesis (Van Valen, 1973). John Maynard Smith made the point that if selection was directional, the population would remain in linkage equilibrium, which neither favors nor disfavors sexual reproduction. But if the environment were to change, sexual reproduction would accelerate adaptation to a new environment (Maynard Smith, 1968). Spatially complex environments, such as described by the Tangled Bank model (Bell 1982), might be expected to favor sexual organisms that produce genetically variable offspring, but for this to explain the maintenance of sex, the environment would need to be constantly changing. Van Valen (1973) proposed the hypothesis that host-parasite interactions could provide continual environmental fluctuations. The theory is that hosts and parasites are in a constant evolutionary arms race: the parasite always needing to invade the host, and the host always trying to evade infection. Often, it is the case that a parasite with genotype A_p is more capable of infecting hosts with genotype A_h, and a parasite with genotype B_p is more capable of infecting hosts with genotype B_h. This can result in cyclic frequency changes in the genotypic composition of the host, as well as the parasite. It should be noted, however, that Van Valen did not initially apply his theory to the maintenance of sex, which was later done by Hamilton (1980). Under these fluctuations in fitness, sex can act to increase the rate of adaptation to the new environment, and can provide a two-fold advantage to a sexual species (Hamilton 1980).

While host-parasite interactions can result in a persistent cyclic arms race, they certainly don't have to, and in many situations they either converge to a steady state, or one of the

genotypes goes extinct (Bell, 1982). Further mathematical analysis has also found that sexual reproduction can only fend off parthenogenesis with very high levels of virulence (May and Anderson, 1983; Howard and Lively, 1994). Attempts to model the effect of recombination on a Red Queen system have found that sex is only beneficial when the cycles have large effects on genotypic fitness, and oscillate as quickly as every two-to-four generations (Peters and Lively, 2000). Most Red Queen interaction models examined so far tend to favor asexuality (Otto and Nuismer, 2004; Salathé et al.; 2008), although a recent experiment with *C. elegans* as the host gave results to the contrary (Morran et. al, 2011).

Another popular theory is that sex can be maintained in a finite population, but not in an infinite one. This theory is a direct descendant of the arguments put forward by Fisher (1930) and Muller (1932; 1964), but with the addition of explicitly taking the size of the population into account. Maynard Smith (1968) used a deterministic model to show that a population in mutation-selection balance will be in linkage equilibrium when reproducing sexually or asexually, but this implicitly assumes that the population size is large enough to ignore stochastic effects. In small simulated populations, however, Hill and Robertson's computer runs (1966) showed this to not be the case. Their simulations often showed that randomly-generated linkage disequilibrium caused linked loci to interfere with each other's response to selection. Felsenstein (1974) then put the pieces together, and used individual-based simulations to demonstrate both Fisher's and Muller's arguments.

This theory works best to explain the maintenance of recombination when selection is weak, and linkage is tight (Otto and Barton, 1997; Otto and Lenormand, 2002). The difficulty with the Fisher-Muller model is that recombination doesn't bring good mutations together any faster, or select for increased levels of recombination unless the effective population size is very

small (i.e. ~1000) (Christiansen et al. 1998; Otto and Barton, 2001). Attempts have been made to account for larger population sizes by either having many linked loci under selection (Keightley and Otto, 2006), or assuming population subdivision (Martin et. al., 2006). Both methods are effective, but it is questionable whether they are effective enough, as the rate of species-wide substitutions is typically far too low to generate appreciable selection for recombination, and the amount of required subdivision is very high (Barton, 2009).

It has been proposed that if alleles do not generally contribute multiplicatively to fitness, but display a negative epistatic fitness curve, sex will act to increase the mean fitness of the population (Kondrashov, 1982). A negative epistatic fitness curve is one where each additional deleterious mutation has an increasingly negative effect on the fitness of the organism, or each additional positive mutation has a decreasingly positive effect (see Figure 1.4). Such a fitness curve changes the genotypic frequencies to be in negative linkage disequilibrium immediately following selection. Kondroshov (1982) showed that with genome-wide negative epistasis, a sexual population's mean fitness could easily overcome the two-fold cost of sex.

The negative epistatic fitness hypothesis has two main theoretical criticisms. The first is that it has been difficult to explain *a priori* why epistatic values should be generally negative, rather than positive or zero. Laboratory experiments attempting to determine the mean level of epistasis within a species have varied greatly in their results (de Visser *et al*. 2007; Kouyos *et. al.*, 2007), but recent direct measurements of the value of epistasis suggest that this could in fact be the case (Chou *et. al*., 2011; Khan *et. al*., 2011). The second criticism is that modifier alleles that select for increased rates of recombination will only increase in frequency provided the effects are weakly negative (Barton, 1995). In other words, this theory can only explain the

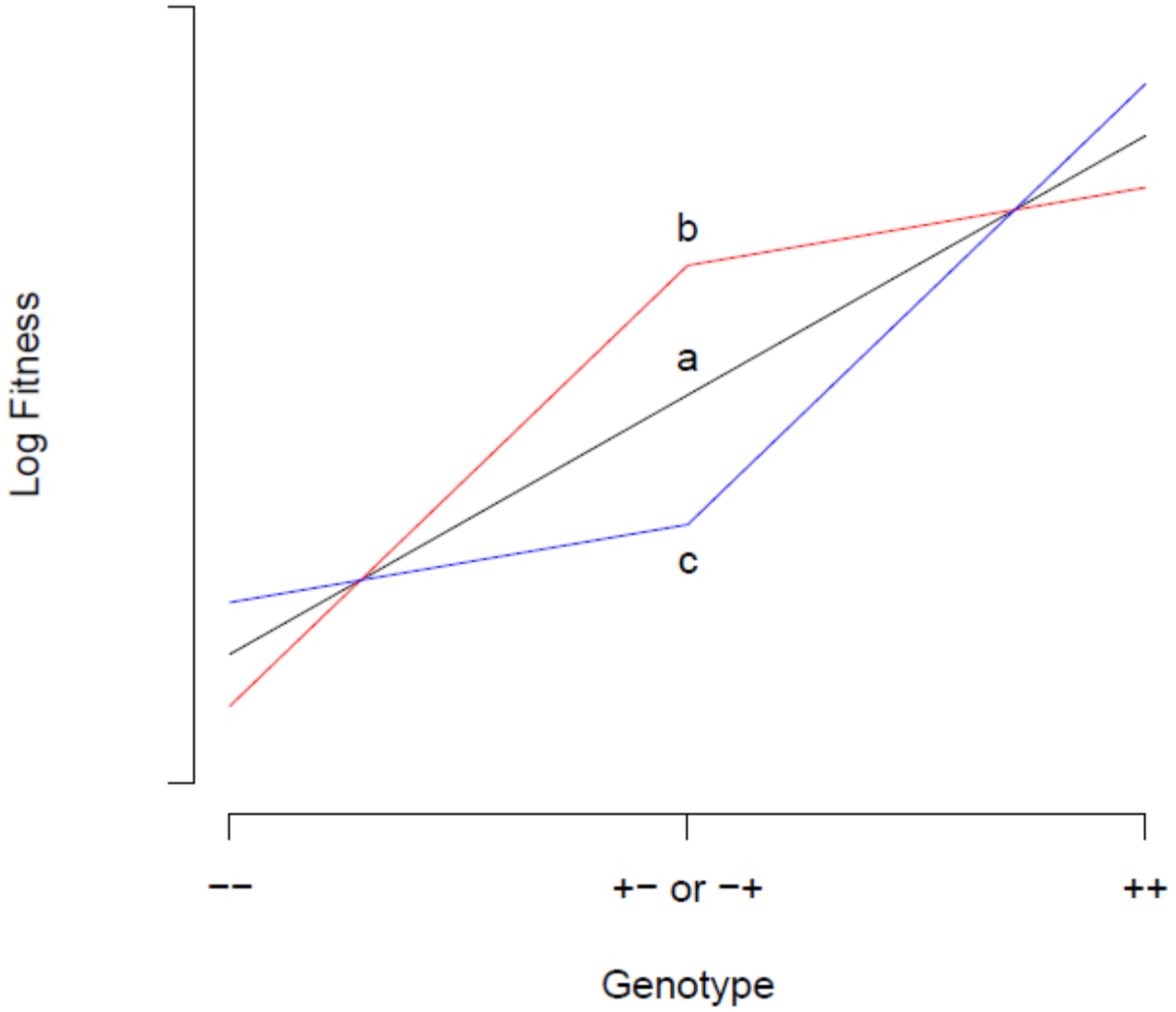


Figure 1.4: The relationship between the shape of a genotypic fitness curve and the value of epistasis.

The genotype of an individual is shown on the x-axis with "+" symbols representing an advantageous allele, and "–" symbols representing a deleterious allele. The log fitness of each genotype is shown on the *y*-axis. a) A non-epistatic fitness curve is one where each additional positive/negative mutation linearly adds to/subtracts from the log fitness of the genotype. It shows as a linear curve when plotted as above. b) A negative epistatic fitness curve is one where each additional deleterious mutation has an increasingly negative effect on the fitness of the organism, or each additional positive mutation has a decreasingly positive effect. In other words, it has a shape that is concave downwards when plotted as above. c) Finally, a positive epistatic fitness curve is the opposite of a negative one, and has a shape that is concave upwards.

maintenance of recombination if allelic interactions are generally weakly negative, and a reason to believe this is somewhat absent.

Provided that the genomic rate of deleterious mutations within a species, *U*, is greater than 1, then it is likely that allelic interactions are generally negatively epistatic. Otherwise, the species will inevitably go extinct by accumulating an unbearable load of deleterious mutations (Kondrashov 1988). Unfortunately, laboratory experiments attempting to determine the rate *U* within a species have given values that are not consistently greater than 1 (de Visser et al. 2007). The results have been all over the map. This doesn't mean that this theory cannot explain the ubiquity of sex, but that it can't be extrapolated as correct using this approach.

Most attempts to model the maintenance of sex have their roots in population genetics, and assume that the genotypes are either advantageous and neutral, or neutral and deleterious. Genotypes therefore affect selection as a whole, making no distinction between fertility and viablility. Each genotype's fitness is absolute, and unaffected by the frequency of other genotypes within the population. Another method is to use a more ecological approach, as in the case of the Red Queen hypothesis, where fitness values are not absolute but are frequency-dependent in some way. The Red Queen hypothesis is a model of interspecies frequency-dependent selection, but there are also models of within-species (aka intraspecies) frequency-dependent selection. As noted by Lewontin (1955), ‘‘it would be strange if what applied to different species did not apply to some extent to different genotypes within the same species.’’ Case and Taper (1986) looked at one such example where they showed an advantage to sexual reproduction in that it provides a greater ability to compete for a wide spectrum of resources. They found that a sexual species can coexist and even supplant an asexual species provided that there is relatively high between-genotype niche differentiation, low environmental variance, and

severe exploitation of resources. Under these conditions, common genotypes heavily deplete their resources creating an environment where a rare genotype has the advantage, and genetic shuffling helps to produce such rare genotypes. It should be noted that this model depends on the existence of different environmental niches (i.e. environmental heterogeneity) in order to explain the maintenance of sex. It has also been shown that the "two-fold cost of sex" is almost never two-fold when the ecological dynamics of growth and competition are taken into account (Doncaster *et. al*., 2000). Doncaster *et. al*. showed that – provided the population was initially sexual – an individual that defected to parthenogenesis would almost never realize anything close to a two-fold advantage. If the sexual population's intrinsic growth rate is sufficiently high, an asexual defector cannot successfully invade the sexual population. The result is a stable coexistence of both reproductive methods.

A simpler approach to that taken by Case and Taper (1986) is to observe the effects of competition on the maintenance of sex without the need for a resource spectrum. Generally speaking, intraspecies competition can be divided into two different types: scramble competition, and contest competition. In scramble competition, resources are somewhat equally distributed among competitors; whereas in contest competition, the "winner takes all" (Brännström and Sumpter, 2005). This way of classifying competition is based on how individuals consume resources. There are other ways of classifying competition, such as by how individuals are competing (examples include interference competition, and exploitation competition), but the scramble/contest dichotomy is the most relevant for our modeling purposes.

Peck and Waxman (2000) showed that if an individual's genotype affects its competitive ability, scramble competition when competing in small groups will result in negative epistasis. In such a case, the mean fitness of the sexual population is significantly higher, and able to resist

invasion from asexuals. They did not, however, consider the effects on a modifier allele. More importantly, their paper provides an intrinsic reason to believe that genotypic interactions should generally follow a negative epistatic fitness curve. My thesis builds on these results by looking at the effects of contest competition on the maintenance of sex. As an overarching goal, we are not so focused on the exact mechanism of competition, but on the fact that competition can happen at the individual level within a species. We wish to know if this fact can be used to explain the maintenance of sex.

1.6 Mathematical population genetics

The study of population genetics has incorporated a wide variety of math techniques over the course of the last 100 years. At the beginning of the 20th century, two main schools of thought on the laws of inheritance were caught in a heated debate. On one side, the biometricians (who were most influenced by Darwin) saw natural selection as a gradual process that acts on traits displaying a continuous spectrum of values. On the other side, the Mendelians saw natural selection as a discrete process that acts on genetic loci, each allele carrying its own distinct value. Fisher (1918) began a reconciliation between these two schools, by showing that a continuous trait such as male human height can be explained using Mendelian genetics when many loci affect the character. Fisher then went on to publish *The Genetical Theory of Natural Selection* (1930) in which he used Mendelian genetics as the underlying model of evolution by natural selection. This reconciliation between genetics, cytology, and evolutionary biology is referred to as the modern synthesis (Huxley, 1942).

Sewall Wright then contributed a "top-down" approach known as quantitative genetics (1920), by taking a continuous trait and breaking the phenotypic variance down into its environmental component and its genetic component. The fraction of the total phenotypic variance that is attributable to genetics is known as heredity. Wright also advanced the ideas of genetic drift, migration, and population subdivision (1931) as factors affecting the genotypic make-up of a population. His general modeling assumptions are now referred to as the Wright-Fisher model. The features of this model are a constant population size, discrete non-overlapping generations, and a diploid organism.

Moran (1962) devised a similar model, but with a haploid organism and overlapping generations. In each time step, one individual dies and one reproduces, thus ensuring that the population size remains constant.

Kimura continued work on the Wright-Fisher model by looking at the change in allele frequencies as a stochastic process using diffusion equations (1964). With these techniques, Kimura was able to calculate the probability of fixation or extinction of an allele for a given population size and selection coefficient (Kimura and Ohta, 1969a; Kimura and Ohta, 1969b).

1.7 Computer-based studies of population genetics

The variety of computer-based techniques used in the study of population genetics has grown significantly as computers have become more powerful. Generally speaking, these techniques can be divided into two categories: equation-based modeling, and individual-based modeling.

Equation-based modeling is where one represents the population as a set of equations, and numerically evaluates the solutions to these equations (Parunak et al., 1998). This is done by substituting biologically relevant parameter values into the set of equations to output an answer. Since population genetic studies often involve observing changes in allele frequencies over successive generations, a popular technique is to iterate a set of equations where each iteration is analogous to a generation. A specific version of such iterative methods is a Markov chain. In a Markov chain, the state of the population in the next generation only depends on its state in the current generation, and that the population transitions from one state to another with a given probability. Markov chains are commonly used to observe stochastic effects, such as when the population size is small and the effects of drift are important, but are irrelevant when observing deterministic effects.

Individual-based modeling (i.e. individual-based simulations) is where one explicitly "creates" every individual in order to observe the dynamics of the system as a whole (Judson, 1994). In this technique, individuals follow certain behavioral rules and interactions, with a built-in element of randomness. Because of this element of randomness, the population needs to be simulated several times before coming to any conclusions. For this reason, individual-based modeling is more computationally intensive than equation-based modeling. On the other hand, this technique has the advantage of being easier to construct without making simplifying assumptions, and can be validated both at the individual level as well as the population level (Parunak et al., 1998).

1.8 Thesis Objectives and Hypotheses

Three key points were made in the literature review:

1. there is an evolutionary advantage to recombination when the population is in negative linkage disequilibrium,
2. a negative epistatic fitness curve can create such negative linkage disequilibrium, and
3. we do not have an explanation, *a priori*, as to why genotypic fitnesses should generally be negatively epistatic.

Peck and Waxman (2000) showed that scramble competition, coupled with resource fragmentation, can provide a negative epistatic fitness curve, as well as an advantage to sex and recombination, but the effects of contest competition have yet to be determined.

The objective of this thesis is to explore the relationship between contest competition and the maintenance of recombination. Specifically, the questions we are interested in are the following:

- When does contest competition create a negative epistatic fitness curve?
- Does contest competition create an environment that is beneficial to sexual reproduction?
- Is contest competition part of a plausible explanation for the maintenance of sex?

The null hypothesis is that intraspecies contest competition will not result in a negative epistatic fitness curve. We ask if contest competition could create an environment that is beneficial to sexual reproduction which would allow sexual individuals to out-compete asexual individuals.

We constructed mathematical models and individual-based simulations as a means to answering these questions. In chapter 2, we explore the relationship between phenotype, competitive selection, and realized fitness in order to discover under what conditions competitive selection will result in negative epistasis. Next in chapter 3 we explore the effects of competition

over repeated generations to see if selection proceeds more rapidly in a sexual population, and we analyze the change in frequency of sexuals when competing with asexuals in a mixed population. Finally in chapter 4, we consider different mappings of phenotype onto fitness to ensure that our findings are robust.

CHAPTER 2

Here we explore the relationship between phenotype, competitive selection, and realized fitness. Specifically, we wish to discover under what conditions competitive selection will result in negative epistasis, thus providing an evolutionary advantage to genetic recombination.

Quantifying the relationships between genotype, phenotypic value, realized fitness and genotype frequencies.

In the classic case of non-competitive selection, the relationship between phenotype and fitness is very simple and indeed it is often assumed that the fitness is simply equal to the phenotypic value. In a competitive situation, however, an individual's fitness is dependent both on its own phenotypic value and also on that of its competitor. Thus the fitness of each type within the competing population is based on its own phenotypic value, along with the phenotypic values of the other types. Since the frequencies of the various types change during the course of selection, the fitnesses are frequency dependent in this case.

2.1 The relationship between genotype and phenotypic value.

Our first goal was to define the relationship between genotype and phenotype. The analysis is limited to two bi-allelic loci, *A* and *B* in a haploid population. The two alleles at the first locus are labeled *a* and *A*, and the two alleles at the second locus are *b* and *B*. The frequency

of the A allele is p_1 and the frequency of the B allele is p_2. Assuming initial linkage equilibrium, the genotypic frequencies for the four possible genotypes are as follows:

Genotype	ab	aB	Ab	AB
Frequency	$(1-p_1)(1-p_2)$	$(1-p_1)p_2$	$p_1(1-p_2)$	p_1p_2

For simplicity, we can write:

$$\begin{aligned}(1-p_1) &= q_1 \\ (1-p_2) &= q_2\end{aligned} \qquad (2.1)$$

Assuming that the effect of allele A on the phenotype is x and the effect of allele B is y, where both x and $y > 0$, and that the effects are multiplicative between loci, we can write the phenotypic values for the four genotypes.

Genotype	ab	aB	Ab	AB
Phenotypic Value (V)	1	$1+y$	$1+x$	$(1+x)(1+y)$

The average phenotypic value of a population can be obtained by summing the individual phenotypic values, weighted by their frequencies within the population.

$$V_{ave} = V_{ab}(1-p_1)(1-p_2) + V_{Ab}p_1(1-p_2) + V_{aB}(1-p_1)p_2 + V_{AB}p_1p_2 \qquad (2.2)$$

As expected, the average phenotypic value increases as the frequency of the A and B alleles increase in the population (see Figure 2.1.1). From the figure we see, however, that the increase is not a linear one. This is because the phenotypic values are multiplicative rather than additive. If we plot the values on a multiplicative scale (by the taking the natural logarithm of each value, then we see a linear increase (see Figure 2.1.2). On this scale, we also notice that the intermediate phenotypic value falls midway between the high and low values.

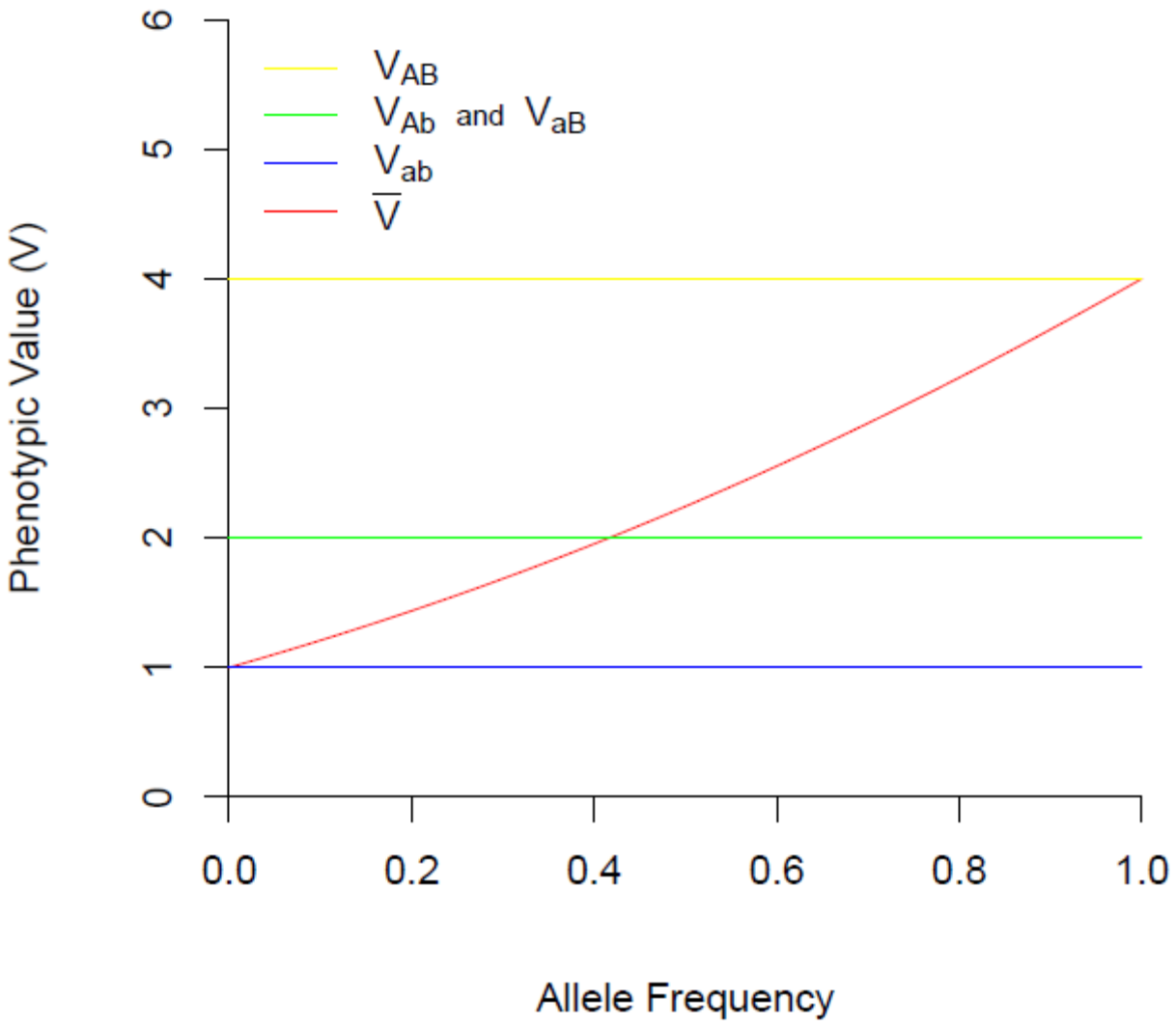


Figure 2.1.1: The relationship between phenotypic value and allele frequency.

The phenotypic value of each genotype is shown, along with the average phenotypic value. In this case, the phenotypic values (*V*) of each of the four genotypes, *ab*, *Ab*, *aB* and *AB* are 1, 2, 2 and 4, respectively. The frequencies of the *A* and *B* alleles are shown on the horizontal axis. For simplicity, the frequencies of the two high-value alleles, *A* and *B*, are equal. Note that although the phenotypic values of each genotype are constant, the average phenotypic value increases as the frequencies of the *A* and *B* alleles increase.

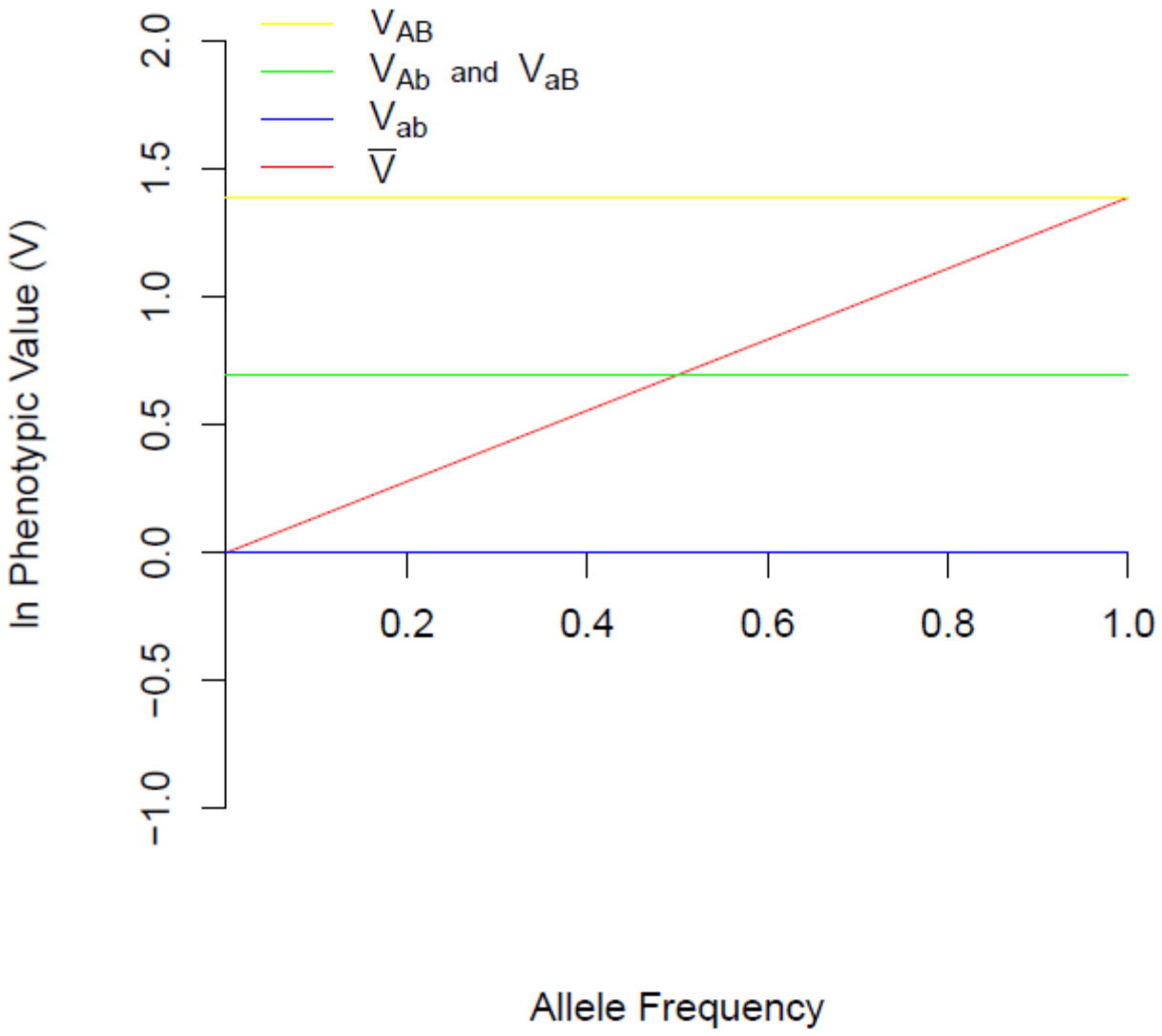


Figure 2.1.2: The relationship between phenotypic value (log scale) and allele frequency. For a description, see Figure 2.1.1. Note that in this case the horizontal lines representing the individual phenotypic values are equidistant and that the average phenotypic value increases linearly.

2.2 The relationship between phenotypic value and realized fitness.

Our second goal was to ask specifically if phenotypic values that follow a simple multiplicative pattern result in realized fitnesses that are also multiplicative. In this case the fitness is based on the outcome of competition between conspecific pairs of individuals. The question of whether fitnesses are multiplicative or not is important for our understanding of the adaptive role of genetic recombination.

In our model, individuals compete in pairs, at random. Competitive success is directly proportional to the phenotypic value of an individual compared to the phenotypic value of its competitor. For example, the probability of success for an individual with a phenotypic value of V_1 in competition with an individual that has a phenotypic value of V_2 is equal to $\frac{V_1}{V_1+V_2}$.

Although the probability of competitive success of an individual is dependent on its phenotypic value, it is not equal to it because it also depends on the frequency of the competing types. The competition matrix is shown in Table 2.1.3 and the expected frequencies of the various competitive interactions are shown in Table 2.2. By multiplying the values in these two Tables, we can estimate the realized fitness of each genotype.

Table 2.1.1 Expected outcome of various pairwise competitive interactions.

Genotype	Genotype of competitor ab	Ab	aB	AB
ab	$\frac{V_{ab}}{V_{ab}+V_{ab}}$	$\frac{V_{ab}}{V_{ab}+V_{Ab}}$	$\frac{V_{ab}}{V_{ab}+V_{aB}}$	$\frac{V_{ab}}{V_{ab}+V_{AB}}$
Ab	$\frac{V_{Ab}}{V_{Ab}+V_{ab}}$	$\frac{V_{Ab}}{V_{Ab}+V_{Ab}}$	$\frac{V_{Ab}}{V_{Ab}+V_{aB}}$	$\frac{V_{Ab}}{V_{Ab}+V_{AB}}$
aB	$\frac{V_{aB}}{V_{aB}+V_{ab}}$	$\frac{V_{aB}}{V_{aB}+V_{Ab}}$	$\frac{V_{aB}}{V_{aB}+V_{aB}}$	$\frac{V_{aB}}{V_{aB}+V_{AB}}$
AB	$\frac{V_{AB}}{V_{AB}+V_{ab}}$	$\frac{V_{AB}}{V_{AB}+V_{Ab}}$	$\frac{V_{AB}}{V_{AB}+V_{aB}}$	$\frac{V_{AB}}{V_{AB}+V_{AB}}$

The table shows the probability that an individual with the genotype listed on the left will win in competition with a genotype listed across the top.

Table 2.1.2 Expected outcome of competitive interactions, expressed in terms of the allelic effects, x and y.

Genotype	Genotype of competitor			
	ab	Ab	aB	AB
ab	$\frac{1}{1+1}$	$\frac{1}{1+(1+x)}$	$\frac{1}{1+(1+y)}$	$\frac{1}{1+(1+x)(1+y)}$
Ab	$\frac{(1+x)}{(1+x)+1}$	$\frac{(1+x)}{(1+x)+(1+x)}$	$\frac{(1+x)}{(1+x)+(1+y)}$	$\frac{(1+x)}{(1+x)+(1+x)(1+y)}$
aB	$\frac{(1+y)}{(1+y)+1}$	$\frac{(1+y)}{(1+y)+(1+x)}$	$\frac{(1+y)}{(1+y)+(1+y)}$	$\frac{(1+y)}{(1+y)+(1+x)(1+y)}$
AB	$\frac{(1+x)(1+y)}{(1+x)(1+y)+1}$	$\frac{(1+x)(1+y)}{(1+x)(1+y)+(1+x)}$	$\frac{(1+x)(1+y)}{(1+x)(1+y)+(1+y)}$	$\frac{(1+x)(1+y)}{(1+x)(1+y)+(1+x)(1+y)}$

In this Table, the phenotypic values are written in terms of the individual allelic effects.

Table 2.1.3 Competition Matrix expressed in terms of x and y, gathering some terms.

Genotype	Genotype of competitor ab	Ab	aB	AB
ab	$\frac{1}{2}$	$\frac{1}{2+x}$	$\frac{1}{2+y}$	$\frac{1}{2+x+y+xy}$
Ab	$\frac{1+x}{2+x}$	$\frac{1}{2}$	$\frac{1+x}{2+x+y}$	$\frac{1}{2+y}$
aB	$\frac{1+y}{2+y}$	$\frac{1+y}{2+x+y}$	$\frac{1}{2}$	$\frac{1}{2+x}$
AB	$\frac{(1+x)(1+y)}{2+x+y+xy}$	$\frac{1+y}{2+y}$	$\frac{1+x}{2+x}$	$\frac{1}{2}$

We can see that the diagonal values are equal to 0.5. This makes sense because an individual competing with another individual of the same genotype and phenotype has a 50% of winning the competition. If we now add the off-diagonal values in corresponding pairs, we see that they all add to 1.0. This also makes sense because an individual's chance of winning a competition equals the opponent's chance of losing. Since these paired values add to 1.0, the average value for each off-diagonal pair is 0.5. This means that the average value of all entries in the matrix is 0.5. This reflects the fact that if $2N$ individuals compete in pairs, there is a total of N survivors.

Table 2.2 Expected frequency of competitive interactions.

Genotype	Genotype of competitor			
	ab	Ab	aB	AB
ab	$(q_1q_2)(q_1q_2)$	$(q_1q_2)(p_1q_2)$	$(q_1q_2)(q_1p_2)$	$(q_1q_2)(p_1p_2)$
Ab	$(p_1q_2)(q_1q_2)$	$(p_1q_2)(p_1q_2)$	$(p_1q_2)(q_1p_2)$	$(p_1q_2)(p_1p_2)$
aB	$(q_1p_2)(q_1q_2)$	$(q_1p_2)(p_1q_2)$	$(q_1p_2)(q_1p_2)$	$(q_1p_2)(p_1p_2)$
AB	$(p_1p_2)(q_1q_2)$	$(p_1p_2)(p_1q_2)$	$(p_1p_2)(q_1p_2)$	$(p_1p_2)(p_1p_2)$

The competitive success of each genotype is determined by a combination of two factors:

1. the outcome of specific pairwise competitive interactions, based on the genotypically-determined phenotypic values of the pair of competing individuals and,
2. the frequency of each type of interaction.

Thus we can calculate the genotype proportions after competition by multiplying the competitive outcomes shown in Table 2.1.1 by their expected frequencies, shown in Table 2.2 and summing for each genotype. This gives the following results.

$$ab: \quad \left(\frac{V_{ab}}{V_{ab}+V_{ab}}\right)(q_1q_2)(q_1q_2)+\left(\frac{V_{ab}}{V_{ab}+V_{Ab}}\right)(q_1q_2)(p_1q_2)+\left(\frac{V_{ab}}{V_{ab}+V_{aB}}\right)(q_1q_2)(q_1p_2)$$
$$+\left(\frac{V_{ab}}{V_{ab}+V_{AB}}\right)(q_1q_2)(p_1p_2)$$

$$Ab: \quad \left(\frac{V_{Ab}}{V_{Ab}+V_{ab}}\right)(p_1q_2)(q_1q_2)+\left(\frac{V_{Ab}}{V_{Ab}+V_{Ab}}\right)(p_1q_2)(p_1q_2)+\left(\frac{V_{Ab}}{V_{Ab}+V_{aB}}\right)(p_1q_2)(q_1p_2)$$
$$+\left(\frac{V_{Ab}}{V_{Ab}+V_{AB}}\right)(p_1q_2)(p_1p_2)$$

$$aB: \quad \left(\frac{V_{aB}}{V_{aB}+V_{ab}}\right)(q_1p_2)(q_1q_2)+\left(\frac{V_{aB}}{V_{aB}+V_{Ab}}\right)(q_1p_2)(p_1q_2)+\left(\frac{V_{aB}}{V_{aB}+V_{aB}}\right)(q_1p_2)(q_1p_2)$$
$$+\left(\frac{V_{aB}}{V_{aB}+V_{AB}}\right)(q_1p_2)(p_1p_2)$$

$$AB: \quad \left(\frac{V_{AB}}{V_{AB}+V_{ab}}\right)(p_1p_2)(q_1q_2)+\left(\frac{V_{AB}}{V_{AB}+V_{Ab}}\right)(p_1p_2)(p_1q_2)+\left(\frac{V_{AB}}{V_{AB}+V_{aB}}\right)(p_1p_2)(q_1p_2)$$
$$+\left(\frac{V_{AB}}{V_{AB}+V_{AB}}\right)(p_1p_2)(p_1p_2)$$

In order to convert these four proportions into genotypic frequencies, they must be normalized by dividing by the average survival rate. This equals $(1-c)$ where c is the culling rate. The culling rate for one full round of pairwise competition is 0.5 because there is only one survivor from each competing pair. Thus, $(1-c)=0.5$ in this case. In general, if the population size remains constant, over several generations, the survival rate is necessarily equal to $1/n$, where *n* is the average number of offspring produced per parent (in this case, the culling rate is $(1-1/n)$).

Once we have the genotypic frequencies after competition, we can calculate the realized fitness of each genotype. The realized fitness of an individual is defined as the number of offspring that survive to maturity. Consequently, in the absence of an intervening round of

recombination, the realized fitness of a genotypic class of individuals can be calculated as the frequency of that genotype after competition, divided by its frequency before competition.

Thus the realized fitnesses of the four genotypes is as follows:

w_{ab}= (Freq.of ab genotype after competition) / (Freq.of ab genotype before competition)

w_{Ab}= (Freq.of Ab genotype after competition) / (Freq.of Ab genotype before competition)

w_{aB}= (Freq.of aB genotype after competition) / (Freq.of aB genotype before competition)

w_{AB} = (Freq.of AB genotype after competition) / (Freq.of AB genotype before competition)

Note that the expression for the genotypic proportions after competition contains the frequency of that genotype before competition as a common factor. This factor cancels out during the calculation of the realized fitness. Therefore, the realized fitness of each of the four genotypes is:

$$
\begin{aligned}
w_{ab} &= \frac{1}{1-c}\left[\left(\frac{V_{ab}}{V_{ab}+V_{ab}}\right)q_1q_2 + \left(\frac{V_{ab}}{V_{ab}+V_{Ab}}\right)p_1q_2 + \left(\frac{V_{ab}}{V_{ab}+V_{aB}}\right)q_1p_2 + \left(\frac{V_{ab}}{V_{ab}+V_{AB}}\right)p_1p_2\right] \\
w_{Ab} &= \frac{1}{1-c}\left[\left(\frac{V_{Ab}}{V_{Ab}+V_{ab}}\right)q_1q_2 + \left(\frac{V_{Ab}}{V_{Ab}+V_{Ab}}\right)p_1q_2 + \left(\frac{V_{Ab}}{V_{Ab}+V_{aB}}\right)q_1p_2 + \left(\frac{V_{Ab}}{V_{Ab}+V_{AB}}\right)p_1p_2\right] \\
w_{aB} &= \frac{1}{1-c}\left[\left(\frac{V_{aB}}{V_{aB}+V_{ab}}\right)q_1q_2 + \left(\frac{V_{aB}}{V_{aB}+V_{Ab}}\right)p_1q_2 + \left(\frac{V_{aB}}{V_{aB}+V_{aB}}\right)q_1p_2 + \left(\frac{V_{aB}}{V_{aB}+V_{AB}}\right)p_1p_2\right] \\
w_{AB} &= \frac{1}{1-c}\left[\left(\frac{V_{AB}}{V_{AB}+V_{ab}}\right)q_1q_2 + \left(\frac{V_{AB}}{V_{AB}+V_{Ab}}\right)p_1q_2 + \left(\frac{V_{AB}}{V_{AB}+V_{aB}}\right)q_1p_2 + \left(\frac{V_{AB}}{V_{AB}+V_{AB}}\right)p_1p_2\right]
\end{aligned}
\tag{2.3}
$$

Substituting with the values shown in Table 2.1.3, we get

$$
\begin{aligned}
w_{ab} &= \frac{1}{1-c}\left(\frac{1}{2}q_1q_2 + \frac{1}{2+x}p_1q_2 + \frac{1}{2+y}q_1p_2 + \frac{1}{2+x+y+xy}p_1p_2\right) \\
w_{Ab} &= \frac{1}{1-c}\left(\frac{1+x}{2+x}q_1q_2 + \frac{1}{2}p_1q_2 + \frac{1+x}{2+x+y}q_1p_2 + \frac{1}{2+y}p_1p_2\right) \\
w_{aB} &= \frac{1}{1-c}\left(\frac{1+y}{2+y}q_1q_2 + \frac{1+y}{2+x+y}p_1q_2 + \frac{1}{2}q_1p_2 + \frac{1}{2+x}p_1p_2\right) \\
w_{AB} &= \frac{1}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}q_1q_2 + \frac{1+y}{2+y}p_1q_2 + \frac{1+x}{2+x}q_1p_2 + \frac{1}{2}p_1p_2\right)
\end{aligned}
\tag{2.4}
$$

We can now plot the relationship between realized fitness and allele frequency, and compare it to the result based on phenotypic values shown in Figure 2.1.1 (see Figure 2.2.1 and Figure 2.2.2).

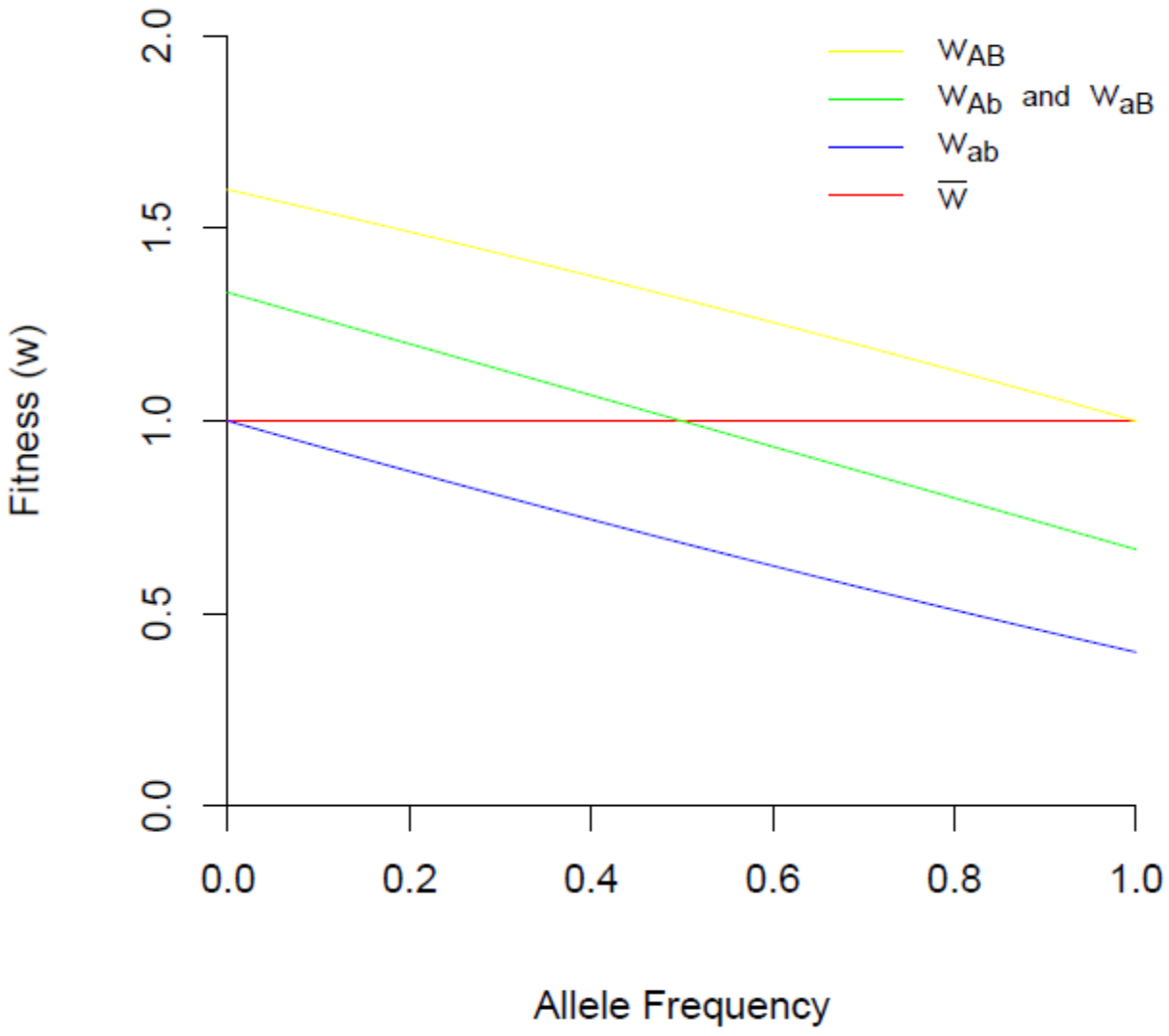


Figure 2.2.1: The relationship between realized fitness and allele frequency.

The realized fitnesses of each genotype (w_{ab}, w_{Ab}, w_{aB} and w_{AB}) are shown, along with the average value. The fitnesses were calculated based on the phenotypic values shown in Figure 2.1.1. The frequency of the *A* and *B* alleles is shown on the horizontal axis. Note that the genotypic fitness values are frequency dependent, and have higher values when the frequency of the *A* and *B* alleles is low, and decreases in value as the frequency of the *A* and *B* alleles increases. The mean fitness of the population on the other hand (shown by the red line) remains constant.

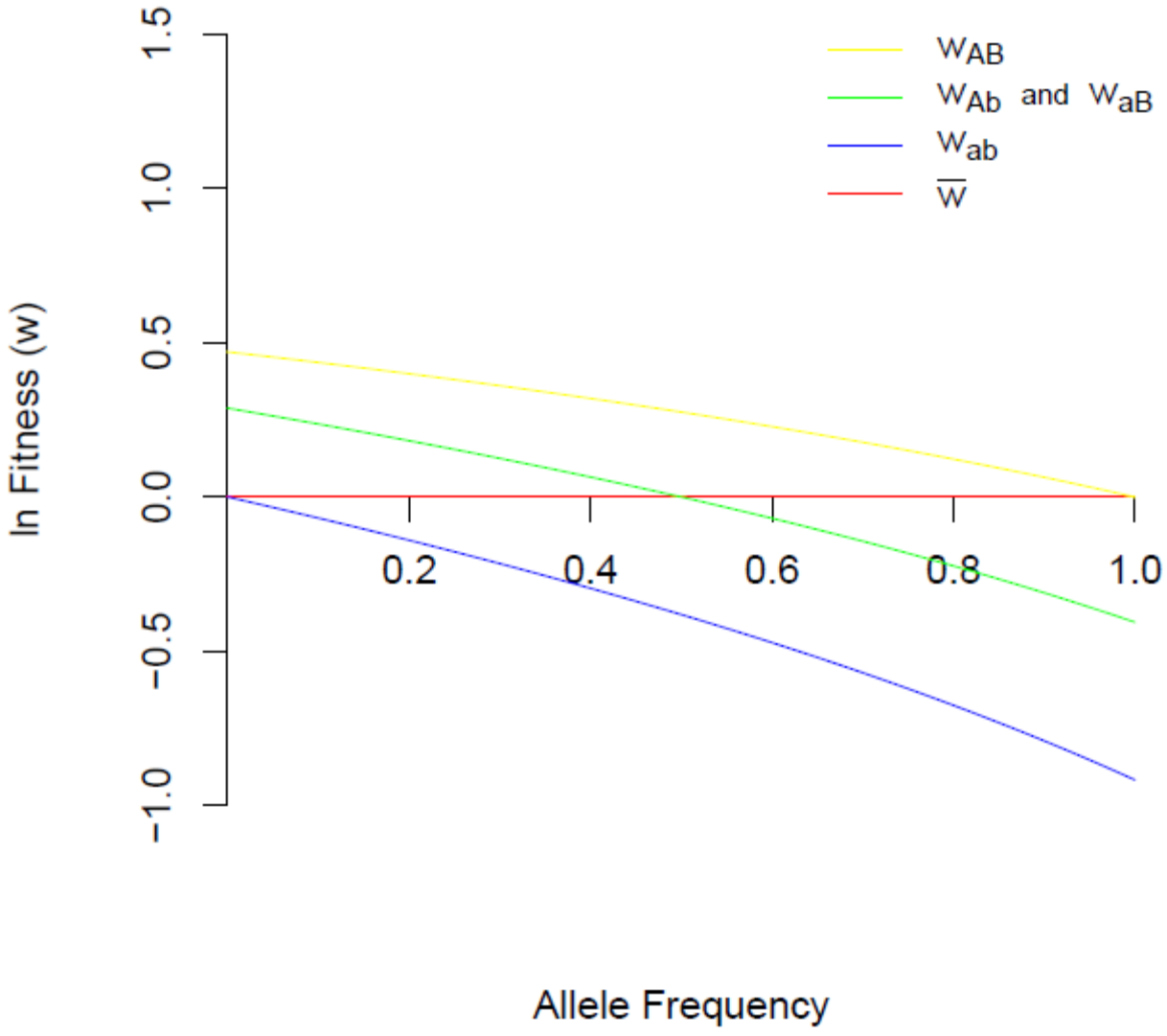


Figure 2.2.2: The relationship between the natural log of the realized fitness (ln *w*) and allele frequency.

For a description, see Figure 2.2.1. Note that, when plotted on a log scale, the fitness values of the intermediate genotypes (shown by the green line) are closer in value to the highest fitness genotype (shown by the yellow line) than they are to the fitness of the lowest fitness genotype (shown by the blue line). The average fitness is shown by the red line.

2.3 Calculating the value of epistasis from the realized fitnesses.

Epistasis is a measure of interaction between the genetic effects at different loci. In the simplest case, if the effects of alleles at different loci are additive, we say that there is no epistasis. In that case, epistasis is a measure of the deviation of the phenotypic values from additivity. This is the definition that is usually used in physiological genetics.

In population genetics, the emphasis is on the genetic effects on fitness rather than the effects on a phenotypic value (such height or weight). In practice, the distinction between phenotype and fitness is usually overlooked and the phenotypic value is used as a convenient proxy for the fitness value. As we show here, however, in the case of competition, phenotype and fitness are related, but not identical, measures.

Epistatic values can be expressed in either an additive or a multiplicative scale. The additive scale is normally used for biomedical applications, whereas the multiplicative scale is more appropriate in population genetics. This is because the genetic effects on fitness multiply over successive generations.

For the two-locus, two-allele case considered here, the additive measure of epistasis is:

$$\varepsilon = w_{ab} + w_{AB} - w_{Ab} - w_{aB} \tag{2.5}$$

The multiplicative measure is :

$$\varepsilon = \ln w_{ab} + \ln w_{AB} - \ln w_{Ab} - \ln w_{aB} \tag{2.6}$$

This is often written as follows.

$$\varepsilon = \ln \frac{w_{ab} w_{AB}}{w_{Ab} w_{aB}} \tag{2.7}$$

The value of multiplicative epistasis in the case of competitive selection can be obtained by substituting the fitness values from equations (2.4) into this equation.

$$\varepsilon = \ln \frac{\left(\frac{q_1q_2}{2} + \frac{p_1q_2}{2+x} + \frac{q_1p_2}{2+y} + \frac{p_1p_2}{2+x+y+xy}\right)\left(\frac{(1+x)(1+y)}{2+x+y+xy}q_1q_2 + \frac{1+y}{2+y}p_1q_2 + \frac{1+x}{2+x}q_1p_2 + \frac{p_1p_2}{2}\right)}{\left(\frac{1+x}{2+x}q_1q_2 + \frac{p_1q_2}{2} + \frac{1+x}{2+x+y}q_1p_2 + \frac{p_1p_2}{2+y}\right)\left(\frac{1+y}{2+y}q_1q_2 + \frac{1+y}{2+x+y}p_1q_2 + \frac{q_1p_2}{2} + \frac{p_1p_2}{2+x}\right)} \qquad (2.8)$$

Note that the normalizing factor, $(1 - c)$ cancels out.

The key question for this study is whether the value of ε given by this equation is negative for all values of p_1 and p_2 between 0 and 1, and for all positive values of x and y. The condition of negativity is satisfied if the numerator is less than the denominator. In order to evaluate the sign of ε, we can simplify and compare the terms in the numerator (Num_1 and Num_2) and the denominator (Dem_1 and Dem_2). First, we write out the four main terms separately.

$$\begin{aligned}
Num_1 &= \frac{1}{2}q_1q_2 + \frac{1}{2+x}p_1q_2 + \frac{1}{2+y}q_1p_2 + \frac{1}{2+x+y+xy}p_1p_2 \\
&= \frac{(1-p_1)(1-p_2)}{2} + \frac{p_1(1-p_2)}{2+x} + \frac{(1-p_1)p_2}{2+y} + \frac{p_1p_2}{2+x+y+xy} \\
Num_2 &= \frac{(1+x)(1+y)}{2+x+y+xy}q_1q_2 + \frac{1+y}{2+y}p_1q_2 + \frac{1+x}{2+x}q_1p_2 + \frac{1}{2}p_1p_2 \\
&= \frac{(1+x)(1+y)}{2+x+y+xy}(1-p_1)(1-p_2) + \frac{1+y}{2+y}p_1(1-p_2) + \frac{1+x}{2+x}(1-p_1)p_2 + \frac{p_1p_2}{2} \\
Dem_1 &= \frac{1+x}{2+x}q_1q_2 + \frac{1}{2}p_1q_2 + \frac{1+x}{2+x+y}q_1p_2 + \frac{1}{2+y}p_1p_2 \\
&= \frac{1+x}{2+x}(1-p_1)(1-p_2) + \frac{p_1(1-p_2)}{2} + \frac{1+x}{2+x+y}(1-p_1)p_2 + \frac{p_1p_2}{2+y} \\
Dem_2 &= \frac{1+y}{2+y}q_1q_2 + \frac{1+y}{2+x+y}p_1q_2 + \frac{1}{2}q_1p_2 + \frac{1}{2+x}p_1p_2 \\
&= \frac{1+y}{2+y}(1-p_1)(1-p_2) + \frac{1+y}{2+x+y}p_1(1-p_2) + \frac{(1-p_1)p_2}{2} + \frac{p_1p_2}{2+x}
\end{aligned} \qquad (2.9)$$

Essentially, we want to show that: $Dem_1Dem_2 > Num_1Num_2$, or put another way, that $Dem_1Dem_2 - Num_1Num_2 > 0$.

Using Mathematica, we can simplify $Dem_1Dem_2 - Num_1Num_2$ down to the form:

$$\frac{xy\{(1+x)(1+y)(2+x+y-2xp_1-2yp_2)+x^2(1+y)p_1^2+(1+x)y^2p_2^2\}}{2(2+x)(2+y)(2+x+y)(2+x+y+xy)} + \frac{x^3y^3(1+x)(1+y)p_1p_2(1-p_1)(1-p_2)}{(2+x)(2+y)(2+x+y)^2(2+x+y+xy)^2} \qquad (2.10)$$

We are assuming that x and y are both greater than zero, and that p_1 and p_2 are both positive with values between zero and one. It should be clear that the second term in this equation is always positive and the denominator of the first term is also positive. In the numerator of the first term, the factor xy is also positive. This leaves us to evaluate whether the second factor in the numerator of the first term, $(1+x)(1+y)(2+x+y-2xp_1-2yp_2)+x^2(1+y)p_1^2+(1+x)y^2p_2^2$, which we will label f, is also positive. We will prove this by breaking it down into four cases.

1. First, note that if $0 < x \leq 1$ and $0 < y \leq 1$, then the term $2+x+y-2xp_1-2yp_2$ would be always non-negative in this case, because neither p_1 nor p_2 can exceed a value of 1. Since the other terms are positive, f will also be positive in this case.

2. Next, let's consider the case where $x; y > 1$.

Since $1-2p_1+p_1^2 = (1-p_1)^2 > 0$,

hence $1-2p_1 > -p_1^2$.

Similarly $1-2p_2 > -p_2^2$.

Then since $x; y > 0$, and rearranging the formula above for f as follows,

$$f = (1+x)(1+y)\big(2+x(1-2p_1)+y(1-2p_2)\big)+x^2(1+y)p_1^2+(1+x)y^2p_2^2$$

$$> 2+x(1-2p_1)+y(1-2p_2)+x^2p_1^2+y^2p_2^2$$

$$> 2-xp_1^2-yp_2^2+x^2p_1^2+y^2p_2^2$$

$$= 2+xp_1^2(x-1)+yp_2^2(y-1)$$

> 0 (because all of the terms are positive given that x and $y > 1$ in this case).

3. In the case where $0 < x \leq 1$ and $y > 1$,

$$f > 2+x(1-2p_1)+y(1-2p_2)+x^2p_1^2+y^2p_2^2$$

$$> 2+x(1-2p_1)-yp_2^2+x^2p_1^2+y^2p_2^2$$

$= \left(2 + x(1 - 2p_1 + xp_1^2)\right) + yp_2^2(y - 1)$

> 0 (because the first term is positive given that $0 < x \leq 1$, and the second term is positive given that $y > 1$).

4. Finally, in the case where $x > 1$ and $0 < y \leq 1$, it can be shown that $f > 0$ similarly to 3, as f is symmetric with respect to x, p_1 and y, p_2.

With the value of f being positive, as well as every other factor in the equation of $Dem_1Dem_2 - Num_1Num_2$, we have now proven that the value of

$$Dem_1Dem_2 - Num_1Num_2 > 0$$

which means that $Num_1Num_2 - Dem_1Dem_2 < 0$, and that the value of epistasis is negative, that is

$$\varepsilon = ln\frac{w_{ab}w_{AB}}{w_{Ab}w_{aB}} < 0.$$

2.4 Calculating epistasis: some numerical solutions.

Although we have shown that the value of epistasis is negative for all positive values of p_1, p_2, x and y, we still need to estimate the value of epistasis over the parameter space. These values are not obvious from the formula itself but we calculate them for given parameter values. For example, in the simple case where the allele frequencies at both loci are 0.5 and the phenotypic effects of both the *A* and *B* alleles are equal to 1, then:

$$Num_1 = \frac{(1-p_1)(1-p_2)}{2} + \frac{p_1(1-p_2)}{2+x} + \frac{(1-p_1)p_2}{2+y} + \frac{p_1p_2}{2+x+y+xy}$$

$$= (0.5)(0.25) + \left(\frac{1}{3}\right)(0.25) + \left(\frac{1}{3}\right)(0.25) + (0.2)(0.25)$$

$$= 0.342$$

$$Num_2 = \frac{(1+x)(1+y)}{2+x+y+xy}(1-p_1)(1-p_2) + \frac{1+y}{2+y}p_1(1-p_2) + \frac{1+x}{2+x}(1-p_1)p_2 + \frac{p_1p_2}{2}$$

$$= (0.8)(0.25) + \left(\frac{2}{3}\right)(0.25) + \left(\frac{2}{3}\right)(0.25) + (0.5)(0.25)$$

$$= 0.658$$

$$Dem_1 = \frac{1+x}{2+x}(1-p_1)(1-p_2) + \frac{p_1(1-p_2)}{2} + \frac{1+x}{2+x+y}(1-p_1)p_2 + \frac{p_1p_2}{2+y}$$

$$= \left(\frac{2}{3}\right)(0.25) + (0.5)(0.25) + (0.5)(0.25) + \left(\frac{1}{3}\right)(0.25)$$

$$= 0.5$$

$$Dem_2 = \frac{1+y}{2+y}(1-p_1)(1-p_2) + \frac{1+y}{2+x+y}p_1(1-p_2) + \frac{(1-p_1)p_2}{2} + \frac{p_1p_2}{2+x}$$

$$= \left(\frac{2}{3}\right)(0.25) + (0.5)(0.25) + (0.5)(0.25) + \left(\frac{1}{3}\right)(0.25)$$

$$= 0.5$$

In this case,

$$\varepsilon = \ln\frac{(0.342)(0.658)}{(0.5)(0.5)}$$

$$\approx -0.105$$

Figures 2.3-2.6 show the value of epistasis over a wide range of the parameter space. We find that the degree of epistasis is more pronounced when the selected alleles have larger phenotypic effects (see Figure 2.5). We also find that for any given effect of allele *A*, increasing the effect of allele *B* produces more negative epistasis, and vice-versa. Figure 2.6 shows that the value of epistasis is highest when the frequencies of both alleles are equal to each other, and lowest when their frequencies are at their most divergent (i.e. when one allele's frequency approaches 1, and the other's frequency approaches 0). Finally, it can be seen from Figures 2.3 and 2.4 that the phenotypic effects of the *A* and *B* alleles are a major factor in determining the value of epistasis, while the frequencies of these alleles are a relatively minor factor. This is especially the case when the phenotypic effects of the *A* and *B* alleles are small (i.e. less than 1). In this case, the value of epistasis is essentially frequency independent.

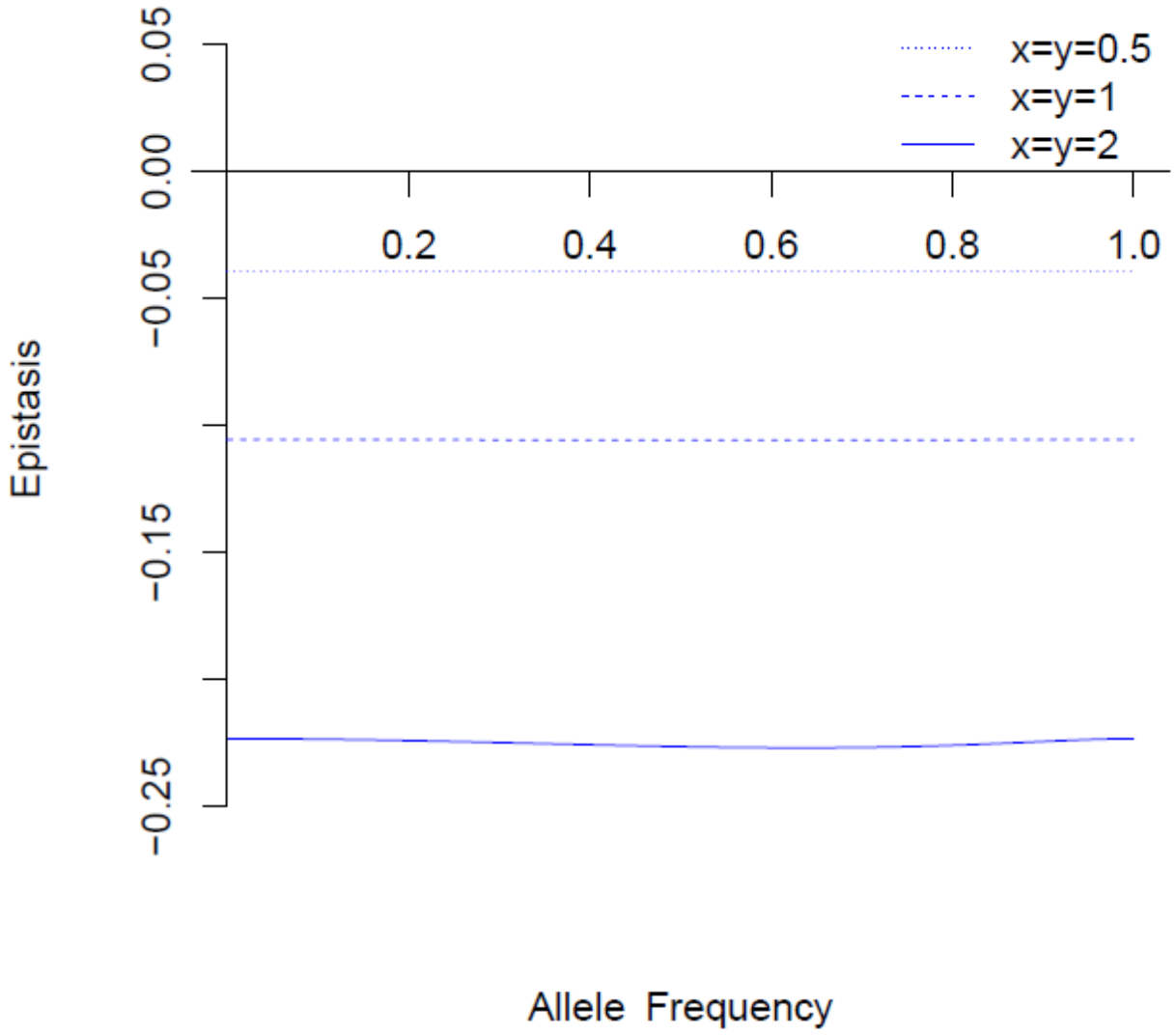


Figure 2.3: The relationship between the value of epistasis (multiplicative) and allele frequency.

The value of epistasis is shown on the vertical axis, and the frequency of the *A* and *B* alleles is shown on the horizontal axis. For simplicity, the frequencies of the high-value alleles, *A* and *B*, are equal; the effect of these alleles on the phenotype, *x* and *y* (respectively), are also equal; and the genotypic frequencies are in linkage equilibrium. Three lines are drawn: where the effects on the phenotype are equal to 0.5, 1, and 2. Note that the value of epistasis is less than zero in all cases. For small phenotypic effects, the value of epistasis is not frequency dependent.

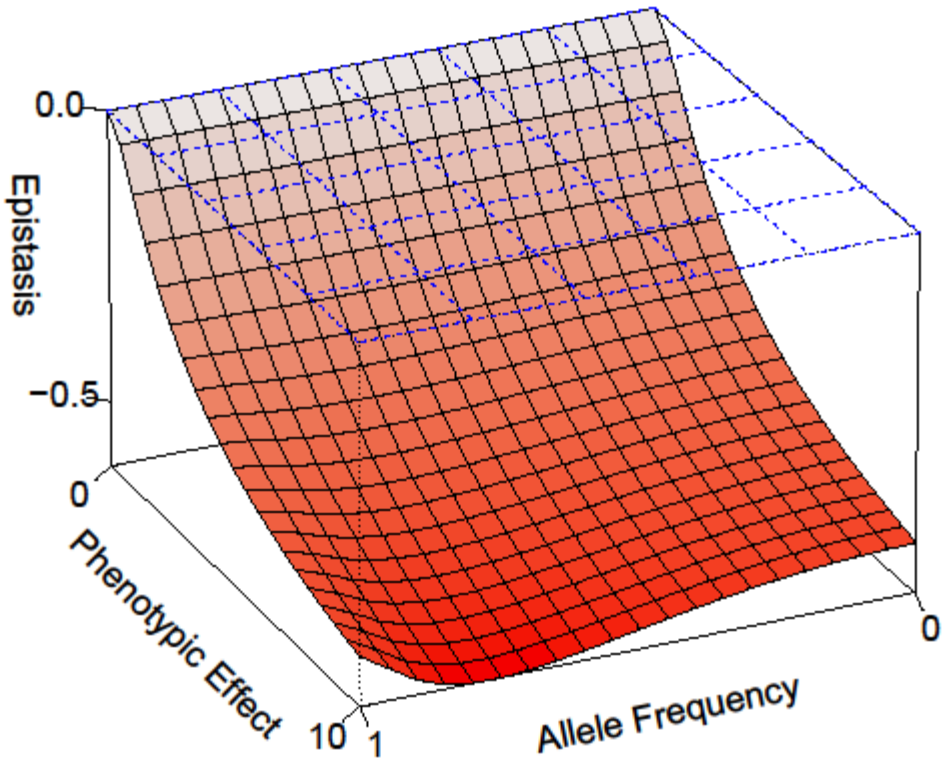


Figure 2.4: The relationship between the value of epistasis (multiplicative scale), phenotypic effect, and allele frequency.

The value of epistasis is shown on the vertical axis, while the frequency of the *A* and *B* alleles (Allele Frequency) and the effect of these alleles on the phenotype (Phenotypic Effect) are shown on the other two axes. For simplicity, the frequencies of the selected alleles, *A* and *B*, are equal; the effects of these alleles on the phenotype, *x* and *y* (respectively), are also equal; and the genotypic frequencies are in linkage equilibrium. Note that the value of epistasis is always negative, and the absolute value of epistasis increases with increasing phenotypic effect.

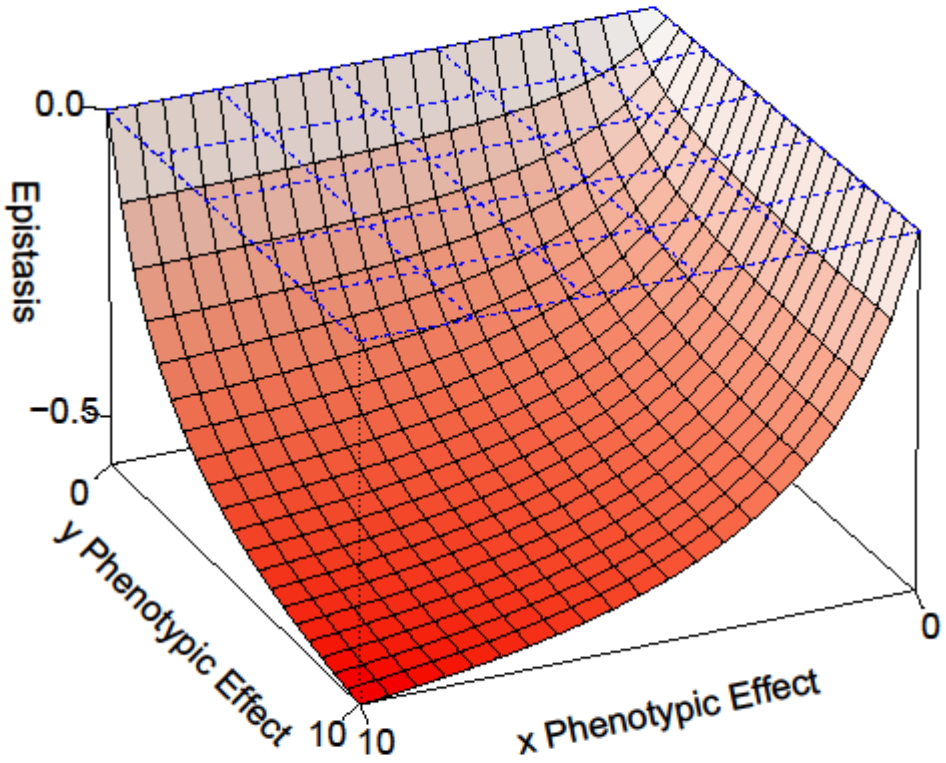


Figure 2.5: The relationship between the value of epistasis (multiplicative), and phenotypic effects.

The value of epistasis is shown on the vertical axis, and the phenotypic effects of the *A* allele (*x*) and the *B* allele (*y*) are shown on the other axes. For simplicity, the frequencies of all four alleles are equal to 0.5, and the population is in linkage equilibrium. The graph shows how the value of epistasis changes when the phenotypic effects of the *A* allele and the *B* allele are not necessarily equal to each other.

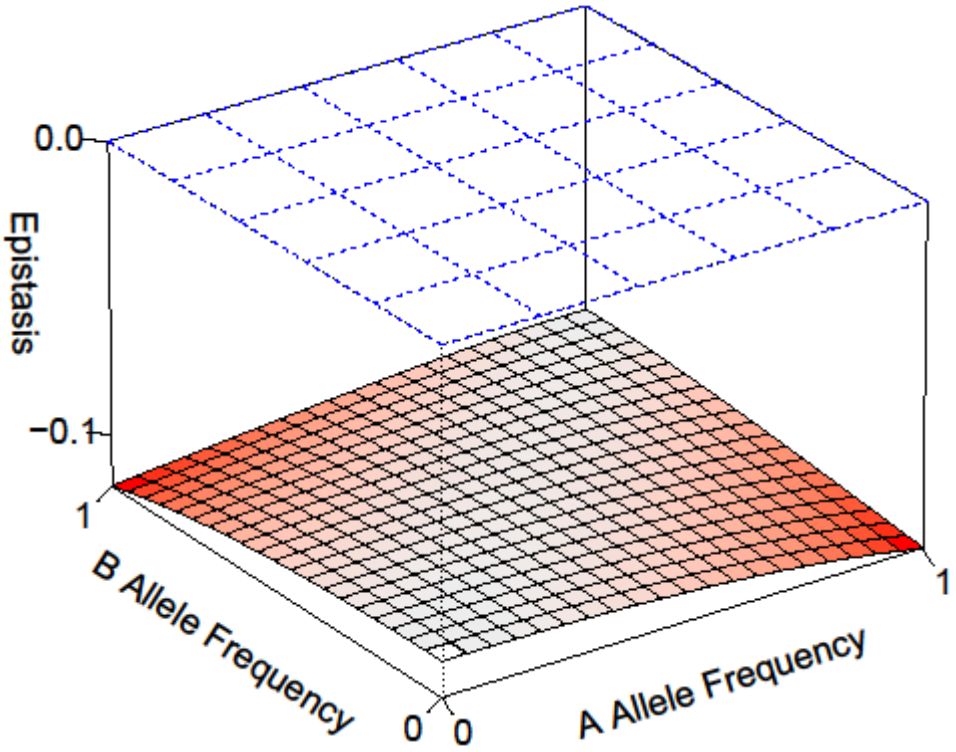


Figure 2.6: The relationship between the value of epistasis (multiplicative scale), and the frequencies of the *A* and *B* alleles.

The value of epistasis is shown on the vertical axis, and the frequencies of the *A* allele and the *B* allele are shown on the other axes. For simplicity, the phenotypic effects of the *A* and *B* alleles are both equal to 1; and the population is in linkage equilibrium. The graph illustrates the change in the value of epistasis when the frequencies of the *A* and *B* alleles are not necessarily equal to each other.

2.5 The effect of initial linkage disequilibrium on the value of epistasis.

So far we have looked at the value of epistasis when the population is initially in linkage equilibrium. We will now examine the change in the value of epistasis when the population is initially in linkage *disequilibrium*.

Numerically, the extent of linkage disequilibrium is often given a value, *D*, which is the difference between the observed and expected frequency of a genotype (Robbins 1918).

$$D = Freq(AB) - p_1 p_2 \tag{2.14}$$

In the two-locus bi-allelic system described above, the value of *D* can also be expressed with the following equation (see Appendix 1 for the derivation).

$$D = Freq(ab)Freq(AB) - Freq(Ab)Freq(aB) \tag{2.15}$$

An alternate equation used to express the value of linkage disequilibrium is known as *Z* (Felsenstein, 1965).

$$Z = \ln \frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)} \tag{2.16}$$

Note that *D* and *Z* always have the same sign, and both are equal to zero when there is linkage equilibrium. The two statistics differ in their range of values. *D* has a maximum range of -0.25 to 0.25, but its range decreases as the frequencies of the *A* and *B* alleles deviate from 0.5. In contrast, *Z* always has a range of -∞ to +∞. See Appendix 3 for more information on the relationship between *D* and *Z*. We use both measures of linkage disequilibrium for the analysis in this section.

Equations (2.4) give the realized fitnesses of each genotype when the population is initially in linkage equilibrium. Note that equations (2.4) express the frequency of each genotype

in terms of its allelic frequencies. If we substitute the genotype frequencies into the fitness equations, they become

$$\begin{aligned}
w_{ab} &= \frac{1}{1-c}\left(\frac{1}{2}Freq(ab) + \frac{1}{2+x}Freq(Ab) + \frac{1}{2+y}Freq(aB) + \frac{1}{2+x+y+xy}Freq(AB)\right) \\
w_{Ab} &= \frac{1}{1-c}\left(\frac{1+x}{2+x}Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1+x}{2+x+y}Freq(aB) + \frac{1}{2+y}Freq(AB)\right) \\
w_{aB} &= \frac{1}{1-c}\left(\frac{1+y}{2+y}Freq(ab) + \frac{1+y}{2+x+y}Freq(Ab) + \frac{1}{2}Freq(aB) + \frac{1}{2+x}Freq(AB)\right) \\
w_{AB} &= \frac{1}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}Freq(ab) + \frac{1+y}{2+y}Freq(Ab) + \frac{1+x}{2+x}Freq(aB) + \frac{1}{2}Freq(AB)\right)
\end{aligned} \tag{2.17}$$

From Appendix 1, we know that the frequencies of each genotype can be expressed in terms of the population's allelic frequencies, and the value of *D*.

$$\begin{aligned}
Freq(ab) &= q_1q_2 + D \\
Freq(Ab) &= p_1q_2 - D \\
Freq(aB) &= q_1p_2 - D \\
Freq(AB) &= p_1p_2 + D
\end{aligned} \tag{2.18}$$

Substituting these values into (2.17) gives the following fitness equations.

$$\begin{aligned}
w_{ab} &= \frac{1}{1-c}\left(\frac{1}{2}(q_1q_2 + D) + \frac{1}{2+x}(p_1q_2 - D) + \frac{1}{2+y}(q_1p_2 - D) + \frac{1}{2+x+y+xy}(p_1p_2 + D)\right) \\
w_{Ab} &= \frac{1}{1-c}\left(\frac{1+x}{2+x}(q_1q_2 + D) + \frac{1}{2}(p_1q_2 - D) + \frac{1+x}{2+x+y}(q_1p_2 - D) + \frac{1}{2+y}(p_1p_2 + D)\right) \\
w_{aB} &= \frac{1}{1-c}\left(\frac{1+y}{2+y}(q_1q_2 + D) + \frac{1+y}{2+x+y}(p_1q_2 - D) + \frac{1}{2}(q_1p_2 - D) + \frac{1}{2+x}(p_1p_2 + D)\right) \\
w_{AB} &= \frac{1}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}(q_1q_2 + D) + \frac{1+y}{2+y}(p_1q_2 - D) + \frac{1+x}{2+x}(q_1p_2 - D) + \frac{1}{2}(p_1p_2 + D)\right)
\end{aligned} \tag{2.19}$$

Here we estimate the value of epistasis over the parameter space p_1, p_2, D, x and y. We have already proven that when D=0, the value of epistasis is always negative. The question is whether the value of epistasis will remain negative when the population is not in linkage equilibrium.

Our approach was to analyze the resulting value of epistasis over a wide range of fixed values of p_1, p_2, x and y; while changing the value of D. In other words, we fixed the frequencies of the *A* and *B* alleles and their respective allelic effects, and varied the value of the initial linkage disequilibrium over its entire range. This was done repeatedly at different fixed values to provide an extensive coverage of the parameter space.

We also analyzed the results in terms of *Z* as our metric of linkage disequilibrium. By knowing the frequencies of the *A* and *B* alleles and the value of *D*, we can obtain the genotypic frequencies using equations (2.18). And by knowing the genotypic frequencies, we can obtain the value of *Z* using equation (2.15). We find that the results are easier to obtain in terms of *D*, but the trends are more apparent in terms of *Z*.

For all the parameter values we tested, the result was always negatively epistatic. The results clearly show that a higher value of initial linkage disequilibrium results in a value of epistasis closer to zero (see Figures 2.7-2.10). In can also be seen that phenotypic effects play a greater role in determining the value of epistasis than initial linkage disequilibrium does (see Figure 2.7 and 2.9). When the phenotypic effects of the *A* and *B* alleles are small (i.e. less than 1), initial linkage disequilibrium does little to change the value of epistasis. Finally, we observe that initial linkage disequilibrium has a greater power to change the value of epistasis when the allelic frequencies of *A* and *B* are close 0.5 (See Figures 2.8 and 2.10).

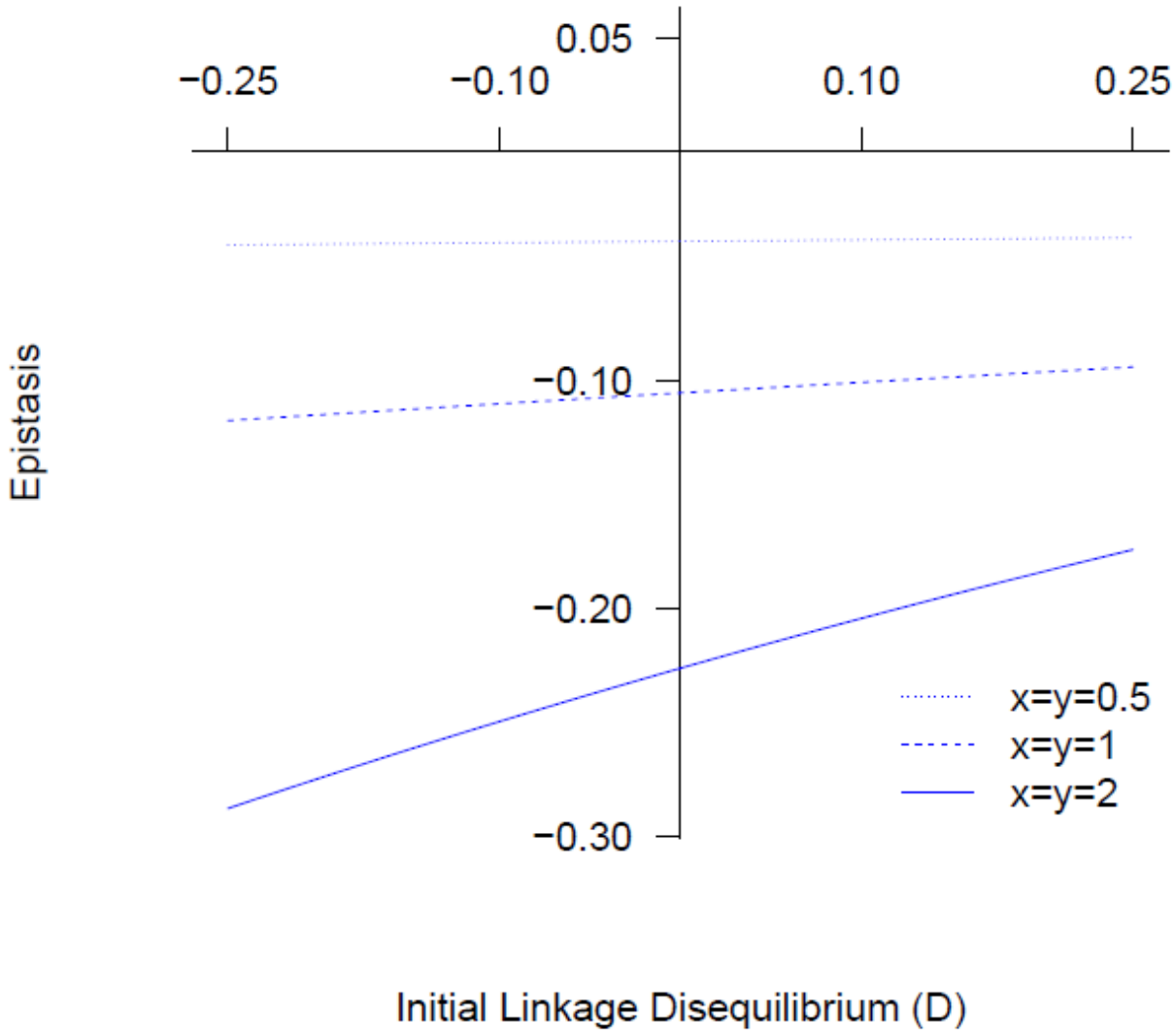


Figure 2.7: The relationship between the initial value of linkage disequilibrium (*D*), and the value of epistasis with different phenotypic effects.

The value of epistasis is shown on the vertical axis, and the initial value of linkage disequilibrium is shown on the horizontal axis. For simplicity, we assume that the frequencies of all four alleles are initially 0.5; and that the effects of the *A* and *B* alleles on the phenotype, *x* and *y*, are equal. Three lines are drawn to show the change in the value of epistasis with different values of the phenotypic effects, *x* and *y*: where they are both equal to 0.5, where they are both equal to 1, and where they both equal 2. Note that an increase in initial linkage disequilibrium increases the value of epistasis, but that the value of epistasis is always negative. Also note that increasing the phenotypic effects decreases the value of epistasis.

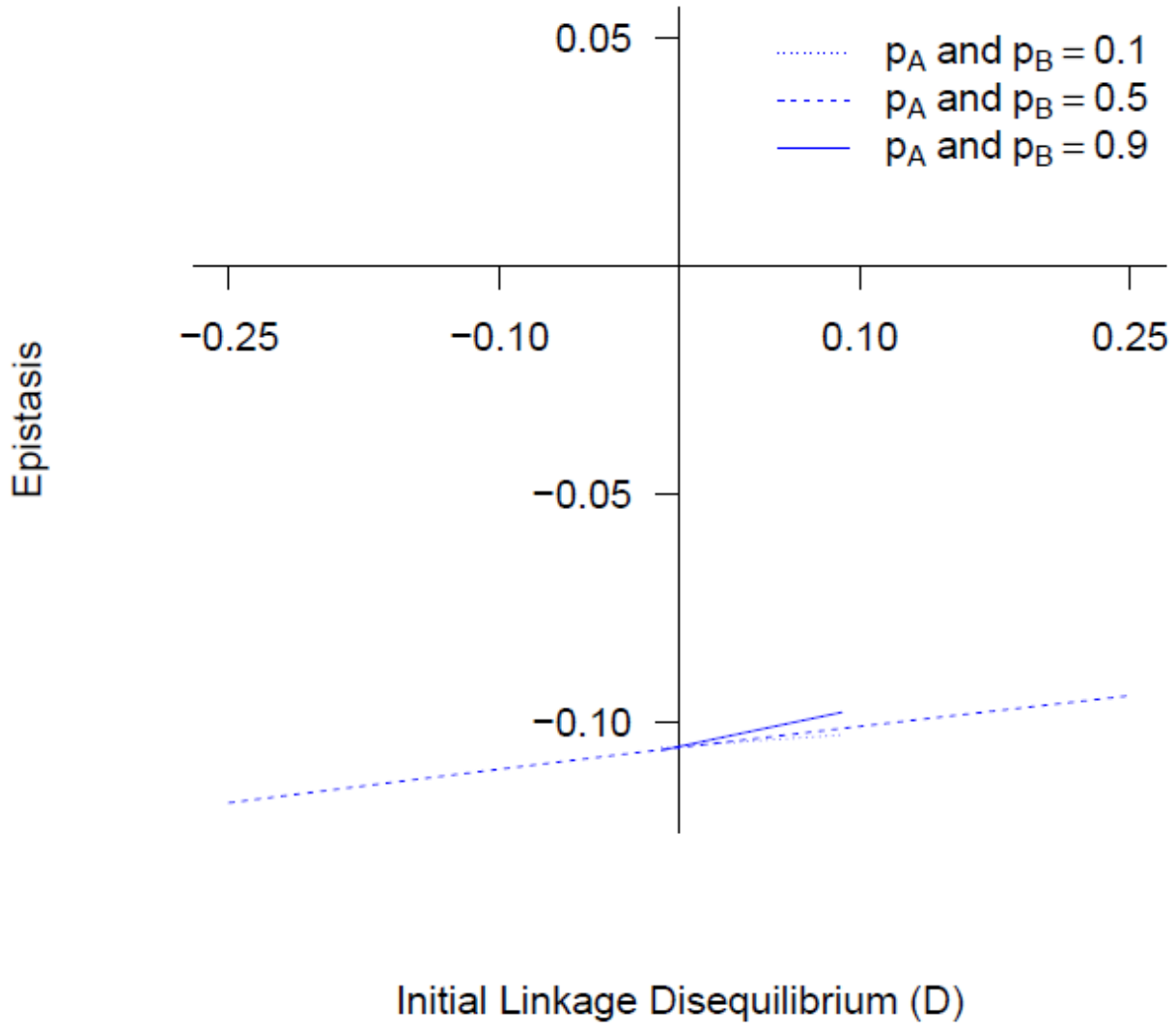


Figure 2.8: The relationship between the initial value of linkage disequilibrium (*D*), and the value of epistasis with different frequencies of the *A* and *B* alleles.

The value of epistasis is shown on the vertical axis, and initial value of linkage disequilibrium is shown on the horizontal axis. For simplicity, we assume that the frequencies the *A* and *B* alleles are initially equal; and that their respective effects on the phenotype, *x* and *y*, are equal to 1.

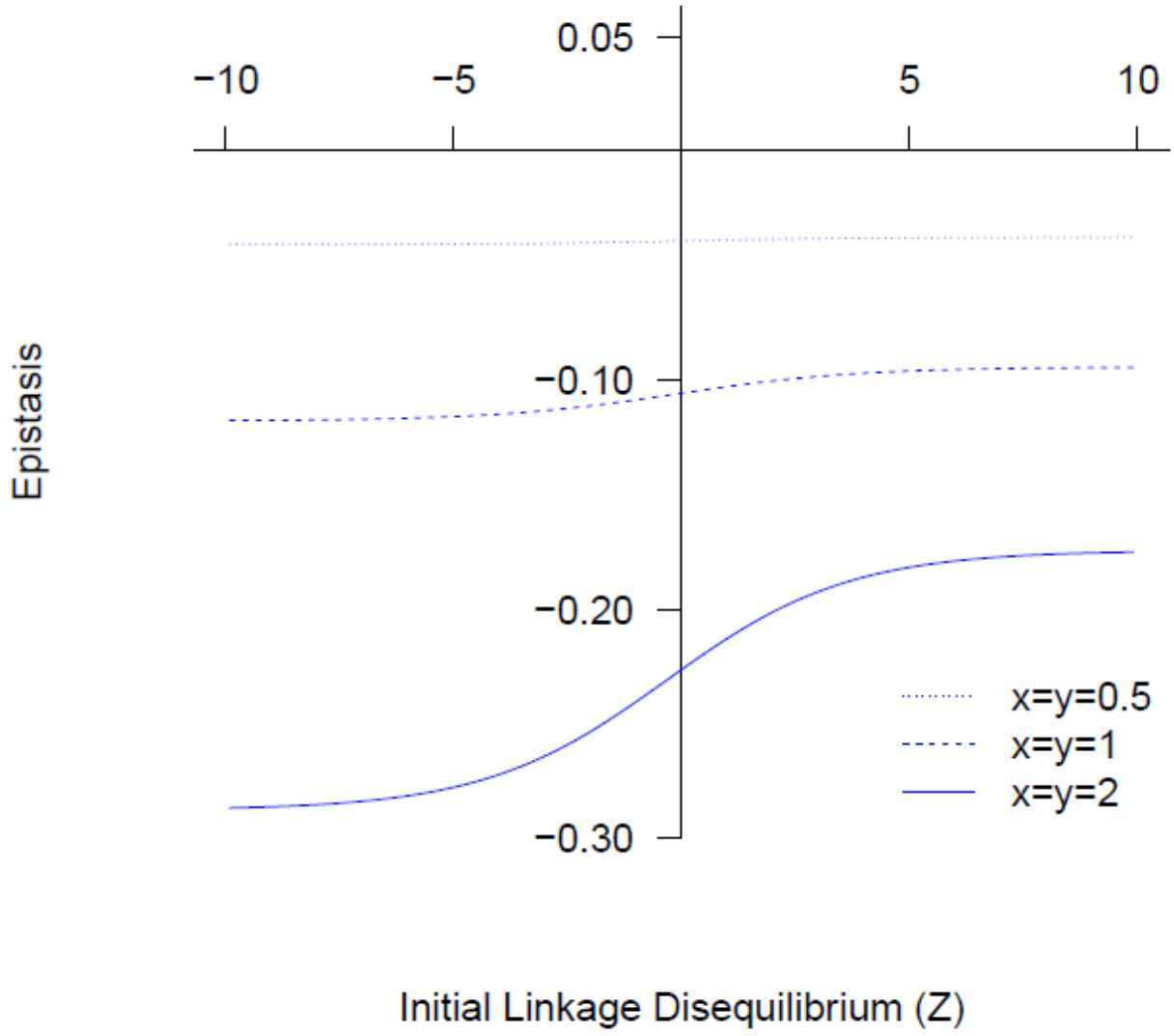


Figure 2.9: The relationship between the initial value of linkage disequilibrium (*Z*), and the value of epistasis with different phenotypic effects.

For a description, see Figure 2.8. *Z* is used to quantify linkage disequilibrium in this case instead of *D*.

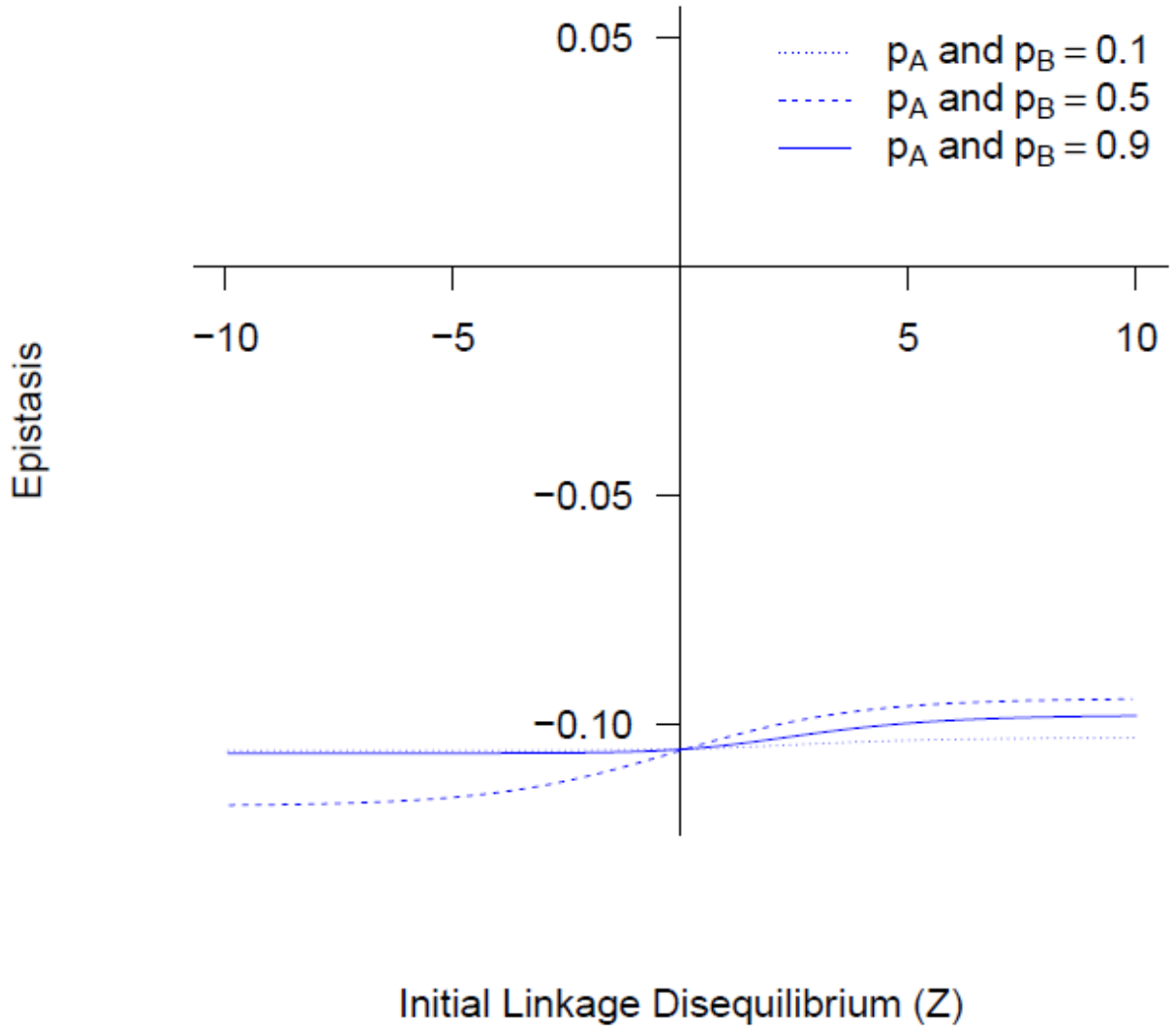


Figure 2.10: The relationship between the initial value of linkage disequilibrium (*Z*), and the value of epistasis with different frequencies of the *A* and *B* alleles.

For a description, see Figure 2.9. *Z* is used to quantify linkage disequilibrium in this case instead of *D*. Note that the value of epistasis has a greater range when the allele frequencies are close to 0.5.

2.6 The effect of negatively-epistatic fitnesses on genotype frequencies.

Now that we have shown that intraspecies competition results in negatively-epistatic fitnesses, our next goal was to determine how these fitnesses affect genotype frequencies. Specifically, we want to know if the resulting genotype frequencies after selection are in negative linkage disequilibrium, and how their frequencies will be altered through random mating and free recombination.

As stated in section 2.2, the frequency of each genotype post-competition can be expressed as follows:

$$\begin{aligned}
Freq(ab)' &= \frac{Freq(ab)}{1-c}\left(\frac{1}{2}Freq(ab) + \frac{1}{2+x}Freq(Ab) + \frac{1}{2+y}Freq(aB) + \frac{1}{2+x+y+xy}Freq(AB)\right) \\
Freq(Ab)' &= \frac{Freq(Ab)}{1-c}\left(\frac{1+x}{2+x}Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1+x}{2+x+y}Freq(aB) + \frac{1}{2+y}Freq(AB)\right) \\
Freq(aB)' &= \frac{Freq(aB)}{1-c}\left(\frac{1+y}{2+y}Freq(ab) + \frac{1+y}{2+x+y}Freq(Ab) + \frac{1}{2}Freq(aB) + \frac{1}{2+x}Freq(AB)\right) \\
Freq(AB)' &= \frac{Freq(AB)}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}Freq(ab) + \frac{1+y}{2+y}Freq(Ab) + \frac{1+x}{2+x}Freq(aB) + \frac{1}{2}Freq(AB)\right)
\end{aligned} \tag{2.20}$$

where $Freq(ab)'$ is the frequency of the *ab* genotype after competition, and similarly for the other genotypes

We observed the value of linkage disequilibrium both after competition and after random mating with free recombination over a wide range of parameter values. When beginning in linkage equilibrium, we found that competition always caused the genotype frequencies to become in negative linkage disequilibrium (see Figures 2.11 and 2.12). The values of linkage disequilibrium (*D* and *Z*) are most negative with strong phenotypic effects of the *A* and *B* alleles, with *Z* being more resilient to changes in allele frequencies. It is also worth noting that the value of *Z* after selection is equal to the value of ε when the population begins in linkage equilibrium (Felsenstein, 1965).

By knowing the frequencies of each genotype post-competition, we can determine the value of *D* using equation (2.15), and we can determine the value of *Z* using equation (2.16).

With random mating and free recombination following competition, the genotype frequencies change as follows (see Appendix 2 for the derivation):

$$\begin{aligned} Freq(ab)'' &= Freq(ab)' - ½D \\ Freq(Ab)'' &= Freq(Ab)' + ½D \\ Freq(aB)'' &= Freq(aB)' + ½D \\ Freq(AB)'' &= Freq(AB)' - ½D \end{aligned} \tag{2.21}$$

where $Freq(ab)''$ is the frequency of the *ab* genotype after recombination, and $Freq(ab)'$ is the frequency of the *ab* genotype before recombination, and similarly for the other genotypes. We can determine the values of *D* and *Z* after recombination by once again plugging the genotype frequencies into equations (2.15) and (2.16), respectively.

From Figures 2.11 and 2.12, we can see that the value of *D* is halved after random mating with free recombination (see Appendix 2), while *Z* is approximately halved. Because the population is in negative linkage disequilibrium after selection, recombination has the effect of bringing the population closer to linkage equilibrium. This increase in the value of linkage disequilibrium, indirectly, means that phenotypic and fitness variances are increased when individuals freely recombine.

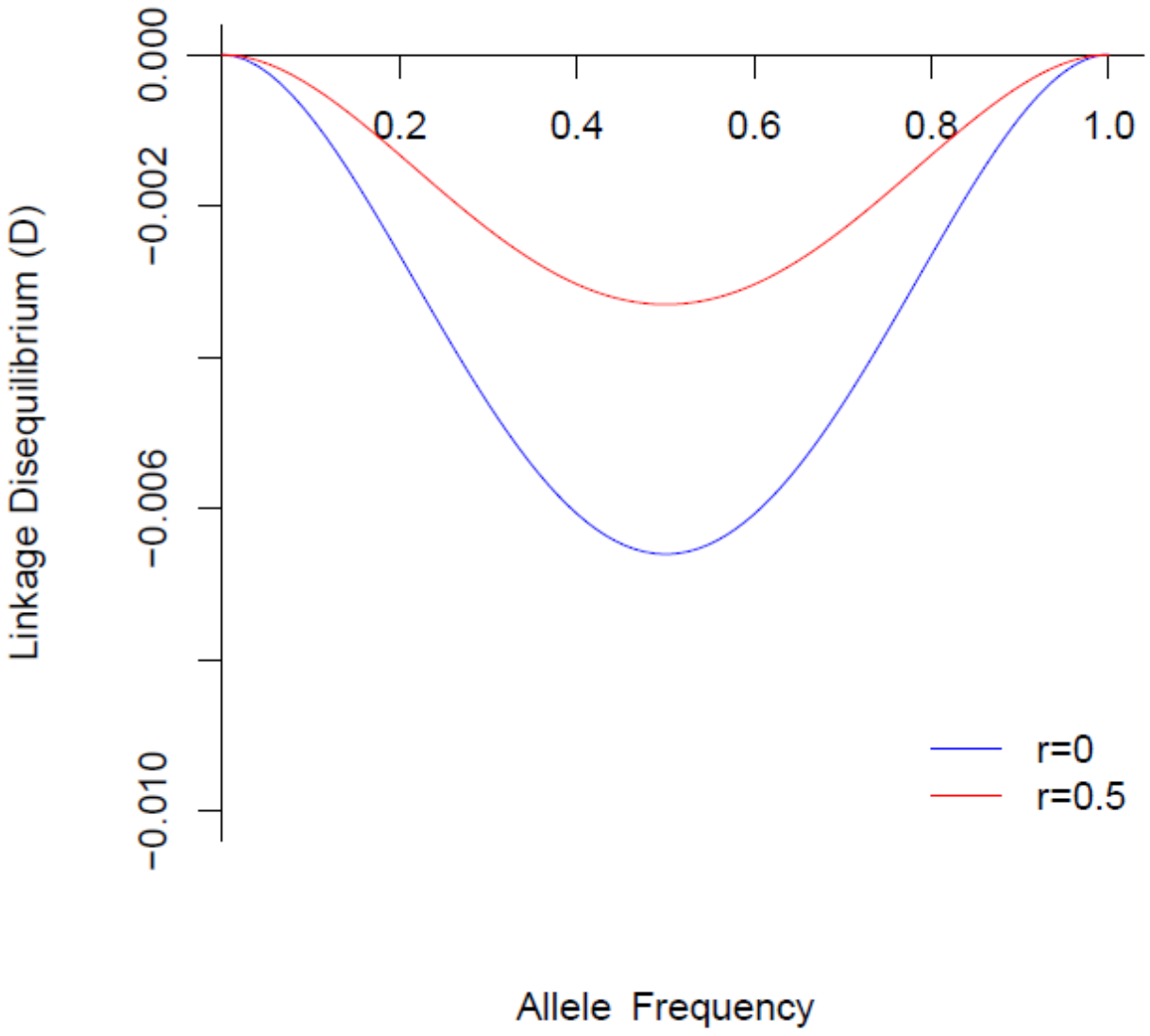


Figure 2.11: The relationship between the value of linkage disequilibrium (*D*) after selection, and allele frequency.

The value of *D* is shown on the vertical axis, while the frequencies of the *A* and *B* alleles after selection (Allele Frequency) are shown on the horizontal axes. For simplicity, the frequencies of the *A* and *B* alleles are assumed to be equal, and the population is initially in linkage equilibrium. Two lines are plotted: the blue line describes the change in the value of *D* at different allele frequencies of *A* and *B* when there is no recombination (r=0), while the red line describes the change in the value of *D* if individuals can recombine freely (r=0.5). Note that the value of *D* is always negative, and has a minimum value when the allele frequencies are equal to 0.5.

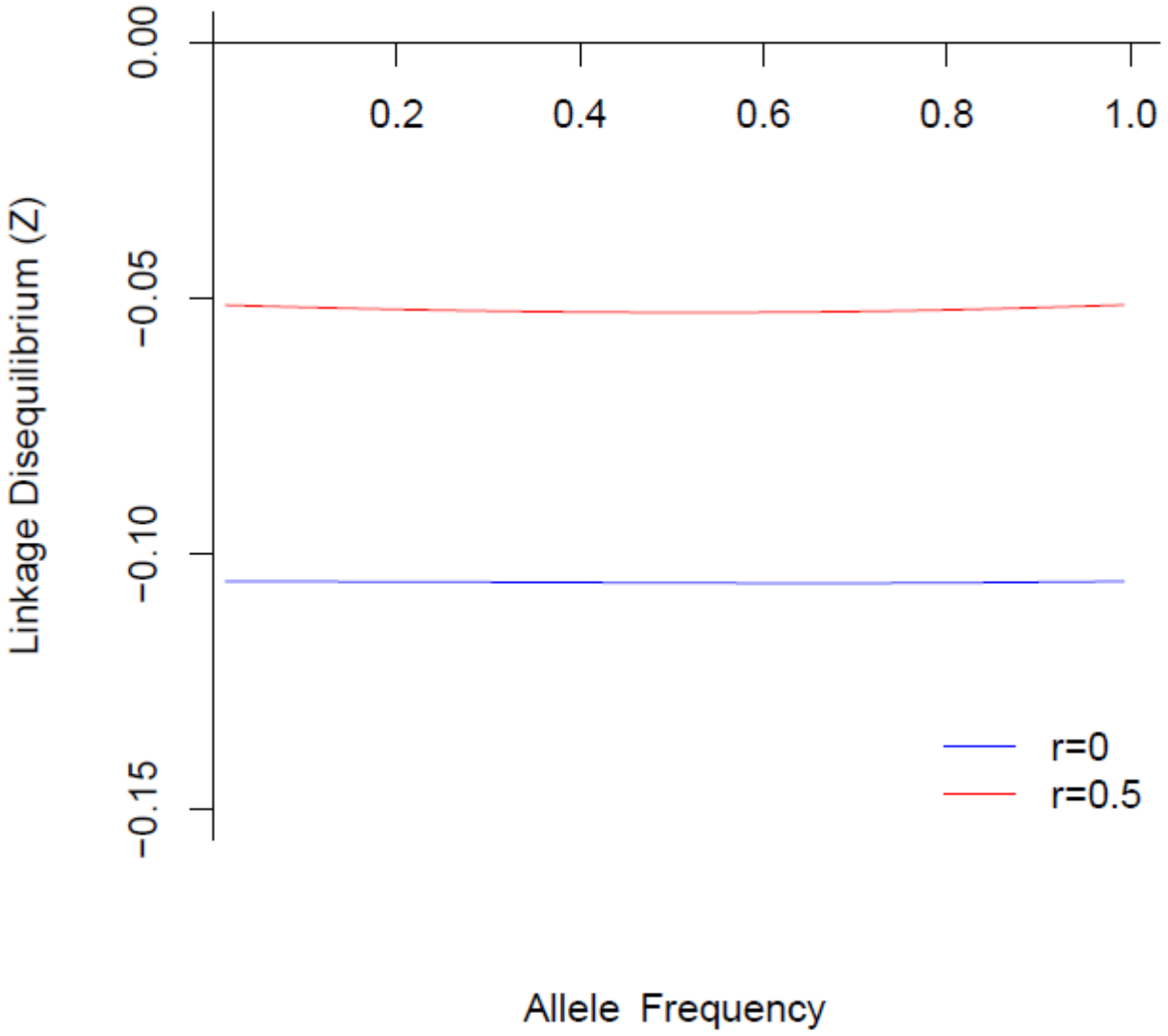


Figure 2.12: The relationship between the value of linkage disequilibrium (*Z*) after selection, and allele frequency.

For a description, see Figure 2.11. *Z* is used to quantify linkage disequilibrium in this case instead of *D*. Note that the value of *Z* is far less frequency dependent than the value of *D*, and that when there is no recombination (r=0), the value of *Z* is equal to the value of epistasis (see Figure 2.3 where x=y=1).

2.7 The effect of genetic recombination on the phenotypic mean and variance.

We have shown that a population initially in linkage equilibrium will be in negative linkage disequilibrium after competition. The value of negative linkage disequilibrium will be halved following a round of random mating with free recombination. Here we discuss how this change in the genotype frequencies affects the mean and variance of the phenotypic values.

We obtain the genotype frequencies after competition with equations (2.20), and their frequencies following a round of random mating with free recombination using equations (2.21). At each of point in time we can compute the mean by summing the individual phenotype values, weighted by their frequencies within the population.

$$V_{ave} = V_{ab}Freq(ab) + V_{Ab}Freq(Ab) + V_{aB}Freq(aB) + V_{AB}Freq(AB) \qquad (2.22)$$

Similarly, we can obtain the variance of the phenotypic values by summing the squared difference between the each genotype's phenotypic value and the phenotypic mean, weighted by its frequency within the population.

$$\begin{aligned} Var(V) = {} & Freq(ab)(V_{ab} - V_{ave})^2 + Freq(Ab)(V_{Ab} - V_{ave})^2 \\ & + Freq(aB)(V_{aB} - V_{ave})^2 + Freq(AB)(V_{AB} - V_{ave})^2 \end{aligned} \qquad (2.23)$$

We computed the phenotypic mean and variance following both competition and random mating over a large region of the parameter space. In all cases tested, we found that the mean is slightly increased following a round of random mating, and the increase is greatest when the frequencies of the *A* and *B* alleles are close to 0.5 (see Figures 2.13 and 2.14). As might be expected, given that the population is in negative linkage disequilibrium, the phenotypic variance is increased after free recombination (see Figure 2.15). This increase is greatest at intermediate frequencies of the *A* and *B* alleles.

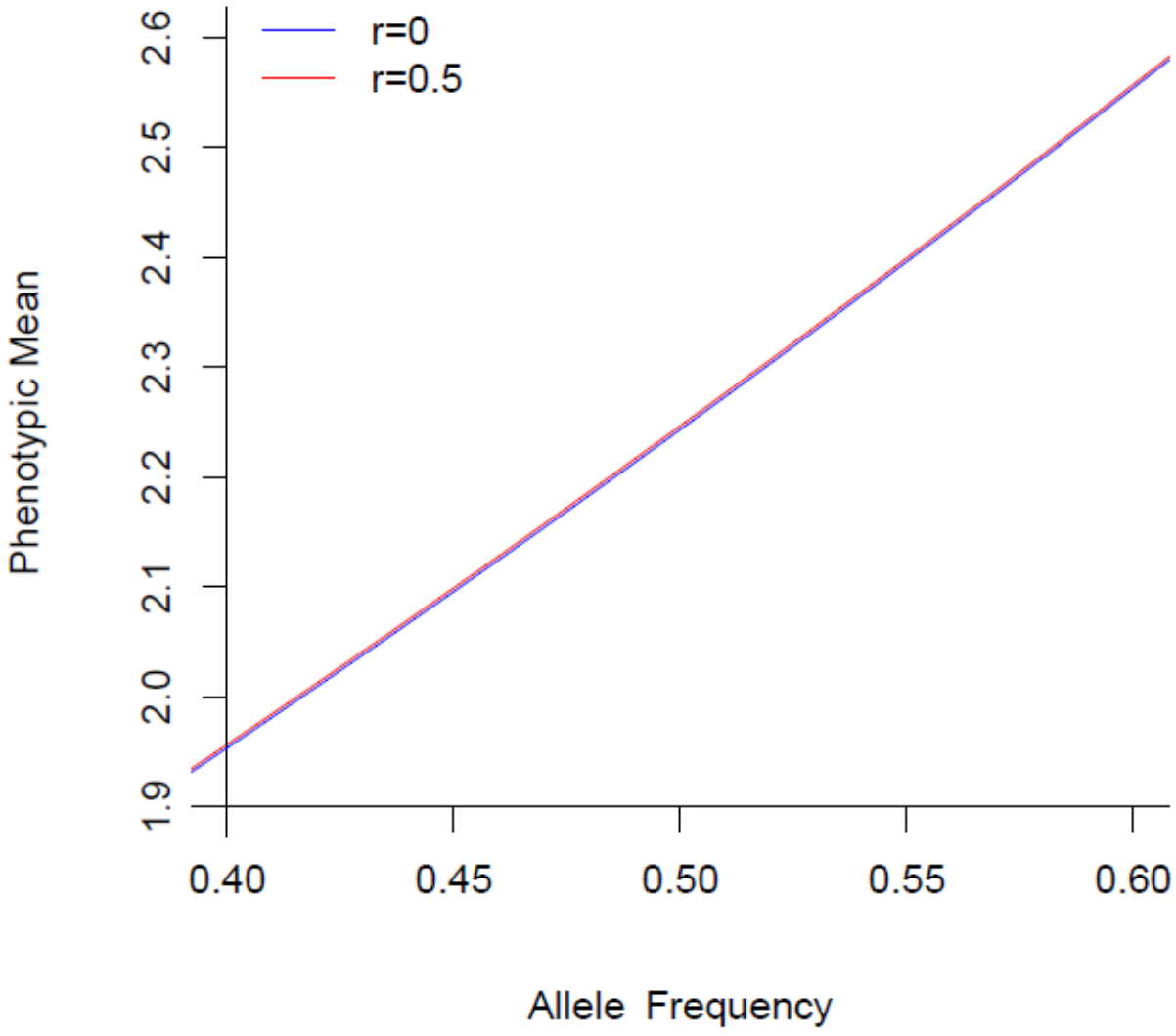


Figure 2.13: The relationship between the phenotypic mean after selection and mating, and allele frequency.

The phenotypic value of each genotype is the same as shown in Figure 2.1.1, and the frequency of the *A* and *B* alleles is shown on the horizontal axis. Two lines are plotted: when there is no recombination after selection (r=0), and when there is free recombination (r=0.5). Note that the mean phenotype after recombination is slightly higher than when there is no recombination.

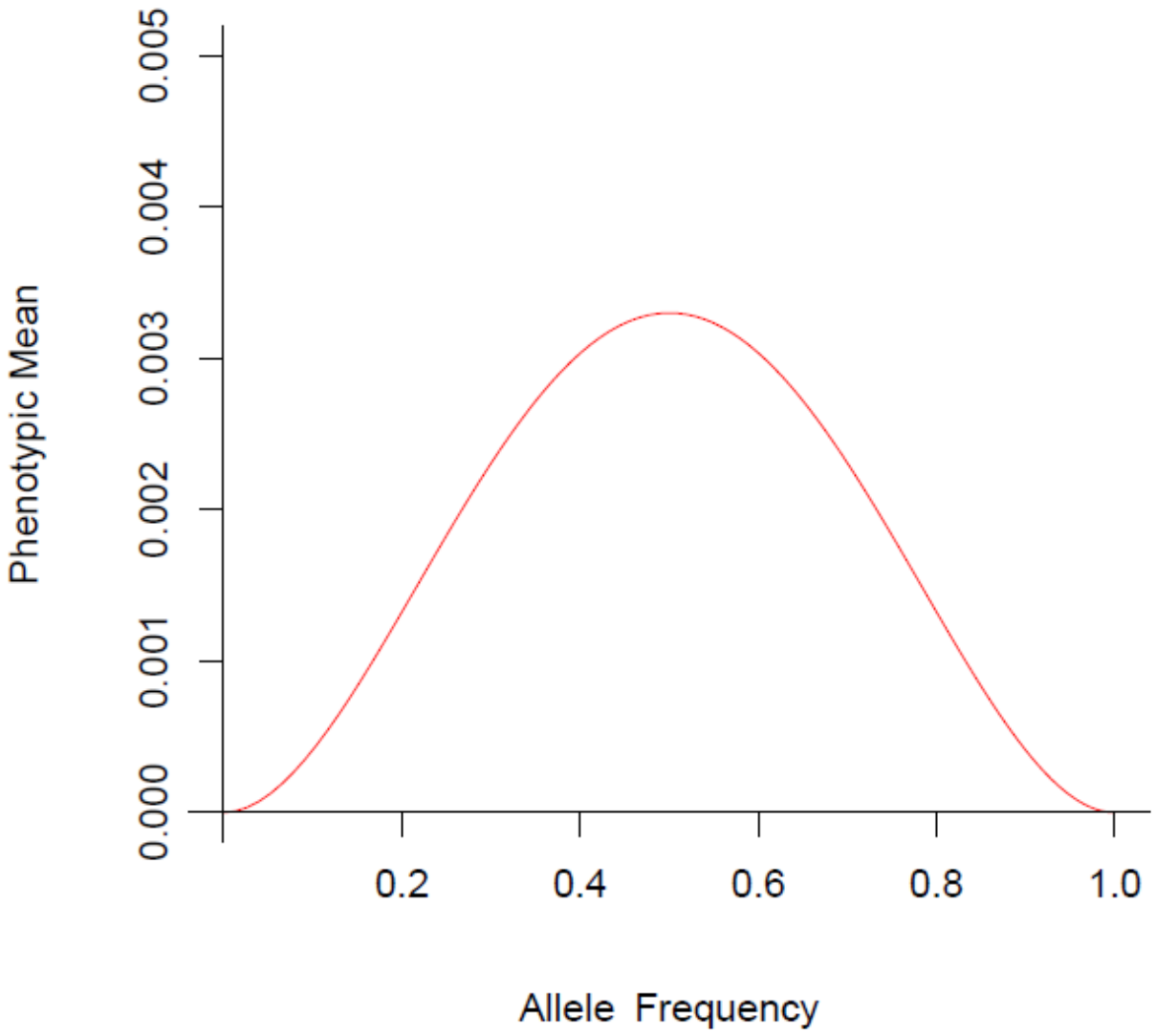


Figure 2.14: The relationship between the difference in the phenotypic mean between free recombinants and non-recombinants, and allele frequency.

For a description, see Figure 2.13. The difference is calculated as the mean phenotype following random mating with free recombination minus the mean phenotype without recombination. Note that the difference is greater than zero over all allele frequencies.

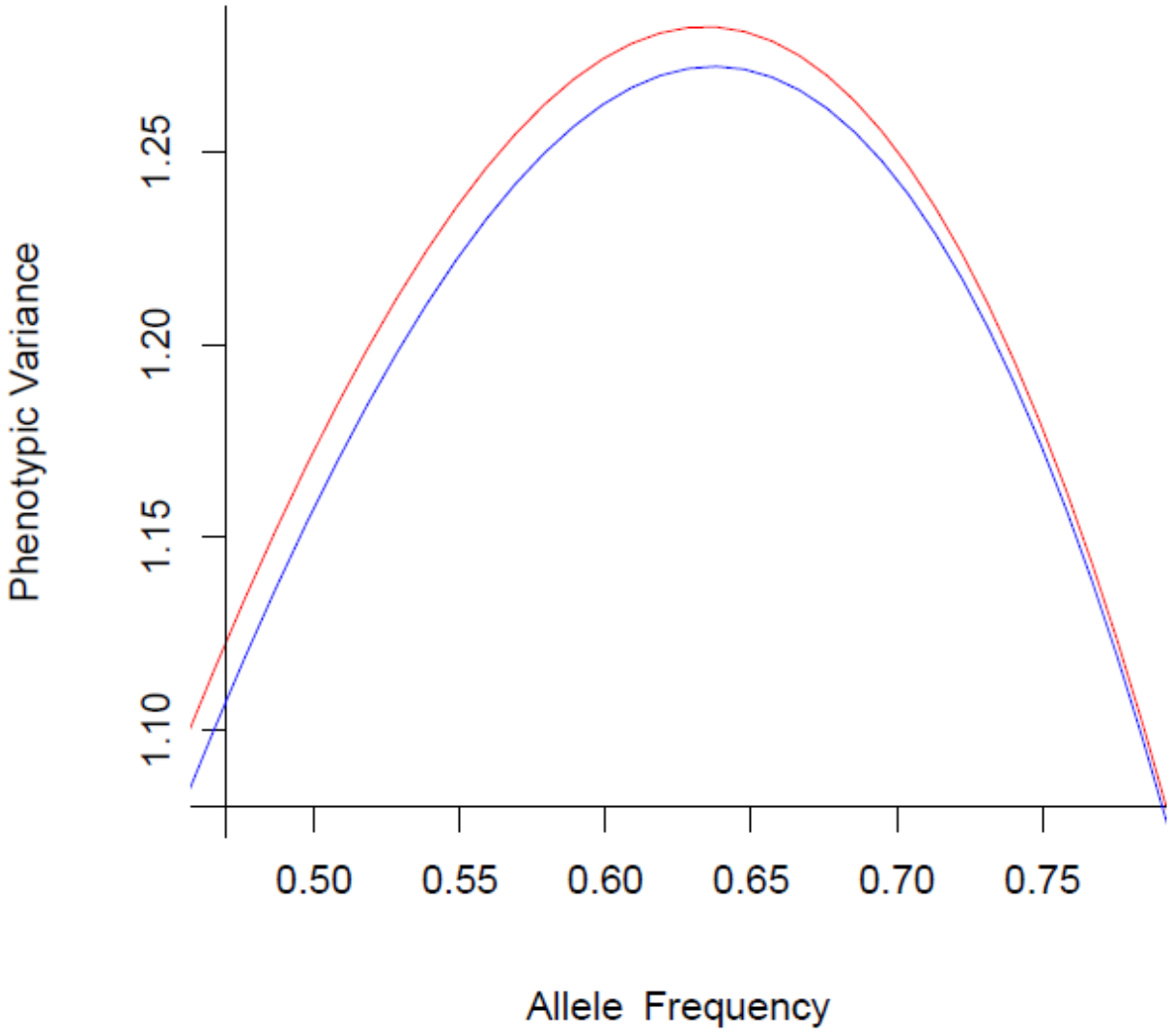


Figure 2.15: The relationship between phenotypic variance and allele frequency.

The phenotypic value of each genotype is the same as in Figure 2.1.1, and the frequency of the *A* and *B* alleles is shown on the horizontal axis. Two lines are plotted: when there is no recombination after selection (r=0), and when there is free recombination (r=0.5). Note that the variance is greater with free recombination.

2.8 The effect of genetic recombination on fitness.

With both the phenotypic mean and variance being increased when individuals randomly mate and freely recombine, we next studied how these changes affect the mean and variance of fitness in the following generation.

We have previously derived the genotypic fitness values when beginning in linkage equilibrium in equations (2.4). We then determined the genotype frequencies following competition using equations (2.20). These frequencies will not change when individuals reproduce in the absence of recombination, as each genotype has the same mean family size. When randomly mating with free recombination, however, the genotype frequencies change according to equations (2.21). Therefore, the genotype frequencies before competing in the next generation are known both in the presence and absence of free recombination. We then determine the genotype fitnesses, now that their frequencies are out of linkage equilibrium, using equations (2.17).

Whether freely recombinant or non-recombinant, we can compute the mean fitness by summing the individual fitness values, weighted by their frequencies within the population. This gives

$$w_{ave} = w_{ab}Freq(ab) + w_{Ab}Freq(Ab) + w_{aB}Freq(aB) + w_{AB}Freq(AB) \qquad (2.24)$$

Similarly, the variance of fitness is calculated by summing the squared difference between each genotype's fitness and the mean fitness, weighted by its frequency within the population, that is

$$\begin{aligned} Var(w) = {} & Freq(ab)(w_{ab} - w_{ave})^2 + Freq(Ab)(w_{Ab} - w_{ave})^2 \\ & + Freq(aB)(w_{aB} - w_{ave})^2 + Freq(AB)(w_{AB} - w_{ave})^2 \end{aligned} \qquad (2.25)$$

As expected, given that the population size is held constant before and after selection, the mean fitness both with and without free recombination is equal to 1 (see Figure 2.16). However,

the variance of fitness is slightly higher when individuals freely recombine (see Figures 2.17 and 2.18). The variance of fitness increases with increasing phenotypic effects of the *A* and *B* alleles, and when the frequencies of these alleles are close to 0.5.

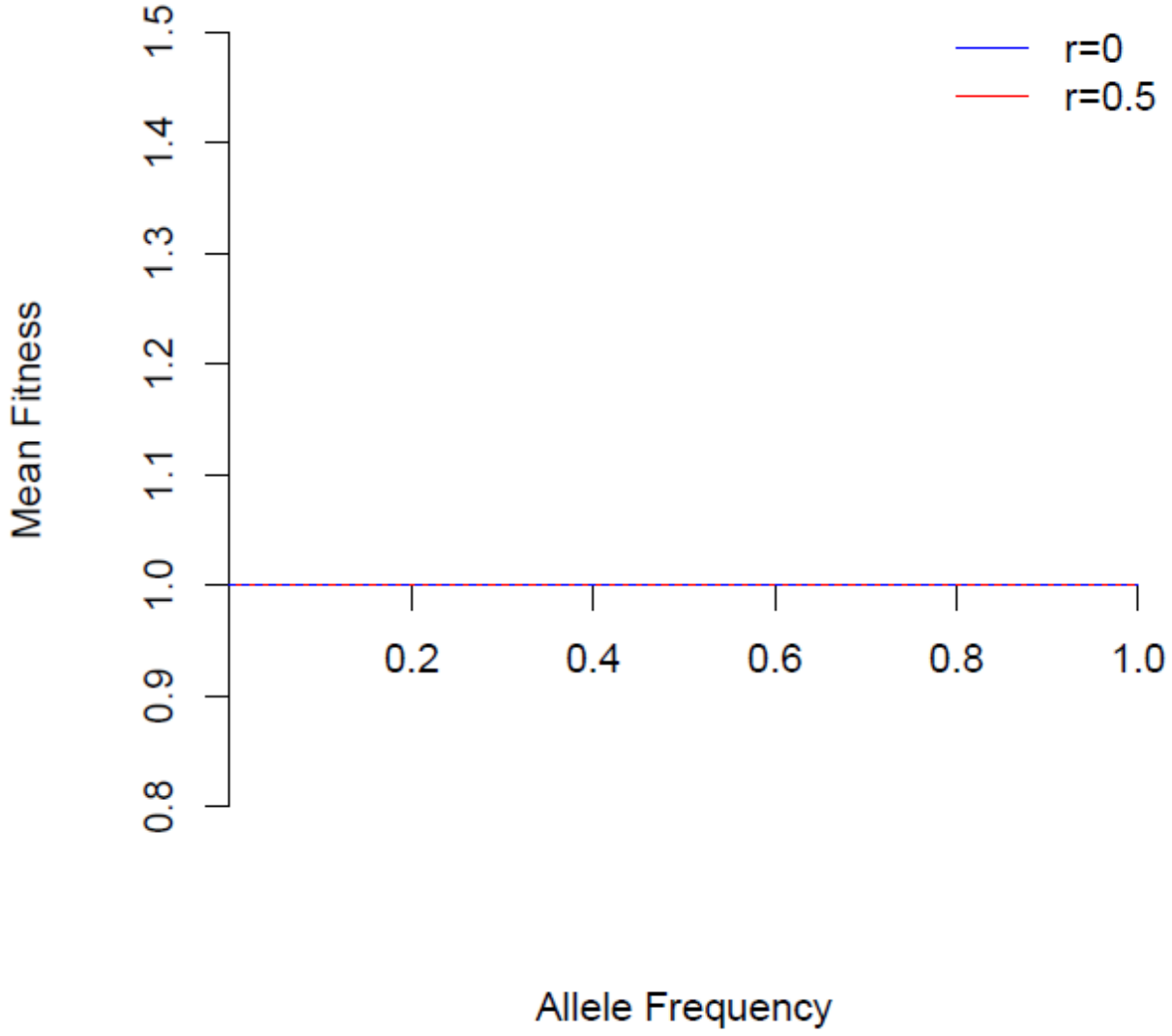


Figure 2.16: The relationship between the mean fitness and allele frequency.

The mean fitness of the population is shown when individuals either do not recombine (r=0), or recombine freely (r=0.5). The frequency of the *A* and *B* alleles is shown on the horizontal axis. In all cases, whether freely-recombinant of non-recombinant, the mean fitness of the population is equal to 1.

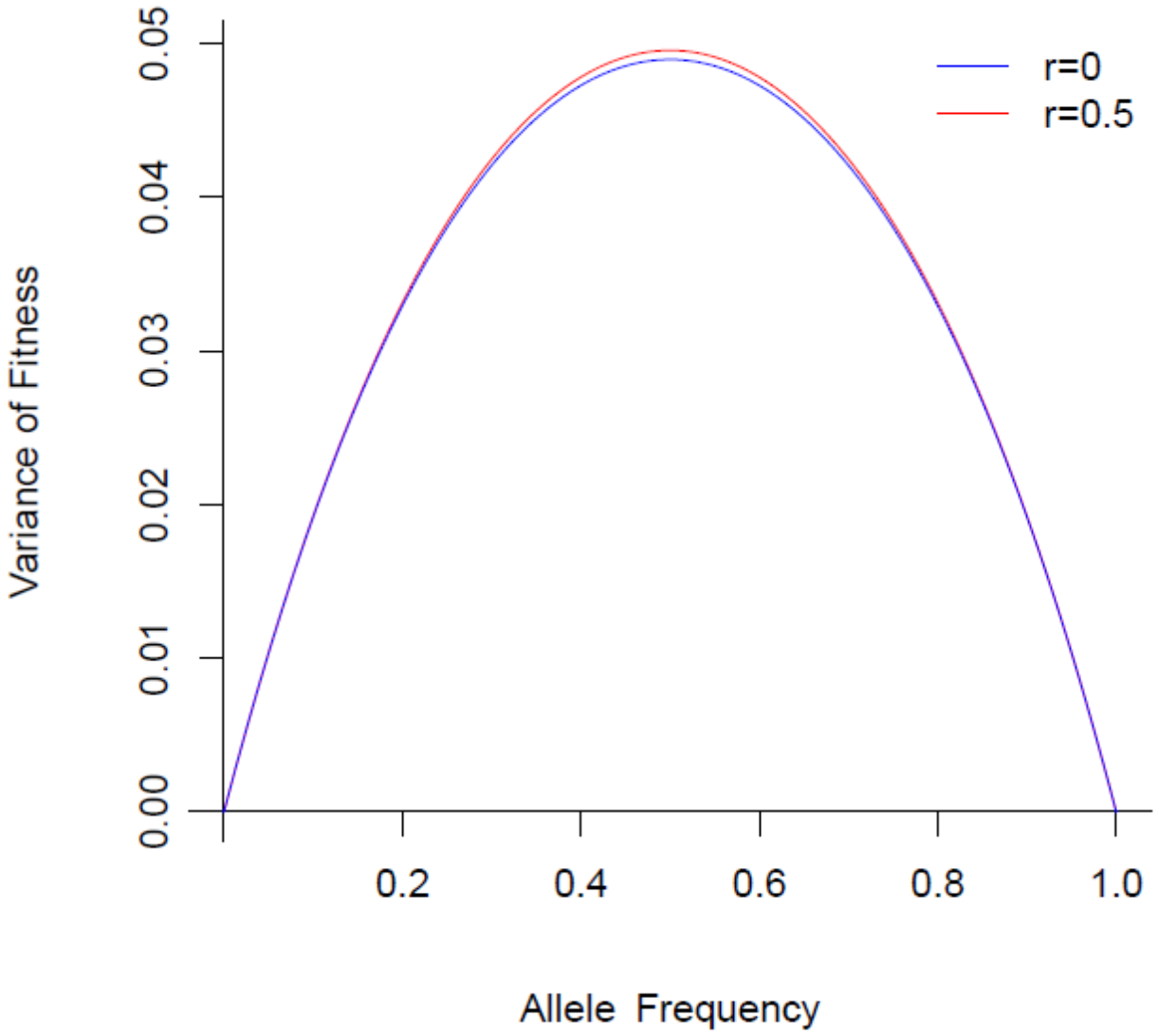


Figure 2.17: The relationship between the variance of fitness and allele frequency.

The variance in fitness is shown when individuals are either non-recombinant (r=0), or recombine freely (r=0.5). The genotypic fitnesses were calculated based on the phenotypic values in Figure 2.1.1. The frequency of the *A* and *B* alleles is shown on the horizontal axis. Note that the variance in fitness is higher with recombination.

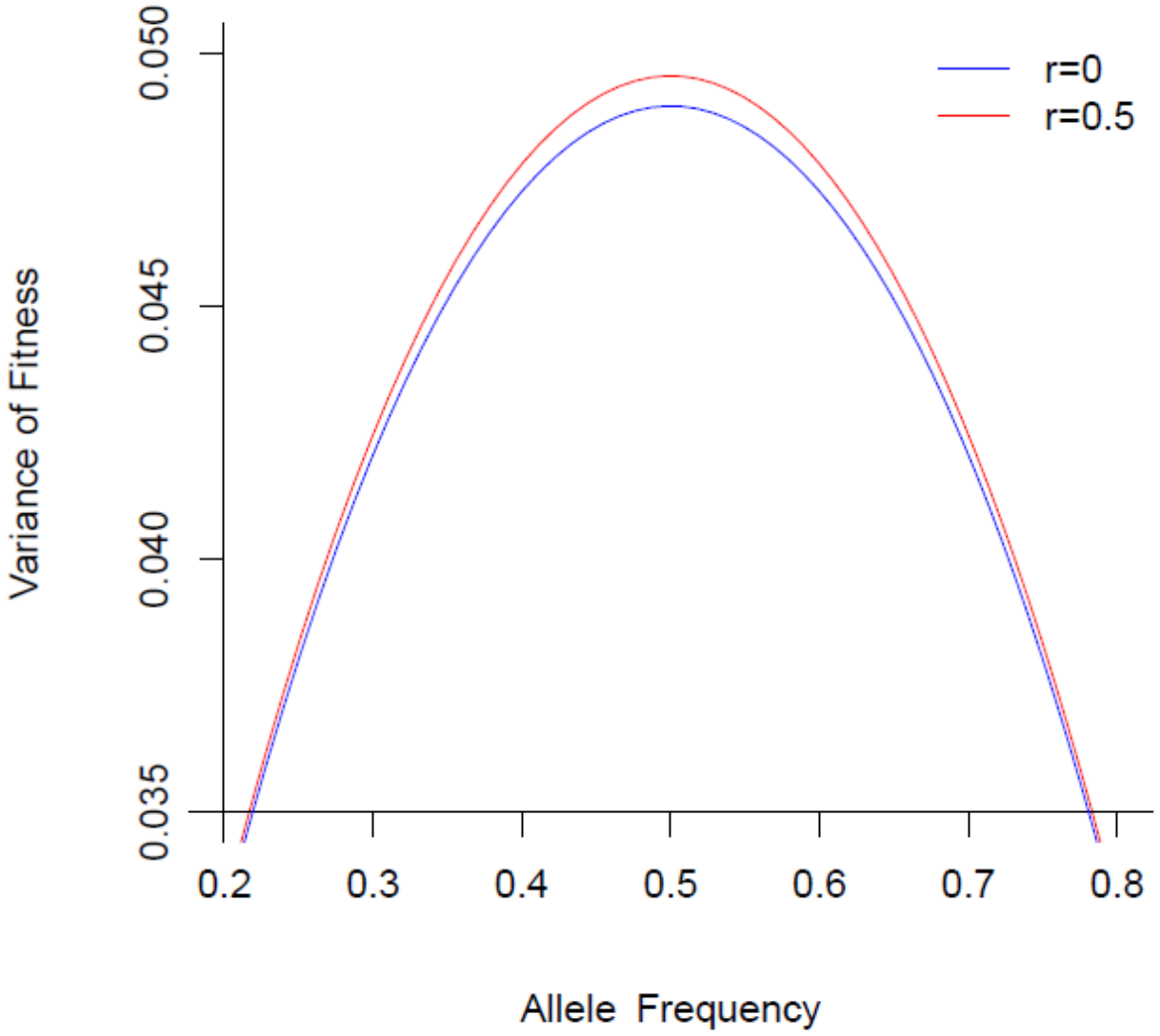


Figure 2.18: The relationship between the variance of fitness and allele frequency when the allele frequencies are close to being equal.

For a description, see Figure 2.17. This Figure highlights the region where the allele frequencies are close to 0.5. The increase in the variance of fitness with free recombination is clearly visible.

2.9 Epistasis at the phenotypic level.

We have shown that multiplicative effects (a.k.a. non-epistatic) on the phenotype result in negative epistasis at the level of fitness. However, phenotypic effects are not always multiplicative. When phenotypic effects are less than multiplicative, we say that there is negative epistasis at the phenotype level; and when the effects are more than multiplicative, we say that there is positive epistasis at the phenotype level. In this section, we study how epistatic effects at the phenotype level affect the value of epistasis at the fitness level.

To show this relationship, we extend our previous model to allow for a deviation from multiplicative effects on the phenotype. We introduce a variable, ϵ, that only affects the phenotypic value of the AB genotype.

Genotype	ab	aB	Ab	AB
Phenotypic Value (V)	1	$1+y$	$1+x$	$(1+x)(1+y)+\epsilon$

Note that when ϵ is equal to 0, there is no epistasis at the phenotype level; when ϵ is negative, there is negative epistasis; and when ϵ is positive, there is positive epistasis.

To illustrate how epistasis at the phenotype level affects epistasis at the fitness level, we will assume that both x and y are equal to 1, and that ϵ equals either -1, 0, 0.5, or 1. The possible phenotypic values of the four genotypes are summarized in the table below.

Table 2.3 The possible phenotypic values of the four genotypes.

Genotype	ab	aB	Ab	AB
Phenotypic Value (V)	1	2	2	3
	1	2	2	4
	1	2	2	4.5
	1	2	2	5

Note that the phenotypic effects are additive in the first row of the table where V_{AB} equals 3.

We insert these phenotypic values into equations (2.3) to obtain the fitness values when beginning with a population in linkage equilibrium. By knowing these values, we can determine the value of fitness level epistasis.

Figure 2.19 shows the results in all four cases of the table above. Overall, we generally find that there is negative epistasis at the fitness level. This result is robust to a wide array of phenotypic effects. In order for intraspecies competition to cause positive epistasis, there needs to be strong positive epistasis at the phenotype level, coupled with the *A* and *B* alleles occurring at high frequencies. Compared to when phenotypic effects are multiplicative, the value of fitness level epistasis is increased when there is positive phenotype level epistasis; and the value of fitness level epistasis is decreased when there is negative phenotype level epistasis. These changes are most pronounced when the frequencies of the *A* and *B* alleles are close to fixation.

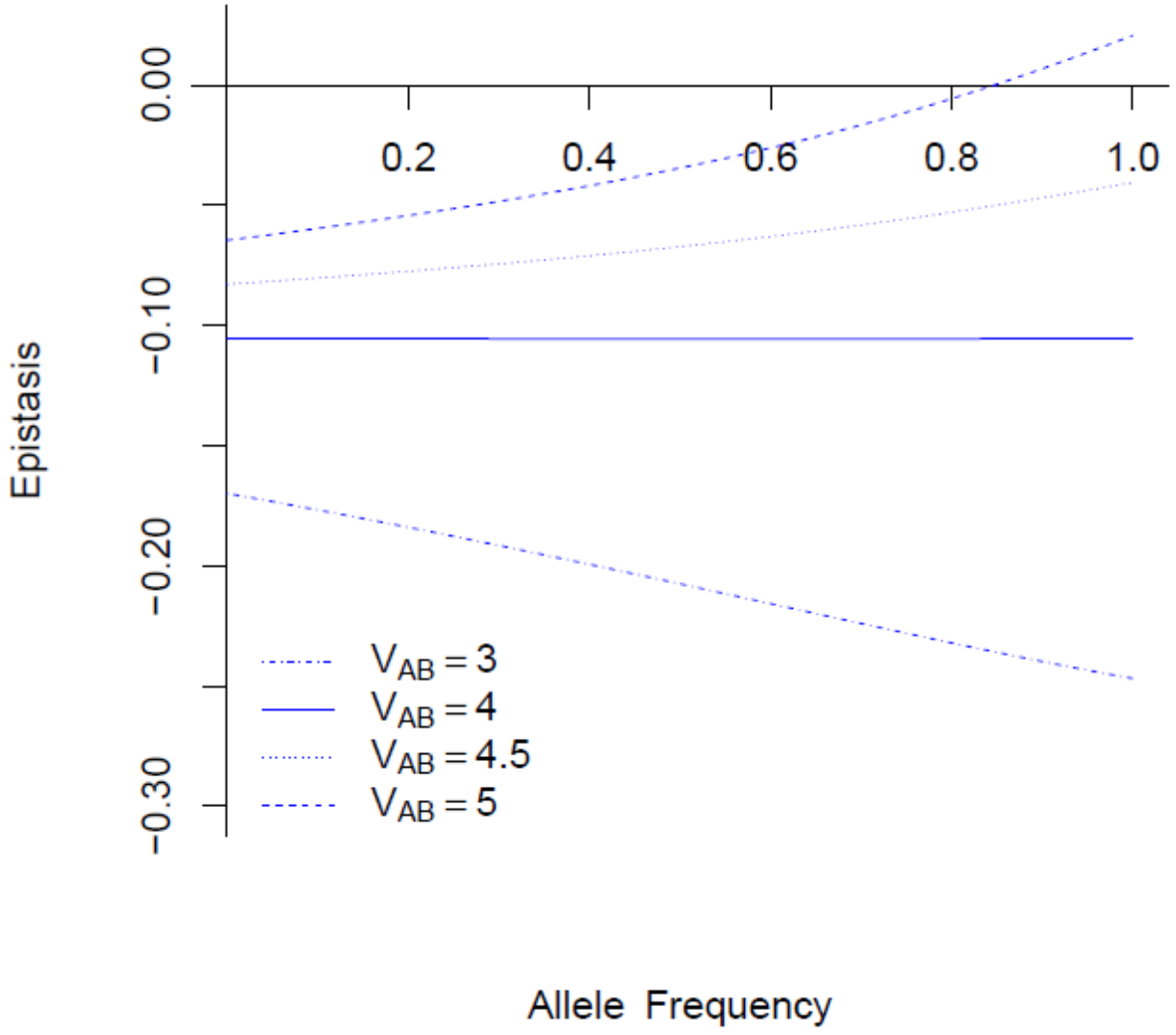


Figure 2.19: The relationship between the value of epistasis (fitness level), allele frequency, and phenotypic epistasis.

The frequency of the *A* and *B* alleles is shown on the horizontal axis, and the value of epistasis at the fitness level is shown on the vertical axis. For simplicity, we assume that $\underline{V_{ab}}=1$, V_{Ab} and V_{aB} are equal to 2, and the population is initially in linkage equilibrium. When $V_{AB}=3$, there is negative epistasis at the phenotype level (and additive phenotypic effects); when $V_{AB}=4$, there is no epistasis at the phenotype level; and when V_{AB} =4.5 or 5, there is positive epistasis at the phenotype level. Note that the value of epistasis at the fitness level is almost always negative.

2.10 Discussion

Together, the results presented here show that intraspecific pairwise competition generates negative epistasis at the level of realized fitness. Intraspecies competition generates negative epistasis regardless of the phenotypic effects and frequencies of the selected alleles (provided that their effects are multiplicative), and regardless of the initial genotype frequencies. This result also holds true when the phenotypic values on which the competition is based are themselves multiplicative; less than multiplicative; and to an extent, more than multiplicative. It is especially interesting to note that positive epistasis at the phenotypic level can translate into negative epistasis at the fitness level.

We have mainly focused on the case where the population is initially in linkage equilibrium, as this is the expected state of a population having undergone random mutations in the absence of selection. The only times we noticed that the population may not be in negative linkage disequilibrium following selection is when beginning in strongly positive linkage disequilibrium, or when phenotypic effects are strongly positively epistatic. In all other cases, recombination acts to reduce the negative linkage disequilibrium that results from the negatively epistatic fitnesses which are generated by competition. Recombination increases the phenotypic mean and variance, increases the variance of fitness, but does not affect the mean fitness of the population. An increase in the genetic variance of fitness, as stated by Fisher (1930), increases the efficiency of natural selection. Therefore, recombination can be advantageous when individuals compete, as it accelerates the rate of evolution.

A key feature of our model is that although the probability of competitive success is dependent on the competitive ability of an individual it is not equal to it. Rather, it is determined by the relationship between the competitive ability of the individual itself and that of its

competitor. And since the frequency of the genotypes changes during the course of the selection process, the average competitive success of a given type also changes in a frequency dependent manner. As pointed out by Milkman (1973), in a competitive selection situation, the phenotype represents the fitness potential rather than the fitness itself. In addition, although the phenotypic values are unconstrained, the realized fitnesses that arise from them as a result of intraspecific competition are constrained. For example, in a population of fixed size living at the carrying capacity, the average realized fitness is by definition equal to 1. This reflects the fact that, overall, in a closed system competitive gains by one type must be offset by competitive losses by other types. As stated by Darwin "a plant which annually produces a thousand seeds, of which on an average only one comes to maturity, may be more truly said to struggle with the plants of the same and other kinds which already clothe the ground". In other words, Darwin saw natural selection as a zero sum game where the number of surviving offspring is no greater than the number of parental individuals, and where competition between conspecific individuals plays a crucial role. This is in contrast to many population genetics models that assign fixed fitnesses to genotypes which then implicitly engage in replication races during which less fit types are diluted out by fitter types rather than being selectively culled from the population.

We have shown that competitive selection creates a situation that could be beneficial to individuals that recombine their genotypes. It should be noted, however, that this conclusion is based on an analysis that is limited to a single generation. While we have shown that intraspecies competition will most often cause negative epistasis, it remains to be seen to what extent this might affect the evolution of sexually recombinants over many generations.

CHAPTER 3

In the previous chapter, we explored the effects of intraspecies competition in an evolutionary context, but limited our analysis to a single generation. By varying the initial conditions, we have shown that intraspecies competition generally results in negative epistasis at the fitness level. This provides a potential evolutionary advantage to sex and recombination. In should be noted, however, that natural selection is not a single generation process.

In this chapter, we explore the effects of competition over repeated generations to see how this might affect the evolution of sexual reproduction. Specifically, we wish to know if recombination allows the selected alleles to increase in frequency at a faster rate. If this is the case, then there is a clear advantage to sexual reproduction, as it would mean that sexuals can evolve at a faster rate than asexuals when under competitive selection.

Iterating competitive selection over several generations.

In this chapter, we explore the effects of intraspecies contest competition over repeated generations. Specifically, we wish to find an answer to the following questions:

1. Does negative epistatic fitness persist generation-over-generation?
2. Do the selected alleles increase in frequency at a faster rate in a sexual population?
3. Does intraspecies competition allow sexual individuals to increase in frequency when competing in a mixed population?

We begin by assuming that the selected alleles have an initially low frequency, and over several generations, they rise to become fixed in the population. We model a population that is either entirely asexual, or entirely sexual and freely recombinant. In both cases, we examine the changes in the value of epistasis generation-by-generation, and observe how this affects the genotype frequencies. We then investigate how these factors influence the rate at which the selected alleles increase in frequency. Finally, we analyze the change in the frequency of sexuals when competing with asexuals in a mixed population.

3.1 Modeling an asexual or sexual population over several generations.

Our first goal was to model the changes in genotypic frequencies generation-over-generation as the frequencies of the *A* and *B* alleles increase in the population. In chapter 2, we derived the change in the frequencies of each genotype under competition (with multiplicative effects on the phenotype), as follows:

$$
\begin{aligned}
Freq(ab)' &= \frac{Freq(ab)}{1-c}\left(\frac{1}{2}Freq(ab) + \frac{1}{2+x}Freq(Ab) + \frac{1}{2+y}Freq(aB) + \frac{1}{2+x+y+xy}Freq(AB)\right) \\
Freq(Ab)' &= \frac{Freq(Ab)}{1-c}\left(\frac{1+x}{2+x}Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1+x}{2+x+y}Freq(aB) + \frac{1}{2+y}Freq(AB)\right) \\
Freq(aB)' &= \frac{Freq(aB)}{1-c}\left(\frac{1+y}{2+y}Freq(ab) + \frac{1+y}{2+x+y}Freq(Ab) + \frac{1}{2}Freq(aB) + \frac{1}{2+x}Freq(AB)\right) \\
Freq(AB)' &= \frac{Freq(AB)}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}Freq(ab) + \frac{1+y}{2+y}Freq(Ab) + \frac{1+x}{2+x}Freq(aB) + \frac{1}{2}Freq(AB)\right)
\end{aligned}
\tag{3.1}
$$

where *Freq(ab)* is the frequency of the *ab* genotype before competition, *x* is the effect of allele *A* on the phenotype, *y* is the effect of allele *B* on the phenotype, *c* is the cull rate, and *Freq(ab)'* is the frequency of the *ab* genotype after competition, and similarly for the frequencies of the other genotypes.

In this section, we consider the cases where either the entire population is asexual, or the entire population is sexual and recombines freely. We assume that the reproductive method has no effect on an individual's family size, or competitive ability. We have also assumed, as in chapter 2, that genotype has no effect on family size. In an asexual population, the offspring of a parent are genetic clones of that parent. Therefore, the genotypic frequencies of the offspring are expected to have the same frequencies as their parents.

In a sexually recombinant population, however, this will not often be the case. As stated in Appendix 2, the genotypic frequencies of the offspring are expected to change as follows with random mating and free recombination:

$$\begin{aligned} Freq(ab)'' &= Freq(ab)' - ½D \\ Freq(Ab)'' &= Freq(Ab)' + ½D \\ Freq(aB)'' &= Freq(aB)' + ½D \\ Freq(AB)'' &= Freq(AB)' - ½D \end{aligned} \tag{3.2}$$

where *Freq(ab)'* is the frequency of the *ab* genotype among the parents, *D* is the value of linkage disequilibrium among the parents, and *Freq(ab)''* is the frequency of the *ab* genotype among the offspring, and similarly for the frequencies of the other genotypes.

Putting the pieces together, we can determine the genotypic frequencies at each point in time in an asexual population by iterating equation (3.1) once per generation, and we can determine the frequencies in a sexual population by iterating equation (3.1) followed by equation (3.2). Therefore, we can determine the frequencies at any point in time provided that we know the genotypic frequencies at time 0, and their phenotypic values.

3.2 The change in fitness epistasis, and linkage disequilibrium over several generations.

In this section, we examine the changes in the value of epistasis and linkage disequilibrium generation-over-generation in both sexual and asexual populations. The key questions for this study are whether the values of epistasis and linkage disequilibrium generally remain negative in each generation.

We begin by assuming that the *A* and *B* alleles have initially low frequencies within the population. We further assume that the population is initially in linkage equilibrium as this is the expected state of the population when exposed to random mutations in the absence of selection. The populations were modeled until the frequencies of the *A* and *B* alleles had become fixed, with the values of epistasis, *D*, and *Z* being computed in each generation.

We computed these values over a large range of the parameter space, with the initial frequencies of the *A* and *B* alleles beginning below 0.5. In all cases, we found that the value of epistasis remains negative over every generation. The value of epistasis changes very little as the frequencies of the selected alleles increase in the population (see Figure 3.1), likely because allelic frequencies are a small factor in determining the value of epistasis (see Figure 2.4).

Because the fitness values are negatively epistatic in each generation, the genotypic frequencies change to build up large amounts of negative linkage disequilibrium in an asexual population (see Figures 3.2 and 3.3). In a sexually recombining population however, recombination acts to increase the value of linkage disequilibrium, bringing its value closer to zero. The value of *D* is most negative when the frequencies of the *A* and *B* alleles are close to 0.5, as this is when *D* has the largest value range (see Figure 3.2). The range in values of *Z*, on the other hand, is less frequency dependent. Figure 3.3 shows that the value of *Z* in an asexual population continually builds up negative linkage disequilibrium. *Z* decreases by the value of

epistasis in every generation (see Figure 3.3). When individuals recombine freely, however, the value of Z quickly levels off to a constant periodic fluctuation (see Figure 3.4). This is because the value of Z decreases following competition by the value of epistasis in that generation. Following random mating with free recombination, however, the value of Z is approximately halved. These two values are of equal and opposite sign when the value of Z after competition is equal to -2ε. Regardless of the type of measurement, there is a far greater build-up of negative linkage disequilibrium in an asexual population than in a sexual population.

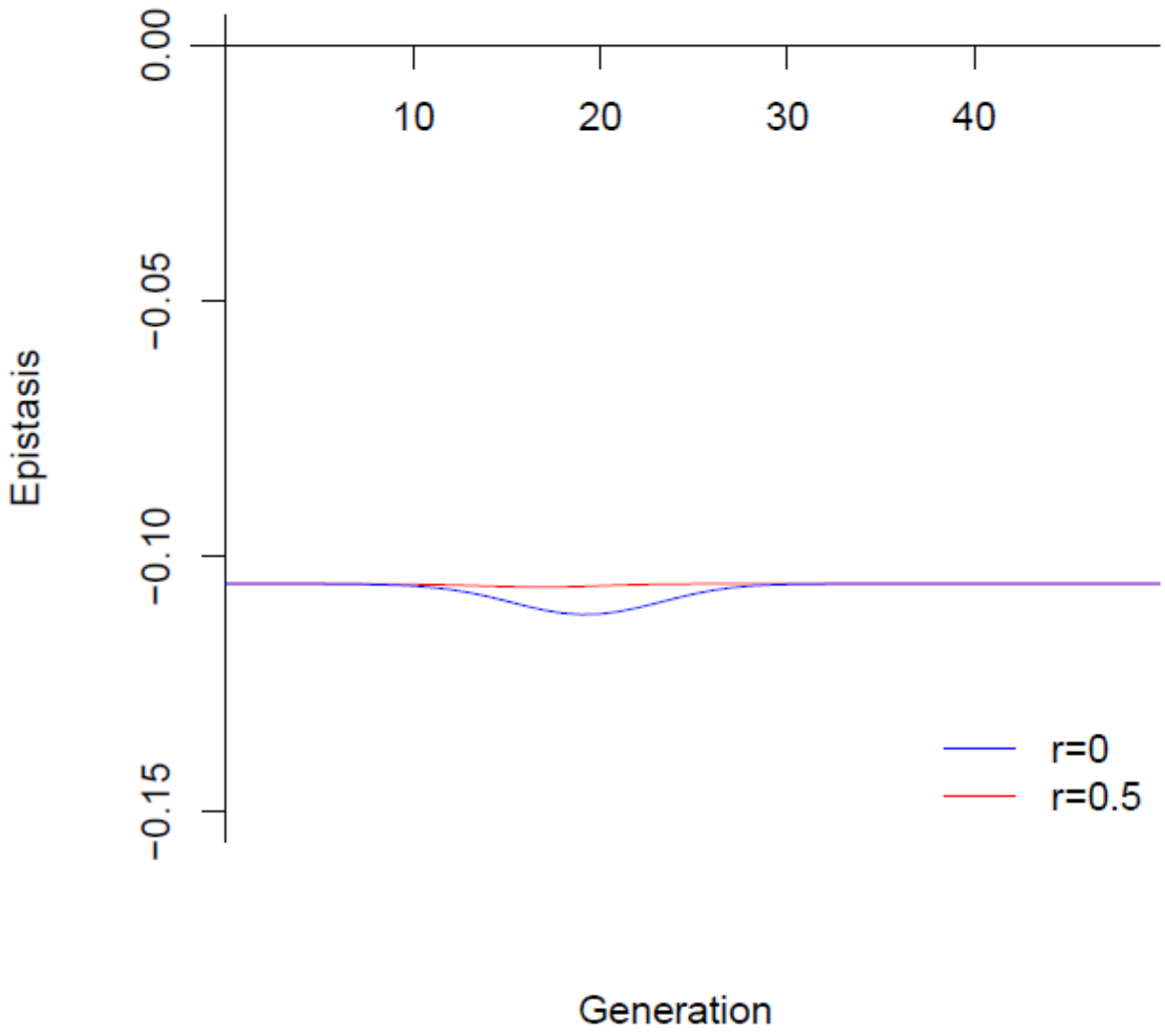


Figure 3.1: The change in the value of epistasis over successive generations.

The change in the value of epistasis is shown when either the population does not recombine (r=0) or recombines freely (r=0.5). In this case, the frequencies of the *A* and *B* alleles have initial values of 0.01, the population is beginning in linkage equilibrium, and the phenotypic values are the same as in Figure 2.1.1. Generations are shown on the horizontal axis, and are counted until the frequencies of the *A* and *B* alleles have risen to fixation. The value of epistasis is relatively constant when individuals are sexual and recombine freely. When individuals do not recombine, the value of epistasis lowers slightly when the genotype frequencies are in strong negative linkage disequilibrium (D) (see Figure 3.2). Note that in both cases, there is negative epistasis in every generation.

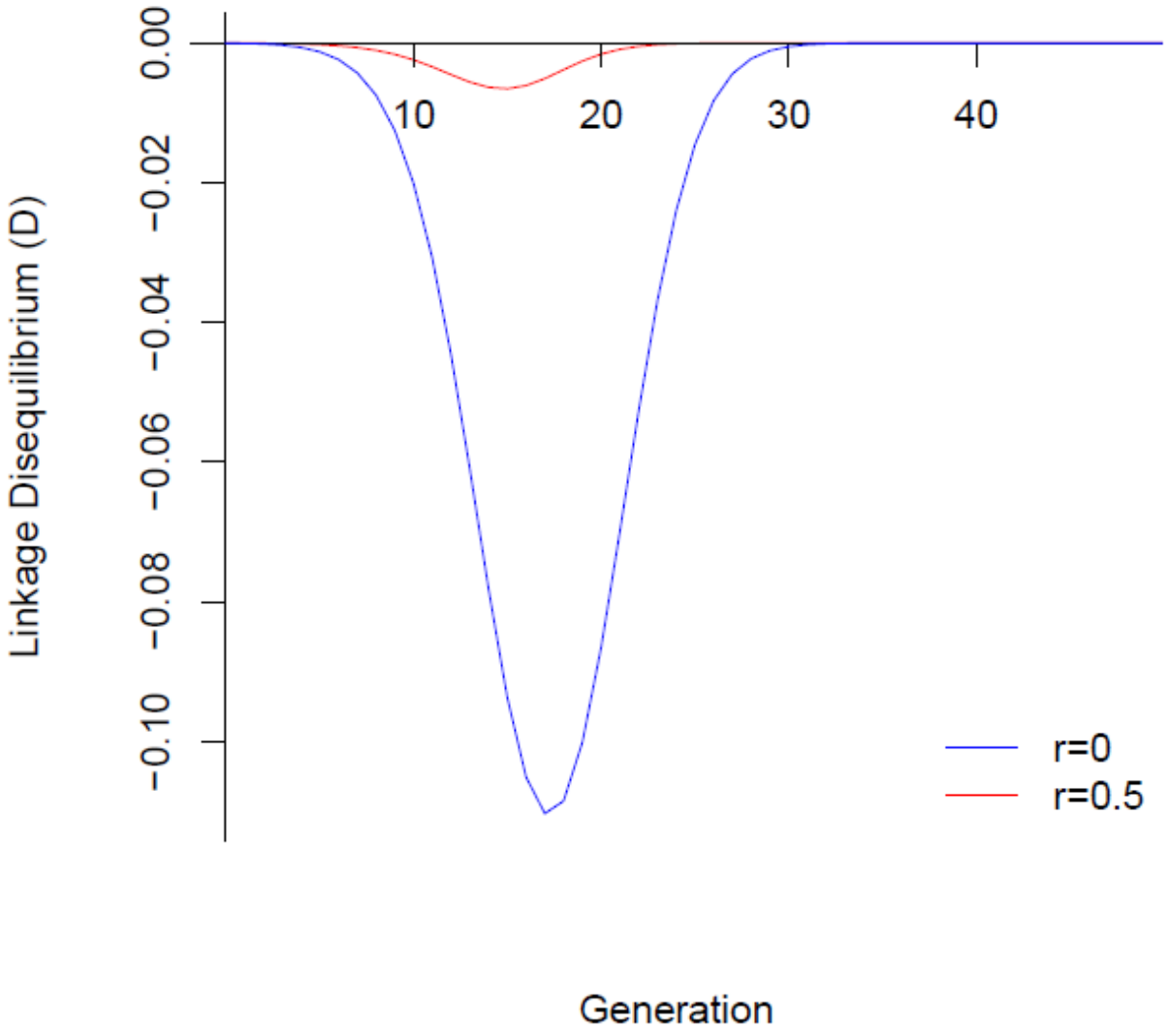


Figure 3.2: The change in the value of linkage disequilibrium (*D*) over successive generations.

The initial genotype frequencies and phenotypic values are the same as described in Figure 3.1. Generations are shown on the horizontal axis, and are counted until the frequencies of the *A* and *B* alleles have risen to fixation. The change in the value of *D* is shown when either the population does not recombine (r=0) or recombines freely (r=0.5). Linkage disequilibrium remains negative as the selected alleles rise to become fixed in the population. The value of *D* is most negative when the allele frequencies are close to 0.5. Note that the value of *D* accumulates less linkage disequilibrium when individuals freely recombine.

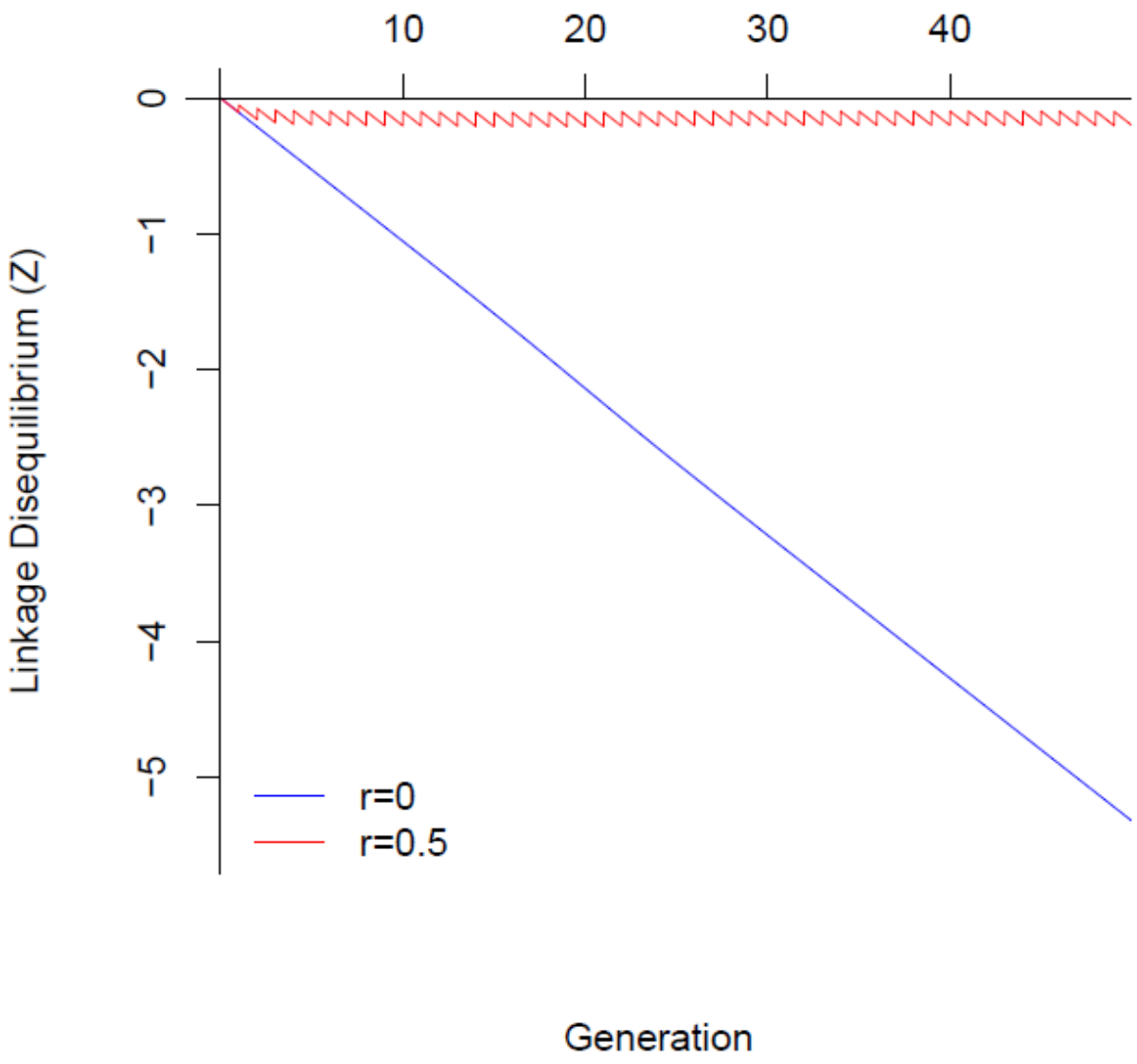


Figure 3.3: The change in the value of linkage disequilibrium (*Z*) over successive generations.

For a description, see Figure 3.2. In this case, linkage disequilibrium is plotted using the value of *Z* instead of *D*. The value of *Z* is plotted twice per generation: after competition, and after random mating. Linkage disequilibrium remains negative as the selected alleles rise to become fixed in the population. Note that when individuals cannot recombine, the value of *Z* decreases in every generation by the value of epistasis. When individuals recombine freely, however, the value of *Z* quickly levels off to a constant periodic fluctuation.

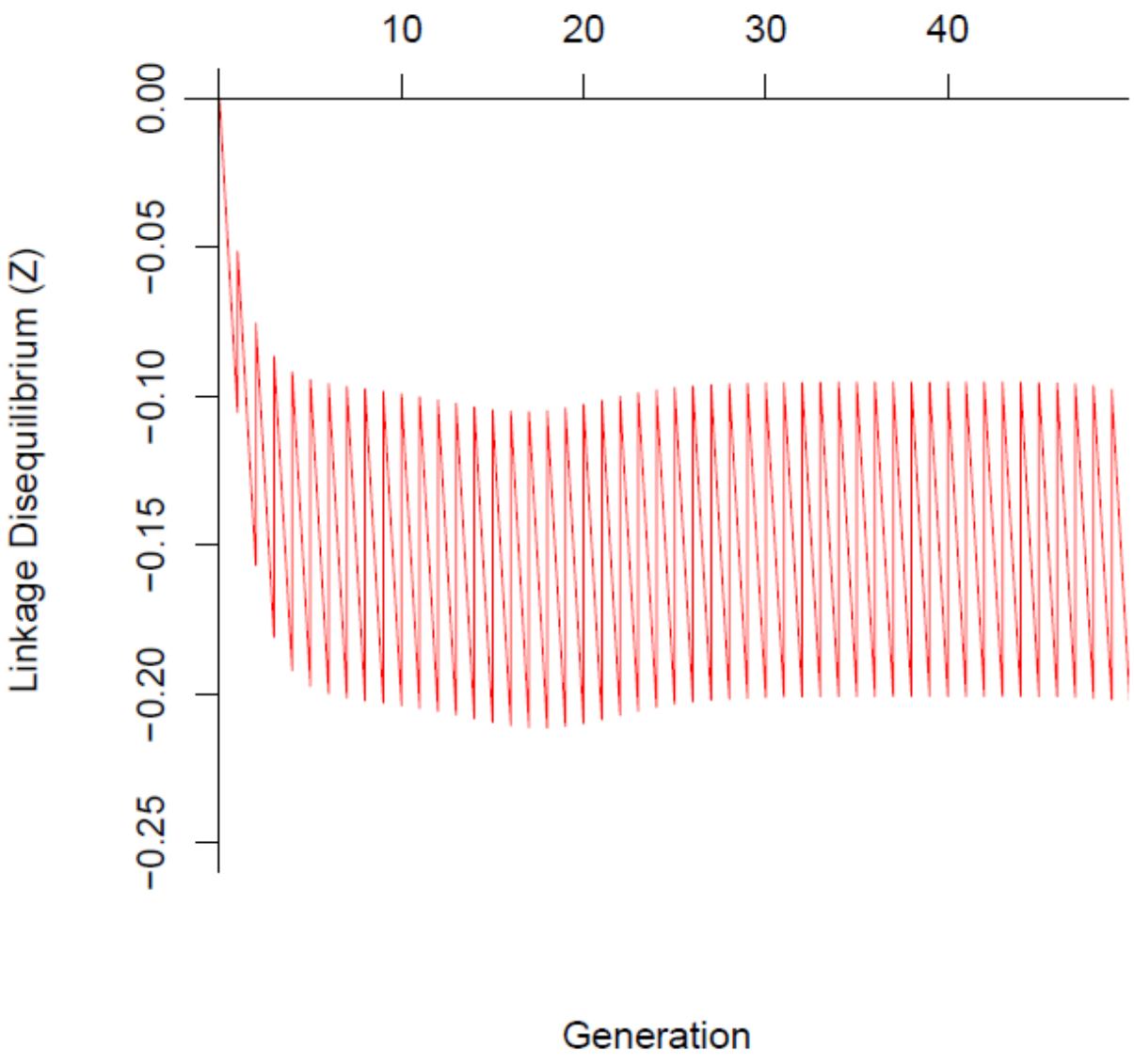


Figure 3.4: The change in the value of linkage disequilibrium (*Z*) over successive generations in a population that recombines freely.

For a description, see Figure 3.3. Here we plot the value of *Z* twice per generation in a sexually recombinant (r=0.5) population: once after selection, and again after recombination. The value of *Z* decreases after selection by the value of epistasis in that generation, and is then halved through the action of recombination. This results in a periodic fluctuation in the value of *Z*.

3.3 Comparing the increase in frequencies of the selected alleles in recombinant and non-recombinant populations.

We have shown that the genotypic fitnesses will be negatively epistatic in each generation as the selected alleles increase in frequency in the population. This results in an asexual population that is increasingly in negative linkage disequilibrium. In a sexually recombinant population, however, there is far less buildup of negative linkage disequilibrium. Here we discuss how these changes in the genotype frequencies affect the increase in frequencies of the *A* and *B* alleles in both sexual and asexual populations.

We begin by assuming, as in section 3.2, that the *A* and *B* alleles have frequencies that are initially below 0.5, and that the population is initially in linkage equilibrium. The sexual and asexual populations were modeled until the frequencies of the selected alleles became fixed. We track the frequencies of the *A* and *B* alleles, and the frequency of the *AB* genotype, comparatively between sexual and asexual populations. We then ran individual-based simulations with the same initial parameters to validate the results. Simulations were run with a population size of 100 000 individuals.

Both in the numerically expected results, and in the individual-based simulations we find that the frequencies of the *A* and *B* alleles increase at a faster rate in a sexual population than in an asexual population (see Figure 3.5). The difference in frequencies is most pronounced when the *A* and *B* alleles begin at very low frequencies, and have strong effects on the phenotype. In both the numerically expected results, and the individual-based simulations we also find that the frequency of the *AB* genotype (i.e. the strongest competitor) increases at a faster rate in the sexual population (see Figure 3.6). When comparing sexuals and asexuals, it is worth noting that

there is a greater disparity between the increase in frequency of the *AB* genotype than between the increase in frequencies of the *A* and *B* alleles (see Figures 3.5 and 3.6).

Taken together with the results in section 3.2, we find that competition creates a negative epistatic fitness curve that persists generation-over-generation. This results in a build-up of negative linkage disequilibrium in an asexual population. There is a reduced build-up in a sexual population with random mating and free recombination. Thus, recombination increases the phenotypic and fitness variance generation-over-generation, as can be noticed by the greater frequency of the *AB* genotype in the sexual population. This increases the action of selection, allowing for the *A* and *B* alleles to increase in frequency at a faster rate in a sexual population.

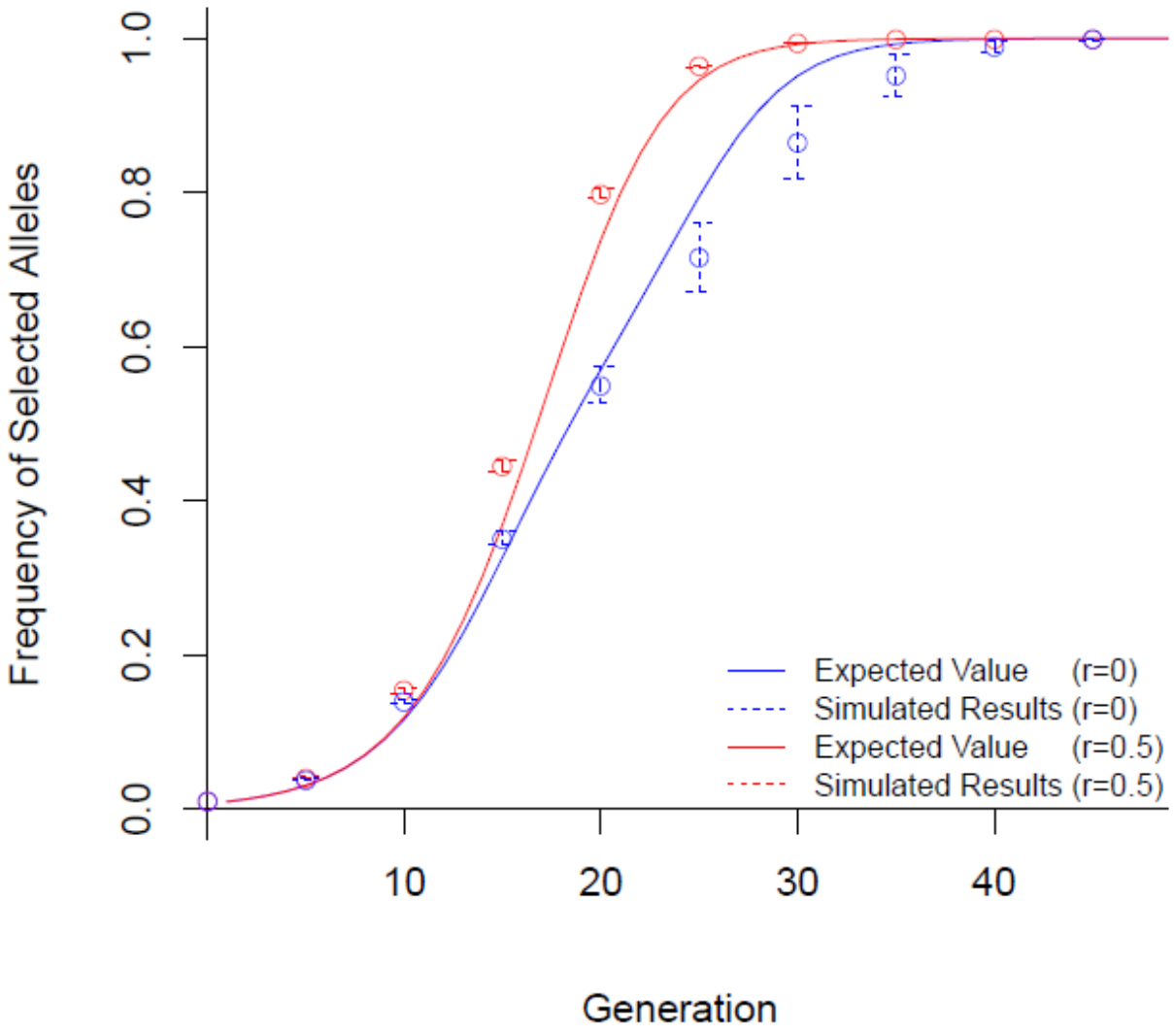


Figure 3.5: The change in the frequency of the selected alleles over successive generations. The frequencies of the *A* and *B* alleles are shown on the vertical axis in two different scenarios: when all individuals do not recombine (r=0), or recombine freely (r=0.5). The initial parameters are the same as in Figure 3.1, and generations are counted until the frequencies of the *A* and *B* alleles are fixed in the population. The numerically expected changes of frequency - in both cases - is shown with a solid line, and were validated by running 10 simulation runs under the same initial parameters. The dots and whiskers show the mean $\mp$1 standard error from the simulation runs, respectively. Note that the frequency of the selected alleles increases at a faster rate with recombination.

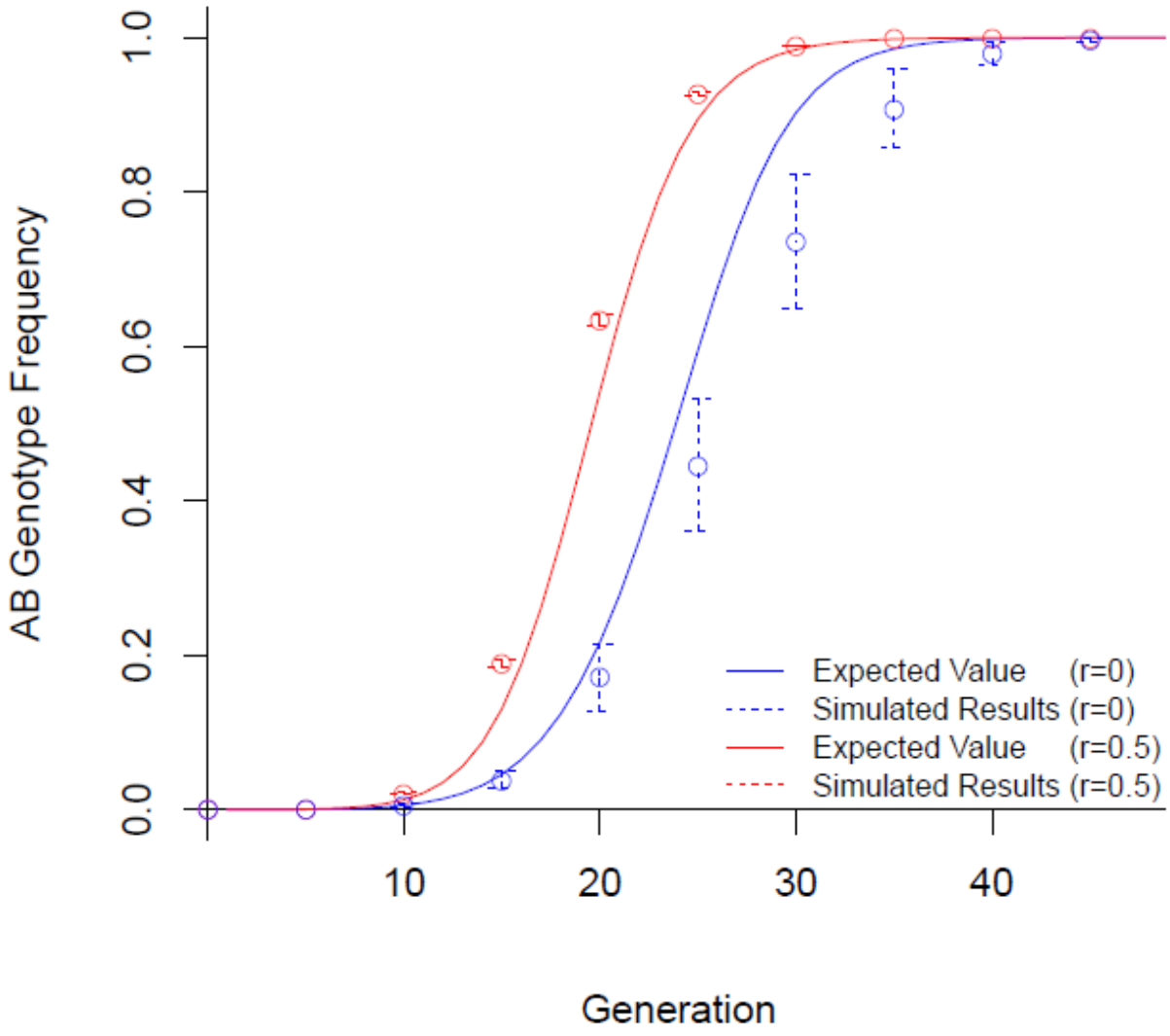


Figure 3.6: The change in the frequency of the AB genotype over successive generations. For a description, see Figure 3.4. In this case, the frequency of the *AB* genotype is plotted instead of the frequency of the selected alleles. Both the numerical and simulated results show that the frequency of the *AB* genotype increases at a faster rate with recombination. When comparing recombinant and non-recombinant cases, note that there is a greater disparity between the frequencies of the *AB* genotypes than between the frequencies of the selected alleles (see Figure 3.5).

3.4 Competing sexuals and asexuals.

In the last section, we showed that the frequency of the selected alleles will increase in frequency at a faster rate in a sexually recombinant population than in an asexual population. In this section, we explore how this might affect the frequency of sexual individuals when competing with asexuals in a mixed population. Specifically, we wish to know if the frequency of sexuals will increase as the frequency of the selected alleles rises to become fixed in the population.

Competition is modeled as before, with reproductive method having no effect on competitive ability. This means that the competitive ability of an *AB* genotyped individual, for example, is the same whether reproducing sexually or asexually. Let *Freq(AB,s)* denote the frequency of *AB* genotyped individuals that reproduce sexually, and *Freq(AB,a)* denote the frequency of *AB* genotyped individuals that reproduce asexually. The frequency of all individuals with the *AB* genotype is therefore *Freq(AB,a)+ Freq(AB,s).* We can express the total frequencies of all four genotypes as follows:

$$\begin{aligned} Freq(ab) &= Freq(ab,a) + Freq(ab,s) \\ Freq(Ab) &= Freq(Ab,a) + Freq(Ab,s) \\ Freq(aB) &= Freq(aB,a) + Freq(aB,s) \\ Freq(AB) &= Freq(AB,a) + Freq(AB,s) \end{aligned} \tag{3.3}$$

These equations simplify how to express the frequencies of each type of interaction since all individuals with the same genotype have the same competitive ability. The frequency of each genotype post-competition can be expressed as follows:

$$\begin{aligned} Freq(ab,a)' &= \frac{Freq(ab,a)}{1-c}\left(\frac{1}{2}Freq(ab) + \frac{1}{2+x}Freq(Ab) + \frac{1}{2+y}Freq(aB) + \frac{1}{2+x+y+xy}Freq(AB)\right) \\ Freq(Ab,a)' &= \frac{Freq(Ab,a)}{1-c}\left(\frac{1+x}{2+x}Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1+x}{2+x+y}Freq(aB) + \frac{1}{2+y}Freq(AB)\right) \end{aligned} \tag{3.4}$$

$$Freq(aB,a)' = \frac{Freq(aB,a)}{1-c}\left(\frac{1+y}{2+y}Freq(ab) + \frac{1+y}{2+x+y}Freq(Ab) + \frac{1}{2}Freq(aB) + \frac{1}{2+x}Freq(AB)\right)$$

$$Freq(AB,a)' = \frac{Freq(AB,a)}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}Freq(ab) + \frac{1+y}{2+y}Freq(Ab) + \frac{1+x}{2+x}Freq(aB) + \frac{1}{2}Freq(AB)\right)$$

$$Freq(ab,s)' = \frac{Freq(ab,s)}{1-c}\left(\frac{1}{2}Freq(ab) + \frac{1}{2+x}Freq(Ab) + \frac{1}{2+y}Freq(aB) + \frac{1}{2+x+y+xy}Freq(AB)\right)$$

$$Freq(Ab,s)' = \frac{Freq(Ab,s)}{1-c}\left(\frac{1+x}{2+x}Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1+x}{2+x+y}Freq(aB) + \frac{1}{2+y}Freq(AB)\right)$$

$$Freq(aB,s)' = \frac{Freq(aB,s)}{1-c}\left(\frac{1+y}{2+y}Freq(ab) + \frac{1+y}{2+x+y}Freq(Ab) + \frac{1}{2}Freq(aB) + \frac{1}{2+x}Freq(AB)\right)$$

$$Freq(AB,s)' = \frac{Freq(AB,s)}{1-c}\left(\frac{(1+x)(1+y)}{2+x+y+xy}Freq(ab) + \frac{1+y}{2+y}Freq(Ab) + \frac{1+x}{2+x}Freq(aB) + \frac{1}{2}Freq(AB)\right)$$

where $Freq(ab,a)'$ and $Freq(ab,s)'$ denote the frequency of *ab* genotyped individuals after competition that are asexual and sexual, respectively, and similarly for the other genotypes. Note that the fitnesses of sexuals and asexuals with the same genotype are equal.

The survivors from competition then go on to reproduce. The sexual individuals mate randomly with each other and freely recombine, and the asexual individuals clone themselves. The frequencies of sexuals and asexuals do not change from parents to offspring as the mean family size of both reproductive methods is the same. The frequencies of the asexual genotypes also do not change among the offspring, but the sexual genotypes can. As stated in Appendix 2, the genotypic frequencies change from parent to offspring in a purely sexual population as follows:

$$\begin{aligned} Freq(ab)'' &= Freq(ab)' - ½D \\ Freq(Ab)'' &= Freq(Ab)' + ½D \\ Freq(aB)'' &= Freq(aB)' + ½D \\ Freq(AB)'' &= Freq(AB)' - ½D \end{aligned} \tag{3.5}$$

where *Freq(ab)'* and *Freq(ab)''* are the frequencies of the *ab* genotype among parents and offspring, respectively, and similarly for the other genotypes; and *D* is the value of linkage disequilibrium among the parents. In a mixed population, however, not all individuals are sexual.

We multiply these values by the frequency of sexuals in the total population to get the frequencies of each sexual genotype. After some simplification, they can be expressed as follows:

$$\begin{aligned} Freq(ab,s)'' &= Freq(ab,s)' - \frac{D_s}{2Freq(s)'} \\ Freq(Ab,s)'' &= Freq(Ab,s)' + \frac{D_s}{2Freq(s)'} \\ Freq(aB,s)'' &= Freq(aB,s)' + \frac{D_s}{2Freq(s)'} \\ Freq(AB,s)'' &= Freq(AB,s)' - \frac{D_s}{2Freq(s)'} \end{aligned} \tag{3.6}$$

where *Freq(s)'* is the frequency of all sexual parents; *Freq(ab,s)'* and *Freq(ab,s)''* are the frequencies of sexual parents and offspring with the *ab* genotype, respectively, and similarly for the other genotypes; and

D_s*=Freq(ab,s)'Freq(AB,s)'-Freq(Ab,s)'Freq(aB,s)'*.

We calculated the expected frequency change of sexual individuals in a mixed population -when beginning in linkage equilibrium- over a wide range of the parameter space, and validated the results with individual-based simulations with a total population size of 100 000. We considered cases where phenotypic effects are multiplicative, negatively epistatic, as well as positively epistatic. We find that the frequency of sexuals generally increases as the selected alleles rise to become fixed in the population (see Figures 3.7 and 3.8). The increase is most pronounced when the frequencies of the selected alleles are initially low, and phenotypic effects are strong. Even when there is positive epistasis at the phenotype level, the frequency of sexuals will still often increase (see Figure 3.8). There needs to be very strong positive epistasis at the phenotype level (i.e. $\epsilon > x$ and y) for the frequency of sexuals to decrease over the course of selection.

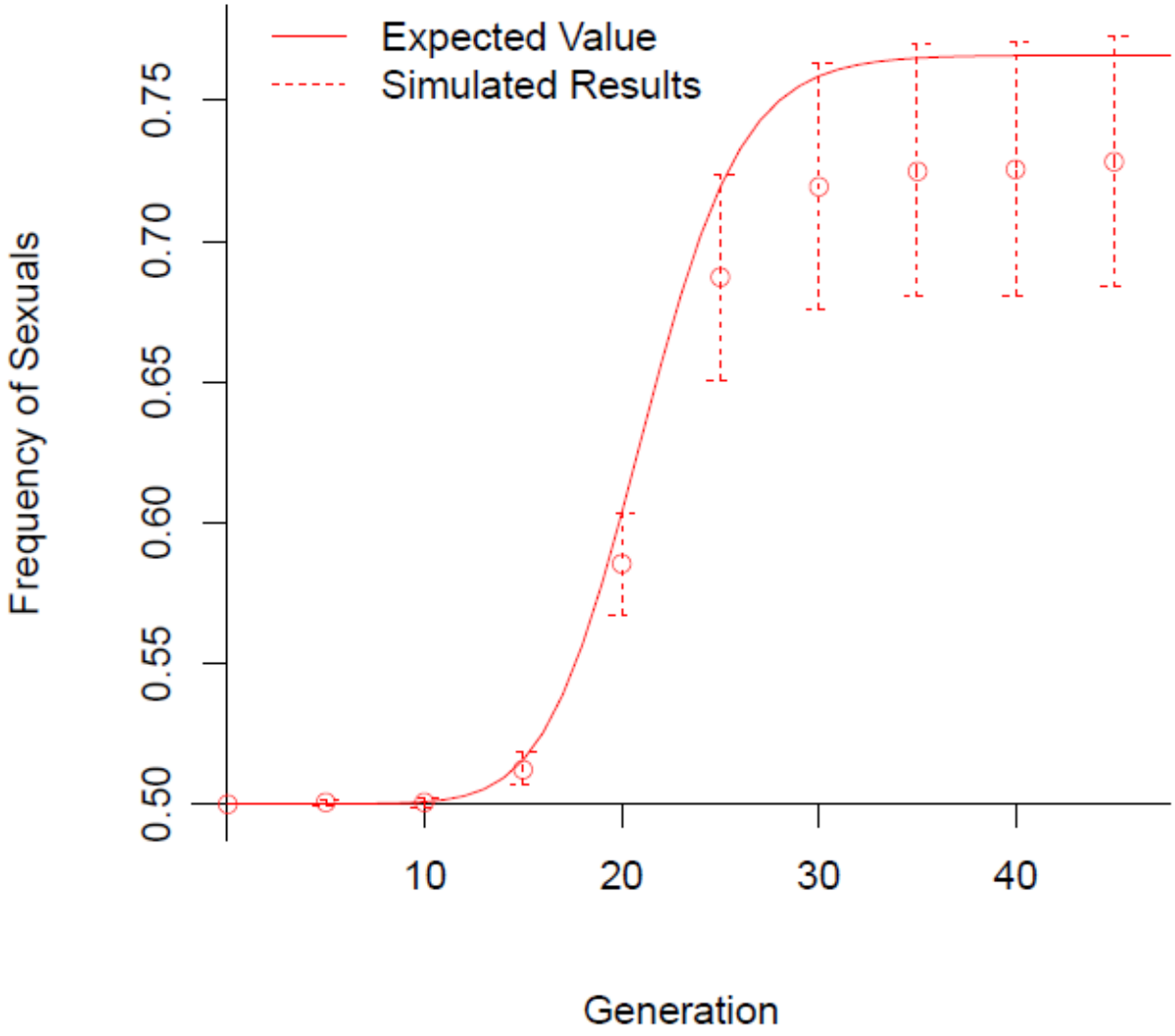


Figure 3.7: The change in frequency of sexuals in a mixed population with multiplicative phenotypic effects.

The initial parameters are the same as in Figure 3.1 with one exception: the population is initially half sexually recombinant (r=0.5), and half asexual. Reproductive method has no effect on competitive ability. Generations are shown on the horizontal axis, and are counted until the frequencies of the *A* and *B* alleles have risen to fixation. The numerically expected frequency of sexuals is shown with a solid line, and was validated by running 10 simulation runs under the same initial parameters. The results are plotted showing the mean ∓ 1 standard error. Note that both the expected and simulated results show that the sexuals increase in frequency.

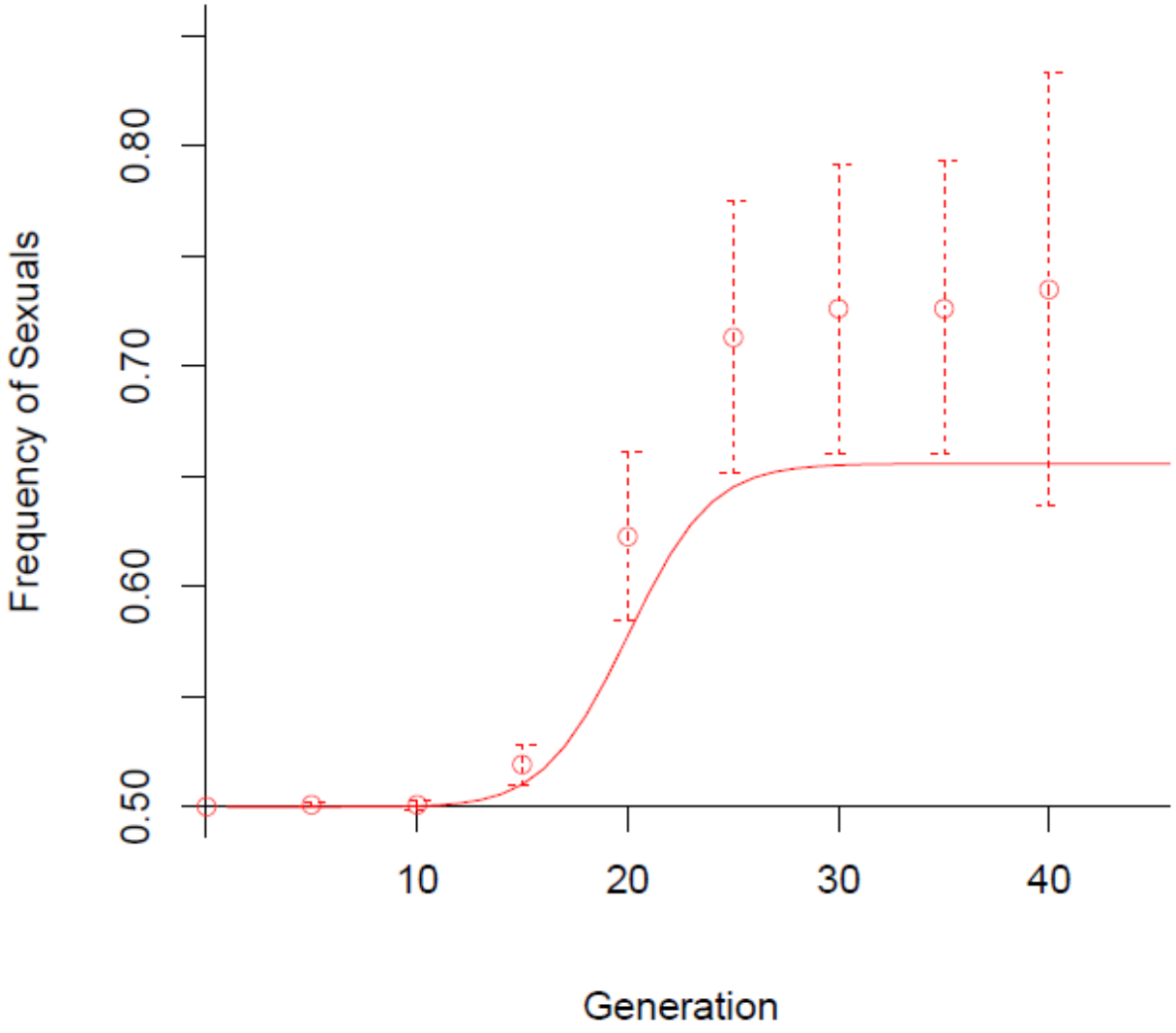


Figure 3.8: The change in frequency of sexuals in a mixed population with positive phenotypic epistasis.

For a description, see Figure 3.7. In this case, the phenotypic values of the four genotypes V_{ab}, V_{Ab}, V_{aB}, V_{AB}, are 1, 2, 2 and 5, respectively. The numerically expected frequency of sexuals is shown with a solid line, and was validated by running 10 simulation runs under the same initial parameters. The results are plotted showing the mean ∓ 1 standard error. Note that both the expected and simulated results show that the sexuals increase in frequency, but the increase is expected to be less than with multiplicative effects on the phenotype (see Figure 3.7).

3.5 Discussion

We have found that intraspecies competition creates a negative epistatic fitness curve that persists generation-over-generation. This results in a build-up of negative linkage disequilibrium in an asexual population, which is alleviated in a sexual population through random mating and free recombination. There is a greater frequency of *AB* genotyped individuals among sexual offspring, which subsequently, are more likely to survive competition. Therefore, the frequencies of the selected alleles increase at a faster rate among sexuals than among asexuals. Sexual individuals increase in frequency at the expense of asexuals when competing in a mixed population. The increase is greatest when the frequencies of the selected alleles are initially low, and their phenotypic effects are strong. We found this to be the case when phenotypic effects are negatively epistatic, non-epistatic, and even somewhat positively epistatic.

While most previous models have focused on selection at the level of fertility and fecundity, our model focuses explicitly on viability selection. Although there is relatively little experimental information available on the relative importance of viability and fertility in determining fitness in nature, a study by Fincke and Hadrys (2001) indicated that fertility can be a poor predictor of overall fitness in insects, suggesting instead that larval survival was the major component.

Taken together, the results suggest that intraspecies competition may help explain the maintenance of sex. Most studies that have attempted to determine the mean level of epistasis within a species have done so using fitness values when raised in the absence of competition (Peck and Waxman, 2000; de Visser et al. 2007; Kouyos *et. al*., 2007). Our results suggest that including competitive interactions into such experiments will alter the results, enough to have a significant impact on their conclusions related to the advantages of sexual reproduction. We have

shown that a sexual population's competitive ability increases at a faster rate, allowing it to increase over generations in a mixed population. In should be noted however, that competitively advantageous alleles need to be continually under selection in order for sex to be maintained. Therefore, one might predict that the levels of sex would be higher among *K*-strategist species than they are among *r*-strategists. We would further predict that the levels of sex should increase among facultative sexual reproducers when forced to compete with one another, as this would constitute an advantageous evolutionary strategy.

We have shown that there is an advantage to sexual reproduction, and we have done so by mapping phenotype onto competitive ability onto fitness. Other mappings are quite possible, however, and it can be argued that a different mapping of phenotype onto fitness may give results that favour asexual reproduction. Therefore, it should be shown that these results are robust before going any further.

CHAPTER 4

So far, we have shown that contest competition favours the maintenance of sex when the phenotypically stronger competitor has odds of winning that are directly proportional to phenotypic values. However, this is not the only mapping that can be made of phenotype onto fitness.

In this chapter, we look into a different mapping of phenotype on fitness, specifically, where the stronger individual *always* wins. Our objective in this section is to provide robustness to our claim that contest competition favours the maintenance of sex.

The maintenance of sex when competitive ability is <u>not</u> proportional to phenotypic value

There are many different ways of mapping phenotype onto competitive ability. Having competitive ability proportional to phenotypic value is only one of them. For this reason, we wish to show that other mappings can also generate negative epistasis and allow for the maintenance of sex. It seems intuitive to believe that either individual has a chance of winning when the two competitors have closely matched phenotypes. But when they are not so closely matched, one can predict the winner with certainty. Therefore, we will now explore a relationship between phenotype and competitive ability that is not proportional, in that the stronger competitor always wins.

First, we develop a genotypic model where the stronger competitor always wins. Next, we show that the values of epistasis and linkage disequilibrium are negative and frequency dependent. We then compare the frequency increases of competitively advantageous alleles in

both sexual and asexual populations to show that the advantageous alleles increase at a faster rate in the sexual population. Finally, we observe a mixed population of sexual and asexual individuals under competitive selection to show how the sexual individuals increase in frequency. These results have been published in the Journal of Heredity (Ackerman et al., 2010).

As in the previous chapters, an individual's competitive ability is dependent both on its own phenotypic value and also on that of its competitor. An individual is guaranteed to win if its competitor has a weaker phenotype, and guaranteed to lose if its competitor has a stronger phenotype. Since the frequencies of the various types change in the course of selection, the fitnesses are still frequency dependent in this model.

4.1 Modeling an asexual or sexual population when the stronger competitor always wins.

We assume a 2-locus, diallelic, haploid population of competing individuals with genotypes *ab*, *Ab*, *aB*, and *AB*. We further assume that the *A* and *B* alleles both equally increase the value of the phenotype (*V*), and that the *AB* genotype has the highest phenotypic value of the four genotypes. Therefore

$$V_{ab} < V_{Ab} = V_{aB} < V_{AB} \tag{4.1}$$

Generations are discrete and non-overlapping. At the beginning of a generation, each individual (*N* total) produces excess offspring, after which the population is halved through random pairwise competition. We assume that each genotype has the same fecundity, meaning that the genotypic frequencies of the offspring are expected to equal that of their parents. When competing, the individual with the higher phenotypic value always wins the competition, while the loser is removed from the population. When both individuals have equal phenotypic values,

they both have equal probabilities of winning. The table below summarizes the outcome of every possible type of pairwise competition.

Table 4.1 Expected outcome of competitive interactions.

	Genotype of competitor			
Genotype	ab	Ab	aB	AB
ab	0.5	0	0	0
Ab	1	0.5	0.5	0
aB	1	0.5	0.5	0
AB	1	1	1	0.5

The table shows the probability that an individual with the genotype listed on the left will win in competition with a genotype listed across the top.

The competitive success of each genotype is once again determined by both the outcome of specific competitive interactions (which are genotypically-determined by the phenotypes of the competitors), and the frequency of each type of interaction. Thus, we can calculate the genotype proportions after competition by multiplying the competitive outcomes shown in Table 4.1 by their expected frequencies, and then summing for each genotype. In order to convert these four proportions into genotypic frequencies, they must be normalized by dividing by the average

survival rate, which is equal to 0.5 as there is only one survivor per competing pair. Therefore, the genotypic frequencies following competition can be expressed as follows:

$$\begin{aligned}
Freq(ab)' &= \frac{Freq(ab)}{0.5}\left(\frac{1}{2}Freq(ab) + 0Freq(Ab) + 0Freq(aB) + 0Freq(AB)\right) \\
Freq(Ab)' &= \frac{Freq(Ab)}{0.5}\left(Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1}{2}Freq(aB) + 0Freq(AB)\right) \\
Freq(aB)' &= \frac{Freq(aB)}{0.5}\left(Freq(ab) + \frac{1}{2}Freq(Ab) + \frac{1}{2}Freq(aB) + 0Freq(AB)\right) \\
Freq(AB)' &= \frac{Freq(AB)}{0.5}\left(Freq(ab) + Freq(Ab) + Freq(aB) + \frac{1}{2}Freq(AB)\right)
\end{aligned} \tag{4.2}$$

where *Freq(ab)* is the frequency of the *ab* genotype among the parents, and *Freq(ab)'* is the frequency of their surviving offspring. The frequencies of the other genotypes are similarly defined. Note that many of the terms in equations (4.2) are equal to zero, and that the sum of the genotypic frequencies among the parents is equal to 1. By removing the null terms and simplifying, we get

$$\begin{aligned}
Freq(ab)' &= Freq(ab)^2 \\
Freq(Ab)' &= Freq(Ab)\big(1 + Freq(ab) - Freq(AB)\big) \\
Freq(aB)' &= Freq(aB)\big(1 + Freq(ab) - Freq(AB)\big) \\
Freq(AB)' &= Freq(AB)\big(2 - Freq(AB)\big)
\end{aligned} \tag{4.3}$$

Once we have the genotypic frequencies after competition, we can calculate the realized fitness of each genotype. As we did in chapter 2, the realized fitness of a genotypic class of individuals is calculated as the frequency of that genotype after competition, divided by its frequency before competition. Thus, the realized fitnesses of the four genotypes are as follows:

$$\begin{aligned}
w_{ab} &= \frac{Freq(ab)^2}{Freq(ab)} \\
w_{Ab} &= \frac{Freq(Ab)}{Freq(Ab)}\big(1 + Freq(ab) - Freq(AB)\big)
\end{aligned} \tag{4.4}$$

$$w_{aB} = \frac{Freq(aB)}{Freq(aB)}\big(1 + Freq(ab) - Freq(AB)\big)$$
$$w_{AB} = \frac{Freq(AB)}{Freq(AB)}\big(2 - Freq(AB)\big)$$

After canceling out common factors, the equations simplify to

$$\begin{aligned} w_{ab} &= Freq(ab) \\ w_{Ab} &= 1 + Freq(ab) - Freq(AB) \\ w_{aB} &= 1 + Freq(ab) - Freq(AB) \\ w_{AB} &= 2 - Freq(AB) \end{aligned} \tag{4.5}$$

As in the previous chapters, we will let p_1, q_1, p_2 and q_2 represent the allelic frequencies of *A*, *a*, *B*, and *b*, respectively. When the population is initially in linkage equilibrium, the fitness equations (4.5) can be expressed in terms of their allele frequencies as follows

$$\begin{aligned} w_{ab} &= q_1 q_2 \\ w_{Ab} &= 1 + q_1 q_2 - p_1 p_2 \\ w_{aB} &= 1 + q_1 q_2 - p_1 p_2 \\ w_{AB} &= 2 - p_1 p_2 \end{aligned} \tag{4.6}$$

We can now plot the relationship between the realized fitness and allele frequency when the population is initially in linkage equilibrium. As can be seen in Figure 4.1, the genotypic fitnesses are heavily frequency dependent. Note that the fitnesses of the *AB* and *ab* genotypes are only dependent on their own frequencies. This makes sense as *AB* genotyped individuals can only lose a competition against their own genotype, while *ab* genotyped individuals can only win against their own genotype. Furthermore, the middle types (*Ab* and *aB*) are more likely to win their competitions and have higher fitness values when mainly competing against weaker individuals (i.e. with the *ab* genotype), and have lower fitness values when mainly competing against stronger individuals (i.e. with the *AB* genotype).

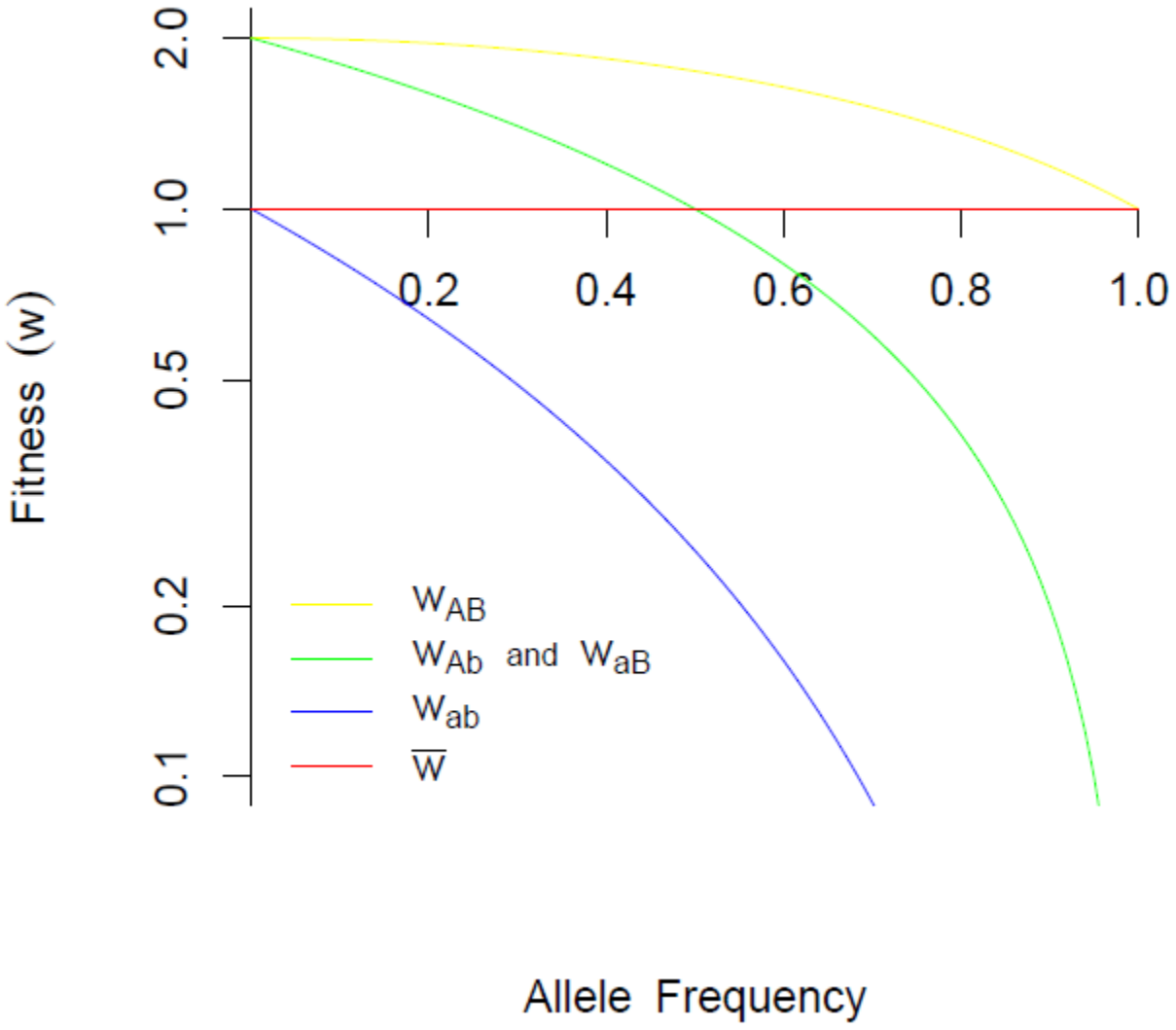


Figure 4.1: The correlation between realized fitness and allele frequency when the total number of surviving offspring is constrained to equal the total number of parents.

The expected realized fitnesses of the individual genotypes were calculated as a function of their own frequency and that of other genotypes in the population of competing offspring using equations (4.6), where w_{AB} represents the fitness of the *AB* genotype, etc. among the competing offspring. See text for an explanation of the formulas.

4.2 Frequency-dependent changes in epistasis and linkage disequilibrium.

We will now examine the values of epistasis and linkage disequilibrium at different allele frequencies when initially in linkage equilibrium. The key question for this study is whether the value of epistasis is negative for all values of p_1 and p_2 between 0 and 1, that is

$$\varepsilon = \ln \frac{w_{ab} w_{AB}}{w_{Ab} w_{aB}} < 0 \tag{4.7}$$

The condition of negativity is satisfied if the numerator is less than the denominator. In order to evaluate the sign of ε, we will begin by substituting the fitness equations (4.6) into the above formula for the value of epistasis. Therefore, we have to show that

$$\ln \frac{q_1 q_2 (2 - p_1 p_2)}{(1 + q_1 q_2 - p_1 p_2)^2} < 0 \tag{4.8}$$

This can be re-written as

$$q_1 q_2 (2 - p_1 p_2) - (1 + q_1 q_2 - p_1 p_2)^2 < 0 \tag{4.9}$$

If we substitute the fact that $q_1 = 1 - p_1$, and $q_2 = 1 - p_2$ into the above equation, and then simplify, we get that

$$-(1 - p_1)^2 - (1 - p_2)^2 - p_1 p_2 (1 - p_1)(1 - p_2) < 0 \tag{4.10}$$

It should be clear that every factor in the above equation has a positive value for values of p_1 and p_2 between 0 and 1, meaning that every term has a negative value. Therefore, the value of epistasis will always be negative. Figure 4.2 shows the value of epistasis when the frequencies of the *A* and *B* alleles are equal to each other. As with when the stronger competitor has better odds of winning, epistatic effects at the fitness level were caused by competitive selection itself, despite the lack of epistasis at the phenotypic level. The resulting realized fitness of each genotype is the result of an interaction between the fitness potential and the frequency of other

competing genotypes in the population. It is the nonlinear mapping of fitness potential onto realized fitness that produces negative epistasis. This is why it is possible to have non-epistatic effects at the phenotypic level, but epistasis at the realized fitness level.

Now that we have shown that intraspecies competition results in negatively epistatic fitness when the strongest competitor always wins, our next goal was to determine if the resulting genotype frequencies after selection are in negative linkage disequilibrium. Furthermore, we would like to know how their frequencies will be altered through random mating and free recombination.

As stated in section 4.1, the frequency of each genotype post-competition can be expressed as follows:

$$\begin{aligned} Freq(ab)' &= Freq(ab)^2 \\ Freq(Ab)' &= Freq(Ab)\big(1 + Freq(ab) - Freq(AB)\big) \\ Freq(aB)' &= Freq(aB)\big(1 + Freq(ab) - Freq(AB)\big) \\ Freq(AB)' &= Freq(AB)\big(2 - Freq(AB)\big) \end{aligned} \tag{4.11}$$

With random mating and free recombination following competition, the genotype frequencies are expected to change as re-stated below:

$$\begin{aligned} Freq(ab)'' &= Freq(ab)' - ½D \\ Freq(Ab)'' &= Freq(Ab)' + ½D \\ Freq(aB)'' &= Freq(aB)' + ½D \\ Freq(AB)'' &= Freq(AB)' - ½D \end{aligned} \tag{4.12}$$

where $Freq(ab)''$ is the frequency of the *ab* genotype after recombination, and $Freq(ab)'$ is the frequency of the *ab* genotype before recombination, and similarly for the other genotypes. We can determine the values of *D* and *Z* after recombination by once again plugging the genotype frequencies into equations (2.15) and (2.16), respectively.

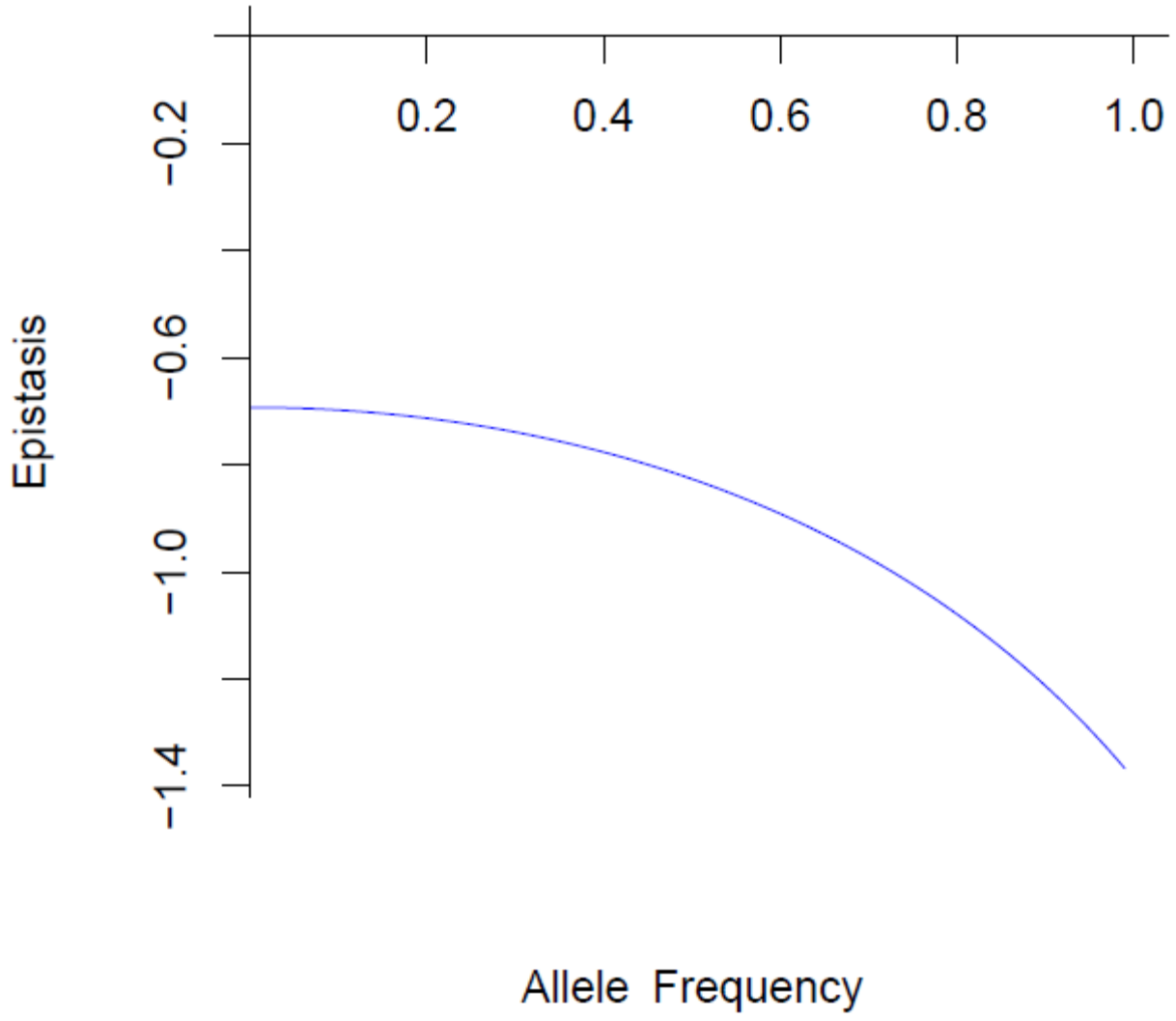


Figure 4.2: The correlation between epistasis and allele frequency, based on the realized fitness values shown in Figure 4.1.

The value of epistasis was calculated using equation (4.7), with the frequency of the *A* and *B* alleles plotted on the *x*-axis. Note that the fitness values that go into the calculation of epistasis are themselves changing in a frequency-dependent manner as shown in Figure 4.1.

From Figures 4.3 and 4.4, we can see that linkage disequilibrium is increased following random mating and free recombination, regardless of how it is measured. This increase in the value of linkage disequilibrium, indirectly, means that phenotypic and fitness variances are increased when individuals freely recombine.

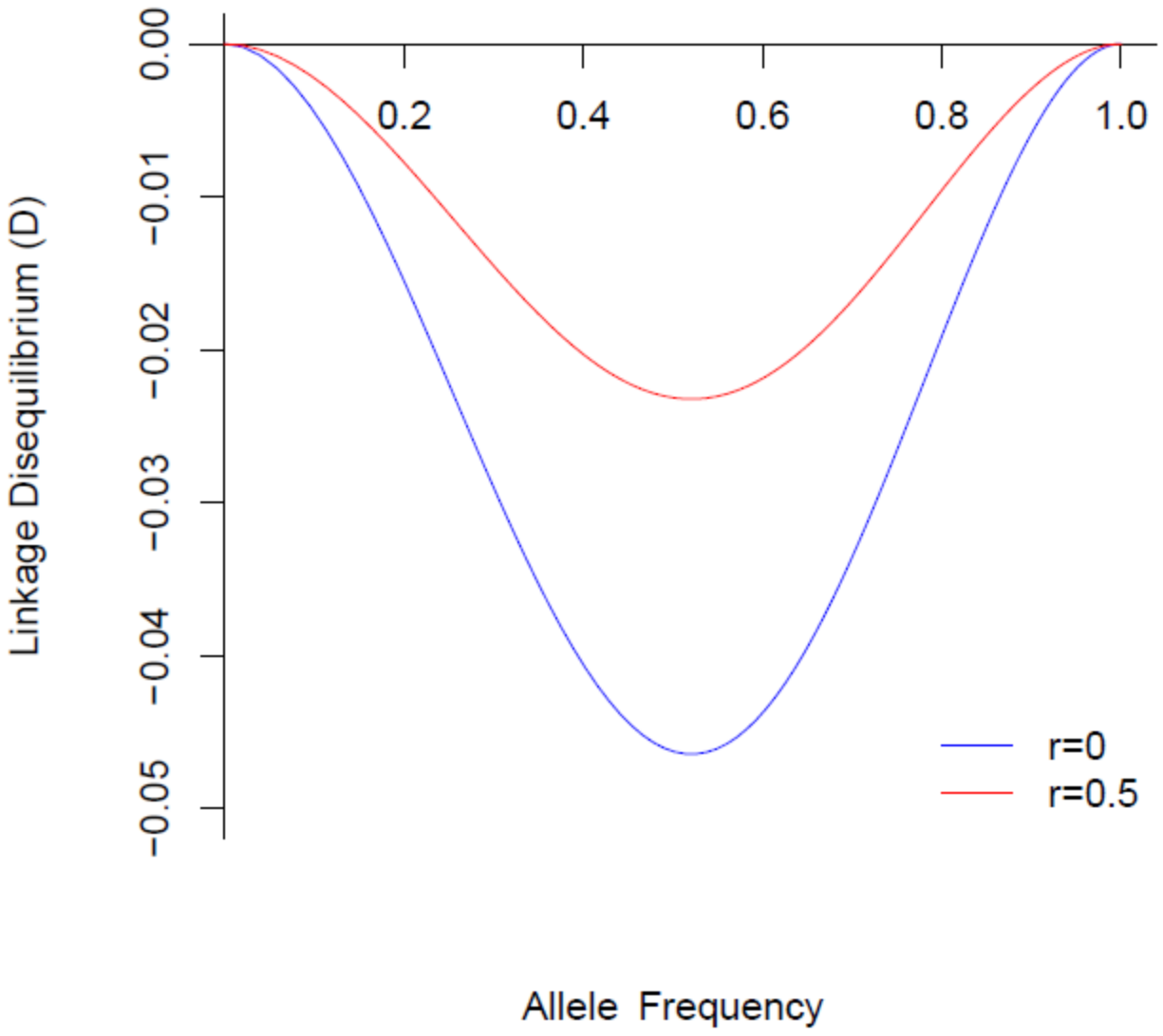


Figure 4.3: The relationship between the value of linkage disequilibrium (*D*) after selection and random mating, and allele frequency.

The value of *D* is shown on the vertical axis, while the frequency of the *A* and *B* alleles after selection is shown on the horizontal axis. For simplicity, the frequencies of the *A* and *B* alleles are assumed to be equal, and the population is initially in linkage equilibrium. Two lines are plotted: the blue line describes the change in the value of *D* when there is no recombination (r=0), while the red line describes the change in the value of *D* if individuals can recombine freely (r=0.5). Note that the value of *D* is always negative, and has a minimum value when the allele frequencies are equal to 0.5.

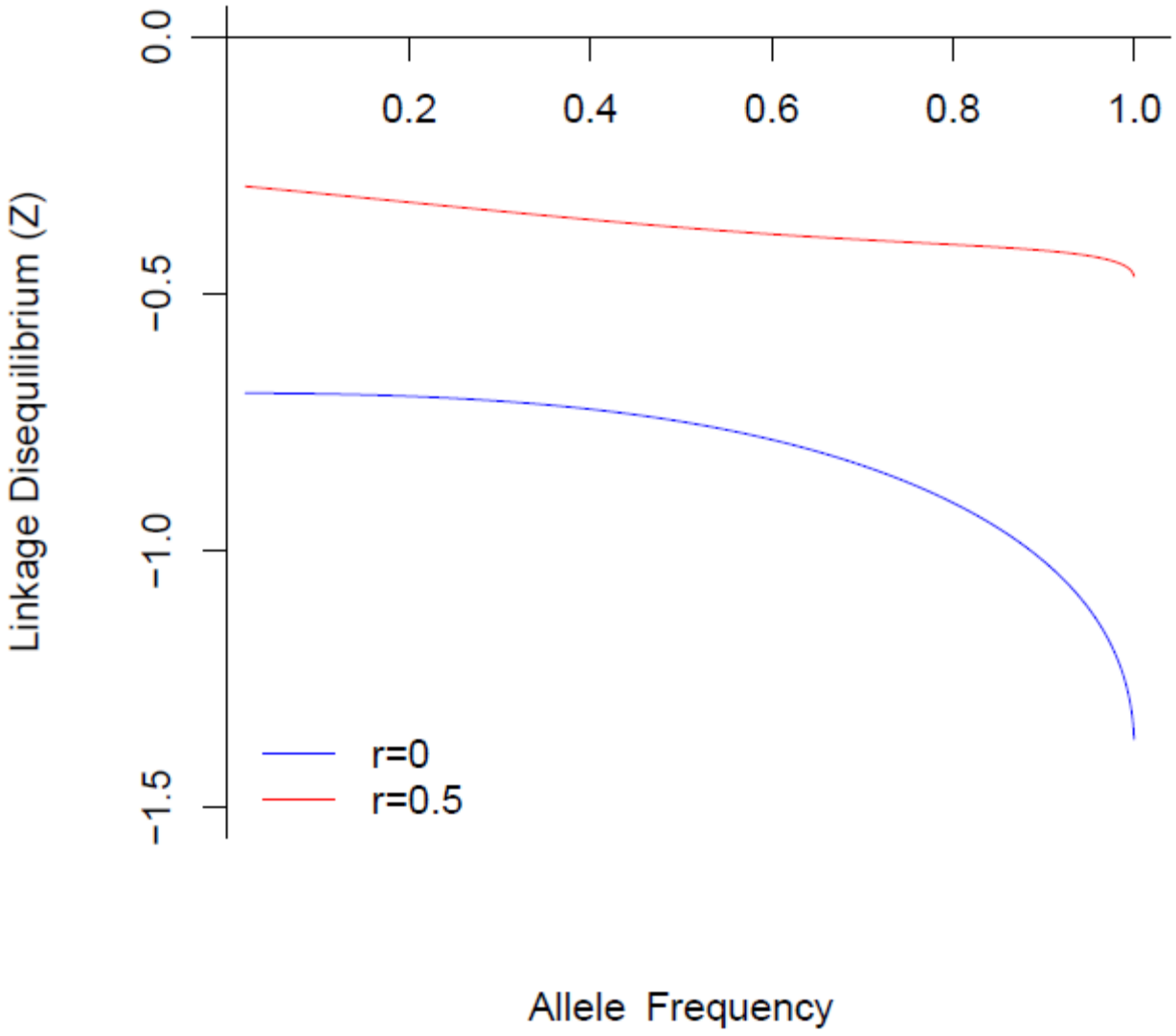


Figure 4.4: The relationship between the value of linkage disequilibrium (*Z*) after selection, and allele frequency.

For a description, see Figure 4.3. *Z* is used to quantify linkage disequilibrium in this case instead of *D*. Note that the value of *Z* increases with free recombination, and is correlated to frequency-dependent epistasis (see Figure 4.2).

4.3 Comparing the increase in frequencies of the selected alleles in recombinant and non-recombinant populations.

Our next goal was to determine whether or not competitively advantageous alleles increase at a faster rate in a sexual population than in an asexual population, provided that the stronger phenotype always wins their competition.

The sexual and asexual populations were modeled with the frequencies of the *A* and *B* alleles being initially equal in both populations, and the genotype frequencies in linkage equilibrium. Generations were computed by iterating equations (4.11) when the population is asexual, and (4.11) followed by (4.12) when the population is sexual. We track of the frequencies of the *A* and *B* alleles, and the frequency of the *AB* genotype, comparatively between sexual and asexual populations until the selected alleles became fixed. We then ran individual-based simulations with the same initial parameters to validate the results. Simulations were run with a population size of 10 000 individuals.

The individual-based simulations gave results that were in strong agreement with the expected frequencies. Compared to when the stronger competitor has better odds of winning (see Figures 4.6 and 3.6), these results show the same trend but with less variance between runs. The simulations that included a round of recombination resulted in a more rapid fixation of the selected alleles, *A* and *B*; as well as of the doubly favored genotype, *AB* (see Figures 4.5 and 4.6). The advantage of recombination is that it reduces the negative linkage disequilibrium that builds up in response to the negatively epistatic competitive selection (see Figure 4.2).

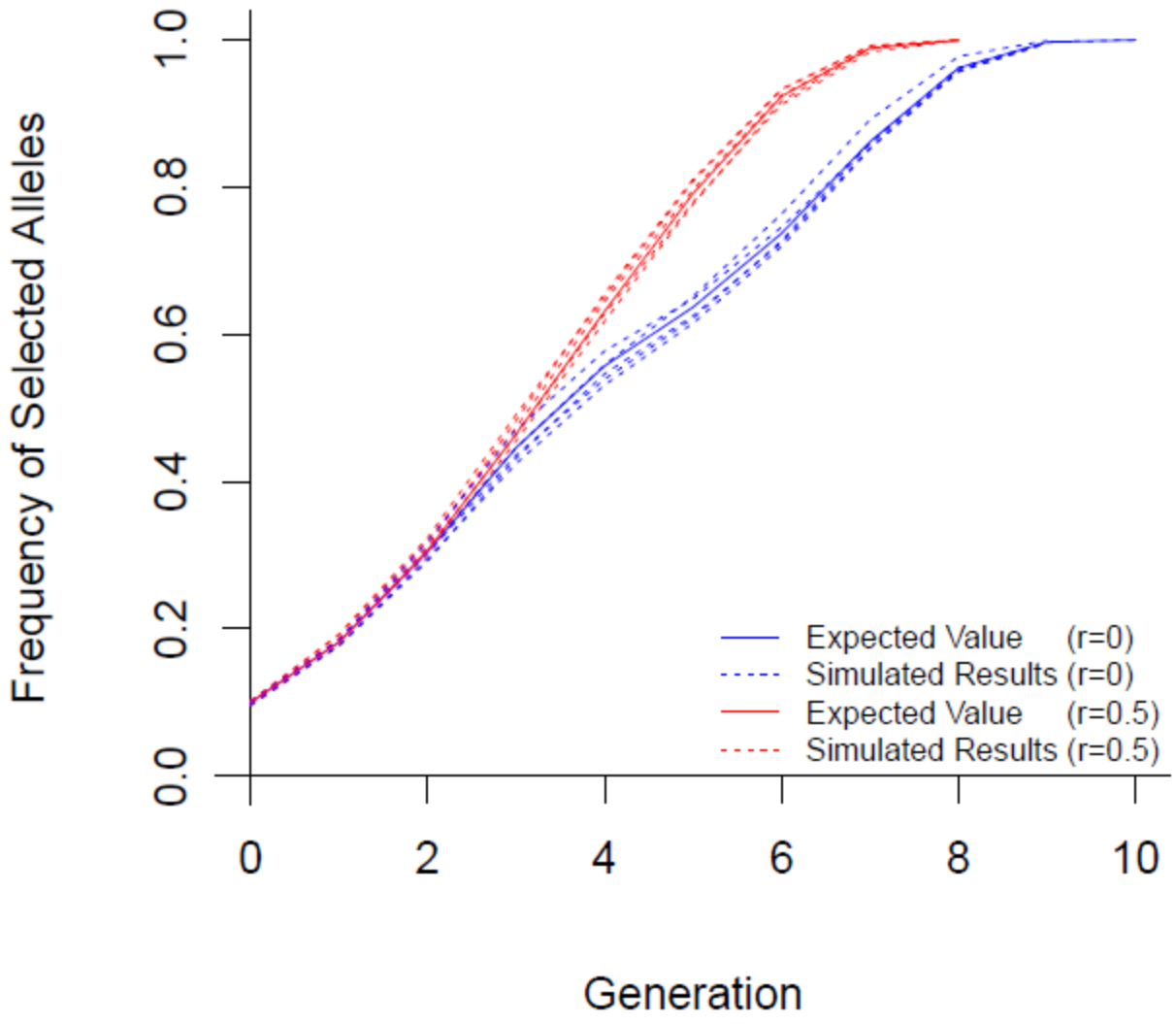


Figure 4.5: The change in the frequency of the selected alleles over successive generations. The frequency of the *A* and *B* alleles is shown on the vertical axis in two different scenarios: when all individuals do not recombine (r=0), or recombine freely (r=0.5). The frequency of the *A* and *B* alleles is initially 0.1, and the population is in linkage equilibrium. Generations are counted until the frequencies of the *A* and *B* alleles are fixed in the population. The numerically expected change of frequency - in both cases - is shown with a solid line. Both results were validated by running 5 simulations under the same initial parameters, and are shown using dotted lines. Note that the frequency of the selected alleles increases at a faster rate with recombination.

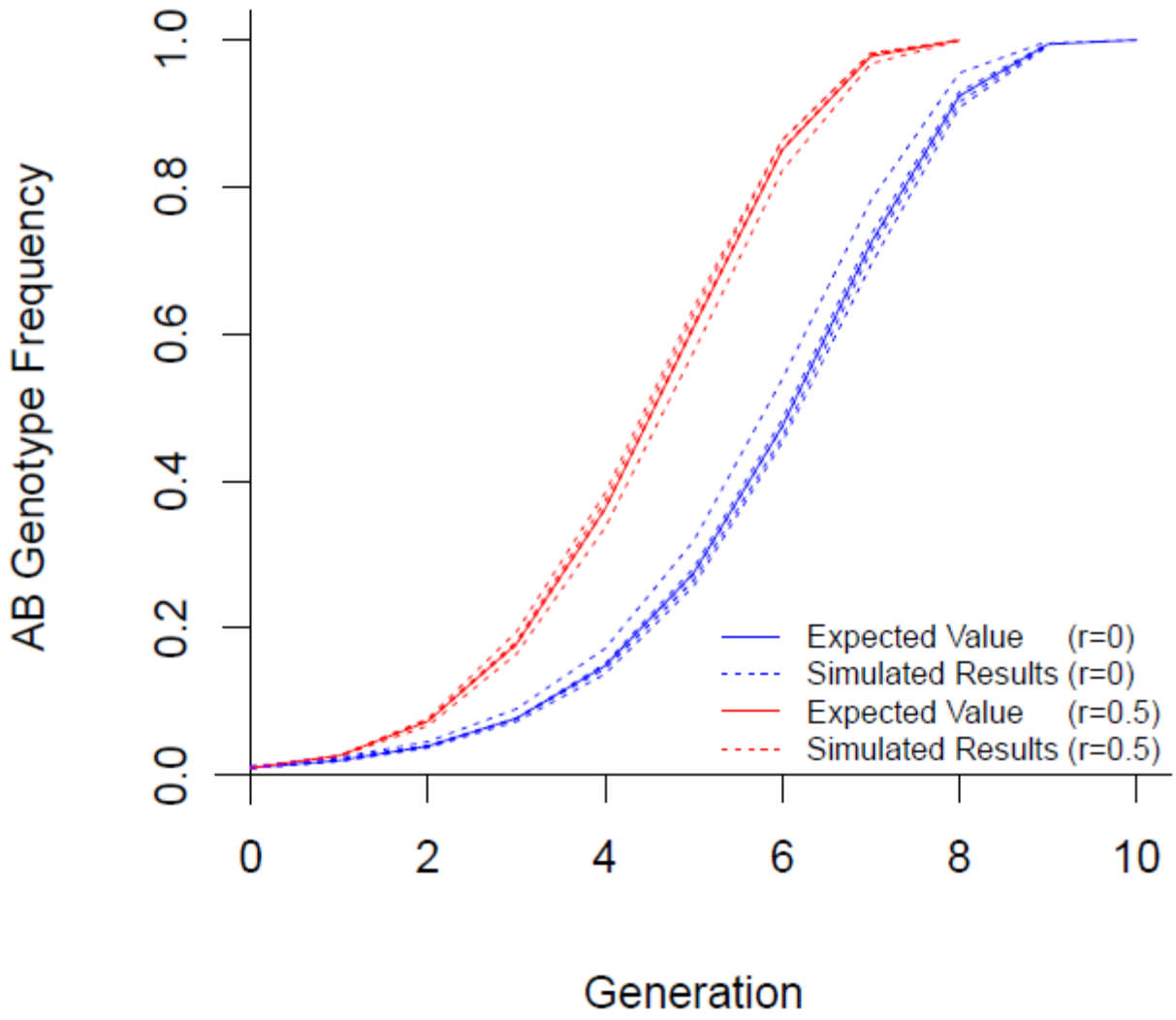


Figure 4.6: The change in the frequency of the *AB* genotype over successive generations. For a description, see Figure 4.5. In this case, the frequency of the *AB* genotype is plotted instead of the frequency of the selected alleles. Both the numerical and simulated results show that the frequency of the *AB* genotype increases at a faster rate with recombination.

4.4 Competing sexuals and asexuals.

The fact that fixation of the selected alleles occurs faster in the recombining population suggests that sexual strains would be at an advantage in a mixed population. We tested this prediction directly by simulating mixtures of sexual and asexual individuals within a single population, provided that the stronger phenotype always wins. Specifically, we want to know if the frequency of sexuals will increase as the frequency of the selected alleles rise to become fixed in the population.

Competition is modeled as before, with reproductive method having no effect on competitive ability. Let *Freq(AB,s)* denote the frequency of *AB* genotyped individuals that reproduce sexually, and *Freq(AB,a)* denote the frequency of *AB* genotyped individuals that reproduce asexually. The frequency of all individuals with the *AB* genotype is therefore *Freq(AB,a)+ Freq(AB,s)*. Continuing this for all four genotypes, we get that:

$$\begin{aligned} Freq(ab) &= Freq(ab,a) + Freq(ab,s) \\ Freq(Ab) &= Freq(Ab,a) + Freq(Ab,s) \\ Freq(aB) &= Freq(aB,a) + Freq(aB,s) \\ Freq(AB) &= Freq(AB,a) + Freq(AB,s) \end{aligned} \tag{4.13}$$

These equations simplify how to express the frequencies of each type of interaction since all individuals with the same genotype have the same competitive ability. The frequency of each genotype post-competition can be expressed as follows:

$$\begin{aligned} Freq(ab,a)' &= Freq(ab,a)Freq(ab) \\ Freq(Ab,a)' &= Freq(Ab,a)\big(1 + Freq(ab) - Freq(AB)\big) \\ Freq(aB,a)' &= Freq(aB,a)\big(1 + Freq(ab) - Freq(AB)\big) \\ Freq(AB,a)' &= Freq(AB,a)\big(2 - Freq(AB)\big) \\ Freq(ab,s)' &= Freq(ab,s)Freq(ab) \end{aligned} \tag{4.14}$$

$$Freq(Ab,s)' = Freq(Ab,s)\big(1 + Freq(ab) - Freq(AB)\big)$$
$$Freq(aB,s)' = Freq(aB,s)\big(1 + Freq(ab) - Freq(AB)\big)$$
$$Freq(AB,s)' = Freq(AB,s)\big(2 - Freq(AB)\big)$$

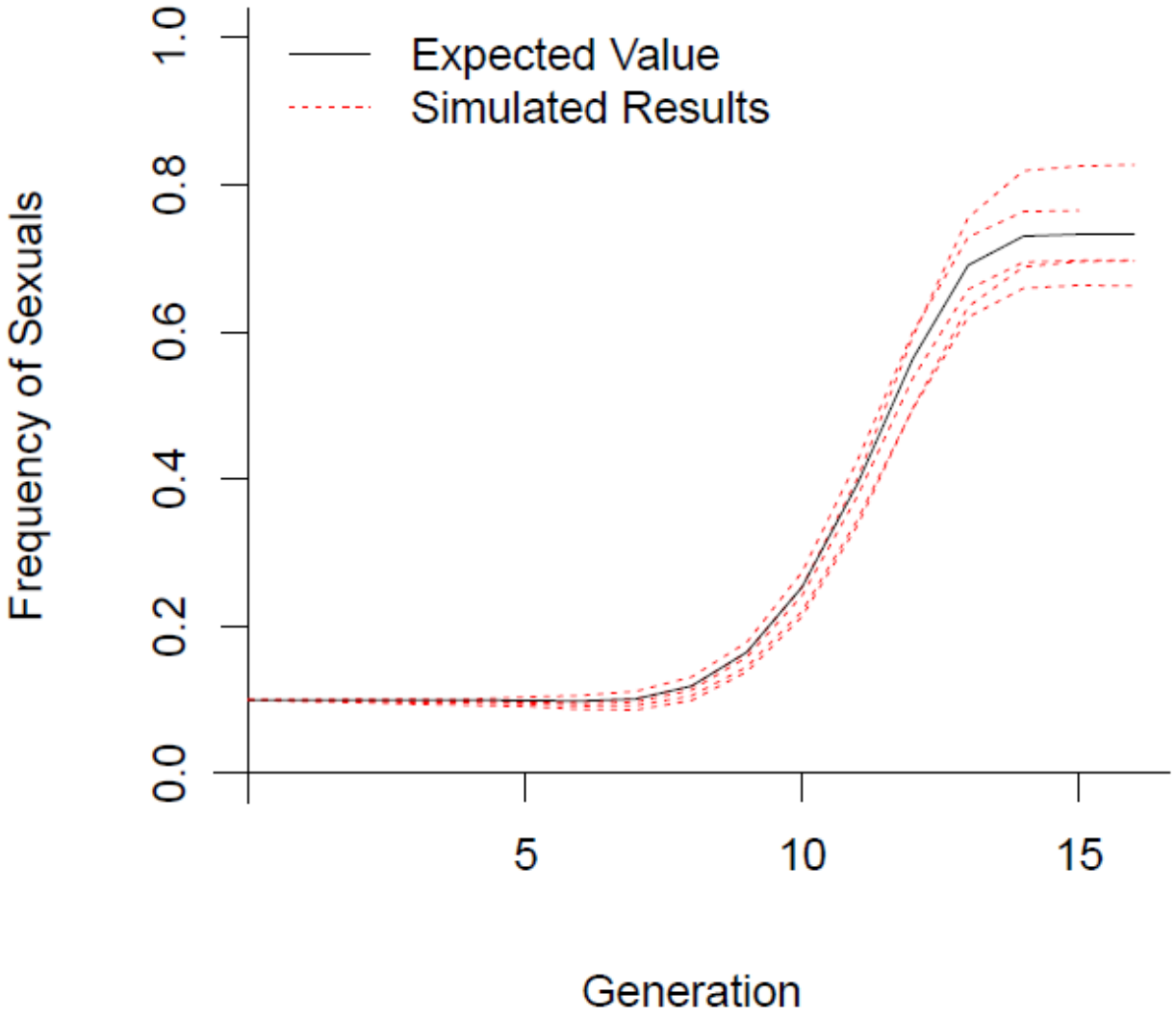
where *Freq(ab,a)'* and *Freq(ab,s)'* denote the frequency of asexual and sexual *ab* genotyped individuals after competition, respectively, and similarly for the other genotypes. Note that the fitnesses of sexuals and asexuals with the same genotype are equal.

The survivors from competition then go on to reproduce. The sexual individuals mate randomly with each other and freely recombine, and the asexual individuals clone themselves. We assume that reproductive method has no effect on the mean family size. Therefore, the frequencies of the asexual genotypes are not expected to change among the offspring. The sexual genotypes, on the other hand, are expected to change frequencies as follows (see section 3.4 for the derivation):

$$\begin{aligned} Freq(ab,s)'' &= Freq(ab,s)' - \frac{D_s}{2Freq(s)'} \\ Freq(Ab,s)'' &= Freq(Ab,s)' + \frac{D_s}{2Freq(s)'} \\ Freq(aB,s)'' &= Freq(aB,s)' + \frac{D_s}{2Freq(s)'} \\ Freq(AB,s)'' &= Freq(AB,s)' - \frac{D_s}{2Freq(s)'} \end{aligned} \tag{4.15}$$

where *Freq(s)'* is the frequency of all sexual parents post-competition; *Freq(ab,s)'* and *Freq(ab,s)''* are the frequencies of sexual parents and offspring with the *ab* genotype, respectively, and similarly for the other genotypes; and D_s=*Freq(ab,s)'Freq(AB,s)'-Freq(Ab,s)'Freq(aB,s)'*.

We calculated the expected frequency change of sexual individuals in a mixed population -when beginning in linkage equilibrium- over a wide range of the parameter space, and validated the results with individual-based simulations with a total population size of 10 000. We find that the frequency of the sexual type does indeed increase during the course of competitive selection (see Figure 4.7). Compared to when the stronger competitor has better odds of winning (see Figure 3.7), the increase in the frequency of sexuals is greater, especially when phenotypic effects are small.



Figuer 4.7: The increase in the frequency of sexual individuals in a mixed population. See Figure 4.5 for a description. Rather than simulating either entirely sexual or entirely asexual populations, here we initiated the simulation with a mixture of types: 10% sexuals and 90% asexuals. The Figure shows that the frequency of the sexuals increased during the course of the simulation. The results for 5 replicate simulations are shown with red lines, and the numerically expected result is shown with a black dashed line. The population size for each replicate was 10 000 individuals which produced 20 000 competing offspring each generation.

4.5 Discussion

We have found that both when the stronger phenotype has better odds of winning, or wins outright, intraspecies competition creates a negative epistatic fitness curve. This changes the genotype frequencies to build-up negative linkage disequilibrium, which is alleviated in a sexual population through random mating and free recombination. In both mappings of phenotype onto competitive ability, the frequencies of the selected alleles increase at a faster rate among sexuals than among asexuals. This allows the sexual individuals to increase in frequency when competing in a mixed population.

Compared to when competitive success is directly proportional to phenotypic value, these results show the same trends, but are more extreme: genotypic fitnesses are more frequency-dependent (especially for the intermediate genotypes, compare Figure 4.1 to Figure 2.2.1), epistatic values are more negative, there is even stronger negative linkage disequilibrium, and sexual individuals show a greater increase in frequency under the same initial parameters. There is a mathematical reason for this, as when phenotypic advantages become very large, the proportional model asymptotically converges towards having the stronger competitor always win (Compare Table 4.1 to Table 2.1.3 where $x=y \rightarrow \infty$). Thus these results show frequency-dependent epistatic effects, as is the case when phenotypic advantages are strong and the stronger competitor has better odds of winning.

We have shown that sexually recombinant individuals will be selected for under contest competition under a variety of mappings of phenotype onto fitness. Other mappings are still plausible, however, and it can be argued that perhaps not all mappings will give negative epistatic results and favour sexual reproduction. That being said, it is promising that these two intuitive mappings do give negative epistatic results. Our findings strengthen the argument that

competitive selection's ability to generate a negative epistatic fitness curve is more likely the rule as opposed to the exception. These results combined with those from the previous chapters help strengthen the argument that competition can explain the maintenance of sex. Under a variety of mappings of phenotype onto fitness, sexual reproduction allows a species to evolve at a faster rate.

CHAPTER 5

General Discussion

Kondrashov (1982) proposed that if alleles do not contribute multiplicatively to fitness, but instead display a negative epistatic fitness curve, then sex will be advantageous because it increases the mean fitness of the population. However, there is no *a priori* reason to believe that fitness values should generally be negatively epistatic, as opposed to positive or zero. Furthermore, the experimental evidence has not been very supportive. A recent database analysis of functional interactions in *Saccharomyces cerevisiae* found that a deviation from multiplicative effects on the phenotype acts as the best predictor of an underlying functional relationship (Mani *et al*., 2008). Therefore, if we assume a direct relationship between phenotype and fitness (a frequent assumption, see Lewontin, 1974), this suggests that zero epistasis is the ideal null hypothesis. Some authors have recently published encouraging results for the presence of negative epistatic fitness by directly measuring the relative fitness of a small set of genotypes (Chou *et al*., 2011; Khan *et al*., 2011). However in general, current reviews suggest that there is no strong experimental support to back up this claim (de Visser *et al*. 2007, Kouyos *et al*., 2007).

Kondrashov (1988) later argued that if the genomic rate of deleterious mutations, U, is greater than 1 in most species, then we can extrapolate that most allelic interactions must be negatively epistatic or the mutation load will be too large. Experimental results have not consistently supported the assumption that U is generally greater than 1 (de Visser *et al*. 2007). Furthermore, long term experiments in *Chlamydomonas reinhardtii* in a benign environment

have found that sexual populations do not evolve to have a greater mean fitness (Renaut *et. al*., 2006).

This lack of evidence in support of ubiquitous negatively epistatic fitness has led some researchers to propose alternative methods of how fitness interactions can promote the maintenance of sex. One approach has been to embrace the variance in types of fitness interactions in the hopes that the negatively epistatic ones will have a greater impact on the maintenance of sex than the positively epistatic ones. This approach has given results that generally favour an asexual mode of reproduction (Otto and Feldman, 1997). Another tactic has been to look at genotypic fitness interactions at multiple loci, and hope that the evolutionary path that is taken on such a rugged fitness landscape will favour the maintenance of sex. This approach might be considered more promising than the former, but so far the theoretical results have been mixed. Otto *et al.* (1994) found that recombination slows the crossing of a fitness valley, but accelerates the ascent to the peak once that valley is crossed. Kondrashov and Kondrashov (2001) found a general disadvantage of sex using individual-based simulations on a fitness landscape allowing for two-dimensional epistasis and no local fitness maxima. Later, Watson and Wakeley (2005) modified this landscape to allow for multi-dimensional epistasis and found conditions where sex does provide an advantage. However, simulations of the specific fitness landscapes found in *Aspergillus niger* favour asexual reproduction (Arjan *et al*., 2009).

Experimental attempts to determine fitness have –generally speaking– not been done in a competitive scenario (Peck and Waxman, 2000). Fitness assays are often accomplished by growing each strain against an identifiable reference strain in a replication-race (Chou *et al*., 2011; Khan *et al*., 2011). This methodology allows for the direct measurement of relative fitness under ideal conditions, but the focus is on fertility selection as opposed to viability selection. In

this study, we argue that the observed level of epistasis will change depending on the growth conditions, with negative epistasis being more likely to be found in competitive environments.

We have developed a model to look exclusively at the effects of viability selection when competing for limited resources. Assuming contest competition in Chapter 2, we showed that multiplicative phenotypic effects give genetic fitness values that are frequency-dependent. In spite of these changes in the fitnesses of individual genotypes, the mean fitness of the population remains constant and equal to 1. During competitive selection, the lowest fitness genotype was more adversely affected than its phenotype might suggest. Consequently, there will be negative epistasis at the fitness level even though there is no epistasis at the phenotype level. We have shown that, when beginning in linkage equilibrium, there will always be negative epistasis at the level of fitness for all phenotypic values; and we showed numerically that this is also true when the population is initially in linkage disequilibrium: be it positive, or negative.

Table 5.1 summarizes various views of the relationship between phenotypic value, absolute fitness (usually assumed to be simply equal to the phenotypic value), relative fitness (where all genotypic fitnesses are normalized relative to the genotype with the maximum fitness), and realized fitness due to pairwise competition (which is used in this study). The Table shows how the absence of epistasis at the phenotypic level results in a corresponding lack of epistasis at the fitness level when fitness is measured either as absolute fitness or relative fitness. But a lack of epistasis at the phenotypic level still results in negative epistasis at the level of realized fitness when we consider pairwise competition.

Table 5.1: The relationship between different measures of fitness, assuming equal genotypic frequencies.

Genotype	Phenotypic value	Absolute fitness	Relative fitness	Realized Fitness
ab	2	2	0.25	0.683
aB	4	4	0.50	1
Ab	4	4	0.50	1
AB	8	8	1.00	1.317
Average fitness		4.5	0.5625	1
Multiplicative epistasis		0	0	-0.106

In this Table, the frequencies of the four genotypes are assumed to be equal to ¼.

Not only did we find there to be a negative epistatic fitness curve with multiplicative effects at the phenotype level, but we also found this to be true when phenotypic effects are additive, when there is negative epistasis at the phenotype level, and to an extent, when there is positive phenotypic epistasis. In general, we found that the population will almost always be in negative linkage disequilibrium after selection, over all allele frequencies.

Tables 5.2.1-5.2.3 summarize how these fitness values change when phenotypic values are not multiplicative. Positive epistasis at the phenotype level has the effect of increasing the mean absolute fitness, while at the same time decreasing the mean relative fitness (compare Tables 5.2.1 and 5.2.2to 5.1). When there is negative phenotypic epistasis, the effects are the opposite: the mean absolute fitness decreases while the mean relative fitness increases (see Table 5.2.3). However, the mean fitness is always equal to 1 with pairwise competition. Furthermore, we can see that the value of epistasis at the fitness level is the same in both absolute and relative terms. But this is not the case with pairwise competition. The range of epistatic values is

dampened, and more importantly, is slightly depressed (see Table 5.2.2 and 5.2.3). Consequently, there is almost always negative epistasis at the fitness level when individuals compete. Very strong positive epistasis at the phenotype level is required for there to be positive epistasis at the fitness level.

Table 5.2.1: The relationship between different measures of fitness, assuming equal genotypic frequencies, and positive epistasis

Genotype	Phenotypic value	Absolute fitness	Relative fitness	Realized Fitness
ab	2	2	0.2	0.667
aB	4	4	0.4	0.976
Ab	4	4	0.4	0.976
AB	10	10	1.0	1.381
Average fitness		5	0.5	1
Multiplicative epistasis		0.223	0.223	-0.034

Table 5.2.2: The relationship between different measures of fitness, assuming equal genotypic frequencies, and strong positive epistasis

Genotype	Phenotypic value	Absolute fitness	Relative fitness	Realized Fitness
ab	2	2	0.167	0.655
aB	4	4	0.333	0.958
Ab	4	4	0.333	0.958
AB	12	12	1.0	1.429
Average fitness		5.5	0.458	1
Multiplicative epistasis		0.405	0.405	0.018

Table 5.2.3: The relationship between different measures of fitness, assuming equal genotypic frequencies, and negative epistasis

Genotype	Phenotypic value	Absolute fitness	Relative fitness	Realized Fitness
ab	2	2	0.333	0.708
aB	4	4	0.667	1.033
Ab	4	4	0.667	1.033
AB	6	6	1.000	1.225
Average fitness		4	0.667	1
Multiplicative epistasis		-0.288	-0.288	-0.207

In the presence of recombination following contest competition, the amount of linkage disequilibrium among loci is reduced, bringing the population closer to linkage equilibrium. This has the effect of increasing the mean phenotype, as well as the phenotypic variance. More importantly, recombination increases the variance of genotypic fitness in the population, which strengthens the action of natural selection.

In Chapter 3, we showed that negative epistatic fitness persists generation-over-generation when competing, regardless of the rate of recombination within the population. We focused primarily on a population that is initially in linkage equilibrium, as this is the expected state of a population exposed to random mutations in the absence of selection. Over successive generations, the amount of negative linkage disequilibrium builds up when not recombining. However, this accumulation is alleviated when recombination is present. The build-up of negative linkage disequilibrium reduces the genotypic variance in the population, which hinders selection. We have shown that the frequencies of the selected alleles increase at a faster rate when recombining, and consequently, so does the frequency of the strongest phenotype.

Therefore, sexual individuals will increase in frequency when competing with asexuals. The only cases where sex is not advantageous are when 1) the population is initially in strong positive linkage disequilibrium, and the selected alleles are close to fixation, or 2) when phenotypic effects show considerable positive epistasis. Both of these cases are more likely the exception than the rule in nature. The selective advantage of recombination is strongest when there is a large difference in competitive ability between genotypes, and when the selected alleles are initially rare.

In Chapter 4, we showed that these results are not specific to a single mapping of phenotype onto competitive ability. When the stronger phenotype always wins, we find that there is always negative epistasis at the fitness level, and we find that this causes negative linkage disequilibrium in the population following selection. Consequently, the stronger phenotypes go to fixation at a faster rate when recombining than without, and allows sexual individuals to increase in frequency when competing against asexual individuals in a mixed population.

This model is a form of rank order selection: where fitness is determined by the ranking of the organism's phenotype relative to those of the rest of the population (Milkman, 1973). The results of these localized competitions give rise to a form of truncation selection: where the individuals that fail to make a certain phenotypic cutoff are culled from the population. The weakest individuals are unable to compete, and consequently have very low fitness values. Meanwhile, individuals with phenotypes that do make the cutoff win their competitions, and have high fitness values. Therefore, as is the case with a direct analysis of truncation selection, we expect to find negative epistatic fitness (Crow and Kimura, 1978).

Taken together, the results suggest that contest competition should create a negative epistatic fitness curve, even if there isn't negative epistasis at the phenotype level. Under a

variety of competitive mappings of phenotype onto fitness, it is expected that individual competition causes negative epistasis at the fitness level. These results provide an *a priori* reason why allelic interactions at the fitness level should generally follow a negative epistatic fitness curve, which is considered to be a missing link in the negative epistatic fitness hypothesis for the maintenance of sex. Our models show that the majority of phenotypic interactions should produce fitness level epistasis for a variety of mappings of phenotype onto fitness, regardless of the frequencies of the selected alleles, and regardless of the initial genotype frequencies. It is especially worth noting that positive epistasis at the phenotype level can translate into negative epistasis at the fitness level.

In the classical Wright-Fisher model of natural selection, it is assumed that an individual's fitness is completely determined by its genotype. While this is an important element of natural selection, in the words of Richard Lewontin (1974) "population genetics, is an essential ingredient, but it is not the entire soup" of evolutionary theory. In practice, it has long been understood that selection is affected by a multitude of factors. For example, modelers of population growth often see selection as being composed of the Malthusian growth rate, which has a genetic component; and the carrying capacity of the environment, for which the genetic component is less clear (Verhulst, 1838). Modelers of intraspecies selection further differentiate the type of selection, as it can be dependent on either the population density, or the current frequencies of other genotypes in the population (Wallace, 1975). To complicate things more, ecologists also find causes of selection in host-parasite interactions, external environmental changes (Benton, 2009), levels of predation, etc.

Our model focuses on a type of frequency-dependent selection, which is implicitly density-dependent as there are assumed to be more offspring than there are resources to

accommodate them. Our model is similar to the frequency-dependent matrix-game model described by Yi and Lessard (2000); with the distinction that we are dealing with a 2-locus haploid organism instead of a 1-locus diploid organism. Wallace (1975) has referred to this particular kind of selection as soft selection: where the most competitive genotypes tend to survive, while other comparatively weaker genotypes are eliminated from the population. Interestingly, the value of epistasis should then be expected to change depending on the dominant form of selection: with no epistasis being favoured when selection is mainly hard, and negative epistasis when selection is mainly soft.

Our hypothesis that sex can be maintained during contest competition was shown to be theoretically plausible. We expect that competitively advantageous alleles will increase in frequency at a faster rate within a sexual population than within an asexual population. This allows the frequency of sexual individuals to increase generation-over-generation when competing in a mixed population.

This being said, there is still work to be done to solidify the theory. While we have shown that there is an evolutionary advantage to sexual reproduction during contest competition, so far, we have only shown the case where the contest is a one-on-one. While this is likely to be the most frequent form of contest competition, it is not the only form. One can imagine an individual-based "soup model" similar to the one presented here where in each contest, three individuals are randomly chosen from the population to compete, or four individuals, or some other number.

The full power of this theory's ability to explain the maintenance of sex has yet to be shown, as it remains to be seen how much more sexuals will be favoured when selection is taking place at many loci. Most recombination-oriented hypotheses require that many loci to be

under selection in order to overcome the so-called "two-fold cost of sex", or selection needs to be very strong (Kondrashov, 1982; West *et al*., 2001; Keightley and Otto, 2006).

If we extend our analysis to three loci (see Appendix 4 for the derivation of pairwise fitness values, and epistasis with three loci), we can see that there is a greater range of genotypic fitness values. The value of pairwise epistasis is slightly dampened, but the difference is equal to the value of 3-loci epistasis (compare Table 5.1 to Table 5.*3*). Therefore, when many loci are under selection, we can expect to find pairwise negative epistasis, as well as higher-order negative epistasis.

Table 5.3: The relationship between different measures of fitness, assuming equal genotypic frequencies, with three loci.

Genotype	Phenotypic value	Absolute fitness	Relative fitness	Pairwise competition
abc	2	2	0.125	0.553
abC	4	4	0.25	0.842
aBc	4	4	0.25	0.842
Abc	4	4	0.25	0.842
aBC	8	8	0.5	1.158
AbC	8	8	0.5	1.158
ABc	8	8	0.5	1.158
ABC	16	16	1.00	1.447
Average fitness		6.75	0.422	1
Pairwise epistasis between A and B		0	0	-0.096
Pairwise epistasis between B and C		0	0	-0.096
Pairwise epistasis between A and C		0	0	-0.096
Higher-order epistasis between A, B and C		0	0	-0.010

Doncaster *et al*. (2000) showed that under ecological conditions of population growth and competition, the two-fold cost of sex is almost never two-fold. Sexuals can resist invasion from an asexual defector, provided that their intrinsic growth rate is sufficiently high. Interestingly, it would appear that ecological conditions both reduce cost of sex, while at the same time increase the advantages of sexual reproduction. It would be interesting to determine the minimum amount of selection required to overcome the two-fold fertility cost when multiple loci are under selection.

Finally, the complete analysis of a theory that attempts to explain the maintenance of sex should address two issues: first is to show that sexuals are favoured over asexuals, and second is to show that a modifier allele that increases the recombination rate is favoured within a sexual population (Otto and Lenormand, 2002). So far we have shown the former, but not the latter. Sexual reproduction can occur with a variety of different recombination rates per chromosome, but the advantage we have shown only exists when the two processes are coupled. Recombination is not always selected for within a sexual population. For instance, Barton's (1995) analysis of an infinite population found that recombination will only be favoured if the value of epistasis fluctuates in sign every few generations, or if there is weak negative epistasis. In a finite population, Barton and Otto (2005) found that recombination will be selected for when the population size is small (i.e. less than 10 000) and there is tight linkage between the recombinant modifier locus and the selected locus (i.e. a recombination rate less than 10 cM). But with larger population sizes and looser linkage, as is the case in humans and many other species, the advantage quickly goes to zero. It remains to be seen under what conditions recombination rates will tend to increase when sexuals compete.

Our results are consistent with those of Peck and Waxman (2000) who found that scramble competition (i.e. when a resource is equally or unequally shared between competitors) can also cause a negative epistatic fitness curve when competing in small groups. They further showed that this allows sexuals to be selected for when competing with asexuals in a mixed population. Taken together, these results suggest that intraspecies competition of any kind will allow for the maintenance of sex, as scramble and contest competition make up the whole of intraspecies competition. However, the fate of a modifier allele is still unclear when individuals compete.

The theory that intraspecies competition causes negatively epistatic fitness is not the only ecological theory that attempts to explain the maintenance of sex. The Tangled Bank hypothesis (Bell, 1982) proposes that sex is advantageous when the environment is constantly changing, and the Red Queen hypothesis (Hamilton, 1980) proposes an advantage when hosts and parasites are caught in a cyclic arms race. Unlike the Tangled Bank, this theory does not require for there to be spatial heterogeneity (Becks and Agrawal, 2010) or a spectrum of resources (Case and Taper, 1986); and unlike the Red Queen hypothesis, it does not require the presence of host-parasite interactions (Hamilton, 1980). But it does require that soft selection be prevalent in nature. These ecological hypotheses propose different causes for the maintenance of sex, but share the commonality that they all require some form of frequency-dependent selection in order for sex to be advantageous. The form of frequency-dependent selection in our model is somewhat different in the sense that the fitness of a genotype is not negatively correlated with its frequency. Specifically, the fittest genotype in the Red Queen and Tangled Bank models is the one with the lowest frequency; but in our model, it is the one with the strongest phenotype. Peck (1993) looked directly at negative frequency-dependent selection where the fertility of an individual is a

decreasing function of its genotype frequency, and came to similar results as both the Tangled Bank and the Red Queen models. Of these two theories, the Tangled Bank models gives results that are more similar to ours in that some form of sex is advantageous when there are more individuals than there are resources to sustain them (Taper and Case, 1986; Sheu and Drossel, 2007). The Tangled Bank model, however, further requires that some resources are overexploited while others are underexploited, but in our model this additional assumption is unnecessary.

Ideally, a theory that fully explains the ubiquity of sex should be applicable wherever sexual reproduction is found in nature. It is possible that intraspecies competition meets this criterion as it is a ubiquitous part of the natural world, and may be found anywhere there is a reproductive excess for the amount of available resources. As noted by Lewontin (1955), ‘‘it would be strange if what applied to different species did not apply to some extent to different genotypes within the same species.’’ A meta-analysis covering 527 field experiments found that when competition was demonstrated and the two forms of competition were differentiated, intraspecific competition was as strong as or stronger than interspecific competition in three-quarters of the experiments (Connell, 1983). Excess fertility increases the likelihood of intraspecific competition and is required for genetic substitutions (Nei, 1971). We have shown that genetic substitutions can occur at a faster rate when the organisms reproduce sexually.

Also, a theory that fully explains the maintenance of sex should not apply for the cases of rare asexually-reproducing species. Therefore, we should expect to find asexual species in areas of low competition: where either the environment is very harsh; or where resources are so plentiful that competing for them is unnecessary (such as in a replication race). In this respect, the relationship between intraspecific competition and recombination may help to explain the

existence of such species. Many of the well-known asexual species live in low competition environments, or have evolved ways of avoiding competition for shared resources. For example, there are several asexual lizards and plants that inhabit desert areas (Kearney, 2003), where conditions are harsh and consequently, have reduced levels of biotic interactions. Dandelions are considered an exemplary *r*-strategist (Gadgil and Solbrig, 1972); they thrive in areas with low competition for sunlight, flowering in the early spring when resources are in high supply, yet demand is relatively low. The *Bdelloid rotifer*'s ability to survive desiccation and remain dormant for extended periods can be seen as an evolved strategy to avoid intense competition when conditions for active life are unfavourable (Ricci, 2001).

In such *r*-selected species, the main component of selection is on fertility and we expected relatively little build-up of negative linkage disequilibrium in response to competitive interactions between conspecific individuals. In the case of *K*-selected species however, such as the plants which "clothe the ground", competition can cause the build-up of negative linkage disequilibrium, thus providing a selective advantage for genetic recombination. Therefore, if sex is maintained through intraspecies competition, then we should expect to find a correlation between the frequency of sexual reproduction and competition. Such a broad experiment may not yet be possible. On a smaller scale, such as a species that is capable of both sexual and asexual reproduction, we might expect to find a correlation between the environment of an organism and the frequency with which it prefers one reproductive method over the other. This is indeed the case in many organisms: many facultative sexual model organisms prefer asexual reproduction when resources are in high supply, and switch to sexual reproduction when starved; and some species like the rotifer *Brachionus calyciflorus* switch to sexual reproduction when living in high densities. If sex is advantageous when individuals compete, then these adaptations are

evolutionarily advantageous, as they maximize the fitness of the individual. That being said, *Brachionus calyciflorus* has been shown to evolve greater sensitivity to a density stimulus when growing in a heterogeneous environment than when growing on either homogeneous high or low quality food sources (Becks and Agrawal, 2010).

This is not to say that intraspecies competition is the definitive answer for maintenance of sex, but that the theory is theoretically and naturally plausible. It is by no means the only reasonable explanation. Like most other genetic variation-centred theories for the evolution of sex, intraspecies competition has the capacity to explain the maintenance of sex, but not its origin.

Our results suggest that epistatic fitness values might change depending on the conditions of the experiment. We hypothesize that experimental results attempting to determine epistasis might find different results when genotypes are grown separately (hard selection) than when grown together (soft selection). Specifically, there might be no epistasis when resources are plentiful, and negative epistasis when resources are scarce.

To our knowledge there is very little experimental work that has been done on intraspecies competition and either fitness level epistasis, or the maintenance of sex. A study focusing on deleterious mutations in the parasitic wasp *Nasonia vitripennis* when grown under competitive conditions found negative epistatic fitness effects for longevity, but not for egg production (Rivero *et. al*., 2003). Another *in silico* experiment was done on the intracellular growth of bacteriophage T7 under a variety of environmental conditions (You and Yin, 2002). They found that mildly deleterious mutations interacted in a negative epistatic fashion in poor-resource environments but showed positive epistasis when grown in rich-resource environments. These results are generally in agreement with our model, as well as our assumptions about the

effects of viability versus fertility selection. The results of Wang *et al*. (2009) are in disagreement at the surface level, as they found that the value of epistasis slightly increases when *Drosophila melanogaster* are grown in a low quality environment as opposed to a high quality environment. However, the same lab also found that selection was harder when *Drosophila* were grown in the low quality environment, not softer (Laffafian *et al*., 2010).

We conclude that intraspecies competition is a plausible explanation for the maintenance of sex. We believe the theory is worth further theoretical and experimental analysis, and that much work still needs to be done before coming to a definitive answer for the enigma of sex. Our simulations suggest that perhaps the advantage of sexual reproduction doesn't lie in fertility selection, but in viability selection. Evolution would be nothing more than a replication race if there was no shortage of resources, which would favour asexual reproduction. But in a competitive situation, it may be better to produce one winner rather than two losers.

REFERENCES

Ackerman S, Kermany AR, Hickey DA. (2010) Finite populations, finite resources, and the evolutionary maintenance of genetic recombination. *Journal of Heredity.* **101:** 135-141.

Arjan, J, De Visser GM., Park SC, Krug J. (2009). Exploring the effect of sex on empirical fitness landscapes. *The American Naturalist*. **174:** 15-30.

Barton NH. (2009) Why sex and recombination? *Cold Spring Harbor symposia on quantitative biology*. **74:** 187-195.

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

Beadle GW, Emerson S. (1935) Further studies of crossing over in attached-X chromosomes of *Drosophila melanogaster*. *Genetics.* **20:** 192-206.

Becks L, Agrawal AF. (2010). Higher rates of sex evolve in spatially heterogeneous environments. *Nature*. **468:** 89-92.

Benton MJ. (2009). The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*. **323:** 728-32.

Bell G. (1982). *The masterpiece of nature: the evolution and genetics of sexuality*. University of California Press.

Bennett JH. (1954) On the theory of random mating. *Annals of eugenics.* **18:** 311-317.

Boulton A, Myers RS, Redfield RJ (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences*. **94:** 8058-8063.

Bernstein H, Byerly HC, Hopf FA, Michod RE. (1985) Genetic damage, mutation, and the evolution of sex. *Science*. 229(4719): 1277-1281.

Bernstein H, Byers GS, Michod RE. (1981) Evolution of sexual reproduction: importance of DNA repair, complementation, and variation. *The American Naturalist*. **117:** 537-549.

Brännström Å, Sumpter DJT. (2005) The role of competition and clustering in population dynamics. *Proceedings of the Royal Society B Biological Sciences*. **272:** 2065-2072.

Cao L, Alani E, Kleckner N (1990). A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. *Cell*. **61:** 1089-1101.

Case TJ, Taper ML. (1986) On the coexistence and coevolution of asexual and sexual competitors. *Evolution*. **40:** 366-387.

Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ. (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science.* **332:** 1190-1192.

Christiansen FB, Otto SP, Bergman A, Feldman MW. (1998) Waiting with and without recombination: the time to production of a double mutant. *Theoretical Population Biology*, **53(3):** 199-215.

Connell JH. (1983) On the prevalence and relative importance of interspecific competition: evidence from field experiments. *The American Naturalist*. **122:** 661:696.

Cordell HJ. (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics.* **11:** 2463-2468.

Creighton HB, McClintock B. (1931) A correlation of cytological and genetical crossing-over in *Zea mays*. *Proceedings of the National Academy of Sciences.* **17:** 492-497.

Crow JF, Kimura M. (1965) Evolution in sexual and asexual populations. *The American Naturalist*. **99:** 439-450.

Crow JF, Kimura M. (1978) Efficiency of truncation selection. *Proceedings of the National Academy of Sciences*. **76:** 396-399

De Visser JA, Elena SF. (2007) The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews Genetics*. **8:** 139-149.

Doncaster CP, Pound GE, Cox SJ. (2000) The ecological cost of sex. *Nature*. **404:** 281-285.

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES. (1987) A genetic linkage map of the human genome. *Cell.* **51:** 319-337.

Felsenstein J. (1965) The effect of linkage on directional selection. *Genetics*. **52:** 349-363.

Felsenstein J. (1974) The evolutionary advantage of recombination. *Genetics*. **78:** 737-756.

Fincke OM, Hadrys H. (2001). Unpredictable offspring survivorship in the damselfly, *Megaloprepus coerulatus*, shapes parental behavior, constrains sexual selection, and challenges traditional fitness estimates. *Evolution.* **55:** 762-772.

Fisher RA. (1918) The correlation between relatives on the supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*. **52:** 399-433.

Fisher RA. (1930) *The genetical theory of natural selection*. Oxford University Press.

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science.* **269:** 496-512.

Gadgil M, Solbrig OT. (1972). The concept of r- and K-selection: evidence from wild flowers and some theoretical considerations. *The American Naturalist*. **106:** 14-31.

Hadany L, Comeron JM. (2008) Why are sex and recombination so common? *Annals of the New York Academy of Sciences.* **1133:** 26-43.

Haldane JBS. (1931) The cytological basis of genetical interference. *Cytologia.* **3:** 54-65.

Hamilton WD. (1980) Sex versus non-sex versus parasite. *Oikos*. **35:** 282-290.

Hardy GH. (1908) Mendelian proportions in a mixed population. *Science.* **28:** 49-50.

Hartl DL, Jones EW. (2005) *Essential genetics: a genomics perspective*. Jones & Bartlett Publishers.

Hedrick PW. (1987) Gametic disequilibrium measures: proceed with caution. *Genetics.* **117:** 331-341.

Hickey DA. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics.* **101:** 519-531.

Hickey DA, Rose MR. (1988) The role of gene transfer in the evolution of eukaryotic sex. In *The evolution of sex*, ed. Michod RE and Levin BR, 161-175. Sunderland, MA: Sinauer.

Hill WG, Robertson A. (1966) The effect of linkage on limits to artificial selection. *Genetical Research*. **8:** 269-294.

Hill WG, Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics.* **38:** 226-231.

Hiraizumi Y. (1971) Spontaneous recombination in *Drosophila melanogaster* Males. *Proceedings of the National Academy of Sciences.* **68:** 268-270.

Holliday R (1964). A mechanism for gene conversion in fungi. *Genetical Research.* **5:** 282-304.

Howard RS, Lively CM. (1994). Parasitism, mutation accumulation and the maintenance of sex. *Nature*. **367:** 554-557.

Huxley J. (1942) *Evolution: the modern synthesis*. Harper & Brothers.

Jeffreys AJ, Kauppi L, Neumann R (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics.* **29:** 217-22.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, and HJ Jacob. (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research.* **14:** 528-538.

Judson OP. (1994) The rise of the individual-based model in ecology. *Trends in Ecology & Evolution.* ***9*:** 9-14.

Kearney MR. (2003). Why is sex so unpopular in the Australian desert? *Trends in Ecology & Evolution.* **18:** 605-607.

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

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. **332:** 1193-1196.

Kimura M, Ohta T. (1969a) The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics.* **63:** 701-709.

Kimura M, Ohta T. (1969b) The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics.* **61:** 763-771.

Kimura M. (1964) Diffusion models in population Genetics. *Journal of Applied Probability*. **1:** 177-232.

Kimura M. (1968) Evolutionary rate at the molecular level. *Nature.* **217:** 624-626.

Kondrashov AS. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genetical Research*. **40:** 325-332.

Kondrashov AS. (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature*. **336:** 435-440.

Kondrashov AS. (1994) The asexual ploidy cycle and the origin of sex. *Nature.* **370:** 213-216.

Kondrashov FA, Kondrashov AS. (2001). Multidimensional epistasis and the disadvantage of sex. *Proceedings of the National Academy of Sciences.* **98:** 12089-92.

Kouyos RD, Silander OK, Bonhoeffer S. (2007) Epistasis between deleterious mutations and the evolution of recombination. *Trends in Ecology & Evolution*. **22:** 308-315.

Laffafian A, King JD, Agrawal AF. (2010) Variation in the strength and softness of selection on deleterious mutations. *Evolution*. **64:** 3232-3241.

Lewontin RC. (1955) The effects of population density and composition on viability in Drosophila melanogaster. *Evolution*. **9:** 27-41.

Lewontin RC. (1964) The interaction of selection and linkage I. General considerations; heterotic models. *Genetics.* **49:** 49-67.

Lewontin RC. (1974). *The genetic basis of evolutionary change*. Columbia University Press, New York.

Lewontin RC. (1988) On measures of gametic disequilibrium. *Genetics.* **120:** 849-852.

Loeschcke V, Christiansen FB. (1984) Evolution and intraspecific exploitative competition II. A two-locus model for additive gene effects. *Theoretical Population Biology.* **26:** 228-264.

Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP. (2008) Defining genetic interaction. *Proceedings of the National Academy of Sciences*. **105:** 3461-3466

Martens K, Rossetti G, Horne DJ. (2003) How ancient are ancient asexuals? *Proceedings of the Royal Society B Biological Sciences.* **270:** 723-729.

Martin G, Otto SP, Lenormand T. (2006) Selection for recombination in structured populations. *Genetics*. **172:** 593-609.

May RM, Anderson RM. (1983) Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society B Biological Sciences*. **219:** 281-313.

Maynard Smith J. (1968) Evolution is sexual and asexual populations. *The American Naturalist*. **102:** 469-473.

Maynard Smith J. (1978) *The evolution of sex*. Cambridge University Press.

Meselson MS, Radding CM (1975). A general model for genetic recombination. *Proceedings of the National Academy of Sciences.* **72:** 358-361.

Milkman R. (1973) A competitive selection model. *Genetics*. **74:** 727-732.

Moran PAP. (1962) *The statistical processes of evolutionary theory*. Clarendon Press.

Morgan TH. (1911) Random segregation versus coupling in mendelian inheritance. *Science.* **34:** 384.

Morran LT, Schmidt OG, Gelarden IA, Parrish RC, Lively CM. (2011) Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science*. **333:** 216-218.

Morton NE, Rao DC, Yee S. (1976) An inferred chiasma map of *Drosophila melanogaster*. *Heredity.* **37:** 405-411.

Muller HJ. (1932) Some genetic aspects of sex. *The American Naturalist*. **66:** 118-138.

Muller HJ. (1964) The relation of recombination to mutational advance. *Mutation Research*. **106:** 2-9.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science*. **310:** 321-324.

Myers S, Freeman C, Auton A, Donnelly P, McVean G (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics*, **40:** 1124-1129.

Narra HP, Ochman H. (2006) Of what use is sex to bacteria? *Current Biology*. **16:** R705-710.

Nei M. (1971). Fertility excess necessary for gene substitution in regulated populations. *Genetics*. **68:** 169-84.

Neumann R, Jeffreys AJ (2006). Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Human Molecular Genetics.* **15:** 1401-1411.

Nicolas AD, Treco NP Schultes, Szostak JW (1989). An initiation site for meiotic gene conversion in the yeast Saccharomyces cerevisiae. *Nature*. **338:** 35-9.

Ohta T. (1982) Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences.* **79:** 1940-1944.

Ohta T, Kimura M. (1969) Linkage disequilibrium due to random genetic drift. *Genetical Research.* **13:** 47-55.

Ohta T, Kimura M. (1971) Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics.* **68:** 571-80.

Otto SP, Barton NH. (1997) The evolution of recombination: removing the limits to natural selection. *Genetics*. **147:** 879-906.

Otto SP, Barton NH. (2001) Selection for recombination in small populations. *Evolution*. **55:** 1921-1931.

Otto SP, Feldman MW. (1997). Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology*. **51:** 134-147.

Otto SP, Feldman MW, Christiansen F. (1994). Some advantages and disadvantages of recombination. *Frontiers in Mathematical Biology*. 198-211.

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

Otto SP, Nuismer SL. (2004) Species interactions and the evolution of sex. *Science*. **304:** 1018-1020.

Otto SP. (2009) The evolutionary enigma of sex. *The American Naturalist*. **174:** 1-14.

Painter TS. (1933) A new method for the study of chromosome rearrangements and the plotting of chromosome maps. *Science.* **78:** 585-586.

Parunak HVD, Savit R, Riolo RL. (1998) Agent-based modeling vs. equation-based modeling: a case study and users' guide. *MultiAgent Systems and AgentBased Simulation*. **1534:** 10-25. Springer.

Peck JR. (1993). Frequency-dependent selection, beneficial mutations, and the evolution of sex. *Proceedings of the Royal Society B: Biological Sciences*. **254:** 87-92.

Peck JR, Waxman D. (2000) Mutation and sex in a competitive world. *Nature*. **406:** 399-404.

Peters AD, Lively CM. (2000) Epistasis and the maintenance of sex. In Wolf JB, Brodie ED, Wade MJ (Eds.), *Epistasis and the Evolutionary Process*. (pp. 99-112). Oxford University Press.

Pineda-Krch M, Redfield RJ (2005). Persistence and loss of meiotic recombination hotspots. *Genetics.* **169:** 2319-33.

Renaut S, Replansky T, Heppleston A, Bell, G. (2006). The ecology and genetics of fitness in *Chlamydomonas*. XIII. Fitness of long-term sexual and asexual populations in benign environments. *Evolution*. **60:** 2272-2279.

Ricci C. (2001). Dormancy patterns in rotifers. *Hydrobiologia.* **446:** 1-11.

Rivero A, Balloux F, West SA. (2003). Testing for epistasis between deleterious mutations in a parasitoid wasp. *Evolution*. **57:** 1698-1703.

Robbins RB. (1918) Some applications of mathematics to breeding problems III. *Genetics.* **3:** 375-389.

Rose MR. (1983) The contagion mechanism for the origin of sex. *Journal of Theoretical Biology.* **101:** 137-46.

Salathé M, Kouyos RD, Bonhoeffer S. (2008) The state of affairs in the kingdom of the Red Queen. *Trends in Ecology & Evolution*. **23:** 439-445.

Scheu S, Drossel B. (2007). Sexual reproduction prevails in a world of structured resources in short supply. *Proceedings of the Royal Society B Biological Sciences*. **274:** 1225-1231.

Sturtevant AH. (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology.* **14:** 43-59.

Sturtevant AH, Bridges CB, Morgan TH. (1919) The spatial relations of genes. *Proceedings of the National Academy of Sciences*. **5:** 168–173.

Sutton WS. (1902) On the morphology of the chromosome group in *Brachystola Magna*. *The Biological Bulletin*. 24-39.

Szathmáry E, Kövér S. (1991) A theoretical test of the DNA repair hypothesis for the maintenance of sex in eukaryotes. *Genetical Research.* **58:** 157-65.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983). The double-strand-break repair model for recombination. *Cell.* **33:** 25-35.

Van Valen L. (1973) A new evolutionary law. *Evolutionary Theory*. **1:** 1-30.

VanLiere JM, Rosenberg NA. (2008) Mathematical properties of the r^2 measure of linkage disequilibrium. *Theoretical Population Biology.* **74:** 130-137.

Verhulst PF. (1838). Notice sur la loi que la population suit dans son accroissement. *Corr. Math. et Phys*. **10:** 113–121.

Vrijenhoek R, Dawley R, Cole C, Bogart J. (1989) A list of the known unisexual vertebrates. In *Evolution and Ecology of Unisexual Vertebrates*, ed. Dawley RM, Bogart JP, **4:** 19-23. New York State Museum.

Wang AD, Sharp NP, Spencer CC, Tedman-Aucoin K, Agrawal AF. (2009). Selection, epistasis, and parent-of-origin effects on deleterious mutations across environments in *Drosophila melanogaster*. *The American Naturalist.* **174:** 863-874.

Wallace B. (1975). Hard and soft selection revisited. *Evolution*. **29:** 465-473.

Watson RA, Wakeley J. (2005). Multidimensional epistasis and the advantage of sex. *Proceedings of the 2005 Congress on Evolutionary Computation*. 2792-2799.

Weismann A. (1891) The significance of sexual reproduction in the theory of natural selection. In *Essays upon heredity and kindred biological problems*. 163-255.

Weinberg W. (1908). ber den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg.* **64:** 368-382.

Welch MD, Meselson M. (2000) Evidence for the evolution of *Bdelloid rotifers* without sexual reproduction or genetic exchange. *Science*. **288:** 1211-1215.

West SA, Lively CM, Read AF. (1999). A pluralist approach to sex and recombination. *Journal of Evolutionary Biology*. **12:** 1003-1012.

Wilkins AS, Holliday R. (2009) The evolution of meiosis from mitosis. *Genetics.* **181:** 3-12.

Wilson EO. (1992) *The diversity of life*. Harvard University Press.

Wilson SR. (1978) A note on assortative mating, linkage and genotypic frequencies. *Annals of Human Genetics.* **42:** 129-130.

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GAT, *et al*. (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. *Science.* **308:** 107-11.

Wright S. (1920) The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences*. **6:** 320-32.

Wright S. (1931) Evolution in Mendelian populations. *Genetics*. **16:** 97-159.

Yi T, Lessard. (2000) Fundamental theorem of natural selection and frequency-dependent selection: analysis of the matrix game diploid model. *Journal of Theoretical Biology*. **206:** 17-25

You L, Yin J. (2002). Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage T7. *Genetics.* **160:** 1273-1281.

Zapata C. (2000) The D' measure of overall gametic disequilibrium between pairs of multiallelic loci. *Evolution.* **54:** 1809-1812.

APPENDICES

Appendix 1

Linkage Disequilibrium

Linkage disequilibrium (D) is defined as the extent to which the frequency of a genotype differs from its expected frequency had the alleles at each locus been combined at random (Robbins, 1918). It describes a situation in which some combinations of alleles occur together more frequently in a population than would be expected had the alleles randomly formed into genotypes. Numerically, it is the difference between the observed and expected frequency of a genotype.

We will assume the same two-locus, two-allele model as we did in chapter 3. Assuming that we have a genotype AB whose frequency is $Freq(AB)$, and where p_1 is the frequency of the A allele in the population, and p_2 is the frequency if the B allele, then we can write the value of linkage disequilibrium as

$$D = Freq(AB) - p_1 p_2 \tag{A1.1}$$

If we already know the value of D, as well as the frequency of each allele, then we can determine the frequency of every genotype in the population. For instance, we can deduce the frequency of AB genotype by rearranging equation (A1.1), that is

$$Freq(AB) = p_1 p_2 + D \tag{A1.2}$$

We can then express the frequency of Ab in terms of D and its allele frequencies as follows:

$$Freq(Ab) = p_1 - Freq(AB)$$

$$= p_1 - (p_1 p_2 + D)$$

$$= p_1(1 - p_2) - D$$

The same reasoning can be applied to the other two genotypes to determine their values. Therefore, the frequency of each genotype, expressed in terms of its allelic frequencies and the value of D, is as follows:

$$\begin{aligned} Freq(ab) &= (1 - p_1)(1 - p_2) + D \\ Freq(Ab) &= p_1(1 - p_2) - D \\ Freq(aB) &= (1 - p_1)p_2 - D \\ Freq(AB) &= p_1 p_2 + D \end{aligned} \tag{A1.3}$$

The value of D can also be expressed using only genotype frequencies as variables. We have

$$D = Freq(ab)Freq(AB) - Freq(Ab)Freq(aB) \tag{A1.4}$$

This statement can be proven by substituting each genotype frequency with its equivalent from equations (6.3):

$$\begin{aligned} &Freq(ab)Freq(AB) - Freq(Ab)Freq(aB) \\ &\quad = ((1 - p_1)(1 - p_2) + D)(p_1 p_2 + D) - (p_1(1 - p_2) - D)\big((1 - p_1)p_2 - D\big) \\ &\quad = p_1 p_2(1 - p_1)(1 - p_2) + p_1 p_2 D + (1 - p_1)(1 - p_2)D + D^2 \\ &\qquad -p_1 p_2(1 - p_1)(1 - p_2) + p_1(1 - p_2)D + (1 - p_1)p_2 - D^2 \\ &\quad = (p_1 p_2 + (1 - p_1)(1 - p_2) + p_1(1 - p_2) + (1 - p_1)p_2)D \\ &\quad = D \end{aligned}$$

This alternate equation for the value of D is useful when we look at the effects of recombination on the frequencies of each genotype.

Appendix 2

Estimating the decrease in Linkage Disequilibrium after one round of random mating

In this section, we will explore how the frequency of each genotype changes as a result of random mating. We start by assuming that there are no other forces acting to change the frequencies of each genotype: no selection, no mutations, and no genetic drift. We will further assume that mating is random across the entire population.

We start by determining the frequency of each mating combination. Under random mating, this is equivalent to fully writing out the expression

$$\begin{aligned}&\big(Freq(ab) + Freq(Ab) + Freq(aB) + Freq(AB)\big)\\ &\text{x}\ \big(Freq(ab) + Freq(Ab) + Freq(aB) + Freq(AB)\big)\end{aligned} \tag{A2.5}$$

Table 1 exhaustively displays a matrix of every possible mate pairing, and the frequency of that pairing.

Table 1 Complete matrix of the frequencies of every mating pairing.

	Frequency of mating type			
	ab	Ab	aB	AB
ab	$Freq(ab)^2$	$Freq(ab)Freq(Ab)$	$Freq(ab)Freq(aB)$	$Freq(ab)Freq(AB)$
Ab	$Freq(Ab)Freq(ab)$	$Freq(Ab)^2$	$Freq(Ab)Freq(aB)$	$Freq(Ab)Freq(AB)$
aB	$Freq(aB)Freq(ab)$	$Freq(aB)Freq(Ab)$	$Freq(aB)^2$	$Freq(aB)Freq(AB)$
AB	$Freq(AB)Freq(ab)$	$Freq(AB)Freq(Ab)$	$Freq(AB)Freq(aB)$	$Freq(AB)^2$

It should be noted that the off-diagonal entries in this table are equivalent. For instance, the entry in the first row and the second column is the same as the entry in second row and the first column. The off-diagonal entries can be added together to get the total frequency of every mating type. There are 10 different mating types in total.

Table 1: The frequency of each mating type when mating is random.

Mating	Frequency of mating (parents)
ab x ab	$Freq(ab)^2$
ab x Ab	$2Freq(ab)Freq(Ab)$
ab x aB	$2Freq(ab)Freq(aB)$
ab x AB	$2Freq(ab)Freq(AB)$
Ab x Ab	$Freq(Ab)^2$
Ab x aB	$2Freq(Ab)Freq(aB)$
Ab x AB	$2Freq(Ab)Freq(AB)$
aB x aB	$Freq(aB)^2$
aB x AB	$2Freq(aB)Freq(AB)$
AB x AB	$Freq(AB)^2$

Next, we will determine the frequency of the progeny produced in each case. We can classify each mating type in the following way, where either

1. both loci are homozygous (for example, AB x AB),

2. one locus is heterozygous (ex. AB x Ab), or

3. both loci are heterozygous (ex. AB x ab).

Both parents must have the same genotype in the first case where both loci are homozygous. All of their offspring will also have this genotype. In the second case, the parents differ in their genotypes at a single locus. With this kind of mating type, recombination has an effect, but its effect is to change the genotype into that of the other parent. The end result is that recombination does not change the genotypic frequencies of their progeny: half will have the genotype of one parent, and half will have the genotype of the other parent. It is only in the third case, where both loci are heterozygous, that recombination will have any effect of the genotypic frequencies of their progeny. In our example mating type, where AB mates with ab, we can see that the gametic genotypes do not change in the absence of a recombination event. When there is a recombination event, the genotypes of the gametes change to Ab and aB. Half of the time there will be a recombination event, and half of the time there won't be when recombination is free. All four genotypes are then equally likely, each one having an expected frequency of ¼. Table 2 summarizes the expected frequency of the progeny produced from each mating type.

Table 2: The expected genotype frequencies among the progeny from each mating type with free recombination.

Mating	Frequency of mating (parents)	Frequency of progeny			
		ab	Ab	aB	AB
ab x ab	$Freq(ab)^2$	1	0	0	0
ab x Ab	$2Freq(ab)Freq(Ab)$	$1/2$	$1/2$	0	0
ab x aB	$2Freq(ab)Freq(aB)$	$1/2$	0	$1/2$	0
ab x AB	$2Freq(ab)Freq(AB)$	$1/4$	$1/4$	$1/4$	$1/4$
Ab x Ab	$Freq(Ab)^2$	0	1	0	0
Ab x aB	$2Freq(Ab)Freq(aB)$	$1/4$	$1/4$	$1/4$	$1/4$
Ab x AB	$2Freq(Ab)Freq(AB)$	0	$1/2$	0	$1/2$
aB x aB	$Freq(aB)^2$	0	0	1	0
aB x AB	$2Freq(aB)Freq(AB)$	0	0	$1/2$	$1/2$
AB x AB	$Freq(AB)^2$	0	0	0	1
Totals (next generation)		$Freq(ab)'$	$Freq(Ab)'$	$Freq(aB)'$	$Freq(AB)'$

We can determine the expected genotype frequencies of the progeny by summing over the columns in this table. For example:

$$Freq(AB)' = \frac{2}{4}Freq(ab)Freq(AB) + \frac{2}{4}Freq(Ab)Freq(aB) + \frac{2}{2}Freq(Ab)Freq(AB) + \frac{2}{2}Freq(aB)Freq(AB) + Freq(AB)^2$$
$$= Freq(AB) - ½D$$

where $D = Freq(ab)Freq(AB) - Freq(Ab)Freq(aB)$

When we perform this summation and simplify for all four genotypes, we get the following as the frequency of each genotype after free recombination:

$$\begin{aligned} Freq(ab)' &= Freq(ab) - ½D \\ Freq(Ab)' &= Freq(Ab) + ½D \\ Freq(aB)' &= Freq(aB) + ½D \\ Freq(AB)' &= Freq(AB) - ½D \end{aligned} \tag{A2.4}$$

Previously, we showed that $Freq(AB) = p_1p_2 + D$. When we substitute the right hand side into our equation for the value of $Freq(AB)'$, we get that

$$\begin{aligned} Freq(AB)' &= (p_1p_2 + D) - ½D \\ &= p_1p_2 + ½D \end{aligned}$$

To summarize, the frequency of the AB genotype was $p_1p_2 + D$ before recombining, and changed to $p_1p_2 + ½D$ after recombining. Therefore, the value of D decreases by a factor of 0.5 in every generation with free recombination.

Appendix 3

Expressing Z in terms of D

We can show the relationship between D and Z as follows.

$$D = \text{Freq(ab)}Freq(AB) - \text{Freq(Ab)}Freq(aB) \tag{A3.1}$$

Dividing across by $\text{Freq(Ab)}Freq(aB)$, we get

$$\frac{D}{Freq(Ab)Freq(aB)} = \frac{Freq(ab)Freq(AB) - Freq(Ab)Freq(aB)}{Freq(Ab)Freq(aB)}$$
$$= \frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)} - 1$$

Then rearranging,

$$\frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)} = \frac{D}{Freq(Ab)Freq(aB)} + 1$$

And taking the log of both sides,

$$\ln\frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)} = \ln\left(\frac{D}{Freq(Ab)Freq(aB)} + 1\right)$$

we find that

$$Z = \ln\left(\frac{D}{Freq(Ab)Freq(aB)} + 1\right) \tag{A3.2}$$

For example, when $D = 0.2$, and $p_1 = p_2 = 0.5$

$$Z = \ln\left(\frac{D}{Freq(Ab)Freq(aB)} + 1\right)$$
$$= \ln\left(\frac{0.2}{0.0025} + 1\right)$$
$$= \ln 81$$
$$\approx 4.39$$

Expressing D in terms of Z

We can also express the value of Z in terms of the value of D in the following way:

$$Z = \ln\left(\frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)}\right) \tag{A3.3}$$

therefore

$$e^Z = \frac{Freq(ab)Freq(AB)}{Freq(Ab)Freq(aB)}$$

Multiplying across by $req(Ab)Freq(aB)$, yields

$$\text{Freq(Ab)}Freq(aB)e^Z = \text{Freq(ab)}Freq(AB)$$

subtracting $\text{Freq(Ab)}Freq(aB)$ from both sides, gives

$$Freq(Ab)Freq(aB)e^Z - \text{Freq(Ab)}Freq(aB) = Freq(ab)Freq(AB) - \text{Freq(Ab)}Freq(aB)$$
$$Freq(Ab)Freq(aB)(e^Z - 1) = Freq(ab)Freq(AB) - \text{Freq(Ab)}Freq(aB)$$

Therefore,

$$Freq(Ab)Freq(aB)(e^Z - 1) = D$$

or

$$D = Freq(Ab)Freq(aB)(e^Z - 1) \tag{A3.4}$$

Figures A3.1 and A3.2 graphically show the relationship between *D* and *Z*. It should be noted that the value of *D* has a maximum range of -0.25 to 0.25, but it is dependent the frequencies of the *A* and *B* alleles. The value of *Z*, on the other hand, always has a range from -∞ to +∞. *D* and *Z* always have the same sign, and both are equal to zero when the population is in linkage equilibrium (see Figure A3.1). Furthermore, there is a linear relationship between the two statistics when the population is close to linkage equilibrium (see Figure A3.2).

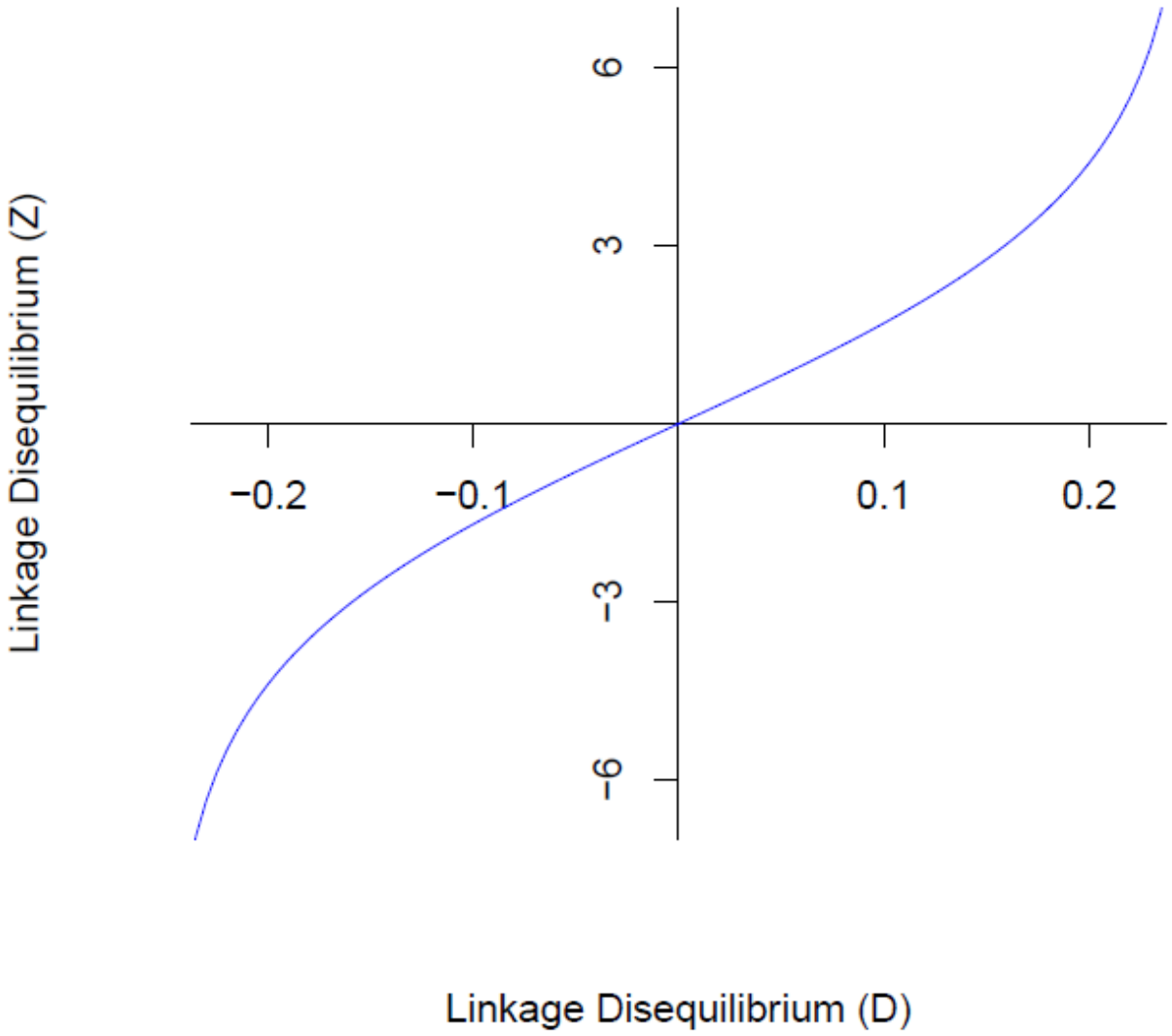


Figure A3.1: The relationship between the value of *D*, and the value of *Z*.

The value of *Z* is shown on the vertical axis, and the value of *D* is shown on the horizontal axis. In this case, the frequencies of all four alleles are equal to 0.5. Note that with these allele frequencies, the range of *D* is between -0.25 and 0.25, and that the range of *Z* is between -∞ and +∞.

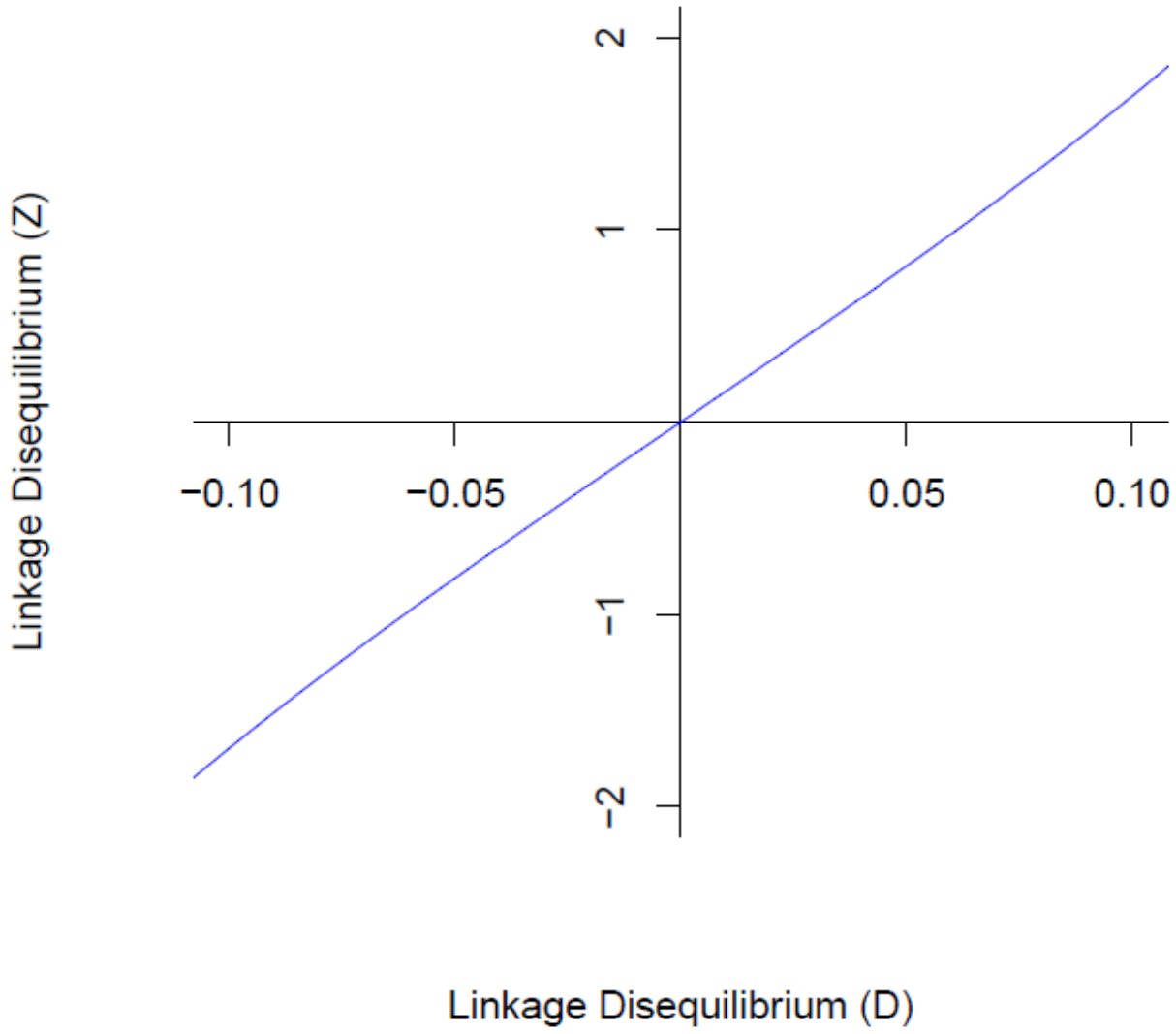


Figure A3.2: The relationship between the value of *D*, and the value of *Z* when the population is close to linkage equilibrium.

For a description, see Figure 3.14. The values of *D* and *Z* are linearly related when the population is close to linkage equilibrium.

Appendix 4

Derivation of fitness values when three loci are under selection

In Chapter 2, we developed a model where the fitness of a genotype can be calculated as the summation of the frequency that this genotype will compete against another genotype, multiplied by the probability that it will win that competition, normalized by the cull rate. Therefore, the only difference between how fitness is calculated when three loci are under selection instead of two loci, is that now there are 8 genotypes instead of 4.

$$w_i = \frac{1}{1-c}\left(\sum_{j=1}^{8} \frac{V_i}{V_i + V_j} Freq(j)\right) \tag{A4.1}$$

i is the genotype of interest, V_i is the phenotypic value of this genotype, V_j is the phenotypic value of the competing genotype (itself included), $Freq(j)$ is the frequency of some genotype, and the c is the cull rate.

Derivation of 3 loci epistasis for fitness

The equation to calculate the value of pairwise epistasis for fitness can be expressed as a deviation from multiplicativity, which is linear on a log scale:

$$\varepsilon_{AB} = ln\, w_{ab} + ln\, w_{AB} - ln\, w_{Ab} - ln\, w_{aB} \tag{A4.2}$$

In order to derive an equation for epistasis at 3 loci, we will begin with the equation above, and extend the model as explained by Hansen and Wagner (2001). We can re-arrange equation (A4.2) to express the fitness of the *AB* genotype as the fitness of the *ab* genotype, plus

the fitness gained by substituting an *a* allele for an *A* allele, plus the fitness gained by substituting a *b* allele for a *B* allele, plus the deviation from linearity known as epistasis.

$$ln\, w_{AB} = ln\, w_{ab} + (ln\, w_{Ab} - ln\, w_{ab}) + (ln\, w_{aB} - ln\, w_{ab}) + \varepsilon_{AB}$$

The equation above simplifies to

$$ln\, w_{AB} = ln\, w_{ab} + ln\frac{w_{Ab}}{w_{ab}} + ln\frac{w_{aB}}{w_{ab}} + \varepsilon_{AB} \quad \text{(A4.3)}$$

This framework can be extended to one where 3 loci are under selection as follows:

$$ln\, w_{ABC} = ln\, w_{abc} + ln\frac{w_{Abc}}{w_{abc}} + ln\frac{w_{aBc}}{w_{abc}} + ln\frac{w_{abC}}{w_{abc}} + \varepsilon_{AB} + \varepsilon_{AC} + \varepsilon_{BC} + \varepsilon_{ABC} \quad \text{(A4.4)}$$

Therefore, the value of epistasis at 3 loci can be calculated with the following equation:

$$\varepsilon_{ABC} = ln\, w_{ABC} - \left(ln\, w_{abc} + ln\frac{w_{Abc}}{w_{abc}} + ln\frac{w_{aBc}}{w_{abc}} + ln\frac{w_{abC}}{w_{abc}} + \varepsilon_{AB} + \varepsilon_{AC} + \varepsilon_{BC}\right) \quad \text{(A4.5)}$$