The Hill-Robertson effect, quasi-truncation selection, and genetic recombination

Qi Zhang

A Thesis

in

The Department

of

Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Biology at

Concordia University

Montreal, Quebec, Canada

May 2007

Your file Votre référence
ISBN: 978-0-494-34471-2
Our file Notre référence
ISBN: 978-0-494-34471-2

Abstract

The Hill-Robertson effect, quasi-truncation selection, and genetic recombination

Qi Zhang

The Hill-Robertson effect and epistasis are often proposed to explain the prevalence of sexual reproduction and recombination. The Hill-Robertson effect occurs in small populations, where a combination of genetic drift and weak natural selection result in negative linkage disequilibrium. In a large population, however, negative epistasis is the major source of negative linkage disequilibrium. The epistasis can be calculated with two models: the additive model and multiplicative model. In populations of any size, recombination reduces the negative linkage disequilibrium, thus enhancing the population's response to selection.

In our research, we examine negative epistasis under quasi-truncation selection. We also study the amount of negative linkage disequilibrium caused by both the Hill-Robertson effect and the negative epistasis under r selection, constant population size, and quasi-truncation selection. In addition, we examine the effect of epistasis calculated with either the additive model or multiplicative model on linkage disequilibrium. We find that non-epistatic selection calculated with the additive model can still generate negative linkage disequilibrium. In large populations, the negative linkage disequilibrium is noteworthily associated with negative epistasis. In

addition, recombination does not always speed fixation even if the average linkage disequilibrium is weak negative.

Acknowledgements

I wish thank to all those who helped me. Without their kind help, I cannot successfully complete my project and thesis. First and foremost, I thank my supervisor Donal Hickey. He started the project and guides me into the research field. Thanks for his encouragement, continued advice, and tolerance. I especially admire his ability to take advantage of some metaphors to illustrate complicated evolutionary theories. I also appreciate other professors' help; I have learned much from their classes and discussion with them.

Secondly, I appreciate the kind help from the members of my lab and friends in biology department. Xiangjia Min gives me suggestion on both my project and daily life. I also enjoy the time spending with Amir, Sol, Wucheng, Sean, and Jessie,. Their lunch stories and jokes make me feel the research work in our lab is not boring. I can still remember the slang, such as "affectionate", learned from Sol.

Last, but not least, I would like to thank my parents' edification and my wife's selfless support and love.

Table of Contents

Abbreviation

D	to indicate the magnitude of linkage disequilibrium
D'	D-prime
e	static epistasis
E	dynamic epistasis
LD	linkage disequilibrium
HRE	Hill-Robertson effect
r	reproductive ability
v	phenotypic value

List of Figures

Introduction

1 Concepts

1.1 Haplotype and phenotype

1.1.1 Haplotype

Haplotype is a set of alleles on a chromosome. Our research studies four two-locus haplotypes ab, Ab, aB and AB. The frequencies of the above four haplotypes are represented by x_1, x_2, x_3, and x_4. A and B are beneficial mutations at the locus 1 and locus 2 respectively. Hence, Ab and aB are intermediately advantageous haplotypes, and AB is the most advantageous haplotype.

1.1.2 Phenotypic value

For any haplotype, we specify an abstract value to quantify phenotype. The abstract value is called phenotypic value. It can be body size, weight, height, and other quantified phenotype(HALLIBURTON 2004). In our study, the intermediate haplotype Ab and aB have the same phenotypic values.

1.2 Fitness

Fitness is the ability of an individual to produce offspring in a particular environment. Absolute fitness (w) is the absolute number of surviving offspring for each genotype

It is calculated as the product of the survival proportion times the average fecundity. To calculate relative fitness (W), the fitness of one haplotype is normalized as W = 1, and the fitnesses of other haplotypes are measured with respect to that haplotype. For convenience, relative fitness W is often written as:

$$W = (1+s)$$

where "s" is called the selection coefficient, and s measures the intensity of natural selection. In our study, w_1, w_2, w_3, and w_4 represent the absolute fitness of ab, Ab, aB, and AB. Meanwhile, W_1, W_2, W_3, and W_4 represent the relative fitness of ab, Ab, aB, and AB.

There are two models to calculate fitness: the additive fitness model and the multiplicative fitness model. In the additive fitness model, the fitness of a two-locus haplotype can be summed up; therefore, if two mutated alleles at different loci appear at the same haplotype, the new fitness is the sum of the two single mutated alleles. On the other hand, in a multiplicative fitness model, the fitness of two loci can be multiplied.

A summary of the two models is as follows:

Additive model

relative fitness of haplotype AB $W_4=1+s_A+s_B$

Since we assume $s_A=s_B=s$ in our study, $W_4=1+2s$

Multiplicative model

relative fitness of haplotype AB $W_4=1*(1+s)*(1+s)=(1+s)^2=1+2s+s^2$

There is only one item difference between the two models, s^2. Considering that new mutations usually bring about small increase or decrease of fitness, i.e. s is often small in actual situation, s^2 is second order small (if s is O, s^2 is O^2). The smaller s^2 item hence can be ignored. So the two models may give the similar results. Maybe sometimes we can regard the same natural world data as an evidence for both additive and multiplicative models.

Moreover, if consider the Log transformation between the additive model and multiplicative model, we see more connection between them.

Apply log to both sides of multiplicative model

Log (relative fitness of AB)=Log(1*(1+s)*(1+s))

Log (relative fitness of AB)=log (1+s)+ log(1+s)

In the transformed space, multiplicative model is changed into additive model

However, the differences between the two models are still noteworthy. We further demonstrated the difference in the Result part 1.1.

1.3 Selection

Some simple selections have constant selection coefficients while complicated selections have non-constant selection coefficients. For the constant selection

coefficient, the environmental effects are often ignored. Absolute fitness of a genotype equals genotype's reproductive ability

1.3.1 r Selection

In r selection, the selection coefficient is constant and the absolute fitness of a genotype equals the genotype's reproductive ability (r) which is constant. In our study, r_1, r_2, r_3, and r_4 represent the reproductive ability of haplotype ab, Ab, aB, and AB respectively. Similarly, we assume that each allele has a constant reproductive rate although its selection coefficient may be variable during selection. In r selection, populations can exponentially grow without limiting the population size. The premise for this idealistic selection is that the carrying capacity of the environment is infinite. However, in the actual situations, organisms often face competition because of the limited available resources.

1.3.2 Truncation selection

Contrary to the r selection, truncation selection describes the situation in which individuals above a threshold phenotypic value are selected as parents for the next generation. Truncation selection is the most efficient form of selection. Truncation is also defined as a process to eliminate individuals with more than a certain number of mutations. (CROW 1997; CROW and KIMURA 1979; KIMURA and CROW 1978).

1.3.3 Quasi-truncation selection

Quasi-truncation selection has been introduced to solve the unreality of the truncation selection model. Instead of abruptly truncating the top ranks in a population, the eliminating strength of quasi-truncation selection gradually increases over a range of mutations (CROW 1997; CROW and KIMURA 1979; KIMURA and CROW 1978).

In our study, quasi-truncation selection has been introduced to solve the unreality of the truncation selection model. In our study, quasi-truncation selection has a constant upper limit of population size, and selects specified phenotypes; therefore, the survival of each haplotype is determined by both the carrying capacity and phenotypic value. Then, the fitness is not only determined by the reproductive ability, but also by the extent of competition among haplotypes in a population and by the environment.

1.3.4 Realized fitness

In truncation selection and quasi-truncation selection, the advantages of the favored phenotypes can vary depending on their frequencies in the populations. To evaluate the fitness in this situation, realized fitness is introduced. The actual advantage at any given time is called the realized fitness.

Phenotypic value is the quantified phenotypes. According to the phenotypic values, the individuals are selected in quasi-truncation selection. However, realized fitness quantifies the result of the above process---the actual survived offspring number of each haplotype in quasi-truncation selection.

Realized fitness is represented by w;

w = the number of one kind of haplotype in the next generation / number of their parents

1.4 Epistasis

Epistasis, interaction between genes, is a deviation from the additivity or multiplicativity in the effect of alleles at different loci. There are two calculating methods for epistasis: epistasis on the additive scale and epistasis on the multiplicative scale. They are derived from the additive fitness and multiplicative fitness respectively. (EWENS 1969; PUNIYANI *et al.* 2004)

If we adopt the additive model, when the fitness value is greater than the sum of fitness at each locus, the phenomenon is called positive epistasis. On the other hand, when the fitness value is less than the sum of fitness at each locus, the phenomenon is called negative epistasis (PHILIPS *et al.* 2000). Similarly, we can define the positive and negative epistasis on the multiplicative scale.

Static epistasis:

In our study, the epistasis calculated with reproductive ability (r) is referred to as static epistasis (e).

Dynamic epistasis

After the truncation selection, the epistasis is calculated with realized fitness using the additive or multiplicative scale. The epistasis calculated with realized fitness is called dynamic epistasis (E).

For the additive scale,

expected realized fitness of genotype AB = $w_1+(w_2-w_1)+(w_3-w_1) = w_2+w_3-w_1$

E (dynamic epistasis) = observed realized fitness-expected realized fitness

$= w_1+w_4-w_2-w_3$

e (static epistasis) = $r_4+r_1-r_2-r_3$ (1)

,where w is the realized fitness.

For the multiplicative scale,

expected realized fitness of genotype AB = $w_1*(w2/w1)*(w3/w1) = w_{2*}w_3 / w_1$

E (dynamic epistasis) = $w_4 - w_2w_3/w_1$

e (static epistasis) = $r_4- r_2r_3/r_1$ (2)

,where w is the realized fitness.

1.5 Linkage disequilibrium

Linkage disequilibrium is a statistical phenomenon in which haplotype frequencies in a population deviate from the values they would ordinarily have if the genes at each locus were combined randomly. The magnitude of LD is often indicated by D and D' (D-prime). There are two loci with two alleles at each locus. Alleles A and a at locus 1, and alleles B and b at locus 2.

D= freq(observed haplotype AB)-freq(A)*freq(B)

where A and B are the alleles at the two loci respectively.

For the biallelic two-locus model, D can also be calculated as:

D= freq(AB)*freq(ab)-freq(Ab)*freq(aB) =x1*x4-x2*x3 (3)

where x1, x2, x3, and x4 respectively represent the frequencies of the four haplotypes. For convenience of comparing linkage disequilibrium in different populations, D' is calculated as below,

D'=D/Dmax (4)

,where Dmax=|max(D)| if D>0

Dmax=|min(D)| if D<0

According to the sign of D (+ /-), the linkage disequilibrium can categorized into positive and negative LD. The positive linkage disequilibrium occurs when an observed haplotype frequency is greater than the expected haplotype frequency. Conversely, the negative linkage disequilibrium occurs when an observed haplotype frequency is smaller than the expected haplotype frequency. The linkage disequilibrium is affected by selection, genetic drift, mutation, and recombination. (FELSENSTEIN 1965; HEDRICK 1987; LEWONTIN 1964; LEWONTIN 1988; MUELLER 2004)

1.6 The Hill-Robertson effect

1.6.1 Genetic drift

Genetic drift is a stochastic process. It is the fluctuation of allele frequencies in a finite population over time. During the process, random events may not always keep the expected frequencies of advantageous genes. In addition, different sampled populations may have varied evolution trajectory. "This evolutionary mechanism, a change in a population's allele frequencies due to chance, is called genetic drift." (CAMPBELL and REECE 2001)

1.6.2 The Hill-Robertson effect

The genetic drift combining with selection may generate negative LD. The phenomenon is called the Hill-Robertson effect. The Hill-Robertson effect covers two features: in small population, the negative linkage disequilibrium (LD) can still appear because of smaller effective population size and selection effect. The LD reduces the effect of selection action upon the loci. (HILL and ROBERTSON 1966)

1.7 The costs and benefits of sexual reproduction

The organisms adopting sexual reproduction have to afford costs; at the same time, the organisms also acquire benefits from the sexual reproduction.

1.7.1 The costs of sexual reproduction

Sexual reproduction generates offspring slower than asexual reproduction because only one of the two parents can give birth to offspring in sexual reproduction and, meiosis is more complicated and energy consuming than mitosis. This is often referred

to as the two-fold cost of sex. In addition, sexual partners are hard to find when the population size is quite small. (MAYNARD SMITH 1978)

1.7.2 The benefits of sexual reproduction

However, the benefits of sexual reproduction and recombination have also been proposed by researchers.

Sexual reproduction and recombination can increase the variation of genetic combination. The varied genetic combination may adapt to different environmental situations and defend against the parasite's attack; the variation may also decrease intra-species competition since individuals with different genetic combination need varied resource to live; (HAMILTON *et al.* 1990; MULLER 1932)

Moreover, the sexual reproduction and recombination may speed up the appearance of new genetic combination. Without recombination, one organism needs two steps to acquire a two beneficial gene combination; however, two single beneficial mutations may be combined with recombination, hence an organism can quickly acquire the combination of two beneficial mutations. In addition, Sex and recombination assist to remove deleterious mutations if two deleterious genes are associated in a single individual. (KONDRASHOV 1988)

1.7.3 The quantification of sexual reproduction

The abstract "reproductive effort" which is the energy consumed during sexual reproduction might be referred to as a general unit of the costs and benefits of sexual reproduction (HIRSHFIELD and TINKLE 1975). However, since reproductive effort does not sufficiently consider the factor of mortality risk, it is not appropriate to quantifying the costs and benefits (DALY 1978; PIANKA and PARKER 1975).

In our study, we focus on how the negative LD is generated in sexual and asexual reproduction. Moreover, we quantify the costs and benefits of recombination through measuring the relative numbers of allelic copies passed to the next generation.

2 Literature review

Sexual reproduction and recombination result in several biological costs. For example, in a sexual population, only female individuals can give birth to the next generation. If we also consider meiosis and fertilization, we see that the sexual reproduction process is more complicated than the asexual reproduction. In addition, to attract mates, organisms may develop some physical characteristics which often jeopardize their own survival. These and other factors (such as energetic costs) mean that sexual reproduction is, in general, less efficient and more costly than asexual reproduction. Despite the obvious costs of sexual reproduction, however, it is still commonly observed in nature.

The recognized benefit of sexual reproduction is genetic recombination, although it has been difficult to define and quantify this benefit. One hypothesis to explain the

paradox of sexual reproduction and recombination is that an uncertain environment favors organisms that adopt sexual reproduction and recombination.(WILLIAMS 1975). Subsequent work has shown that such a benefit depends on a negative autocorrelation between successive environmental conditions. Antagonistic coevolution between parasites and their hosts would result in this type of fluctuating selection (BARTON 1995) although it is difficult to see how this can be generalized to other ecological situations.

The hypothesis that there is a benefit for recombination in a changing environment leads to the question of whether recombination may then be unnecessary in a stable environment. The following model has been proposed to explain the advantage of recombination in a small population. In a small population, there exists almost no individuals that have the fittest possible genotype, combining most of the beneficial mutations at different genetic loci, or the chance is very small when these beneficial mutations first appear. Recombination joins beneficial alleles (or deleterious alleles) into one individual.(PHILIPS *et al.* 2000)

Another possible explanation for sexual reproduction and recombination is the existence of negative epistasis. (see definition in methods section below) (CORDELL 2002; EWENS 1969; PHILIPS *et al.* 2000). Negative epistasis describes the situation where the effects of selection at multiple loci are less than the sum or product of their effects at individual loci. This type of selection alters the genotype frequencies in the population, generating negative linkage disequilibrium. The advantage of genetic

recombination is that it breaks down this negative linkage disequilibrium and restores the statistical independence of allele frequencies at different loci. In this way, genetic recombination can result in the more efficient elimination of multiple deleterious mutations, thus solving the problem known as Muller's ratchet.(KONDRASHOV 1984; KONDRASHOV 1988; MULLER 1950)

It has been shown that, in small populations, negative linkage disequilibrium can be built up by the joint effect of random genetic drift and selection, and this effect is often referred to as the "Hill Robertson Effect" (HRE) (Hill and Robertson 1966). Even in a large population, where genetic drift is less important, this effect is still noteworthy if we consider a large number of genetic loci (Keightley and Otto 2006) or sub-divided populations (Martin et al. 2005). In this case also, recombination helps break down the accumulated negative linkage disequilibrium.

The simplest models in population genetics assume constant selection coefficients for alleles and assume that there are no epistatic interactions between alleles at different loci. While such models might be realistic for exponentially growing populations in an unlimited environment, they are not realistic for populations that are limited by finite resources (HARTL and CLARK 1989). In such cases, truncation selection may provide a more realistic description of the situation (CROW 1997; CROW and KIMURA 1979). Strict truncation selection – which is often used in artificial breeding programs - involves selecting a breeding population only from individuals whose phenotypic value is above a specified cut-off value. Quasi-truncation selection, or competitive

selection, also involves selecting individuals with the highest phenotypic value but, rather than specifying a given cut-off phenotypic value, the number of selected individuals is determined by the carrying capacity of the environment. Thus, the probability of being selected (the "realized fitness") depends not only on the phenotype of the individual itself but also on the phenotypes of other individuals in the population. While truncation selection is both efficient and biologically realistic, it results in the buildup of negative linkage disequilibrium between beneficial alleles at different loci. This would impede the progress of selection. In such cases, recombination can restore the population to linkage equilibrium (BARTON 1995; KOUYOS *et al.* 2006; OTTO 2001)

3 Research objectives

For my thesis research, I investigate the effects of competitive selection in natural populations. Rather than using real populations, I simulate populations of various sizes, and with two genetic loci under selection. The relationship between genotype and phenotype will be constant, but the fitness of these phenotypes will depend on the phenotypic composition of the population at a given generation. I focus particularly on the changing values of the realized fitnesses as the population evolves under selection. This will allow me to calculate the value of epistasis between loci and to investigate the associated buildup of linkage disequilibrium. The effects of recombination are evaluated by comparing the results of simulations that include recombination to those that do not.

The research objectives are listed as below:

1 Compare the additive model with the multiplicative model

2 Study the Hill-Robertson effect with multiplicative scale

3 Study the more realistic quasi-truncation rather than truncation selection

4 Measure dynamic epistasis with realized fitness

5 Compare the LDs caused by the Hill-Robertson effect and epistasis

6 Study the impact of recombination on fixation rate

Methods

To achieve the above research objectives, the following methods are applied.

1. Simulation environment

In our study, programs written in Java language are developed to implement Monte Carlo methods. The J2SE 1.5, chart generating software jFreeChart 0.9.2, and XML parser Dom4j-1.6.1 are applied. The programs run on an Ultra Sparc client-server system installed with a Solaris 9.0 operating system.

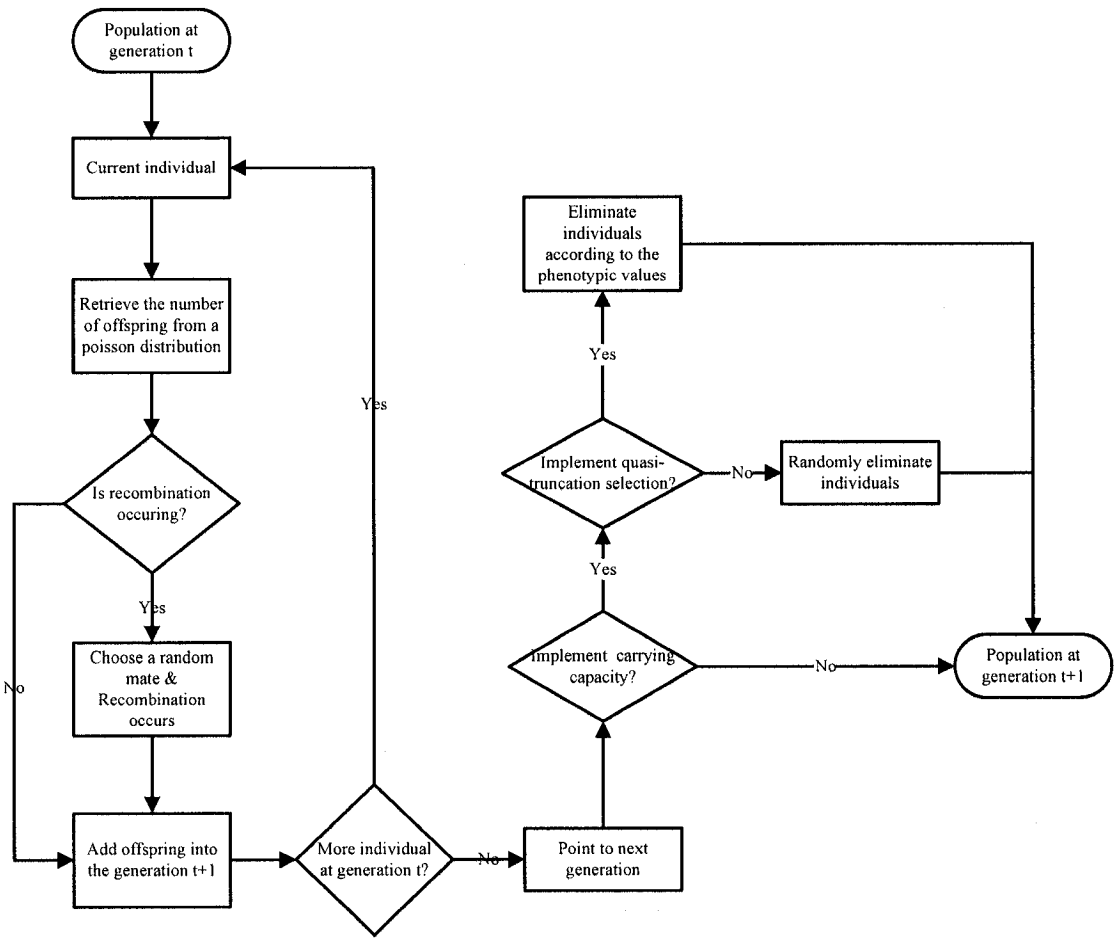
2. Initial conditions

For massive simulation, initial parameters are specified in a configuration file. In this study, four haplotypes ab, Ab, aB, and AB with specified initial fitness are chosen. The initial epistasis calculated with the reproductive ability (r) is zero. The phenotypic values (v) are specified and kept constant during evolution (v4>v2, v3; v2, v3>v1). In addition, the initial haplotype frequencies are chosen to keep the initial linkage equilibrium (The initial D is zero). We also assume that no new mutation

occurs, and the recombination rate is constant during evolution. Generation time, initial population size and carrying capacity are specified at the beginning.

3 Evolutionary simulations

3.1 Flow charts



3.2 Pseudocode

```
Input       haplotype frequency, initial population size, end population size,
            generation, reproductive ability, phenotypic value, recombination rate, Is
            truncation selection or not,
Initialize a population with the initial haplotype frequencies;
While (current generation<= initial total generation) do
  While (there exist individuals not processed yet) do
    Select an individual and calculate its offspring number according to Poisson
    distribution, the mean value of Lamda=reproductive ability;

    If (recombination is true) then
        Randomly select an individual for recombination;
    Endif
    Add offspring to next generation;
  Endwhile
  If (implement carrying capacity) then
        If (implement truncation selection) then
            Eliminate individuals according to phenotypic values;
        Else
            Randomly eliminate individuals to keep constant population size;
        Endif
  Endif
```

generation-1

Endwhile

3.3 Stochastic simulation

Stochastic process is simulated with random number generator. The generated random number is uniformly distributed between 0 and 1.

For the number of offspring, Poisson distribution is used. If X is the number of events occur randomly in a time interval and λ is the mean number of events in the time period,

$$X = \max\{n: U_1 \cdot U_2 \cdot \ldots \cdot U_n \geq exp(-\lambda)\}$$

U is the probability of corresponding event. Then, the following algorithm can be used:

```
Set a = exp(-λ), r = 1, X = -1
While(r > a) {Set r = r • U, X = X + 1)}
Return X
```

In our study, the X is the number of offspring. Hence, the number of offspring center on the value of reproductive ability (This is the mean value of λ).(J. H. AHRENS and U. DIETER 1973; KNUTH 1969)

3.4 Reproduction and recombination

In each generation, the number of offspring reproduced by a single individual is calculated by its reproductive ability (r) which is initially specified. The variance of

the offspring number, as mentioned in the last Section 3.3, is calculated with a value from a Poisson distribution.

If recombination occurs, the mating partner will be randomly selected from the population. Then the two parents cross over: reciprocally exchange alleles at loci; then, two new haplotypes are formed which different from their parents.

3.5 Population size

The initial population size is specified by input. In the following generations, the population size is the sum of offspring in the previous generation. Population size can fluctuate since the variance of offspring number (See 3.3). Once the population size reaches the initially specified carrying capacity, randomly selected individuals in the population will be eliminated to keep the population size below the carrying capacity. However, in quasi-truncation selection, the individuals with high phenotypic values will have high priorities to reproduce offspring and transfer their genes to the next generation even if the population size already reaches the carrying capacity.

4 Data analysis

For the small population, change of output is not sensitive to input parameters since the stochastic effect is overwhelmingly obvious. For large populations, the output change with the input parameters in a more predictable way since stochastic effect is less obvious in these large populations.

For the small populations (population size=100), the fluctuation of LD is considerable, we simulate 300 replicates to fully study the fluctuation. While the population size increase, the fluctuation decrease correspondingly. Hence, we decrease the replicates. When simulating the large population (population size=10,000), the each replicates approximately reach the same result. The table below uses the standard deviation of 10 random samples to demonstrate the fluctuation of LD in different size of populations under quasi-truncation selection.

	Population size=100	Population size=1000	Population size=10,000
Standard deviation	.0141	0.0031977	0.00016501

Considering the deterministic behavior is more obvious in larger populations, we choose the following simulation strategies:

For each set of initial parameters, we run 300 replicates for the small populations (population size=100), 100 replicates for medium populations (population size=1000), and 10 replicates for the large populations (population size=10,000).

The realized fitness, haplotype frequencies, allele frequencies, D value (linkage disequilibrium), and E value (dynamic epistasis), calculated in each generation, are stored in output files for further analysis. These output results are statistically analyzed in the situations of r selection, constant population size, and quasi-truncation selection.

Results

With a two-locus model, computer simulations for the large populations and small populations are shown in the figures below. In the simulation of the large population, we focus on the deterministic effect since the stochastic effect is not obvious. On the contrary, in the simulation of the small populations, we show the large stochastic effects

1 Simulation of stochastic process in large populations

1.1 Constant population size

From Figure 1, we can see that when the additive model is applied, the LD is negative although the initial epistasis and linkage disequilibrium is zero. Conversely, when multiplicative model is applied, the LD is around zero when the initial epistasis equals zero. With high reproductive rate (reproductive rate=3) of allele A and B, the negative LD calculated with the additive model is noteworthy.

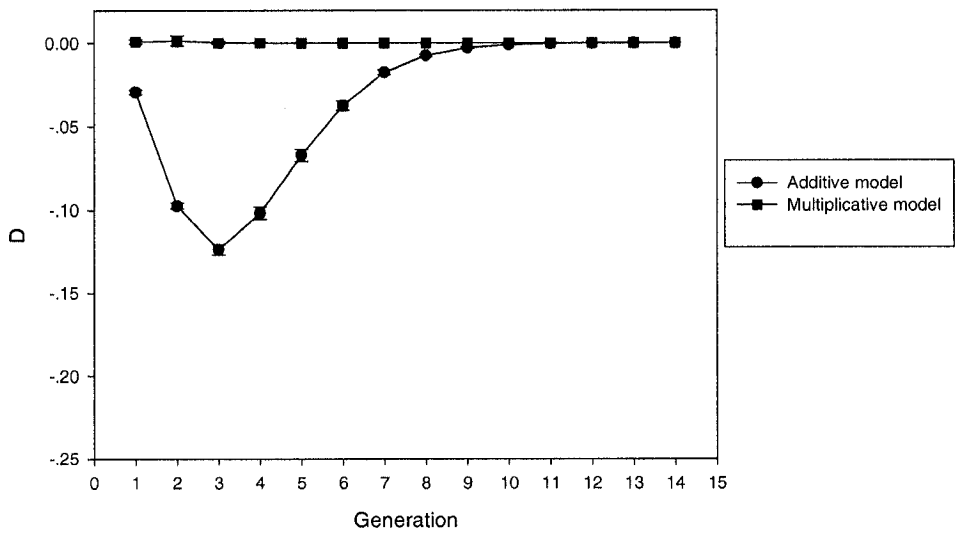
1.2 Quasi-truncation selection (large populations)

In our study, the individuals with higher phenotypic values are chosen to reproduce, hence keeping the constant population size. In large populations under quasi-truncation selection, even all the four haplotypes have the same reproductive ability, negative LD is still obvious (See Figures 2A). Since we use the same reproductive ability, this result clearly shows the effect of an environment on the negative LD. If an environment favors one specific phenotype, the individuals with the phenotype

have higher propriety to reproduce while other individuals without the phenotype have few chances to reproduce. Therefore, this leads to the difference of realized fitness between the individuals with higher and lower phenotypic values.

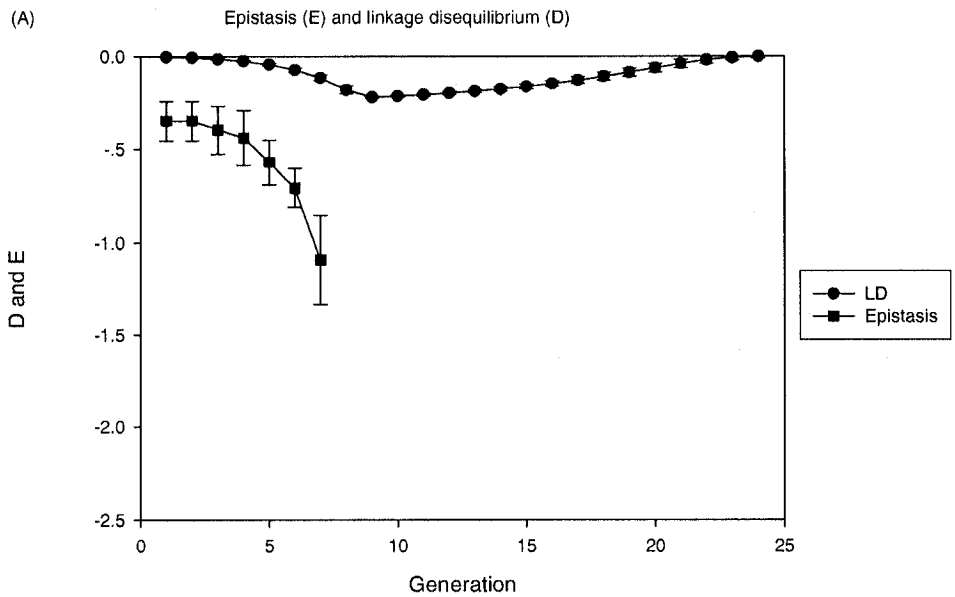
In Figure 2A, both epistasis and D are negative before the haplotype with the lowest phenotypic value (haplotype ab) vanishes (See Figure 2B).For the comparison of linkage disequilibrium between the different populations, we also show the D' and epistasis in Figure 3. From Figure 4, we obtain the correlation coefficient R=0.9983, and the coefficient of determination R Square=0.9966. Both of them demonstrate the linear relationship between the E and D.

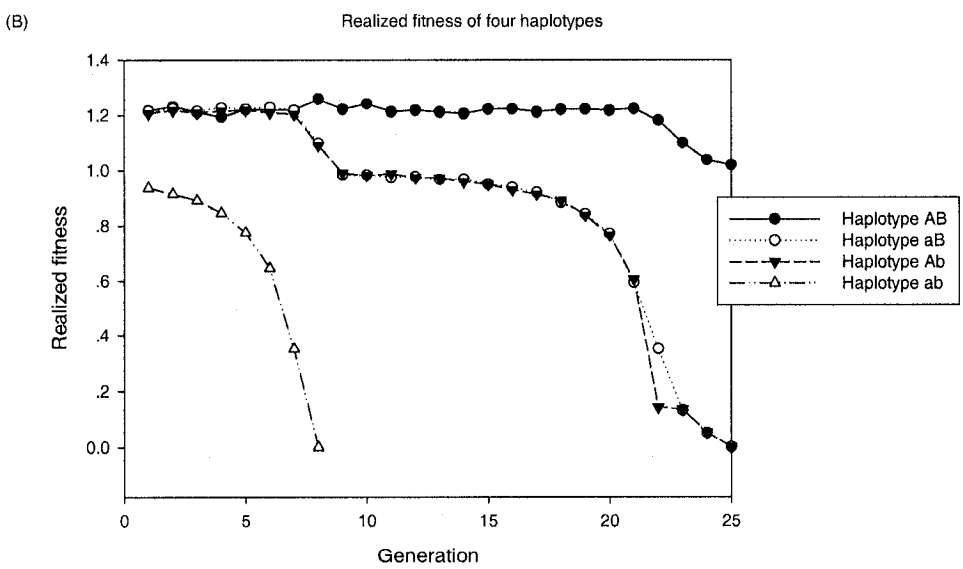
Figure 1 The impact of additive and. multiplicative model on the Hill-Robertson effect



Y axis represents the average D determined from 10 replicate runs. The series additive model plots D value under initial non-epistasic selection which is calculated with the additive scale, and series multiplicative model plots D value under initial non-epistasic selection which is calculated with the multiplicative scale. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01, respectively. The reproductive rate for both the beneficial allele A and allele B is 3. Carrying capacity is 10,000, and the initial population size is also 10,000.

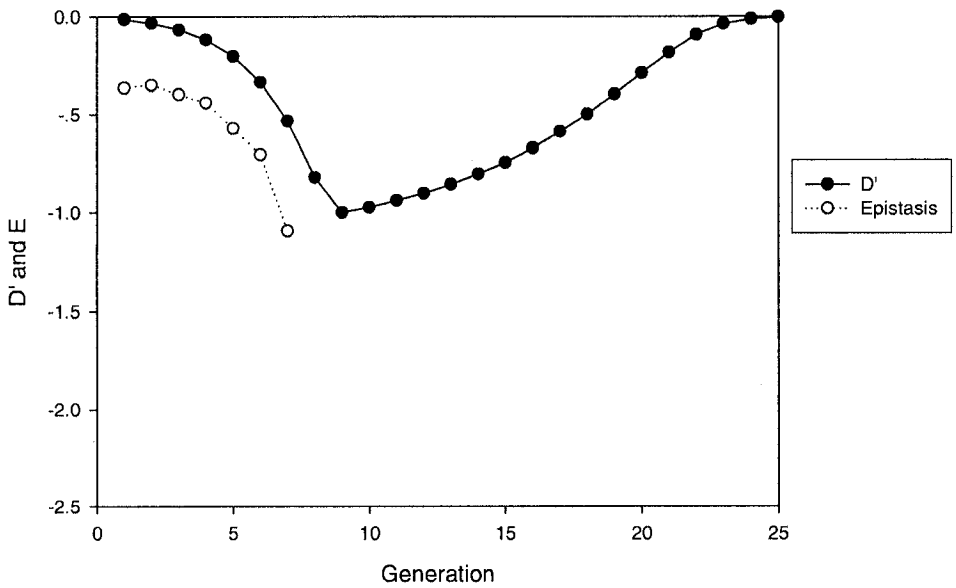
Figure 2 Epistasis and linkage disequilibrium under quasi-truncation selection





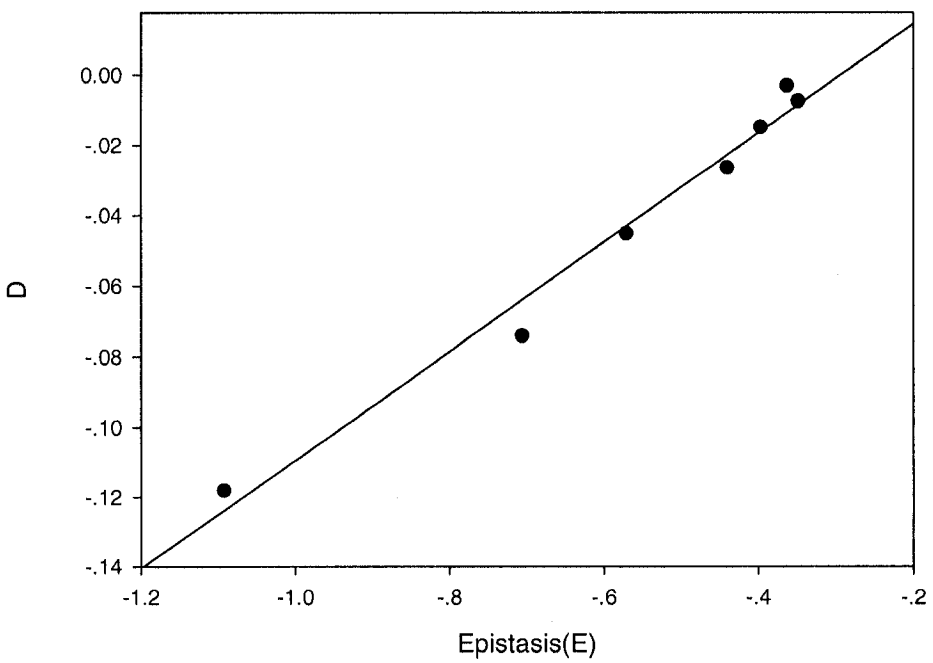
In Figure A, Y axis represents the average D and E determined from 10 replicate runs. In Figure B, Y axis represents the average realized fitness of four haplotypes AB, aB, Ab, and ab determined from the same 10 replicate runs. The series LD plots D values, and series Epistasis plots E values calculated with the multiplicative model. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01, respectively. The reproductive ability of the four haplotypes ab, aB, Ab, and AB is 1.05. Carrying capacity is 10,000, and the initial population size is also 10,000.

Figure 3 D' and Epistasis under quasi-truncation selection



To compare linkage disequilibrium of different allele frequencies, we also show the D' and E in this figure. This figure uses the same data as that in figure 2.

Figure 4 The regression between Epistasis and linkage disequilibrium under quasi-truncation selection



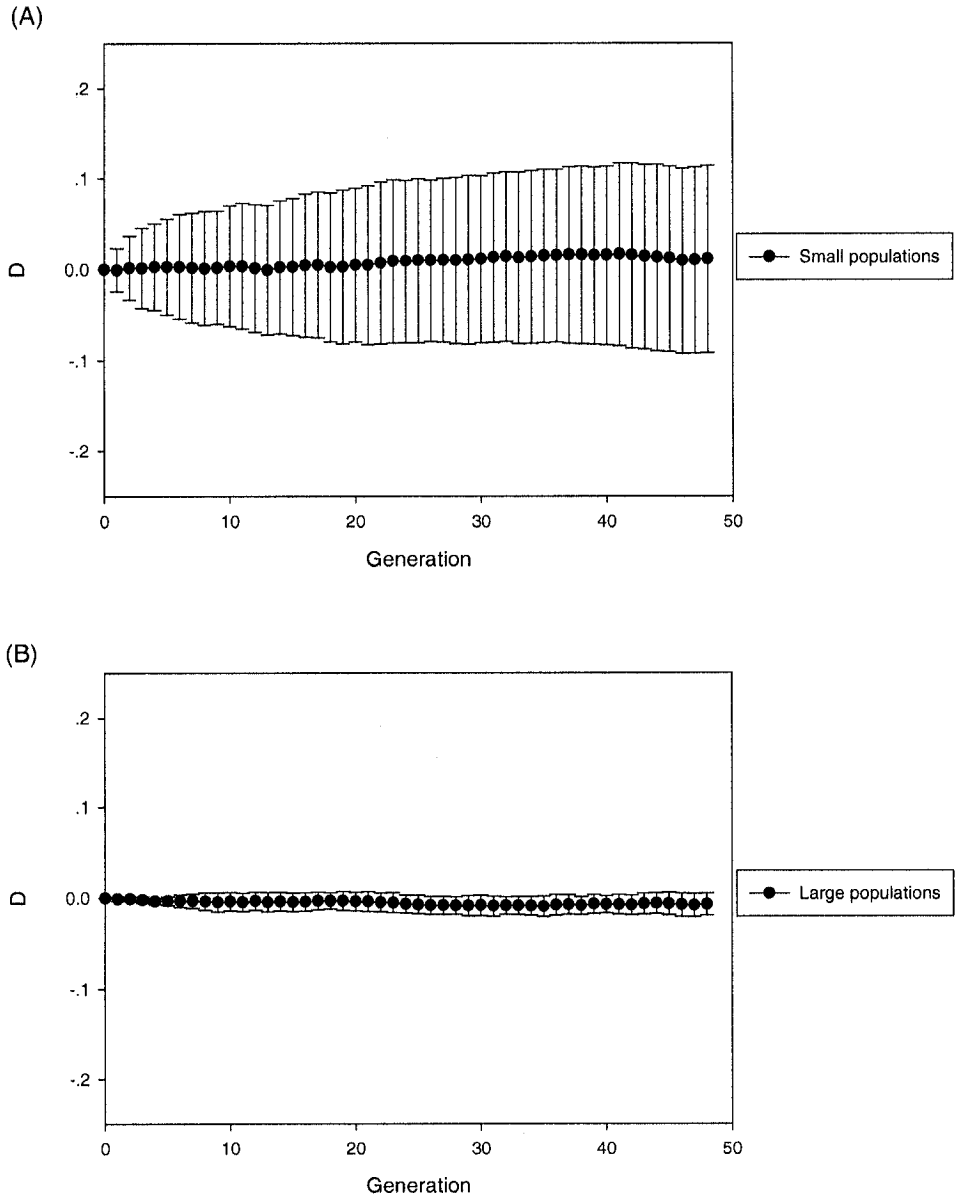
Epistasis E and linkage disequilibrium lie on X axis and Y axis respectively. This figure uses the same data as that in figure 2. The correlation coefficient R equals 0. 0.9983, and the coefficient of determination R Square equals 0.9966

2 Simulation of stochastic processes in small populations

2.1 Genetic drift (no selection)

The LD generated by genetic drift (All the haplotypes have equal reproductive ability) is presented in Figure 5. Without selection, the mean values of D in each generation determined from the replicate runs is around zero because genetic drift generates totally equal positive LD and negative LD (See Figure 5 A and 5B). To examine the variance of LD generated by genetic drift, we also show the standard deviation in Figure 5. The standard deviation decreases along with the increase of population size since the effect of genetic drift is reduced in large populations. This trend is clearly seen in Figure 5A and Figure 5B.

Figure 5 Genetic drift and linkage disequilibrium (without selection)



The frequencies of the four initial haplotypes are 0.25. There is no selection upon the haplotypes (All the haplotypes have the same reproductive ability) (A) Small population size is 100, and the average of D is determined from 300 replicate runs. (B) Large population size is 10,000. The average of D is determined from 10 replicate runs. Both figures (A) and (B) show standard deviation.

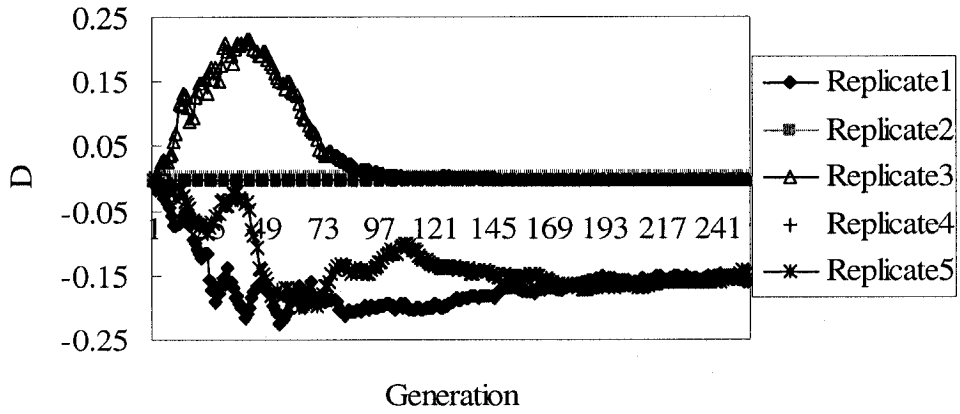
2.2. The Hill-Robertson effect (HRE)

To compare the HRE with genetic drift, we also simulate the Hill Robertson effect in small populations. Under selection, the magnitude of D is less than that generated by only genetic drift with the same population size. However, the average of D caused by the HRE is overall negative since some of the positive LD is reduced when selection is introduced (see Discussion 2.2 for further explanation). The average of D is still smaller than the magnitude of D. The reason is that some negative LDs are still offset by positive LDs. In the figures below, we focus on the HRE in the case of either r selection or constant population size.

2.2.1 r selection

In Figure 6, five randomly chosen results of replicate runs under r selection are shown. In Figure 6, the replicate 3 reaches carrying capacity (Population size equal 10,000) at generation 86. Both the replicate 1 and replicate 5 reach carrying capacity at generation 128. Before the above replicates reach carrying capacity, they are under r selection. From Figure 6A, we see that noteworthy negative LD in the replicate 1 and 5 is generated while considerable positive LD in the series 3 is generated. Although the average D of many replicate runs is not obviously negative, noteworthy negative LD are actually accumulated in some trajectories (replicate 1 and 5) under r selection.

Figure 6 Linkage disequilibrium under r selection (The Hill-Robertson effect)



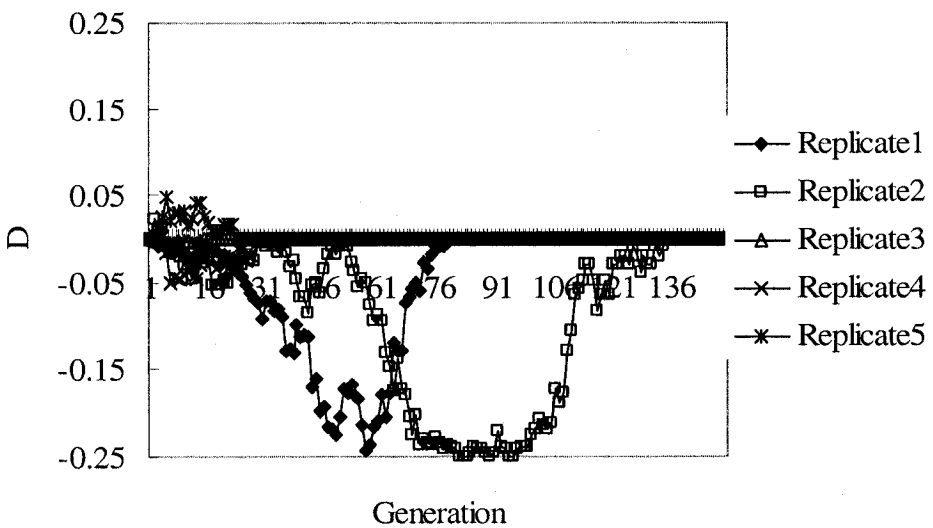
A series of D values determined from 5 replicates lies on the Y axis, respectively. The multiplicative model is applied. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01, respectively. The reproductive rate for both the beneficial allele A and allele B is 0.05. Carrying capacity is 10,000, and the initial population size is 100.

2.2.2 Constant population size

In figure 7, five randomly chosen results of replicate runs with constant small population size are shown. With the weak selection (reproductive rate=0.05), the series 1 and 2 in Figure 7 show noteworthy negative D values although the average D values of the ten replicate runs are not considerably negative. In the figure, population size does not always remain constant (equal to carrying capacity). Because of the obvious stochastic effect in small populations, sometimes population size is below the carrying capacity even though the reproductive rate is greater than zero. We also try to keep the population size constant by increasing initial population fitness and selection strength; however, with the stronger selection, negative LD generated by the HRE is less noteworthy (See multiplicative model in Figure 1). In actual situations, both the strictly constant population size with the strong selection and fluctuating population size around the carrying capacity with the weak selection may exist. Our study shows the HRE is more considerable under weak selection (See Discussion section for further explanation).

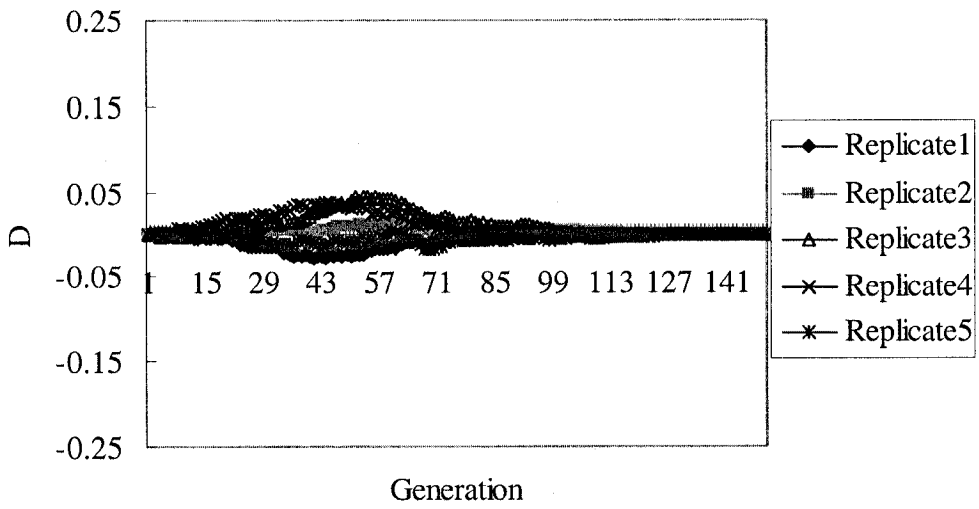
In Figure 8, five randomly chosen results of replicate runs with constant large population size are shown. Contrary to Figure 7, both the positive and negative linkage disequilibrium is not noteworthy.

Figure 7 Linkage disequilibrium in constant small populations (The Hill-Robertson effect)



A series of D values determined from 5 replicate runs lie on the Y axis. The multiplicative model is applied. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01 respectively. The reproductive rate for both the beneficial allele A and allele B is 0.05. Carrying capacity is 100, and the initial population size is 100.

Figure 8 Linkage disequilibrium in constant large populations (The Hill-Robertson effect)

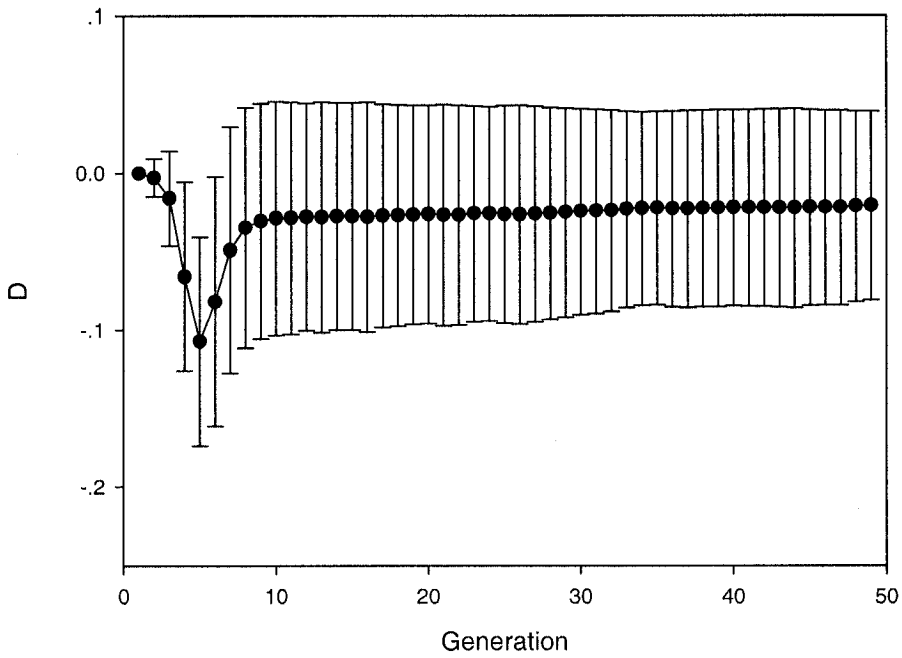


A series of D values determined from 5 replicate runs lie on the Y axis. The multiplicative model is applied. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01, respectively. The reproductive rate for both the beneficial allele A and allele B is 0.05. Carrying capacity is 10,000, and the initial population size is 10,000.

2.3 Quasi-truncation selection (small populations)

Comparing Figure 9 with Figure 2, we see that the average of negative D becomes noteworthy when population size increases. However, in small populations (population size=100), the negative LD lasts a longer time than that in large populations (population size=10,000). In addition, in some replicate runs, the negative D is noteworthy in small populations because of the considerable stochastic effect in small populations. The same reason may explain that the standard deviation in Figure 9 is greater than that in Figure 2. Although the stochastic effect may accumulate negative LD, the stochastic effect overall offsets the quasi-truncation work. However, we see the average values of negative LD have the similar trajectory as that in Figure 2. Around the time one haplotype with the lowest phenotypic value (haplotype ab)vanishes, D reaches the minimum value (See Discussion 1.3 for further details).

Figure 9 Linkage disequilibrium (D) in small populations under quasi-truncation selection



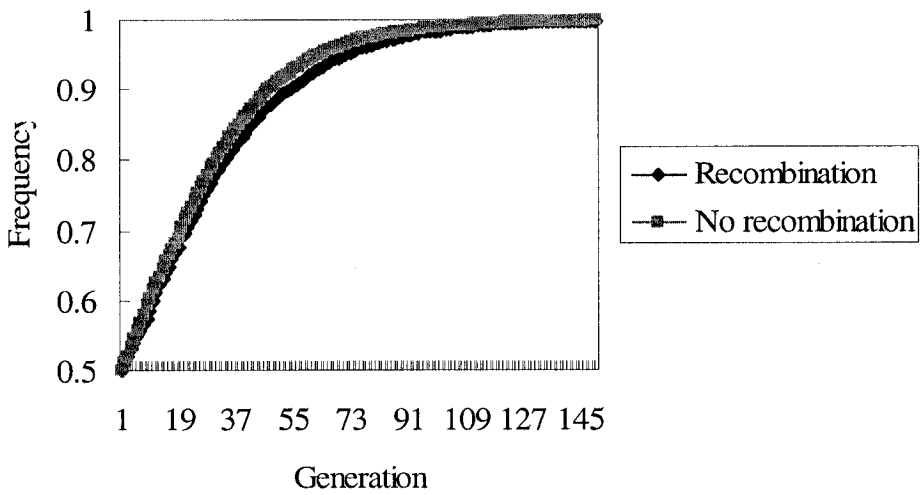
On the Y axis is the average D determined from 300 replicate runs. The D value is calculated with the multiplicative model. The initial frequencies of the four haplotypes ab, aB, Ab, and AB are 0.81, 0.09, 0.09, and 0.01 respectively. The reproductive rate for both the beneficial allele A and allele B is 0.5. Carrying capacity is 100, and the initial population size is also 100.

3 Recombination

Figure 10 shows the effect of recombination on fixation of allele B when overall weak negative linkage disequilibrium is generated by the Hill-Robertson effect. In Figure 10, we see that recombination overall decreases the average frequencies of beneficial allele B and at the same time, delays the fixation of allele B. In fact, recombination has a similar effect on another beneficial allele A

Conversely, Figure 11 shows the effect of recombination on fixation of allele B when noteworthy negative linkage disequilibrium is generated by quasi-truncation selection. In Figure 11A, we see that recombination increases the average frequencies of beneficial allele B and at the same time, speeds the fixation of allele B. Recombination has a similar effect on another beneficial allele A. The inflexion point around generation 8 in Figure 11A corresponds to the vanishing of haplotype ab shown in Figure 11B. After the extinction of haplotype ab, the increase of frequencies of allele B is slow.

Figure 10 The frequencies of allele B with weak negative D generated by the Hill-Robertson effect



The average fixation time of Allele A

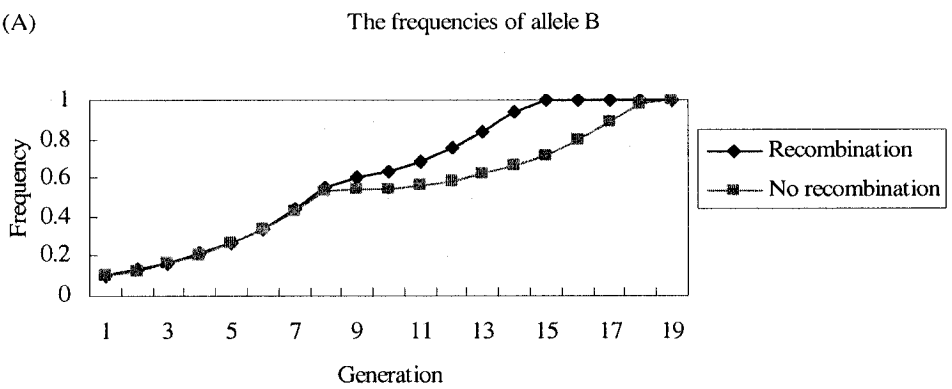
Allele A	Without recombination	With recombination
Average fixation time (generation)	109.7	112.8

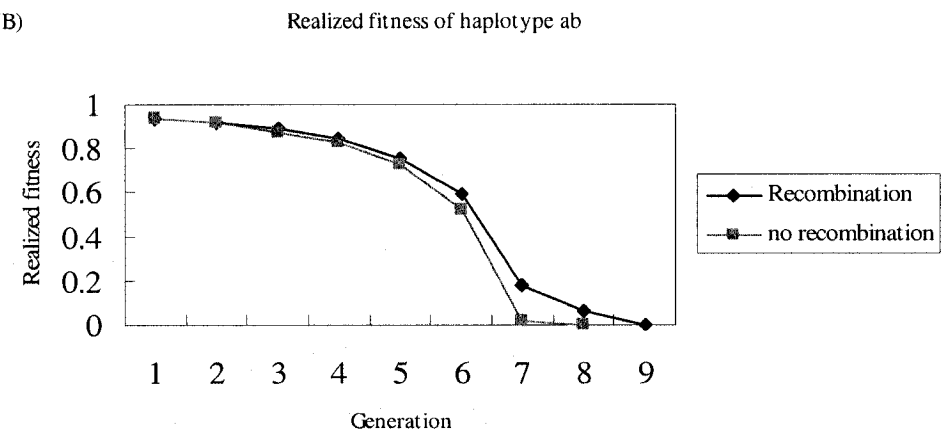
The average fixation time of Allele B

Allele B	Without recombination	With recombination
Average fixation time (generation)	106.32	114.27

The frequencies of the four initial haplotypes are 0.25 respectively. The multiplicative model is applied. The reproductive rate for both the beneficial allele A and allele B is 0.05. Population size is 1,000. The average allele frequencies of allele B and D values are determined from 100 replicate runs. Probability of recombination is 0.2

Figure 11 The frequencies of allele B with noteworthy negative D under quasi-truncation selection





The average fixation time of Allele A and B

Allele A and B	Without recombination	With recombination
Average fixation time (generation)	17.4	14

The frequencies of the four initial haplotypes ab, Ab, aB, and AB are 0.01, 0.09, 0.09, and 0.81 respectively. The reproductive rate for both the beneficial allele A and allele B is 0.05. Population size is 10,000. The average D values and allele frequencies of

allele B in Figure A are determined from 10 replicate runs. The average realized fitnesses of haplotype ab are also determined from the same 10 replicate runs. Probability of recombination is 0.2

Discussion

In Results section, we demonstrate the stochastic process in large and small populations, and we also show that recombination reduces linkage disequilibrium and affects fixation time. In the following parts, we will generalize deterministic models to approximately describe the evolution processes in large populations, and we also discuss noteworthy stochastic effect in small populations. Furthermore, we will discuss the relation between recombination and fixation.

1 Deterministic models

Deterministic models ignore random effects. The recursive mathematical equation of a dynamic system is resolved directly. In our study, mathematical models for r selection, constant population size, and quasi-truncation selection are generalized.

1.1 r selection

In r selection, the frequency of any generation can be tracked through the initial fitness (reproductive ability) and the initial frequencies of haplotypes. Since the growth of the population is not restricted, the realized fitness equals a haplotype's reproductive ability. Hence, the epistasis (E) calculated with realized fitness equals zero, if the initial epistasis (e) calculated with the reproductive ability equals zero.

The frequency of t'th generation is

$$x_1(t) = \frac{x_1 r_1^{\ t}}{x_1 r_1^{\ t} + x_2 r_2^{\ t} + x_3 r_3^{\ t} + x_4 r_4^{\ t}}$$

This can be generalized as the following equation:

$$x_i(t) = \frac{x_i r_i^{\ t}}{\sum_{k=1}^{k=4} x_k r_k^{\ t}} \tag{5}$$

where x_i represents the frequency of the i'th haplotype. (The details of derivative are in the Appendix A)

Consequently,

$$D(t) = \frac{x_4 r_4^{\ t} x_1 r_1^{\ t} - x_2 r_2^{\ t} x_3 r_3^{\ t}}{(\sum_{k=1}^{4} x_k r_k^{\ t})^2} \tag{6}$$

The initial epistasis calculated with the multiplicative model is zero, that is,

$r_4 = r_2 * r_3 / r_1$ or $r_4 * r_1 = r_2 * r_3$

It is easy to see that D(t)=0; however, if the epistasis is calculated with the additive model, D(t) is always negative although the epistasis stays at zero (See Appendix A for proof).

1.2 Constant population size

Unlike the populations under r selection, the population size is kept constant. However, to keep the constant population size, the eliminated individuals are randomly selected according to their haplotype frequencies in the whole population. Hence, each haplotype still has the same frequency as that under r selection. The realized fitness of the haplotypes is:

$$w_i(t) = \frac{N * x_i^{t+1}}{N * x_i^{t}} = \frac{x_i^{t+1}}{x_i^{t}} = \frac{\frac{x_i r_i^{t+1}}{\sum_{k=1}^{k=4} x_k r_k^{t+1}}}{\frac{x_i r_i^{t}}{\sum_{k=1}^{k=4} x_k r_k^{t}}} = \frac{r_i \sum_{k=1}^{k=4} x_k r_k^{t}}{\sum_{k=1}^{k=4} x_k r_k^{t+1}} \quad (7)$$

,where i represents the i'th haplotype, and t represents the t'th generation (see Appendix B for proof).

Using the additive or multiplicative model, dynamic epistasis (E) calculated with realized fitness will remain zero, if the initial epistasis (e) is zero. (See Appendix B for proof)

Since with the constant population size, the frequencies of the haplotypes are the same as that under r selection, then the equation (5) can still be applied in this case:

$$x_i(t) = \frac{x_i r_i^{t}}{\sum_{k=1}^{k=4} x_k r_k^{t}}$$

Similarly, D calculated with the equation (6) can also be used in this case, and D stays at zero while epistasis calculated with multiplicative model is zero. Moreover, D will be negative while epistasis with additive model keeps at zero. (See Appendix B for proof)

1.3 Quasi-truncation selection

Like the cases of constant population size discussed in the above1.2, quasi-truncation selection also keep the constant population size; however, to keep the constant population size, the elimination of individuals are associated with the phenotypes of

individuals. The individual with larger phenotypic values will have high priority to reproduce.

Quasi-truncation selection can be mathematically described with piecewise (stepwise) functions.

Frequencies

Before the growth of haplotypes aB and Ab is restricted (truncated), the frequencies x_2, x_3, and x_4 are the same as those under r selection.

$$x_2(t) = x_2 r_2^{\,t}$$

$$x_3(t) = x_3 r_3^{\,t}$$

$$x_4(t) = x_4 r_4^{\,t} \tag{8}$$

However, the frequency of x_1 is determined by x_2, and x_3 and x_4.

Since $x_1(t) = 1 - x_2(t) - x_3(t) - x_4(t)$,

$$x_1(t) = 1 - x_2 r_2^{\,t} - x_3 r_3^{\,t} - x_4 r_4^{\,t} \tag{9}$$

It is obvious that $x_1(t)$ is monotonically decreasing.

After the growth of haplotypes aB and Ab is restricted (truncated),

$$x_1(t) = 0$$

$$x_2(t) = x_3(t) = (1 - x_4 r_4^{\,t})/2$$

$$x_4(t) = x_4 r_4^{\,t} \tag{10}$$

Fitness

Before the growth of haplotype aB and Ab is restricted (truncated), the realized fitness w_2, w_3, and w_4 are the same as those under r selection.

$$w_1(t) = \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}}$$

$$w_2(t) = r_2$$

$$w_3(t) = r_3$$

$$w_4(t) = r_4 \qquad (11)$$

After the growth of haplotypes aB and Ab is restricted (truncated), the realized fitness w_4 is the same as that under r selection.

$$w_1(t) = 0$$

$$w_2(t) = w_3(t) = \frac{(1 - x_4 r_4^{t+1})/2}{(1 - x_4 r_4^{t})/2} = \frac{1 - x_4 r_4^{t+1}}{1 - x_4 r_4^{t}}$$

$$w_4(t) = r_4 \qquad (12)$$

Epistasis (with additive scale)

Before the growth of haplotypes aB and Ab is restricted (truncated),

$$E(t) = w_4(t) + w_1(t) - w_2(t) - w_3(t)$$

$$E(t) = r_4 + \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}} - r_2 - r_3$$

$$E(t) = \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}} - r_1 \qquad (13)$$

E is negative since $w_1(t) < r_1$ (See Appendix C for proof)

After the growth of haplotypes aB and Ab is restricted (truncated),

$$E(t) = w_4(t) + 0 - w_2(t) - w_3(t)$$

$$E(t) = r_4 - 2 * \frac{1 - x_4 r_4^{t+1}}{1 - x_4 r_4^{t}} \tag{14}$$

Epistasis (with multiplicative scale)

Before the growth of haplotypes aB and Ab is restricted (truncated), the realized fitness w_2, w_3, and w_4 are the same as those under r selection.

$$E(t) = w_4(t) - w_2(t)w_3(t) / w_1(t)$$

$$E(t) = r_4 - r_2 r_3 \left(\frac{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}}{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}\right) \tag{15}$$

E is always negative. (See Appendix C for proof).

After the growth of haplotypes aB and Ab is restricted (truncated),

$$E(t) = w_4(t) - w_2(t)w_3(t) / 0)$$

Hence, E(t) cannot be calculated.

Linkage disequilibrium

Before the growth of haplotypes aB and Ab is restricted (truncated),

$$D(t) = x_1(t)x_4(t) - x_2(t)x_3(t) \tag{16}$$

D(t) is always negative, and monotonically decreasing (See Appendix C for proof)

After the growth of haplotypes aB and Ab is restricted (truncated),

$$D(t) = 0 * x_4(t) - x_2(t)x_3(t) = -x_2(t)x_3(t) \quad (17)$$

D(t) is always negative, and monotonically increasing.

We can see that negative D reaches minimum value around the time when haplotype Ab and aB start to be truncated.

From the deterministic models and simulations, we see that quasi-truncation selection is remarkably important to LD generation since quasi-truncation selection is associated with negative epistasis.

However, quasi-truncation selection cannot be easily predicted with initial parameters although a broken line approximation for truncation selection was proposed. In the approximation, the fitness function is linearly proportion to the phenotypic values.

$w(v)= kv$

,where k is a constant, and w is a fitness function. The fitness increases while phenotypic values increase. The extreme case of the approximation is the strict truncation selection(CROW 1997; CROW and KIMURA 1979).

However, in actual situations, fitness does not follow a fixed pattern of strict truncation selection or some intermediate alternatives. The fitness is dynamically changing. What patterns are taken is determined by the environmental carrying capacity and the level of competition between the individuals. Our study on quasi-

truncation selection focus on more realistic cases: realized fitness has different functions during varied evolutionary stages (As discussed in the above paragraphs). We can examine an evolutionary trajectory: at the beginning time, at least some organisms with intermediate phenotypic values can reproduce and the reproduced offspring still survive in the environment although the organisms with the least favored phenotypes have less fitness regardless of their reproductive ability. In addition, the organism with the most favored phenotypes cannot dominate the environment instantly, since the environment still has carrying capacity to be filled. How long time dose the process take is determined by the reproductive ability of the favored organism and the carrying capacity of the specific environment. All in all, during the evolutionary process, the fitness cannot be described with only strict truncation selection or a simple approximation.

Our study gives the mathematical models for the special case of quasi-truncation selection in which the different evolutionary stages are respectively described with piece-wise functions (See Deterministic models in Discussion section) in which the two intermediately advantageous genotypes have the same reproductive rates and phenotypic values. With some modifications, the models can be further expanded to include more general cases—the two intermediately advantageous genotypes have different reproductive ability and corresponding different phenotypic values.

In addition, under quasi-truncation selection, our study shows that the change of dynamic epistasis measured with realized fitness corresponds to the change of LD for

large populations while other papers only discussed the relationship between initial epistasis which is determined by initial fitness and LD. Moreover, this study shows noteworthy negative epistasis and negative LD simply resulting from competition for the limited resource even in a special case of each haplotype with an equal reproductive rate (See Figure 2).

1.4 Epistasis and linkage disequilibrium

Our study, first time, shows the relation between dynamic epistasis and LD although other papers mention that the negative epistasis may generate negative LD; however, they use instead the static epistasis(OTTO 2001).

Not only are both negative epistasis and LD generated under quasi-truncation selection, but also linear relationship between epistasis and LD under quasi-truncation selection is directly demonstrated in our simulations (See Figure 3). However, epistatic selection does not always correspond to LD (See 1.3 in Discussing for details).

Although epistasis cannot always correspond to LD, still deserving notice is the fact that the formula for the non-epistatic selection calculated with multiplicative model and LD match each other gracefully.

$w_1 * w_4 = w_2 * w_3 \quad \leftarrow\rightarrow \quad x_1 * x_4 = x_2 * x_3$

When epistasis with the multiplicative model is zero, D will be zero. This may also have the implication of the difference of the additive model and the multiplicative model.

1.5 The additive model vs. the multiplicative model

In their origin paper, Hill and Robertson used additive model instead of multiplicative model to demonstrate the negative LD generated by the combination of genetic drift and selection. The HRE has been used without pointing out that the additive model itself can also generate negative LD(HILL and ROBERTSON 1966).

From the above discussion on deterministic models and the simulations for the r selection, constant population size, and quasi-truncation selection, we see that, under the additive scale, even the non-epistatic selection can still generate negative LD (See Figure 1 in Results section). The reason is that the non-epistatic selection calculated with additive scale is still negative epistatic with multiplicative scale. Moreover, the generated negative LD is noteworthy when the selection is strong (Reproductive rates are large). Although multiplicative calculation can be transformed to the additive model by logarithmic calculation, in this study, their effect on LD is different.

2 Stochastic Processes

2.1 Genetic drift

Genetic drift (without any selection) can generate LD with a large magnitude (See Figure 5). However, LD generated by genetic drift only is not always negative in multi

replicate runs. The positive and negative LDs have equal opportunity to occur. For the larger population size, the stochastic effect is not so noteworthy.

Although genetic drift itself was studied by researchers(KIMURA 1955; KIMURA 1957), I would like to argue the point that how important is genetic drift, on its own, to the evolution of sex? If we say that the mechanisms giving rise to the sexual reproduction and maintaining sex are different(CROW 1999), quasi-truncation selection may play an important role in the origin of sex since quasi-truncation selection can generate negative LD quickly. Then, a strategy such as sexual reproduction to counterbalance the negative LD generated by quasi-truncation selection is necessary. What is more, the significance of genetic drift to the maintenance of sexual reproduction cannot be ignored. Once the organisms developed the sexual reproduction, they cannot suddenly change back to asexual mechanism. To maintain sexual reproduction is not as hard as to initiate it. If only the advantages of sex equal the disadvantages of sex or little higher, the sexual reproduction would be preserved.

Now papers usually argue that how large the magnitude of multi-locus association against two-fold cost (ILES *et al.* 2003; KEIGHTLEY and OTTO 2006). I would like to emphasize the impact of genetic drift on maintaining sex, which was ignored sometimes. To further clarify that, we should notice that most new mutations are neutral. Genetic drift rather than HRE and quasi-truncation selection is easy to be found an underlying force for generating LD.

2.2 Hill-Robertson effect

Keightley and Otto (2006) simulated the Hill-Robertson effect on multi-locus with deleterious mutation. However, the negative LD was not directly demonstrated since it used a modifier locus to show the relation between HRE and recombination. In our study, how the negative LD is generated by the HRE is shown. In addition we focus on the beneficial mutation rather than deleterious mutation.

Also different from the simulation in Hill and Robertson's work (1966), we use multiplicative scale, instead of additive scale. According to our above discussion, the non-epistatic selection calculated with the additive model still generates negative LD. To eliminate the confounding effect caused by the additive model, we, hence, simulate the Hill-Robertson effect with the multiplicative model. Our simulation shows, with multiplicative scale, selection can reduce positive LD, hence accumulating negative LD since positive LD accelerates a fixation rate. In Figure 6 and 7, we see that the evolutionary trajectory with positive LD will quickly reach fixation. In the last generations, most of remaining trajectories have negative LD. Hence, HRE will build up negative LD which makes the maintenance of sexual reproduction preferable. Although the average negative LD in our study is not noteworthy, the negative LD in some runs of simulation is quite large.

2.3 Stochastic processes and quasi-truncation selection

While the stochastic processes such as HRE and genetic drift generate considerable negative LD, the stochastic processes overall devalue the effect of the deterministic process on generating negative LD in small populations (See Figure 9). However, in some replicates, the negative LD caused by the collaboration of stochastic effect and quasi-truncation selection is considerable.

3 Recombination and fixation

In the above discussion, we see that LD may be caused by both deterministic and stochastic processes. Recombination is dispensable to break up the generated LD and accelerate evolutionary process.

The other papers use recombination modifier locus to show how the HRE and negative epistasis affect the recombination (BARTON and OTTO 2005; KEIGHTLEY and OTTO 2006; KOUYOS *et al.* 2006). In our study, to illustrate how negative LD is generated, we separately simulate recombination and non-recombination. We use fixation time to evaluate the effect of recombination on evolution (See Figure 10 and 11). For genetic drift and the HRE, we see that the average fixation time is not shortened by recombination. One reason is the average of LD generated by genetic drift and HRE is not noteworthy. Another reason is that at the beginning generations, positive LD is almost as noteworthy as negative LD since weak selection cannot reduce positive LD in short time. Recombination may reduce the positive LD in the beginning generations and hence, decreases the beneficial allele's chances of fixation. However, for some evolutionary trajectories, negative LD is noteworthy. The value of

D even is around -0.24 which is almost near the minimum of negative LD for tow loci (D= -0.25). For these evolutionary trajectories, recombination may accelerate fixation process. Under quasi-truncation selection, we see the recombination decrease the fixation time.

4 Conclusion

Our research shows that non-epistatic selection with additive model still generates negative linkage disequilibrium. The Hill-Robertson effect may generate significant negative LD under multiplicative scale. The more realistic quasi-truncation selection can generate negative LD in large populations. In quasi-truncation selection, dynamic epistasis measured with realized fitness correlates to LD. Stochastic effect overall devaluate negative LDs caused by negative epistasis. Recombination may shorten fixation time.

In summary, populations may build up negative LD in many ways: genetic drift, the HRE, quasi-truncation selection, and even non-epistatic selection with additive scale. To eliminate the accumulated LD, some sorts of mechanisms are needed, which is a good reason why recombination was adopted by organisms and continue to be employed until now.

Literature Cited

BARTON, N., and S. P. OTTO, 2005 Evolution of random drift. Genetics **169:** 2353–2370.

BARTON, N. H., 1995 A general model for the evolution of recombination. Genet. Res. **65:** 123-144.

CAMPBELL, N. A., and J. B. REECE, 2001 The evolution of populations, pp. 445-462 in *Biology*. Pearson Education, Inc., San Francisco.

CORDELL, H. J., 2002 Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human Molecular Genetics **11:** 2463-2468.

CROW, J. F., 1997 The high spontaneous mutation rate: Is it a health risk. PNAS **94:** 8380-8386.

CROW, J. F., 1999 The omnipresent process of sex. J. Evol. Biol **12:** 1023-1025.

CROW, J. F., and M. KIMURA, 1979 Efficiency of truncation selection. PNAS **76:** 396-399.

DALY, M., 1978 The cost of mating. The American Naturalist **112:** 771-774.

EWENS, W. J., 1969 Two locus behavior, pp. 97 in *Population genetics*, edited by B. M.S.

FELSENSTEIN, J., 1965 The effect of linkage on directional selection. Genetics **52:** 349-363.

HALLIBURTON, R., 2004 Quantitative genetics, pp. 525-591 in *Introduction to population genetics*. Pearson Prentice Hall, Upper Saddle River.

HAMILTON, W. D., R. AXELROD and R. TANESE, 1990 Sexual reproduction as an adaptation to resist parasites. PNAS **87:** 3566-3573.

HARTL, D. L., and A. G. CLARK, 1989 Ecological complications, pp. 230-231 in *Principles of population genetics*. Sinauer associates, Inc. publishers.

HEDRICK, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331-341.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

HIRSHFIELD, M. F., and D. W. TINKLE, 1975 Natural selection and the evolution of reproductive effort. Proc. Nat. Acad. Sci. **72:** 2227-2231.

ILES, M. M., K. WALTERS and C. CANNINGS, 2003 Recombination can evolve in large finite populations given selection on sufficient loci. Genetics **165:** 2249-2258.

J. H. AHRENS, H., and G. U. DIETER, 1973 Computer Methods for Sampling from Gamma, Beta, Poisson and Binomial Distributions. Computing **12:** 223-246.

KEIGHTLEY, P. D., and S. P. OTTO, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature **443:** 89-92.

KIMURA, M., 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Sprig Harb. Symp. quant. Biol **20:** 33--55.

KIMURA, M., 1957 Some problems of stochastic processes in genetics. Ann. math. Statist. **28:** 882-901.

KIMURA, M., and J. F. CROW, 1978 Effect of overall phenotypic selection on genetic change at individual loci. PNAS **75:** 6168-6171.

KNUTH, D. E., 1969 *Art of computer programming, volume 2: seminumerical algorithms.* Addison-Wesley Pub. Co., Upper Saddle River, NJ

KONDRASHOV, A. S., 1984 Deleterious mutations as an evolutionary factor. I. The advantage of recombination. Genet. Res. **44:** 199–218.

KONDRASHOV, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336:** 435-440.

KOUYOS, R. D., S. P. OTTO and S. BONHOEFFER, 2006 Effect of varying epistasis on the evolution of recombination. Genetics **173:** 589-597.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models Genetics **49:** 49-67.

LEWONTIN, R. C., 1988 On measures of gametic disequilibrium. Genetics **120:** 849-852.

MARTIN, G., S. P. OTTO and T. LENORMAND, 2005 Selection for recombination in structured populations. Genetics **172:** 593-609.

MAYNARD SMITH, J., 1978 *The evolution of sex* Cambridge University Press, New York

MUELLER, J. C., 2004 Linkage disequilibrium for different scales and applications. Briefings in bioinformatics **5:** 355-364.

MULLER, H. J., 1932 Some genetic aspects of sex. Am. Nat. **8:** 118–138.

MULLER, H. J., 1950 Our load of mutations. Amer. J. Hum. Genet. **2:** 111-176.

OTTO, S. P., AND N. BARTON, 2001 Selection for recombination in small populations. Evolution **55:** 1921–1931.

PHILIPS, P. C., S. P. OTTO and M. C. WHITLOCK, 2000 Beyond the average: the evolutionary importance of gene interactions and variability of epistatic effects., pp. 20-38 in *Epistasis and the evolutionary process*, edited by J. B. WOLF, E. D. B. III and M. J. WADE. Oxford University Press.

PIANKA, E. R., and W. S. PARKER, 1975 Age-specific reproductive tactics. The American Naturalist **109:** 453-464.

PUNIYANI, A., U. LIBERMAN and M. W. FELDMAN, 2004 On the meaning of non-epistatic selection. Theoretical population biology **66:** 317-321.

WILLIAMS, G. C., 1975 An important question, it's easy answer, and the consequent paradox, pp. 3-14 in *Sex and evolution*, edited by R. M. MAY. Princeton University Press, Princeton, New Jersey.

Appendix

A. r selection

N represents the initial population size, and N(t) represents the population size of t'th generation. x1, x2, x3, and x4 represent the initial haplotype frequencies. Xi(t) represent the haplotype frequency in the t'th generation. r represents the reproductive ability of the i'th haplotype.

$$N(1) = N * x_1 * r_1 + N * x_2 * r_2 + N * x_3 * r_3 + N * x_4 * r_4$$

$$N(2) = N * x_1 * r_1^{\,2} + N * x_2 * r_2^{\,2} + N * x_3 * r_3^{\,2} + N * x_4 * r_4^{\,2}$$

.....

$$N(t) = N * x_1 * r_1^{\,t} + N * x_2 * r_2^{\,t} + N * x_3 * r_3^{\,t} + N * x_4 * r_4^{\,t} \qquad (18)$$

Hence,

$$x_1(t) = N * x_1 * r_1^{\,t} / N(t)$$

$$x_1(t) = \frac{x_1 r_1^{\,t}}{x_1 r_1^{\,t} + x_2 r_2^{\,t} + x_3 r_3^{\,t} + x_4 r_4^{\,t}}$$

Similarly, we can derive $x_2(t)$, $x_3(t)$, and $x_4(t)$.

They can be expressed as:

$$x_i(t) = \frac{x_i r_i^{\,t}}{\sum_{k=1}^{k=4} x_k r_k^{\,t}} \qquad (5)$$

where x_i represents the frequency of the i'th haplotype.

According to the formula D(t)= x1(t)*x4(t) - x2(t)*x3(t)

$$D(t) = \frac{x_1 r_1^{\ t}}{\sum_{k=1}^{k=4} x_k r_k^{\ t}} * \frac{x_4 r_4^{\ t}}{\sum_{k=1}^{k=4} x_k r_k^{\ t}} - \frac{x_2 r_2^{\ t}}{\sum_{k=1}^{k=4} x_k r_k^{\ t}} * \frac{x_3 r_4^{\ t}}{\sum_{k=1}^{k=4} x_k r_k^{\ t}}$$

Consequently,

$$D(t) = \frac{x_4 r_4^{\ t} x_1 r_1^{\ t} - x_2 r_2^{\ t} x_3 r_3^{\ t}}{(\sum_{k=1}^{4} x_k r_k^{\ t})^2} \qquad (6)$$

Theorem 1.

If the initial epistasis calculated with the multiplicative model is zero, and initial D is zero, then D(t)=0

Proof.

If the initial epistasis calculated with the multiplicative model is zero, that is,

$r_4 = r_2 * r_3 / r_1$ or $r_4 * r_1 = r_2 * r_3$

then,

$$D(t) = \frac{x_4 r_4^{\ t} x_1 r_1^{\ t} - x_2 r_2^{\ t} x_3 r_3^{\ t}}{(\sum_{k=1}^{4} x_k r_k^{\ t})^2} = \frac{x_1 x_4 (r_1 r_4)^t - x_2 x_3 (r_2 r_3)^t}{(\sum_{k=1}^{4} x_k r_k^{\ t})^2} = \frac{x_1 x_4 (r_1 r_4)^t - x_2 x_3 (r_1 r_4)^t}{(\sum_{k=1}^{4} x_k r_k^{\ t})^2}$$

Since D=x1*x4-x2*x3=0,

D(t)=0

Theorem 2.

If the epistasis is calculated with the additive model, D(t) is always negative although the epistasis stays at zero

Proof.

If initial epistasis e=0 then,

r1+r4=r2+r3

Since initial D= x1*x4-x2*x3=0,

$$D(t) = \frac{x_4 r_4^{\,t} x_1 r_1^{\,t} - x_2 r_2^{\,t} x_3 r_3^{\,t}}{(\sum_{k=1}^{4} x_k r_k^{\,t})^2}$$

$$\frac{x_1 x_4 (r_1 r_4)^t - x_1 x_4 (r_2 r_3)^t}{(\sum_{k=1}^{4} x_k r_k^{\,t})^2} < 0 \Leftrightarrow r_1 r_4 < r_2 r_3$$

$$\Leftrightarrow r_1(r_2 + r_3 - r_1) < r_2 r_3 \qquad (19)$$

$$\Leftrightarrow r_1 r_2 + r_1 r_3 - r_1^{\,2} < r_2 r_3$$

$$\Leftrightarrow r_1 r_2 - r_1^{\,2} < r_2 r_3 - r_1 r_3$$

$$\Leftrightarrow r_1(r_2 - r_1) < r_3(r_2 - r_1)$$

Since $r_2 - r_1 > 0$

$$\Leftrightarrow r_1 < r_3$$

In this study, r1 is less than r3, hence,

$$D(t) < 0$$

B. Constant population size

Unlike the populations under r selection, the population size N is kept constant. However, each haplotype still has the same frequency as that under r selection.

$$x_i(t) = \frac{x_i r_i^{\,t}}{\sum_{k=1}^{k=4} x_k r_k^{\,t}} \qquad (5)$$

Hence the conclusions for D in Appendix A can also be applied to D in a constant population size.

The realized fitness of the i'th haplotypes is:

$$w_i(t) = \frac{N * x_i^{\,t+1}}{N * x_i^{\,t}} = \frac{x_i^{\,t+1}}{x_i^{\,t}} = \frac{\dfrac{x_i r_i^{\,t+1}}{\sum_{k=1}^{k=4} x_k r_k^{\,t+1}}}{\dfrac{x_i r_i^{\,t}}{\sum_{k=1}^{k=4} x_k r_k^{\,t}}} = \frac{r_i \sum_{k=1}^{k=4} x_k r_k^{\,t}}{\sum_{k=1}^{k=4} x_k r_k^{\,t+1}} \qquad (7)$$

,where i represents the i'th haplotype, and t represents the t'th generation

Theorem 3. Calculated with either the additive or multiplicative model, E(t)=0 if initial e=0

Proof.

If the additive model is used, then r1+r4=r2+r3

Considering equation (7), then

$$E(t) = w_4(t) - [w_2(t) + w_3(t) - w_1(t)] = [r_4 - (r_2 + r_3 - r_1)]\frac{\sum_{k=1}^{k=4} x_k r_k^{\,t}}{\sum_{k=1}^{k=4} x_k r_k^{\,t+1}}$$

$$=0$$

Similarly, we can prove that E(t) calculated with multiplicative model is also zero.

C. Quasi-truncation selection

Before the growth of haplotype aB and Ab is restricted (truncated), the realized fitness w_2, w_3, and w_4 are the same as those under r selection.

$$w_1(t) = \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}}$$

$$w_2(t) = r_2$$

$$w_3(t) = r_3$$

$$w_4(t) = r_4 \qquad (11)$$

Epistasis (with additive scale)

Before the growth of haplotypes aB and Ab is restricted (truncated),

$$E(t) = w_4(t) + w_1(t) - w_2(t) - w_3(t)$$

$$E(t) = r_4 + \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}} - r_2 - r_3$$

$$E(t) = \frac{1 - x_2 r_2^{t+1} - x_3 r_3^{t+1} - x_4 r_4^{t+1}}{1 - x_2 r_2^{t} - x_3 r_3^{t} - x_4 r_4^{t}} - r_1 \qquad (13)$$

Since r1, r2, r3, and r4 are greater than 1, hence

$$x_2 r_2^{t+1} + x_3 r_3^{t+1} + x_4 r_4^{t+1} > x_2 r_2^{t} + x_3 r_3^{t} + x_4 r_4^{t}$$

then,

$$1 - (x_2 r_2^{t+1} + x_3 r_3^{t+1} + x_4 r_4^{t+1}) < 1 - (x_2 r_2^{t} + x_3 r_3^{t} + x_4 r_4^{t})$$

$$\frac{1 - (x_2 r_2^{t+1} + x_3 r_3^{t+1} + x_4 r_4^{t+1})}{1 - (x_2 r_2^{t} + x_3 r_3^{t} + x_4 r_4^{t})} < 1$$

Considering under our quasi-truncation selection, the growth of haplotype ab is always restricted; hence, haplotype ab cannot achieve a realized fitness greater than its reproductive ability. i.e., w1(t)<r1

E(t) is negative since $w_1(t)<r_1$.

Epistasis (with multiplicative scale)

Before the growth of haplotypes aB and Ab is restricted (truncated), the realized fitness w_2, w_3, and w_4 are the same as those under r selection.

$$E(t) = w_4(t) - w_2(t)w_3(t) / w_1(t)$$

$$E(t) = r_4 - r_2 r_3 \left(\frac{1 - x_2 r_2^{\,t} - x_3 r_3^{\,t} - x_4 r_4^{\,t}}{1 - x_2 r_2^{\,t+1} - x_3 r_3^{\,t+1} - x_4 r_4^{\,t+1}}\right) \qquad (15)$$

Similar to the proof for the additive model, since $w_1(t)<r_1$

$$E(t) = r_4 - r_2 r_3 / w_1(t) < r_4 - r_2 r_3 / r_1 = 0$$

E(t)= is negative.

Before the growth of haplotypes aB and Ab is restricted (truncated),

$$D(t) = x_1(t)x_4(t) - x_2(t)x_3(t) \qquad (16)$$

According to the equation (8),

$$x_2(t) = x_2 r_2^{\,t}$$

$$x_3(t) = x_3 r_3^{\,t}$$

$$x_4(t) = x_4 r_4^{\,t} \qquad (8)$$

,and r2, r3, r4 are greater than 1, hence,

$$x_2(t) > x_2$$

$$x_3(t) > x_3$$

$$x_4(t) > x_4$$

Then,

$$x_1(t) = 1 - x_2(t) - x_3(t) - x_4(t) < 1 - x_2 - x_3 - x_4 = x_1$$

$$\Rightarrow x_1(t) < x_1$$

$$\Rightarrow x_1(t) < x_1 r_1^{\ t}$$

With the multiplicative model,

$$D(t) = x_1(t)x_4 r_4^{\ t} - x_2 r_2^{\ t} x_3 r_3^{\ t} < x_1 r_1^{\ t} x_4 r_4^{\ t} - x_2 r_2^{\ t} x_3 r_3^{\ t} = x_1 x_4 (r_1^{\ t} r_4^{\ t} - r_2^{\ t} r_3^{\ t}) = 0$$

Hence, D(t) is always negative, and monotonically decreasing

Similarly, we can prove that with the additive model.

D. Programming code

See the code package on the server:

http://bioinformatics.concordia.ca/~qzhang/QiZhangThesis