The evolution of recombination rates caused by recurrent deleterious mutations

by
Daniel Paul Falush

A thesis submitted for the degree of
Doctor of Philosophy of the University of London

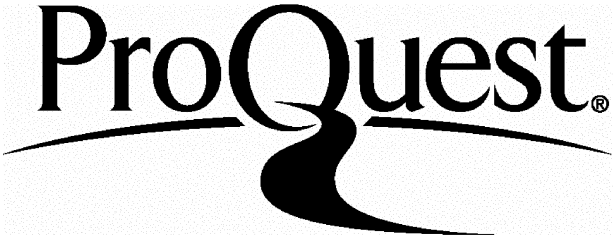
The Galton Laboratory

Department of Biology

University College London

4 Stephenson Way

London NW1 2HE



ABSTRACT

The exchange of genetic material through recombination underlies the evolutionary process. However, the factors which determine the evolution of the rate of exchange are poorly understood. In the thesis, I concentrate on the evolution of recombination rates caused by recurrent deleterious mutations. I have developed an approach to modelling recombination evolution which overcomes many of the limitations of previous methods. My main innovation is to make a simplifying assumption; I do not consider genes in linear arrays on chromosomes. Instead of treating genes as beads on a string, I treat them as beads in a pot.

An important aspect of the variation in recombination is the variation between the sexes. For example, in humans, the female genetic map is 60% longer than the male genetic map. I use my genes-in-a-pot model to investigate the evolution of sex differences in recombination. I show that sex differences in recombination may evolve as a result of sex differences in the strength of selection. My results suggest that, in humans, the genetic map is longer in females because selection against deleterious mutations is stronger in males. Additionally, I show that, in mammals, male recombination rates should depend primarily on longevity while female recombination rates should show stronger dependence on sex differences in selection. The limited empirical data is consistent with the pattern I suggest.

In the final chapters, I deal with some objections to my model of recombination evolution. A first objection is that an assuption of the model, synergistic interactions between mutations, is not biologically applicable. I show that synergism is likely to evolve in sexual species like mammals, which have high mutation rates. A second objection is that recombination is an incidental consequence of the crossing over mechanism which causes it. I show that selection on recombination is necessary to explain important features of crossover distribution.

ACKNOWLEDGMENTS

The Ph.D. was funded by the Wellcome foundation.

I would additionally like to thank my supervisor Andrew Pomiankowski for continued encouragement despite the prolonged absence of a concrete result. Others at UCL who have helped greatly are John Atwood (computing), Patrice David (proof reading some very challenging documents), Michelle Dres (proof reading the final draft), Chris Jiggins (showing how it should be done) and Sue Povey (much discussion on recombination).

Outside UCL I particularly appreciate the help and stimulation provided by Austin Burt, Marc Feldman, Maj Hultén, Alexey Kondrashov, Yoh Iwasa, the elusive Jochim, Richard Nichols, Sally Otto, Rosie Redfield and Stuart West.

The rapid rewrite would have been impossible without the assistance of Chris Jiggins, Nick Isaac, Owen Rose and Lacy Rumsey. I greatly appreciate their help. Additionally, I would like to thank Yuuma, for sleeping nights, and Mari, for selfless support during the day.

TABLE OF CONTENTS

FIGURES

TABLES

TABLE OF SYMBOLS

I have ommited the symbols used in describing the three locus model (Section 1.6.1) and the interval halving algorithm (Section 2.7). These symbols are not referred to in other sections of the text.

Symbol	Usage
U	The mutation rate per diploid genome per generation.
$\overline{w}$	The mutational load (defined in section 1.1.1).
$w(n)$	The fitness of an individual with n mutations.
α, β	Parameters of the Gaussian fitness function $w(n) = \exp-(\alpha n + \frac{1}{2}\beta n^2)$.
$\overline{n}$	Mean number of mutations per diploid individual, measured before selection.
V	Variance in mutation number of diploid individuals, measured before selection.
Δx	(Chapter 1) Change in quantity x in a single generation.
Δx	(Chapter 2,3) Change in quantity x caused by selection.
δr	The amount by which recombination is modified.
$\frac{\delta \overline{n}}{\delta r}$	The difference in mean mutation number between individuals heterozygous for the modifier and individuals from the unmodified population.
$\frac{\delta V}{\delta r}$	The difference in variance in mutation number between individuals heterozygous for the modifier and individuals from the unmodified population.
$d(\delta r)$	The gradient of selection on a recombination modifier which alters recombination by δr.
d	The gradient of selection on a recombination modifier of small effect.
r	The recombination rate between loci. Under genes-in-a-pot recombination, r is the recombination rate between all pairs of loci.
k	The truncation point of the fitness function during computer simulation; individuals with k or more mutations have zero fitness. Gametes with k or more mutations are not tracked.
R	The proportion of a pot inherited from one of the parental potsunder genes-in-a-pot recombination. The other parental pot contributes 1-R of its genetic material.
f_t	The frequency of the modifier at generation t.

Continued overleaf.

TABLE OF SYMBOLS - CONTINUED

A haplont is the genetic material provided by one gamete. An individual has two haplonts, one from his or her mother and one from his or her father.

Symbol	Usage
M	Quantity in maternal haplont.
P	Quantity in paternal haplont.
$\overline{M}$	Mean number of mutations in maternal haplont.
V_M	Variance in mutation number in maternal haplont.
D_M	The reduction in variance of maternal haplonts due to linkage disequilibrium, $V_M - \overline{M}$.
X^b	Quantity X measured before selection.
X^a	Quantity X measured after selection.
X_f	Quantity X measured in females.
X_m	Quantity X measured in males.
d_m	Gradient of selection on a modifier of recombination which modifies recombination in males alone. Other symbols for the effect of sex-specific modification are adapted analogously.
$B(i,R)$	Binomial distribution with mean Ri and variance $R(1-R)i$.
$\Re$	The potential for recombination to increase the variance in mutation number. Under genes-in-a-pot recombination, the variance in mutation number of recombined gametes is $\frac{1}{2}r\Re$ higher than the variance in mutation number of similar gametes produced without recombination. $\Re = 2Cov_{MP} - D_M - D_P - (M-P)^2$.
hs	The effective selection against each mutation $\frac{\bar{n}^b - \bar{n}^a}{\bar{n}^b}$.
T,e	Parameters in the fitness function $w(n) = 1 - \left(\frac{n}{T}\right)^{1/(1-e)}$. When $e = 1$, the function gives truncation selection, such that individuals with less than T mutations have equal fitness, and individuals with T or more mutations have zero fitness.
V_G	The genetic variance of a trait.
V_M	The increase in genetic variance of a trait caused by a single generation of mutation.

Chapter 1

A pot model of recombination

Abstract

This chapter develops a model for calculating selection on recombination modifiers. I introduce a specific recombination mechanism, named "genes-in-a-pot". This mechanism allows the strength of selection on modifiers to be calculated exactly. The model incorporates epistatic interactions between deleterious mutations at many loci. Results are compared with the approximations of Charlesworth (1990) and Barton (1995). Both of these approximations are derived assuming weak selection. The comparison shows that the weak selection approximation is significantly less accurate in predicting the strength of selection on recombination modifiers than in predicting the properties of the unmodified population.

1.1 General background

1.1.1 The mutational load

In order for organisms to replicate, the genetic material needs to be duplicated, and this inevitably entails occasional errors. A few of these errors will increase the fitness of the individuals in which they occur. In other cases the new variant will not differ significantly from the original. In many cases, however, the new variant will disrupt the function of the genetic material and thereby reduce the fitness of the carrier.

The mechanisms for DNA replication have evolved to become very accurate. The rate at which an individual base pair is replicated incorrectly in each duplication in lower eukaryotes is of the order of 10^{-10} (Drake 1991). Nevertheless, the rate of error is significant in complex organisms such as *Homo sapiens* because the genetic material undergoes O (10^2) cell divisions on average in being passed from one generation to the next (Li et al. 1996). *Homo sapiens* has O (10^5) genes (Fields et al. 1994). In yeast the average gene has O (10^3) base pairs (Bird 1995). An order of magnitude estimate of the total deleterious mutation rate per individual per generation, U, can be obtained by multiplying these numbers together. 10^{-10} errors/basepair/cell division * 10^2 cell divisions/generation * 10^5 genes * 10^3 base pairs per gene gives an order of magnitude estimate of 10^0 deleterious mutations per generation. For a more rigorous estimation of U using a genomic approach see Kondrashov and Crow (1993).

The rate of deleterious mutation is significant because the great majority of deleterious mutations must be removed by selection if the species is not to suffer slow deterioration and eventual extinction. A deleterious mutation rate of order 1 or greater becomes problematic for the survival of any species because of the resultant 'mutational load'. In an asexual population, the members of the fittest class must on average produce at least e^U surviving individuals if the size of the fittest class is not to shrink (Kimura and Maruyama 1966). A continuous reduction in the size of the fittest class would lead to the inexorable deterioration of the species and its eventual extinction. This rate of reproduction becomes implausible if the mutation rate is large. The mean fitness $\overline{w}$ is defined as the reciprocal of the average rate of production of surviving individuals produced by the fittest class of individuals in a population of constant size. The load is 1 - $\overline{w}$.

An example of a natural population which is able to survive despite a high load is the crested newt *Triturus cristatus carnifax* (Macgregor and Horner 1980). The newt is only able to develop successfully if it is heteromorphic for chromosome 1. This

necessity entails a load of at least 0.5, because half the individuals produced by any mating are homomorphic. Of course, this ignores the mutational load, which must come in addition. How this newt species can bear such a large additional load, when other species do not, is unknown. A possible explanation is that the carcasses of homomorphic newts serve to nourish their heteromorphic brothers and sisters. In this case the newt would not necessarily suffer a great disadvantage compared to other species. The example shows that a high load is not necessarily ruinous for a species, provided it has a reasonably high fecundity. However a high mutational load is certain to be ruinous in a species such as *Homo sapiens* where the maximal fecundity of females is quite limited (Crow 1989).

In a sexual species the mutational load may be less than $1-e^{-U}$ provided that the 'fitness function', which gives the relationship between mutation number and mean fitness, is synergistic. The fitness function $w(n)$ is synergistic if fitness decreases increasingly quickly as the mutation number, n, increases. Synergism creates linkage disequilibria between deleterious mutations because synergistic selection reduces the variance in mutation number by more than it reduces the mean. Recombination (mediated through sexual reproduction) reduces these linkage disequilibria and restores variance. The additional variance ensures that the individuals which are removed by selection have, on average, more than one additional mutation than the individuals which survive (Muller 1958). Hence the mutational load is reduced. It has been suggested that the effect of recombination on these linkage disequilibria may have caused recombination to evolve (Feldman, Christiansen, and Brooks 1980; Kondrashov 1984; Charlesworth 1990; Barton 1995).

1.1.2 Heritable variation in recombination rate

The possibility for recombination to evolve has been demonstrated by a number of studies which have shown heritable variation in the recombination rate. Several authors have performed direct selection experiments on the recombination rate e.g. Chinnici (1971a,b), Shaw (1972). These experiments have demonstrated the potential for both higher and lower recombination to evolve. See Brooks (1988) for a review.

Other experiments have demonstrated that increased recombination can evolve in response to directional selection on another trait (see Korol and Iliadi 1993 for a review). The most elegant of these experiments was performed by Korol and Iliadi (1993), who selected for negative and positive geotaxis (direction of movement in a maze) in *Drosophila melanogaster*. They found a substantial increase in recombination rates in response to both selection regimes. Average recombination in the geo^{-} line was

increased by 66% compared to the control while recombination in the *geo*$^{+}$ line was increased by 78%. In both cases, the rate of increase of recombination varied between chromosomal regions, with some regions showing little or no increase in recombination, and other regions showing large increases (up to fourfold).

Directional selection experiments are relevant to the hypothesis that recombination evolves as a result of the deleterious mutations, because the purging of deleterious mutations represents continuous directional selection to increase fitness. Indeed, the similarity between directional selection on a quantitative trait and selection against deleterious mutations has been mathematically formalised by Barton (1995). In both cases, recombination may be beneficial because it increases additive genetic variance. In the case of directional selection, the response to selection is hastened. In the case of selection against deleterious mutations, the efficiency of selection is increased.

Additional evidence that recombination may evolve comes from differences between species (Burt and Bell 1987; True, Mercer and Laurie 1995) and the sexes (Trivers 1988). This evidence will be discussed in the course of the present thesis.

1.1.3 Modifier theory

Evolutionary geneticists have attempted to understand the evolution of recombination using modifier theory (Nei 1967; Feldman, Otto, and Christiansen 1996). The principle behind this body of theory is that recombination can evolve as a result of altering the genetic associations of genes. For example, recombination can create a chromosome containing the genes *AB* by recombining a chromosome containing *aB* with a chromosome containing *Ab*. Modifier theory supposes that there is a third modifier locus *M* which controls the recombination rate between the "selected" A and B loci and any number of other selected loci. It is further assumed that the variants at locus *M* are themselves selectively neutral. Because the modifier locus is selectively neutral, its evolution depends on altering the genetic combinations, and hence the linkage disequilibria, at the selected loci.

Modifier models can be used to ask a number of questions about the evolution of recombination as a result of linkage disequilibria that build up through selection on deleterious mutations. Is there strong enough selection on recombination to alter crossing over frequency or chromosome number? If so, does it act to increase or decrease the frequency or number? Has selection on recombination led to differences between species in either crossing over frequency or chromosome number? Does the modification of recombination frequency lead to a significant benefit for the species, i.e. by reducing its mutational load?

This chapter introduces a novel modifier model. The model is used in later chapters to make empirical predictions about patterns of sex difference in crossing over in animals (chapter 2) and to ask about the effect of sex differences in mutation rate on the overall level of crossing over (chapter 3).

1.2 A critique of other approaches

The modifier model that is introduced here is, at first sight, rather peculiar. So before introducing it I will discuss the strengths and limitations of other polygenic approaches.

To formulate models of selection on recombination it is necessary to find a way of representing the effect of both selection and recombination on the distribution of alleles. The models discussed here also incorporate mutation. This is, in itself, unproblematic provided it is assumed that mutations arise independently of each other and independently of gene frequencies at other loci. This assumption is quite reasonable because mutation is thought to be a random process (Cavalli-Sforza and Lederberg 1955). This assumption will be relaxed slightly in chapters 2-3, which allow sex differences in the mutation rate.

1.2.1 Charlesworth's approximation

One approach was introduced by Bulmer (1980) and subsequently developed by Charlesworth (1990) to model the evolution of recombination modifiers as a result of the presence of deleterious mutations. Bulmer was concerned to represent the effect of selection and recombination on a quantitative trait. In order to facilitate this he made two assumptions. The first is that there are an infinite number of loci. The second is that the trait in question has a normal distribution. In order to apply this method to selection on deleterious mutations, Charlesworth (1990) assumed that the number of mutations carried by an individual is a quantitative trait and that the number of mutations each individual or gamete carries follows a normal distribution. Each mutation is assumed to have an identical effect on fitness. Mutation adds an average of U mutations per diploid individual per generation, increasing the mean and variance in the mutation number of individuals by U.

The fitness function is assumed to be Gaussian with $w(n) = \exp-(\alpha n + \frac{1}{2}\beta n^2)$, where n is the number of mutations an individual contains. α determines the fitness effect of each mutation, acting independently, and β determines the strength of epistasis. When $\beta = 0$, the fitness function is multiplicative (no epistasis). When

$\beta > 0$, the fitness function is synergistic. When $\beta < 0$ there is diminishing returns epistasis- the opposite of synergism. In a sexual population selection is assumed to fall on diploid individuals.

1.2.1.1 Calculation of equilibrium values of the unmodified population

The choice of a Gaussian fitness function allowed Charlesworth to write an expression for the change in mean mutation number in each generation $\Delta\bar{n}$ caused by selection and mutation, in terms of U, α and β, and the mean, $\bar{n}$, and variance, V, in mutation number in individuals before selection acts. This equation is valid both for a sexual and for an asexual population. For a population at mutation/selection balance, $\Delta\bar{n} = 0$.

In order to represent the effect of selection and recombination on the distribution of mutations, Bulmer made two assumptions. First, that the fresh linkage disequilibria created by selection in a single generation are the same in coupling (within a haplont inherited from one parent) as in repulsion (between the haplonts inherited from the two parents). Second, he assumes that selection creates the same quantity of linkage disequilibria between all pairs of loci.

Both of these assumptions are justified provided selection is weak relative to recombination. In this case, the linkage disequilibrium which builds up between any pair of loci is quite small. If selection is strong relative to recombination then significant linkage disequilibria may build up between loci, depending on their linkage. Note that the first assumption is a special case of the second. The first assumption is convenient because it ensures that the effect of selection on linkage disequilibria in a given generation is independent of the effect of recombination on the linkage disequilibria in the same generation.

Under Bulmer's assumptions, the effect of selection and recombination in each generation can be considered separately. The effect of selection on the linkage disequilibria can be calculated using normal theory and depends on β and V.
The effect of recombination is simply to reduce the linkage disequilibrium between any pair of genes by a proportion which is the recombination rate between them.

Bulmer described his approach as 'rather heuristic' and it does have some limitations. First, the approach breaks down if there is no recombination, because the model predicts infinite linkage disequilibrium. Secondly, it ignores the possibility of covariance between maternal and paternal haplonts caused by non-random mating. Thirdly, it ignores the possibility of changes in linkage disequilibria that result from

differences in the genetic composition of maternal and paternal haplonts. I discuss the effect of such differences in chapter 3.

The first of the limitations listed above was of immediate concern to Charlesworth because he wished to calculate mutation frequencies for realistic genetic maps. For short genetic maps, the recombination rate between many pairs of loci is small. Therefore Charlesworth modified Bulmer's equation to allow the effect of selection to depend on the level of preexisting linkage disequilibria between loci. In Bulmer's equations, the only effect of selection on the linkage disequilibria is through a reduction of V. Charlesworth argued that selection has another effect, reducing the existing linkage disequilibria between each pair of loci by removing some of the mutations between which linkage disequilibria occurs. He therefore added an additional term which allows for this. Charlesworth's argument is plausible. However, it has not been demonstrated that the additional term provides an accurate generalisation of Bulmer's equations to stronger selection.

Charlesworth was able to calculate the equilibrium level of linkage disequilibrium $V - \bar{n}$ in terms of V, $\bar{n}$ and β by integrating his modified form of Bulmer's equation over the genetic map. This integral can be calculated numerically for an arbitrary genetic map. Combining the equation for $V - \bar{n}$ with the equation obtained by setting $\Delta\bar{n} = 0$ he was able to solve to find the values of $\bar{n}$, $\bar{w}$ and V.

1.2.1.2 Selection on recombination modifiers

Charlesworth went on to consider selection on recombination modifiers placed in an arbitrary position on the chromosome. The modifier undergoes selection because of its effect on the distribution of mutations. Charlesworth assumed:

(1) that the modifier is rare.

(2) that the modifier has a small effect.

Because the modifier is rare, it always appears in heterozygous condition. Charlesworth calculated the asymptotic values of the mean and variance in mutation number for an individual heterozygous for the modifier, using methods analogous to those used to calculate $\bar{n}$ and V for the unmodified population as described above. These values were in turn used to calculate the asymptotic rate of increase of the modifier. The assumption that the modifier has a small effect allows second-order difference terms between the modified and unmodified populations to be dropped. The equations are rather fiddly because it is necessary to account for the position of each pair of recombining loci with respect to the modifier. As before, Charlesworth considered selection and recombination separately, before adjusting the level of linkage

disequilibria to allow for the effect of selection on linkage disequilibria which arise as a result of the change in mutation frequencies.

1.2.2 Barton's approximation

Barton and Turelli (1991) and Turelli and Barton (1994) developed a general method of representing multilocus selection and recombination which was extended by Barton (1995) to allow calculation of selection on recombination modifiers. I shall give a brief summary of the features of their method that are most pertinent to the present discussion of modifier models. The method allows gene frequencies to be calculated provided that they rapidly approach an invariant state, which is named 'quasi-linkage equilibrium' or QLE. For models of selection on deleterious mutations, a QLE is attained provided selection is weak relative to recombination. In this case, the level of linkage disequilibria rapidly reaches an equilibrium. The method can be extended to calculate selection on recombination modifiers provided that the frequency of the modifiers itself rapidly attains an invariant state. The rapid attainment of this invariant state is ensured when the modifiers have a small effect on the recombination rate.

Barton (1995 appendix 3) used the QLE method to analyse the problem addressed by Charlesworth. Barton provided expressions for $\Delta\bar{n}$ and V. The former is identical to that of Charlesworth. The latter agrees when selection is weak relative to recombination. Note that, when selection is weak relative to recombination, the additional term that Charlesworth adds to Bulmer's original equations can be neglected.

Barton assumed that the modifier has a small effect, and made a number of additional assumptions, each based on the premise of weak selection. Under these assumptions Barton was able to provide simple expressions for $\frac{\delta\bar{n}}{\delta r}$ and $\frac{\delta V}{\delta r}$. $\delta\bar{n}$ and δV are the difference in the mean and variance in mutation number between individuals heterozygous for the modifier and the rest of the population. δr is the amount by which the modifier alters recombination. These values can be used to calculate the gradient of selection on the modifier d.

1.2.3 Accuracy

Accuracy is an important issue for both approximations, although for different reasons. Charlesworth intended his approximation to be quite generally applicable. Barton claimed that his approximation is accurate at the limit of weak selection. This claim is valid, as long as the algebra is correct. However, the usefulness of his

approximation is dependent on how quickly the approximation becomes inaccurate away from the unrealistic limit.

Charlesworth (1990, Table 3) and Barton (1995, Table 3) checked their respective approximations for $\bar{n}$, $\bar{w}$ and V using the exact values calculated for free recombination (r=0.5). Charlesworth (1990, Table 2) additionally checked his approximations for no recombination (r=0). He also (1990, Tables 2,3) checked the validity of the normal approximation for r=0 and r=0.5 by calculating the skew and kurtosis of the exact distribution of mutations.

However, the only check on the accuracy of the approximations of $\frac{\delta\bar{n}}{\delta r}$ and $\frac{\delta V}{\delta r}$ and d has been a comparison between the two approximations (Barton 1995 table 3). This check is not adequate. Despite their different derivations, the approximations of Barton and Charlesworth are similar; in both cases the initial premise is weak selection. Therefore a comparison of the results of the two approximations does not provide a good check on the accuracy of either. Hence the need for an alternative approach. Here I develop a method which allows calculation of the strength of selection on recombination modifiers without any assumptions about the strength of selection. The values obtained are used to check the accuracy of Barton and Charlesworth's approximations.

1.3 The "genes-in-a-pot" approach

The model developed in this thesis differs from that of Charlesworth in two key respects. First, the assumption that mutations are normally distributed is dropped. Second, a specific recombination mechanism, called "genes-in-a-pot", is assumed. This approach allows the exact effect of selection and recombination on the distribution of mutations to be modelled.

There is no linkage of genes onto linear chromosomes. Instead, genes mix freely in "pots". A pot can therefore be represented by the number of mutations it contains. It is assumed that a gamete has only one pot, so that a diploid individual has two pots, one from his or her mother and one from his or her father. The population is assumed to be infinite. During haploid phases of the life cycle the population can be represented by the frequency distribution of mutations per pot. During diploid phases it can be represented by the matrix of frequencies of individuals with x mutations in their first pot, and y mutations in their second pot. The mutation frequencies of male and female individuals are represented and calculated separately. Individuals are diploid and are formed by random pairing of male and female gametes. There are three stages in each generation: selection, recombination and mutation.

(1) Selection

In this thesis, selection is assumed to fall on diploid individuals. The number of gametes each individual produces is a function of the number of mutations that he or she carries. For the purposes of simulation, it is assumed that the maximum number of mutations per haploid genome is less than a given value k. Any fitness function can be used, but that selected is truncated, so that individuals with k or more mutations have zero fitness. The fitness function accurately approximates the equivalent continuous fitness function provided that the frequency of individuals with k or more mutations is negligible. The fitness functions are normalised by dividing the frequency of individuals with each number of mutations by the mean fitness. Consequently, the mean number of offspring per individual is two.

(2) Recombination and reduction

The distribution of mutant genes in gametes depends on the number of mutants carried by the paternal and maternal chromosomes and the level of recombination. In genes-in-a-pot recombination, two types of gametes are produced with equal frequencies. One gamete type inherits each paternal mutation with probability R and each maternal mutation with probability $1-R$. In the other type, the probabilities are reversed. The value of R is assumed to be the same for every individual. The value of R determines the degree of mixing of maternal and paternal genes, and hence the level of recombination. Effectively,

all loci are equally linked and the recombination rate between each pair of loci is $r = 2R(1-R)$. Genes-in-a-pot recombination is a notional recombination mechanism, which involves a proportion R of the genetic material from one parental pot being poured into the gametic pot. The gametic pot is then filled up using 1-R of the genetic material from the other parental pot. Genes-in-a-pot recombination results in an equal degree of linkage disequilibrium between every pair of loci.

(3) Mutation

I use the same approach as Charlesworth's infinitesimal scheme, where the mutation rate at each locus is independent of the number of mutations already present in the individual. Each gamete acquires a random number of mutations from a Poisson distribution. Consequently, the number of new mutations introduced to each pot does not depend on the number of mutations already in the pot. However, pots with a finite number of mutations could also be simulated by assuming that the mutation rate per pot depends on the number of pre-existing mutations. New mutations are random and independent of one another, so the assumption that mutation occurs after recombination does not alter the distribution of mutations in gametes.

1.3.1 Simulation of recombination modifiers

Charlesworth and Barton both assume that the modifier alters recombination in every generation. In my model, the modifier only has an effect on recombination in a single generation. This deviation from the population recombination rate will be termed a modification event. The frequency changes of the modifier gene are then followed in the generations subsequent to the modification event.

The effect of the modifier is calculated as follows:

(1) The properties of the unmodified population are calculated at mutation/selection balance.

(2) Before the modification event, the distribution of mutations in individuals with the modifier is the same as in the unmodified population. During the modification event, the recombination rate between all pairs of genes is equal to $r+ \delta r$. The distribution of mutations in those gametes which carry the modifier immediately after the modification event is calculated assuming this value. Immediately after the modification event, the modifier is assumed to have a frequency of f_0, where 0 refers to the zeroth generation.

(3) Subsequent to the modification event, gametes containing the modifier are paired with gametes from the unmodified population. The individuals that are formed by this

process undergo mutation and selection in the same way as members of the unmodified population. The recombination rate between all the selected loci is assumed to be r, as for the unmodified population. For a "linked" modifier, the recombination rate between the modifier and the selected loci is also assumed to be r. Modifiers with different degrees of linkage to the selected loci can also be simulated, but only linked modifiers are used in this thesis, so the possibility will not be discussed in detail here.

(4) The frequency changes of the modifier gene are followed for a number of generations subsequent to the modification event. Calculation is terminated when the frequency of the modifier ceases to change significantly. In practice, this occurs after a small number of generations. Selection and mutation alter the frequency distribution of mutations in the gametes with which the modifier segregates. This pushes the frequency distribution of mutations towards that of the unmodified population at mutation/selection equilibrium. At the same time the modifier recombines with gametes from the unmodified population. As a result, the distribution of mutations that accompany the modifier approaches the distribution of the unmodified population very quickly. As long as the distribution of mutations differs from that of the unmodified population, the frequency of the modifier changes. Once the distribution reaches that of the unmodified population, there is no further change in frequency.

For modifiers of small effect, the effect of the modifier in one generation is independent of the effect of the modifier in other generations (Barton 1995). This result allows the output of the present model to be compared directly with the approximations of Charlesworth (1990) and Barton (1995) (henceforth C & B). The value of the selection gradient of the modifier, d (C & B), is equal to the frequency change caused by a single modification event, $\frac{f_\infty - f_0}{f_0 \delta r}$ (the present model). The value of $\frac{\delta \bar{n}}{\delta r}$ and $\frac{\delta V}{\delta r}$ for a modifier of recombination each generation (C & B) is equal to the cumulative total of the difference in mean and variance in mutation number between individuals with the modifier and individuals from the unmodified population, summed over each generation after a single modification event (the present model).

There are two advantages to following the effects of a single modification event rather than calculating the asymptotic properties of a modifier which acts each generation. The first is that it is slightly quicker computationally. The second is that it permits calculation of the effect of a modification event on mutation frequencies in the

generations subsequent to the event. This information can be helpful in understanding the selective pressures on the modifier.

1.3.2 Accuracy

Any error in the calculation of the mutation frequencies of the unmodified population will lead to a change in frequency of the modifier each generation. As a result, the accuracy of the selection gradient d is limited by the accuracy with which the properties of the unmodified population are calculated. The problem is not serious, and is solved by calculating $\overline{w}$ to a higher degree of accuracy than d.

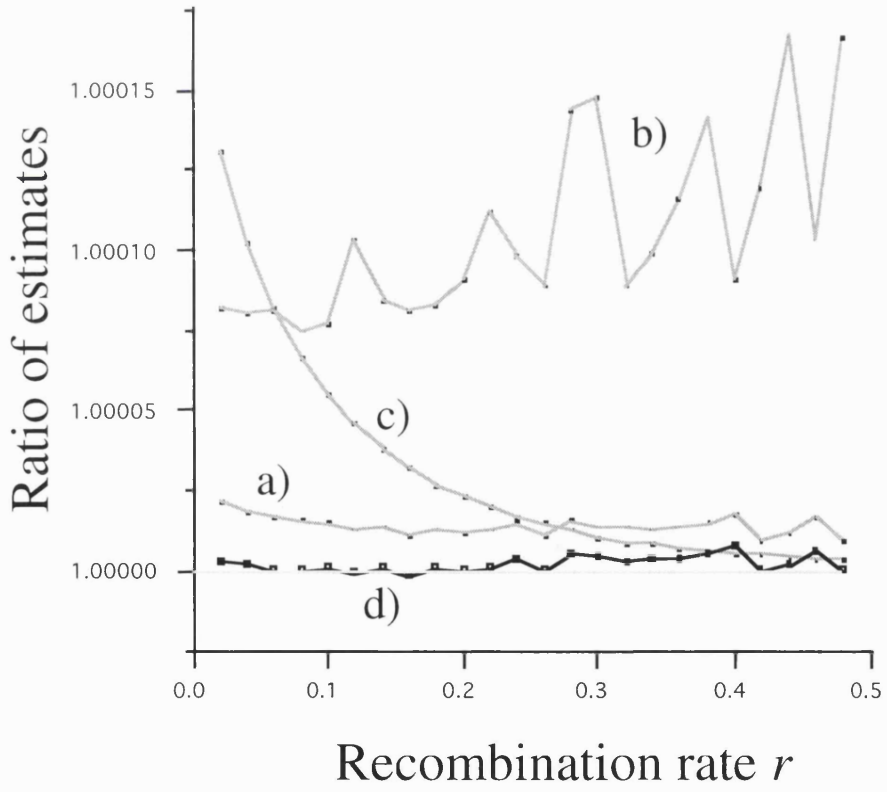
I use a scheme in which $\overline{w}$ is calculated one hundred times more accurately than d. Calculation of the mutation frequencies of the unmodified population is terminated when $\overline{w}$ changes by less than 10^{-14} in each generation. Calculation of modifier mutation frequencies is terminated when the modifier frequency changes by less than 10^{-12} each generation. These numbers are chosen so that each iterative calculation of $\overline{w}$ and d is itself reasonably accurate; the computer used is accurate to approximately one in 10^{16} in each calculation.

The accuracy with which d is calculated can be checked by comparing the estimates of d provided by different values of δr. When the above scheme is used, all errors caused by inaccuracy of calculation should be largest when δr is small. When δr is large, the modifier may cease to behave like a modifier of small effect. In this case, the exact values of d show some dependence on δr. Therefore, d should show least dependence on δr for intermediate values of δr.

Figure 1.1 illustrates the effect of altering δr. The figure shows the ratio of estimates of d for different values of δr. A deviation from 1 indicates a difference in estimates. Calculations were performed assuming $w(n) = 1 - \frac{n}{10}$, U=1, with a linked modifier. The figure shows that values of δr of 10^{-4} and 10^{-5} give very similar estimates of d, differing by less than 0.003% for all values of r (Figure 1.1a). $\delta r = 10^{-6}$ gives a less similar estimate of d, because of inaccuracy caused by early termination of modifier frequency (Figure 1.1b). $\delta r = 10^{-3}$ also gives a less similar estimate of d when linkage is tight (r is small, Figure 1.1c). This is because d shows significant dependence on δr when linkage is tight and δr is large.

It is also possible to check the assumption that modification events in different generations have independent effects. To do this, the fitness effect of a single modification is compared with the fitness effect per modification of modification in ten successive generations. For modifiers of size 10^{-5} the interaction between modifier events in different generations is very small (Figure 1.1d).

Figure 1.1 Test of simulation accuracy
$d(\delta r_1)$ is the selection gradient of a modifier that alters recombination by δr_1. (a), (b) and (c) each represent $\frac{d(\delta r_1)}{d(\delta r_2)}$ for fixed values of δr_1 and δr_2, plotted against the recombination rate of the unmodified population *r*. A deviation from 1 implies a difference in estimates of *d* given by different values of δr. Calculations were performed assuming $w(n) = 1 - \frac{n}{10}$, *U*=1, with a linked modifier.
(a) shows $\frac{d(10^{-4})}{d(10^{-5})}$. (b) shows $\frac{d(10^{-5})}{d(10^{-6})}$. (c) shows $\frac{d(10^{-3})}{d(10^{-4})}$.
(d) shows the estimates of *d* produced when the modifier alters recombination in ten successive generations, divided by the estimate of *d* given when the modifier alters recombination only once. Recombination is modified by $\delta r = 10^{-5}$ in each generation.



1.3.3 Validity

It is important to have some check on validity when using computer simulations to produce results. The best check is a comparison with exact results, where possible, or approximations where not. The values of $\bar{w}$, $\bar{n}$ and V calculated using the program and given in Table 1.1a agree with the exact values which were calculated by Charlesworth (his Table 3). There are two exceptions, which may be typographical errors. A comparison of my modifier results with those of Charlesworth (1990) and Barton (1995) is made in the next section. However, because these results are crucial to the validity of my model, I briefly summarise the relevant points here. Values calculated for the effect of a modifier of small effect differed from the approximate results in a plausible and consistent fashion. The deviation is modest when the mutation rate is low, as expected. Further, there is a consistent pattern to the differences between the three methods; e.g. for $\frac{\delta\bar{n}}{\delta r}$ my results are about $\frac{5}{3}$ as far away from those of Barton as those of Charlesworth are for all mutation rates. Therefore I conclude that my results are valid.

In the absence of analytical results to compare with the results produced by the model when the sexes differ, I have sought to ensure the validity of the results with as many internal consistency checks as possible. The most important checks are the following:

(1) The modifier frequency should stabilise after modified recombination.
The modifier fitness will not stabilise after a modified recombination event unless, after modified recombination, modifier mutation frequencies are calculated in an equivalent way to the frequencies in the unmodified population. So the program will not produce a finite value for the modifier frequency unless both are calculated correctly or, alternatively, both are calculated incorrectly but in an equivalent manner. Therefore, as a matter of good programming practice, I programmed the modifier frequency calculations without reference to the code for calculating the mutation frequencies in the unmodified population.

(2) Consistency with previous versions of the model.
After any new feature of the model is introduced (e.g. two sexes), the model is run without the feature (e.g. with the sexes identical) to ensure that it produces the same results as before.

(3) Mirror image checks.
The model should produce the same results when the parameters for males and females are swapped.

These internal consistency checks do not guard against the possibility of conceptual error. But they make it improbable that the program produce incorrect results simply because of programming error.

1.4 Comparison of three methods

The value of selection on a recombination modifier which reduces the recombination rate from 0.5 to 0.5-δr between all pairs of loci is calculated by both Charlesworth and Barton using their respective approximations. They do not need to specify a recombination mechanism in order to make these calculations. In fact, the mechanism which ensures that all pairs of loci are equally linked is genes-in-a-pot recombination.

The genes-in-a-pot model can be used to calculate values for the unmodified population and selection on recombination modifiers without assumptions about either the distribution of mutations or the strength of selection. The calculations are exact, excepting constraints imposed by the accuracy with which the computer makes calculations (see section 1.3.2 above). These exact values can be compared with the values calculated by Barton and Charlesworth's approximations. The comparison (Table 1.1) is made for standard selection parameters, which means that both selection and epistasis are quite weak.

Barton provides estimates of $\bar{w}$ and $\bar{n}$ which are quite accurate, even when the mutation rate is high (Table 1.1a). The estimates of V-$\bar{n}$ (a measure of linkage disequilibrium) are less accurate, being about 25% too high when U=2. The estimates for selection on the modifier, compared to genes-in-a-pot, are less accurate still (Table 1.1b), d being about 50% too high when U=2. Nevertheless, the estimates are reasonably accurate when U is low.

Table 1.1 Comparison of three methods
Comparison of results obtained using the genes-in-a-pot method with the approximations of Charlesworth and Barton. The fitness of individuals carrying *n* mutations is $w(n) = \exp(-0.002n - 0.0004n^2)$. There is free recombination. The approximations of Charlesworth and Barton are taken from Barton 1995, Table 3. *U* is the mutation rate per diploid individual per generation. (a) shows mean fitness ($\bar{w}$), the mean number of mutations per individual before selection ($\bar{n}$), and the reduction in variance in the number of mutations due to linkage disequilibria (V-$\bar{n}$). (b) shows the selection gradient on the modifier *d*, and the difference in mean ($\frac{\delta\bar{n}}{\delta r}$) and variance ($\frac{\delta V}{\delta r}$) between individuals heterozygous for the modifier and the rest of the population. The genes-in-a-pot modifier calculations are made using a value for δr of -0.00001; Charlesworth uses a value of δr of -0.01; Barton's approximation is valid for modifiers of small effect.

(a)	genes-in-a-pot			Charlesworth (1990)			Barton (1995)		
U	$\bar{w}$	$\bar{n}$	V-$\bar{n}$	$\bar{w}$	$\bar{n}$	V-$\bar{n}$	$\bar{w}$	$\bar{n}$	V-$\bar{n}$
2.0	0.329	50.8	-1.60	0.333	50.6	-1.6	0.343	49.79	1.98
1.5	0.433	43.6	-1.21	0.436	43.4	-1.2	0.445	42.84	-1.47
1.0	0.571	35.1	-0.82	0.571	35.0	-0.8	0.578	34.63	-0.96
0.5	0.752	24.1	-0.41	0.750	24.2	-0.4	0.753	24.03	-0.46
0.1	0.940	9.90	-0.07	0.938	10.1	-0.1	0.938	10.05	-0.08

(b)	genes-in-a-pot			Charlesworth (1990)			Barton (1995)		
U	*d*	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$	*d*	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$	*d*	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$
2.0	0.00181	-0.0337	0.912	0.00234	-0.0443	1.170	0.00271	-0.0490	1.170
1.5	0.00101	-0.0231	0.704	0.00126	-0.0293	0.881	0.00143	-0.0319	0.881
1.0	0.00042	-0.0133	0.484	0.00050	-0.0163	0.589	0.00056	-0.0174	0.587
0.5	0.00007	-0.0051	0.249	0.00007	-0.0059	0.289	0.00008	-0.0062	0.290
0.1	-0.00001	-0.0005	0.047	-0.00001	-0.0005	0.052	-0.00001	-0.0005	0.052

When *U* is high, Charlesworth's estimates for $\bar{w}$, $\bar{n}$ and V-$\bar{n}$ are considerably more accurate than those of Barton (Table 1.1a). Further, the estimates for $\bar{w}$, $\bar{n}$ and V-$\bar{n}$ remain quite accurate when there is no recombination (Table 2 of Charlesworth 1990, not reproduced here), while the approximation of Barton breaks down. However the estimates for modifier parameters do not show a similar improvement over those of Barton (Table 1.1b); the value of *d* is about 3/5 as inaccurate, compared to genes-in-a-pot, as that of Barton for *U*>1, whereas the value of $\bar{w}$ is only about 1/4 as inaccurate.

The value of $\frac{\delta V}{\delta r}$ estimated by the two approximations is virtually identical. The lack of difference between the estimates of $\frac{\delta V}{\delta r}$ contrasts with the significant differences in the estimates for the amount of variance lost through linkage disequilibria (V-$\bar{n}$). Both estimates of $\frac{\delta V}{\delta r}$ are quite inaccurate, compared to the value obtained using genes-in-a-pot, even when U is low.

Charlesworth's estimates are little more accurate than those of Barton because his modifier calculations do not adequately represent the effect of stronger selection. Recall that, to allow for stronger selection, Charlesworth adjusted the weak selection estimates for linkage disequilibria in order to allow for the effect of selection on mutation frequencies. In the modifier calculations, Charlesworth adjusted the estimates for linkage disequilibria of both individuals from the unmodified population and individuals heterozygous for the modifier. However he did not adjust the difference between individuals with and without the modifier. An adjustment is necessary because the effect of modifying recombination depends on the level of linkage disequilibria. The absence of any adjustment explains why Charlesworth's estimates of $\frac{\delta V}{\delta r}$ are very close to those of Barton.

Charlesworth's estimate for the strength of selection on recombination modifiers does not appear to give a substantial improvement over that of Barton. There is little evidence that it is accurate where Barton's approximation breaks down, i.e. for stronger selection or real linkage maps.

This conclusion is unfortunate, because it means that there are no reliable estimates of the strength of selection on realistic recombination modifiers caused by fitness interactions between deleterious mutations. I do not have any immediate solution to this problem. Therefore, throughout the rest of the thesis, many of the estimates of the strength of selection on recombination will be relative; that is to say the strength of male recombination will be compared to the strength of selection on female recombination, and so on.

1.5 Have epistatic interactions between deleterious mutations moulded the recombination rate?

Charlesworth used his approximation to assess the effect of recombination on population mean fitness and of selection on recombination modifiers assuming a fairly realistic genetic map. Calculations were performed for 'standard selection parameters' estimated from a mutation accumulation experiment performed by Mukai (1969) on the fruitfly *Drosophila melanogaster*. The experiment found weak but significant synergistic epistasis with each mutation having a small effect on fitness (as measured by viability in homozygous condition). The estimates for α and β that are obtained are $\alpha = 0.002, \beta = 0.0008$ (Crow 1970; Charlesworth 1990).

Charlesworth found that mean fitness is several percent higher for organisms with many chromosomes than for organisms with few. However the map length of each chromosome has much less effect on mean fitness. The only exception to this occurs if either the map length of each chromosome is very short or the number of chromosomes is small.

Additionally, Charlesworth calculated the strength of selection on recombination modifiers. *Drosophila* has three approximately equal-sized chromosomes and one very small one, with a sex-averaged map length of 1.4 Morgans (2.8 crossovers per haploid genome) (Lindsley and Zimm 1992). When U=1, this gives an estimate for the strength of selection on a modifier that increases the number of crossovers per chromosome by one of about $10^{-3} - 10^{-4}$. Selection of this intensity is probably significant in a species with a large population size such as *Drosophila melanogaster*, but is weak enough to be overwhelmed by other undetectably small factors selecting on crossing over.

Charlesworth's estimates of the strength of selection on recombination modifiers can be criticised on a number of counts:
(1) The estimates ignore variability in the interactions between individual pairs of mutations. Variability in the strength of interactions will reduce selection to increase recombination (Otto and Feldman 1997).
(2) The validity of Mukai's empirical estimates have been strongly questioned, most notably by Keightley (1996). For a recent defense of Mukai's results see Crow (1997).
(3) The accuracy of Charlesworth's theoretical estimates has been shown to be questionable in Section 1.4.

Nevertheless, I will assume that Charlesworth's qualitative conclusions are correct, so that there is weak, but possibly significant selection to increase recombination in *D. melanogaster*. The actual level of recombination that is observed may have arisen because of selection to increase recombination rate caused by epistatic interactions between mutations. Once a certain level of recombination was reached, selection to increase recombination became weak enough that other factors, such as drift, became important. The balance of these factors has determined the overall level of recombination.

Is this conclusion compatible with the hypothesis that epistatic interactions between deleterious mutations have moulded recombination rates in other species?
The map length of *D. melanogaster*, at 1.4 Morgans, is very short compared to other obligately sexual organisms. For example, the map length of humans is 37 Morgans (Dib et al. 1996). The hypothesis that selection has moulded recombination rates implies that this must be because either:

(1) epistasis is weaker in *D. melanogaster* than in other obligately sexual organisms.
(2) the mutation rate is lower in *D. melanogaster* than in other obligately sexual organisms.
(3) both (1) and (2) are true.

In sexual organisms, the mutation rate and the strength of epistasis are connected in two important ways. First, the level of epistasis alters the mutational load. In species with stronger synergistic epistasis than *D. melanogaster*, the mutational load will be lower for a given mutation rate than it would be in *D. melanogaster*. This increases the mutation rate that is consistent with the survival of these species. Secondly, I argue in chapter 4 that the strength of epistasis evolves as a response to the number of deleterious mutations that individuals in the population carry. If this is the case, the strength of epistasis is likely to be higher in organisms with a higher mutation rate. These two factors together make it plausible that other species have a significantly higher mutation rate and significantly stronger synergistic epistasis than *D. melanogaster*. Therefore it is plausible that the longer map lengths of other species reflect much stronger selection to increase recombination.

1.6 Direction of selection on recombination modifiers

The conditions for the invasion of modifiers increasing recombination have been investigated by Feldman et al. (1980), Kondrashov (1984), Charlesworth (1990), Barton (1995) and Otto and Feldman (1997) and reviewed by Feldman et al. (1996). Here I shall summarise their findings.

1.6.1 The three-locus model

Feldman et al. (1980) and Otto and Feldman (1997) analysed a three-locus model. In the model there are two selected loci *A* and *B* and a modifier locus *M* arranged in a linear sequence. In the simpler haploid version of the model the viabilities of the loci *AB*, *Ab*, *aB* and *ab* are 1, 1-*s*, 1-*s* and $(1-s)^2+\varepsilon$ respectively. *s* determines the fitness effect of the *a* and *b* loci acting independently and ε determines the strength of epistasis between them. When $\varepsilon < 0$, selection is synergistic. When $\varepsilon > 0$ diminishing returns apply. In each generation, a fraction μ of the *A* and *B* alleles mutate to *a* and *b* respectively. The *M* locus has two alleles *M* and *m*, which cause different rates of recombination between the *A* and *B* alleles. Recursions can be written for the change in frequency of each of the eight haplotypes arising from the action of mutation, selection and recombination. It is assumed that the allele *M* is fixed, and the conditions for the invasion of *m* are studied. If selection is sufficiently strong, relative to mutation, then the *M* haplotypes attain mutation/selection balance.

When *m* is rare, squared terms for *m* haplotypes can be ignored, making the recursions for *m* linear. The condition for *m* to invade is that the leading eigenvalue of the linearised recursions for the four haplotypes which contain *m* is greater than 1. The size of this eigenvalue was studied by evaluating the characteristic polynomial $f_H(\lambda)$ at $\lambda = 1$. If epistasis is weak relative to selection, the direction of selection on recombination is determined by the sign of epistasis. For synergistic epistasis, recombination is selected to increase, whereas for diminishing returns epistasis ($\varepsilon > 0$), recombination is selected to decrease.

For stronger epistasis, the analysis is more complicated. Recombination reduction is always selected for under diminishing returns epistasis. However, under synergistic epistasis, free recombination is only stable if epistasis is weak relative to selection. Assuming mutation is weak relative to selection free recombination is stable provided $0 > \varepsilon > -\frac{s^2(3-s)}{1-s}$ (Otto and Feldman 1997, Equation 3). Selection on diploid individuals can also be considered; under equivalent assumptions, the condition for the stability of free recombination is the same (Otto and Feldman 1997).

1.6.2 A polygenic approach

Kondrashov (1984) analysed a polygenic model which is, in essence, similar to the pot model presented here. Kondrashov calculated the effect of recombination assuming a realistic crossing over mechanism. However, he assumed at each stage that the linkage disequilibrium between each pair of loci is identical. Thus, although his chromosomes are described as strings, they behave like pots.

Kondrashov considered selection on an unlinked recombination modifier. As was the case for the three-locus model, he found that free recombination was stable to invasion by a modifier reducing recombination, provided there was weak synergistic epistasis. For stronger synergistic epistasis, the direction of selection depends on the mutation rate. For a low mutation rate, there is selection for reduced recombination when r=0.5. For a high mutation rate, free recombination is stable. Kondrashov presented an argument which suggests that the direction of selection on recombination might depend on the mutational load. He considered the case of truncation selection. Truncation selection implies that individuals with less than T mutations have equal fitness, and individuals with T or more mutations had a fitness of zero. Kondrashov pointed out that the immediate effect of a modification event on the frequency of the modifier would depend on mean fitness. Assuming that the distribution of mutations is normal, a modifier would suffer an immediate decrease in frequency if the load was less than 0.5. Otherwise, the modifier would immediately increase in frequency. Simulation results (Kondrashov 1984, and genes-in-a-pot, not presented) have confirmed that, in the extreme case of truncation selection, free recombination is stable provided $\overline{w} < 0.7$. Under standard selection parameters, the comparable figure is $\overline{w} < 0.85$.

1.6.3 Reconciliation of the three-locus model with polygenic approaches

Otto and Feldman (1997) reconciled the results of the three-locus model with the results obtained using polygenic approaches. In the three-locus model the direction of selection on recombination shows little dependence on the mutation rate. Instead, it is dependent on the strength of epistasis relative to the strength of selection. This contrasts with Kondrashov's observation that the mutation rate is critical to the direction of selection on recombination when epistasis is strong. Nevertheless, Otto and Feldman were able to relate the selection on recombination in a polygenic setting to the results of the three-locus model. They defined "apparent epistasis" and "apparent selection" according to the marginal effects on fitness of increasing the mutation number by one. Defined in this way, apparent epistasis reduces, relative to apparent

selection, as the mutation number of individuals increases. Therefore, for high mutation rates, apparent epistasis is weak, relative to apparent selection, and free recombination is evolutionarily stable.

1.6.4 The QLE method

Barton (1995) used the QLE method to investigate the conditions under which modifiers of recombination are selected to increase. He derived an expression (1995, Equation 18) for the selection on a modifier increasing recombination which is valid provided that selection and epistasis are both weak, and that the modifier has a small effect on recombination. The expression has two terms. The first refers to the immediate effect of modified recombination on fitness. Recombination breaks up combinations of genes that have survived selection in previous generations. Therefore, the resulting combinations of genes are less fit than the original ones. This term always selects against the modifier. The second term refers to the effect of recombination on additive genetic variance. The sign of this term depends on epistasis. Under diminishing returns epistasis, the modifier reduces additive genetic variance, and the second term is also negative. Under synergistic epistasis, a modifier allele increasing recombination increases variation and becomes associated with genotypes with very high fitness. Therefore, the second term is positive in this case.

The magnitude of the second term depends on the linkage of the modifier to the loci whose linkage disequilibria are reduced. For this reason, the first term becomes relatively larger, compared to the second, when linkage is loose. The relative size of the two terms also depends on the apparent strength of epistasis; when apparent epistasis is weak, the first term is small. Therefore, for high mutation rates, only the second term is important.

The QLE method differs from the others considered here in that it does not assume a fitness function. The assumption of a fitness function is unrealistic in two important respects. First, a fitness function entails all mutations having an equal effect on fitness. Secondly, a fitness function entails that the strength of epistasis between all pairs of loci is the same. Otto and Feldman (1997) applied Barton's expression in order to assess the effect of variable epistasis on selection on recombination. They found that variability in the size of the pairwise epistasis between mutations contributes to the size of the first term, and therefore selects against recombination. Variability in epistasis will be particularly important to selection on recombination if the mutation rate is low. Empirical evidence on the variability of epistasis is discussed in Section 4.2.1.

1.6.5 Summary

In a polygenic model, when the mutation rate is high, the direction of selection on recombination is largely determined by the sign of average epistasis. When average epistasis is negative (synergism), there is selection for increased recombination. When average epistasis is positive (diminishing returns), there is selection for decreased recombination.

For lower mutation rates, the conditions for selection on increased recombination are more stringent. Free recombination is only stable if average epistasis is weakly synergistic and the variance in epistasis is low. However, for a tightly linked modifier, modifiers increasing recombination will invade provided average epistasis is synergistic. Therefore, when the mutation rate is low, there may be an intermediate recombination rate between modifier and selected loci which ensures that modified recombination is selectively neutral. In this case, a modification event causes an immediate decrease in frequency of the modifier. However, the increase in the additive genetic variance in fitness caused by the event ensures that in later generations the modifier becomes associated with genotypes with fewer mutations than average. The frequency of the modifier increases, so that, overall, the modifier undergoes no change in frequency.

Chapter 2

Sex differences in recombination

Abstract

Trivers (1988) speculated that sex differences in recombination rates are caused by sex differences in selection. I investigate the theoretical basis for this claim, working on the assumption that selection to increase recombination results from synergistic interactions between deleterious mutations. I find that there is stronger selection to increase recombination in the sex in which selection on deleterious mutations is weaker. The magnitude of the effect depends on the sex difference in fitness effect of each mutation. In consequence, selection for sex differences in recombination can only be significant if the mutations responsible for selection on recombination each have a large effect.

2.0 Introduction

This chapter is chiefly concerned with one theoretical explanation for sex differences in recombination. I apply my genes-in-a-pot model in order to assess whether sex differences in selection on deleterious mutations can explain sex differences in recombination. In order to set this explanation in context, I first review the empirical evidence for sex differences in recombination (Section 2.1). Next, I review the existing theoretical explanations (Section 2.2). At the end of the section (2.2.5), I describe the questions my model results are intended to address. Section 2.3 contains the mathematical preliminaries necessary to interpret the model results. With these preliminaries completed, I am finally ready to present my model results (Section 2.4). The results show that selection for sex differences in recombination can only be significant under certain biological conditions. Section 2.5 reviews the empirical data relevant to the question of whether or not these conditions are actually fulfilled in nature.

2.1 Evidence concerning sex differences in recombination

Data on sex differences in recombination comes from two sources. The first source, chiasma data, is obtained by counting the number of chiasmata on meiotic chromosomes. Chiasmata hold the chromosomes together and are thought to be the physical manifestations of crossovers (Tease and Jones 1978). For the purposes of sex difference studies, the biggest limitation on the quality and reliability of chiasma data is that it is hard to obtain good meiotic preparations from oocytes in many species, especially mammals, due to the large size of the egg and the narrow developmental window within which each oocyte yields a good preparation (M. Hulten, G. Jones pers. com.). As a result, the literature contains about 20 times as many reports of chiasma data in males as in females (Burt, Bell, and Harvey 1991). Additionally, the chiasma data from females may be less accurate than that obtained from males, because the quality of the available meiotic preparations is typically poorer. Poor preparations may result in the miscounting of chiasmata. This inaccuracy is hard to quantify.

The second source, genetic data, is becoming rapidly more plentiful. In order to make genetic maps, the recombination rates between markers must be recorded. Enlightened authors report male and female recombination rates between each pair of markers separately. Currently, extensive genetic maps are only available for a handful of higher organisms. There are a large number of species for which less extensive maps exist. However, making an objective survey of this data would be difficult because of the great variation in the quantity of data available for each species. Objectivity is important because data from a few loci does not necessarily give a good indication of the overall sex difference in recombination. In species where extensive genetic maps do exist there

is pronounced variation in the level of sex differences in recombination across the chromosome. For example, in humans, the male recombination rate is higher than the female recombination rate close to the telomeres, but is much lower than the female rate in the interior of the chromosome (Fain et al. 1996; Hultén et al. 1978).

The level of variability observed in the sex difference in recombination between loci makes any conclusions based on a few loci unreliable. For example, Dunn and Bennett (1967) surveyed the available genetic data from the mouse, *Mus musculus,* and concluded that the direction of the sex difference in crossing over differed according to chromosome. This conclusion is contradicted by a study of chiasma distribution in mouse chromosomes performed by Lawrie et al. (1995) who found the same pattern observed in human chromosomes; more chiasmata near to the telomeres in male meiosis, but higher recombination overall in females meiosis. The same pattern is observed for every chromosome. The source of the discrepancy between the conclusions of the two studies is that the loci surveyed by Dunn and Bennett do not span a large enough proportion of the chromosome to give a reliable indication of the overall level of crossing over on each chromosome. This conclusion is confirmed by the more extensive mouse genetic maps currently available, which show similar sex differences in the number of crossovers on each chromosome (Dietrich et al. 1996).

Trivers (1988) made two claims about empirical sex differences in recombination. First, he claimed that the recombination rate tends to be higher in males than in females. Secondly, he claimed that there tends to be more recombination in the heterogametic sex. Trivers provided several examples and counterexamples for both generalisations. For the reasons given above, Trivers's genetic evidence difficult to evaluate. I will summarise of the findings of a comprehensive review of the chiasma data performed by Burt et al. (1991), compiled from chiasma data from two sexes for 54 species (excluding humans). Their principal findings were as follows:

(1) Despite pronounced and consistent sex differences in chiasma frequency in many species, there is no overall sexual dimorphism in recombination rates in animals. There is no overall trend in any of the animal taxa in which both sexes have a chiasmate meiosis for either sex to have more chiasmata.

(2) In species with sex chromosomes, only the heterogametic sex is ever achiasmate. However, in species where both sexes have a chiasmate meiosis, there is no association evident between chiasma number and gamety. There are hermaphroditic species which are achiasmate in one sex.

(3) Within each taxon, the male and female chiasmata frequencies of each species are correlated.

2.2 Possible explanations for sex differences in recombination

2.2.1 Gamety

Haldane (1922) proposed that recombination tends to be reduced in the heterogametic sex, as a result of selection for reduced recombination in the sex chromosomes. Gamety has clearly played some role in the evolution of sex-specific recombination. At the very least, selection to maintain recombination on the sex chromosomes of the homogametic sex has prevented the evolution of achiasmate meiosis in that sex in any species studied. However, in species where both sexes show crossing over, no overall association with gamety is evident. Therefore the circumstances in which gamety is important to sex differences in recombination seem to be relatively infrequent.

2.2.2 Physiological factors

Meiosis proceeds quite differently in males and females, so that it is possible that there may be some sex-specific cost or benefit associated with crossing over. For example, in humans the level of crossing over has been found to influence the probability of nondisjunction in females (Koehler et al. 1996). Another possibility is that sex differences in the level of crossing over occur as a result of differences in the amount of double-stranded DNA damage (Bernstein, Hopf, and Michod 1988). However, the absence of consistent sex differences in crossing over in any of the animal taxa surveyed by Burt et al. (1991) implies that there are no consistent and selectively important sex specific factors of this kind in these taxa.

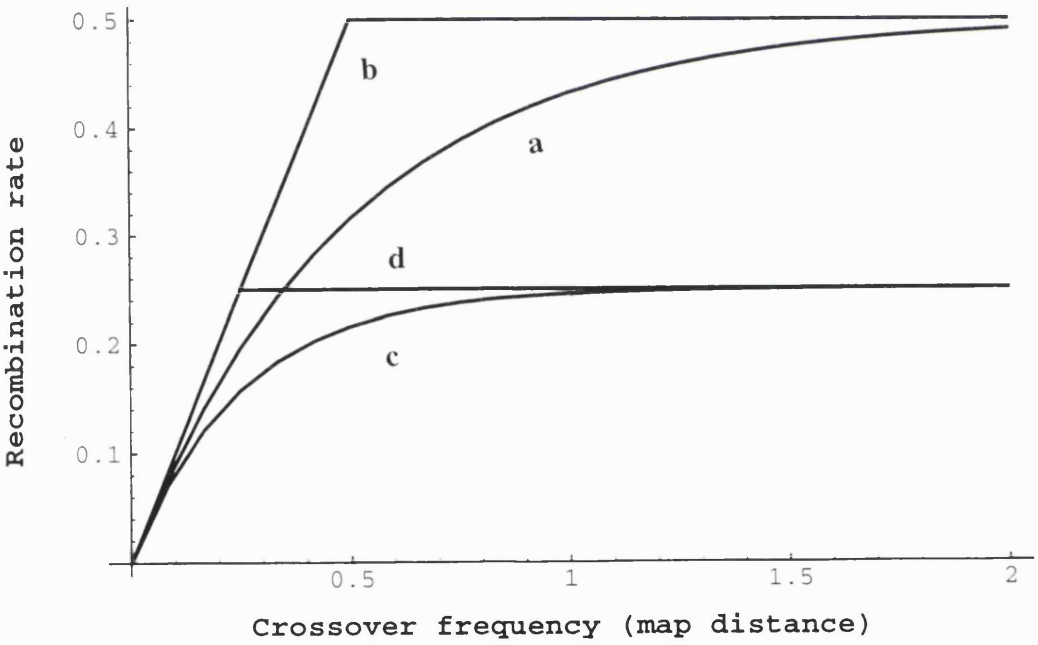
2.2.2.1 Diminishing returns of chiasmata

The relationship between chiasma frequency and recombination rate is non linear. After the first chiasma, additional chiasmata give diminishing returns, because they may negate the recombinational effect of the first. If two crossovers occur on the same strand of DNA, then the loci flanking the two crossovers are not recombined. The relationship between the chiasma frequency in an interval and the recombination rate in that interval depends on the strength of chiasma interference (Figure 2.1a,b). I discuss chiasma interference in more detail in Chapter 5.

Sex differences in recombination reduce the level of recombination achieved by chiasmata. For example, if recombination is restricted to one sex, then the recombination rate between genes cannot exceed 0.25 (Figure 2.1c,d). Clearly, the diminishing returns of chiasmata may constrain the sex difference in recombination that

evolves. If there is selection for a high recombination rate then this favours recombination in both sexes. This selection will constrain evolution of recombination in response to selection for sex differences in recombination.

Figure 2.1 Relationship between the recombination rate of a pair of genes and the frequency of crossing over between them.
Maximal chiasma interference implies that chiasmata are always the same genetic distance apart. No interference implies that chiasmata are randomly positioned with respect to each other.
(a) no chiasma interference, with the sexes identical.
(b) maximal chiasma interference, with the sexes identical.
(c) no chiasma interference, with crossing over restricted to one sex.
(d) maximal chiasma interference, with crossing over restricted to one sex.
The figure shows that the average recombination rate between genes for a given frequency of crossing over depends on the strength of interference and the sex difference in recombination. Under maximal chiasma interference, the average recombination rate between genes is the same regardless of the sex difference in recombination, provided r<0.25. Above r=0.25 the recombination rate is higher if the sexes are identical. In the absence of interference, the recombination rate is always higher if the sexes are identical.



Diminishing returns may be particularly important if crossing over is costly. Evidence that crossing over may indeed be costly comes from an experiment performed by Magni and Borstel (1962). Magni and Borstel found that the mutation rate in meiosis was between six and twenty times higher than the mutation rate in mitosis. The most likely explanation is that the double-strand breaks involved in crossover formation expose the DNA to damage, some of which cannot be repaired. If this is the case, crossing over imposes a mutagenic cost in yeast. It there are sex differences in

crossing over, then more chiasmata will be needed to achieve the same level of recombination. So large sex differences are unlikely to evolve if crossing over is costly.

2.2.3 Drift

It is possible that genetic drift affects sex differences in recombination. For example, the hypothesis that the direction of the sex difference in recombination is determined by drift is consistent with the pattern observed by Burt et al. (1991)

A stronger role for drift is suggested by the hypothesis that selection determines only the average chiasmata frequency, whilst the proportion of the chiasmata which occur in each sex is determined by drift. As drift is neutral, all proportions of sex bias are equally likely, including all crossing over in males or all crossing over in females (Burt, Bell and Harvey 1991). Such extreme biases are impossible because at least one chiasmata per chromosome is necessary to ensure nondisjunction. Burt et al. attempt to allow for this by considering excess chiasma frequency (the number of chiasmata per meiosis minus the haploid chromosome number). This adjustment is only adequate provided that interference is maximal, so that one chiasma per chromosome, on average, is enough to ensure disjunction (see discussion of interference in Section 2.4.3 and in Section 5.6). Despite the adjustment, the sex difference in crossing over that is observed is smaller than expected under the hypothesis (Burt et al. 1991). One explanation would be the diminishing returns of chiasmata, as discussed in 2.2.2.1 above. Alternatively, male and female recombination rates may be similar because of mutational bias in modifiers of recombination. Males and females share the same genes, so that most mutations are likely to affect traits similarly in males and females (Rice 1984).

For example, Halliday and Arnold (1987) have suggested that high levels of remating are of no direct benefit to females but instead persist because of selection for high levels of mating in males. Concurring with this, a genetic correlation has been found between male and female mating speed in *Drosophila* (Stamenkovic-Radek, Partridge, and Andjelkovic 1992).

The existence of sexually antagonistic genes, which increase fitness in one sex but decrease fitness in the other, suggests that similarities between the sexes may persist in evolutionary time despite being maladaptive. These genes are likely to be common if mutations for sex differences are rare. To give a simple example, one can imagine a species in which the optimal size for males is different from the optimal size for females. In the absence of genetic variation for sex differences, both sexes are likely to

evolve to be an intermediate size. Mutations with a similar effect on size in males and females will be sexually antagonistic.

Rice (1992, 1996) found evidence for the occurrence of sexually antagonistic genes. In a first experiment, he used genetic markers to ensure that autosomal chromosomes segregated as if they were sex chromosomes determining females(Rice 1992). After the chromosome had been restricted to females for a number of generations, males with the chromosome had decreased fitness compared to controls. In a second experiment, Rice (1996) relaxed selection on female fitness by ensuring that the female genetic material was cycled out of the experimental line after being used for a single generation. He found that male fitness increased compared to controls. This experiment can be interpreted as showing that, before the experiment began, the phenotype of males was sub-optimal because their genes had undergone selection as part of the female genotype in previous generations.

2.2.4 Sex differences in selection

Trivers (1988) proposed that sex differences in recombination arise from sex differences in selection. Trivers's contention is that selection may act more strongly in one sex than in the other. So, if, for example, selection is stronger in males then: "the genes and combinations of genes being passed in males will be superior on average, compared to genes passed in females" (Trivers 1988). In this case,"[i]nsofar as the actual combinations in which a male's genes appear are important to their success, then he will be selected to reduce rates of recombination (compared to females) in order to preserve these beneficial combinations."

Trivers's contention is hard to evaluate because he does not specify the nature of the genetic variation which might lead to one combination of genes being better than another. A contrary interpretation of Trivers's contention was made by Burt et al. (1991):

> Many theorists believe that the main function of recombination is to reduce linkage disequilibrium. As two potentially important sources of linkage disequilibrium are selection and drift, one might expect that the sex experiencing the more intense selection, or otherwise having the higher variance in reproductive success, should have more recombination. This prediction is exactly opposite to that made by Trivers.

Burt et al.'s argument is somewhat more explicit than that of Trivers. But they do not spell out a mechanism by which a sex difference in the level of linkage disequilibrium might lead to a sex difference in recombination.

2.2.5 The aim of this chapter

The aim of this chapter is to establish whether sex differences in selection can, in fact, cause sex differences in recombination. The verbal arguments of Trivers (1988) and Burt et al. (1991) are not adequate. Further progress is dependent on explicit models of the evolution of recombination. I investigate sex differences in selection on recombination modifiers. As in Chapter 1, I assume that there are epistatic fitness interactions between deleterious mutations which occur throughout the genome. In each generation, selection creates linkage disequilibria because of these interactions. These disequilibria are broken down by recombination. When interactions are predominantly synergistic, modifiers increasing recombination can invade under a set of conditions which are described in Section 1.6. Under these conditions, the function of recombination , or, to be more precise, the factor which causes selection to maintain recombination, is the reduction of linkage disequilibrium. Thus the model is of the type alluded to by Burt et al. (1991).

An accurate representation of the effect of selection in each generation is obviously critical to the analysis. I therefore use the genes-in-a-pot simulation method described in Chapter 1. The method assumes a simplified recombination mechanism, but accurately represents the effect of selection on the distribution of mutations in diploid individuals.

I wish to establish whether there is significant selection for sex differences in recombination. If so, I would like to establish the size of expected sex differences in recombination. I divide my analysis into three sections:

(i) Is there a sex difference in the effect of recombination on linkage disequilibria?

(ii) If there is a sex difference in the effect of recombination, can this account for sex differences in the strength of selection on recombination modifiers?

(iii) How large are predicted sex differences in recombination ?

Once these questions have been answered, it should be possible to derive empirical tests to establish whether sex differences in recombination are indeed caused by sex differences in selection on deleterious mutations.

2.3 Preamble to model results

2.3.1 Notation incorporating sex differences

Before presenting model results, it is necessary to introduce some additional notation from Chapter 1. Differences between the sexes entail differences between maternal and paternal haplonts. I introduce a notation which allows for this possibility. $\overline{M}$ and $\overline{P}$ are the mean number of mutations in the maternal and paternal haplonts. V_M and V_P are the variance in mutation number of the haplonts. In the absence of linkage disequilibria, $V_M = \overline{M}$. $D_M = V_M - \overline{M}$ is the additional variance in mutation number due to linkage disequilibria of the maternal haplont. Under synergistic epistasis D_M is negative at mutation /selection balance. $D_P = V_P - \overline{P}$ is the corresponding quantity for the paternal haplont. Cov_{MP} is the covariance between maternal and paternal haplonts. Other symbols are retained from previous chapters: $V = V_M + V_P$, $\bar{n} = \overline{M} + \overline{P}$, $V - \bar{n} = D_M + D_P$.

The notation is somewhat less specific than in Charlesworth (1990) and Chapter 1. Charlesworth refers to the size of each quantity before selection. Here I am interested in describing the change in the distribution of mutation number caused by selection. Therefore I use the notation in a more flexible manner. The quantity X is followed by a superscript b when measured before selection, and superscript a when measured after. I define $\Delta X = X^a - X^b$. Note that this is a different use of Δ to that of Charlesworth, who refers to the change in quantities as a result of the combined effects of mutation, selection and recombination. Before selection, males and females are identical. After selection males and females may differ. A quantity X is followed by a subscript m when measured in males, and a subscript f when measured in females.

Modifier notation is adapted straightforwardly from the previous chapter. d_m is the gradient of selection on a modifier of recombination in males and d_f is the equivalent gradient in females. $d_m + d_f = d$. Other quantities are adapted analogously, e.g. $(\frac{\delta \bar{n}}{\delta r})_m$ is the difference in mean mutation number between individuals containing a male modifier and the unmodified population, divided by the amount by which the modifier alters recombination in every generation.

2.3.2 The effect of recombination on the variance in mutation number

When genes-in-a-pot recombination is assumed, the only information required to describe a pot is the number of mutations it contains. Therefore, the effect of recombination can be described by its effect on the distribution of mutations.

Charlesworth (1990) found that the distribution of mutations remained close to normal, provided selection was fairly weak, even when r=0. This finding implies that most of the linkage disequilibria which build up are attributable to a reduction in variance in mutation number.

Genes-in-a-pot recombination when R<1/2 produces two types of gamete. The first type receives most of its genetic material from its grandmother (i.e. the mother of the individual whose chromosomes are undergoing recombination). I shall refer to the number of mutations in this pot as $X1$. The second type has $X2$ mutations and receives its genetic material primarily from its grandfather.

The mean number of mutations of a proportion R of a haplont X, which may be either maternal or paternal, is

$$\mathrm{E}\left(X_i B(i,R)\right) = R\overline{X}. \tag{2.1}$$

where $B(i,R)$ is from a binomial distribution with mean Ri and variance $R(1-R)i$.

The variance in number of mutations of R of a haplont is

$$E\left((X_i B(i,R))^2\right) - R^2\overline{X}^2 = R(1-R)\overline{X} + R^2\sum_i i^2 X_i - R^2\overline{X}^2$$
$$= RV_X - R(1-R)D_X \tag{2.2}$$

Therefore the means and variances of X1 and X2 are

$$\overline{X1} = R\overline{P} + (1-R)\overline{M},$$
$$\overline{X2} = (1-R)\overline{P} + R\overline{M},$$
$$V_{X1} = RV_P + (1-R)V_M + R(1-R)(2Cov_{MP} - D_P - D_M),$$
$$V_{X2} = (1-R)V_P + RV_M + R(1-R)(2Cov_{MP} - D_M - D_P).$$

The variance in mutation number of the gametes produced by genes-in-a-pot recombination is:

$$\mathrm{Var}\left(\frac{1}{2}\text{ of X1 and }\frac{1}{2}\text{ of X2}\right)$$
$$= \mathrm{E}\left[\left(\frac{1}{2}X1 + \frac{1}{2}X2\right)^2\right] - \left(\frac{1}{2}\overline{X1} + \frac{1}{2}\overline{X2}\right)^2$$
$$= \frac{1}{2}V_{X1} + \frac{1}{2}V_{X2} + \frac{1}{4}\left(\overline{X1} - \overline{X2}\right)^2$$
$$= \frac{1}{2}(V_P + V_M) + \frac{1}{4}(2Cov_{MP} - D_P - D_M) + \frac{1}{4}(1-2R)^2\left[(\overline{P} - \overline{M})^2 - (2Cov_{MP} - D_P - D_M)\right]$$

Therefore recombination increases the variance in mutation number of recombined gametes, compared to unrecombined gametes by

$$\frac{1}{2}r\left[2Cov_{MP} - D_P - D_M - (\overline{M} - \overline{P})^2\right] \qquad (2.3).$$

The amount by which recombination can increase the variance is crucial to my analysis, so I will represent $\left[2Cov_{MP} - D_P - D_M - (\overline{M} - \overline{P})^2\right]$, its potential to do so, by the symbol $\Re$.

2.4 Model results

In this chapter, I only consider cases where the mean number of mutations in maternal and paternal gametes is approximately the same [$(M-P)^2 << D_P + D_M$]. This would be true if either (1) sex differences in the strength of selection and the mutation rate are both modest, or (2) selection and mutation occur predominantly in the same sex. In the latter case sex differences in selection and mutation have counterbalancing effects and therefore create little difference in mutation number between maternal and paternal gametes. The effect of differences between maternal and paternal gametes is discussed in Chapter 3.

In the following sections, I describe how my model results answer the three questions asked in Section 2.2.4.

2.4.1 Is there a sex difference in the effect of recombination on linkage disequilibria?

Recombination reduces linkage disequilibria by $\frac{1}{2}r\Re^a$. Before selection, males and females have the same genotype, so $\Re_m^b = \Re_f^b$. Therefore, sex differences in the effect of recombination must arise through a difference between $\Delta\Re_m$ and $\Delta\Re_f$.
I ignore sex differences for a moment, and describe the factors influencing the magnitude of $\Delta\Re$ when the sexes are identical.

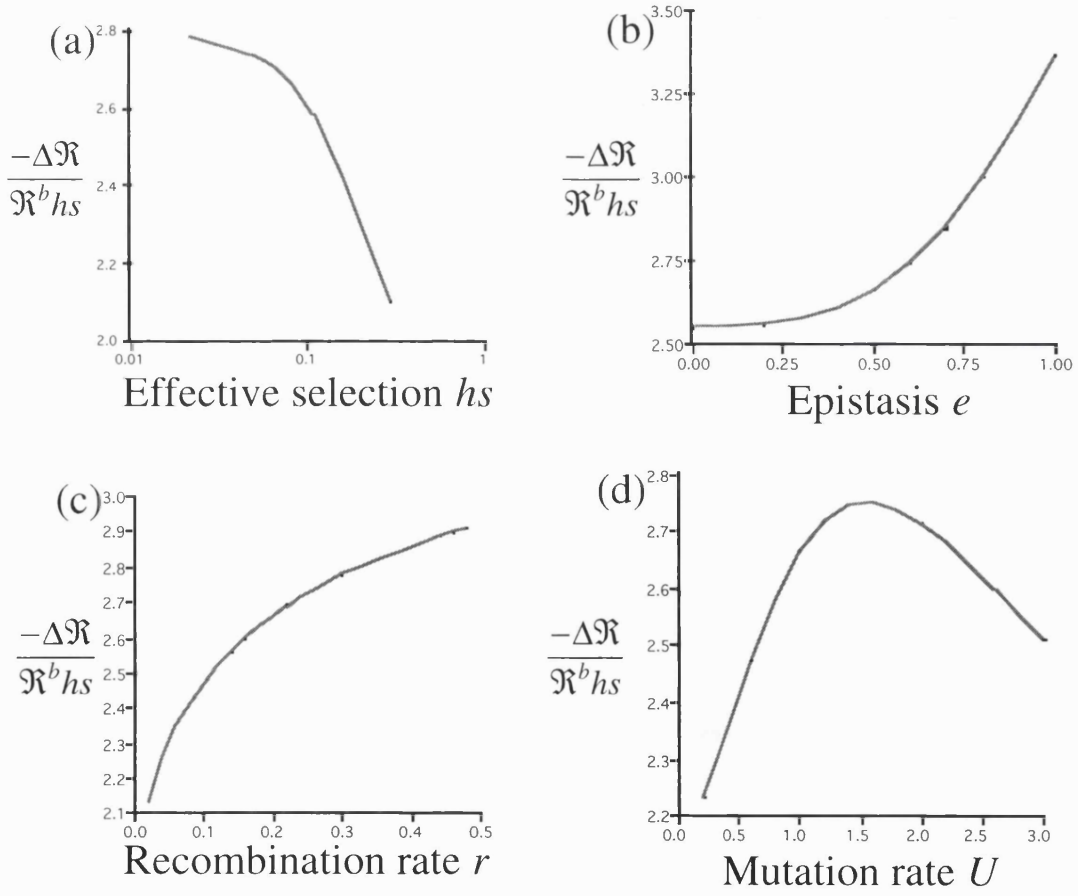
Before selection acts there are no repulsion linkage disequilibria ($Cov_{MP}^b = 0$), because individuals are assumed to be formed by the random pairing of male and female gametes. So $\Re^b = -D_M^b - D_P^b$. Selection creates both coupling and repulsion linkage disequilibria, which have opposing effects on $\Delta\Re$. Recombination breaks up coupling linkage disequilibria, but converts repulsion linkage disequilibria to linkage disequilibria within a haplont. $\Delta\Re$ depends on the difference between these two effects.

In the absence of preexisting genetic associations, selection will create an equal quantity of coupling and repulsion linkage disequilibria, because the effect of selection is determined by the number of mutations, regardless of the haplont within which they occur. In this case, $\Delta\Re = 0$, and recombination has no effect. However, if there are preexisting genetic associations, then these linkage disequilibria will determine the subsequent effects of selection. As stated in Section 1.2.1.1, Charlesworth (1990) argued that selection has two opposing effects on linkage disequilibria, increasing linkage disequilibria by changing V and decreasing them by reducing $\bar{n}$. The first of these effects is independent of preexisting linkage disequilibria and therefore applies to both coupling and repulsion disequilibria, with no net effect on $\Delta\Re$. However, the second effect is dependent on preexisting disequilibria and therefore only applies to coupling. Consequently $\Delta\Re$ is determined by the reduction in $\bar{n}$.

Charlesworth (1990) argued that preexisting linkage disequilibria will be reduced by a proportion which is approximately equal to twice the effective selection on each mutation. The effective selection hs is defined as $\frac{-\Delta\bar{n}}{\bar{n}^b}$. The rationale for Charlesworth's argument is that if linkage disequilibria are predominantly pairwise, reducing the mutation frequencies at each locus by a factor 1-hs would reduce linkage disequilibrium by $(1-hs)^2 \cong 1-2hs$. Given that $\Delta\Re$ is determined by the reduction in $\bar{n}$, this gives $\Delta\Re \cong -2hs\Re^b$, so that $\frac{-\Delta\Re}{\Re^b hs} \cong 2$.

In Figure 2.2 I test whether this estimate of $\Delta\Re$ is robust to changes in the level of epistasis, the level of linkage disequilibria and the selection on each mutation. In these graphs the quantity $\frac{-\Delta\Re}{\Re^b hs}$ is plotted. For all of the parameter values shown $2.0<\frac{-\Delta\Re}{\Re^b hs}<3.5$. Therefore the estimate is reasonably accurate over all parameter values.

Figure 2.2 Test of the accuracy of the estimate of $\Delta\Re$: $\Delta\Re \cong -2hs\Re$. The quantity $\frac{-\Delta\Re}{\Re^b hs}$ is plotted for a variety of parameter values. The sexes are identical. The estimate implies $\frac{-\Delta\Re}{\Re^b hs}$=2. Here I use the fitness function $w(n) = 1 - \left(\frac{n}{T}\right)^{1/(1-e)}$, where n is the number of mutations each individual contains. T determines the fitness effect of each mutation. Individuals with T or more mutations have zero fitness. e determines the strength of epistasis. When e=0, there is additive epistasis, a weak form of synergistic epistasis. When e=1 there is truncation selection. Truncation selection implies that all individuals with less than T mutations have an equal fitness of unity, whereas individuals with T or mutations have zero fitness. Truncation selection is the strongest form of synergistic epistasis. This fitness function incorporates stronger epistasis than the Gaussian fitness function and is more convenient in a simulation framework. Each graph incorporates T=40, e=0.5, U=1.0, r=0.2. In (a) T is varied. $\frac{-\Delta\Re}{\Re^b hs}$ is plotted against the effective selection on each mutation hs. In (b) the strength of epistasis e is varied. In (c) the recombination rate r is varied. In (d) the mutation rate U is varied.



In detail, the figure illustrates:

(a) *hs* varies by about a factor of 10, whereas $\frac{-\Delta\Re}{\Re^b hs}$ varies by about a quarter.
Therefore the effect of selection on $\Re$ is small if the effective selection against each mutation is small.

(b) $\frac{-\Delta\Re}{\Re^b hs}$ is higher for stronger epistasis. So $\Re$ is more strongly affected if selection is strongly synergistic than if selection is weakly synergistic, provided *hs* is kept constant.

(c) $\frac{-\Delta\Re}{\Re^b hs}$ is close to 2 when linkage between all loci is tight (*r* is small). So the estimate of $\Delta\Re$ remains good when there are substantial linkage disequilibria.

(d) $\frac{-\Delta\Re}{\Re^b hs}$ does not show linear dependence on *U*.

The investigation of the effect of selection on $\Re$ leads me to the following conclusions:

(1) Selection reduces the effect of recombination on linkage disequilibrium. Selection always increases the linkage disequilibria between haplonts (repulsion) by more than it increases the linkage disequilibria within a haplont (coupling). Repulsion linkage disequilibria reduce the effect of recombination.

(2) The proportion by which selection reduces $\Re$ is approximately twice the effective selection against each mutation. The amount by which selection reduces $\Re$ is larger if epistasis is strong. $\Re$ is reduced by very little if selection is weak, regardless of the strength of epistasis.

In this section, I have attempted to establish whether there is a sex difference in the effect of recombination on linkage disequilibria. I have demonstrated that $\Delta\Re$ is dependent on a reduction in $\bar{n}$. Consequently, sex differences in $\Delta\Re$ will depend on sex differences in the effective selection. One can go further and say that, provided epistasis is similar in males and females, the sex difference in the effect of recombination on linkage disequilibria after selection $\frac{\Re_f - \Re_m}{\Re^b}$ should approximately

be equal to minus 2 times the sex difference in the effective selection on each mutation, $-2(hs_f - hs_m)$.

So $\Re_f - \Re_m \cong -2(hs_f - hs_m)\Re^b$ (2.4)

2.4.2 Can the sex difference in the effect of recombination account for sex differences in the strength of selection on recombination modifiers?

The genes which descend from males have a very similar future to the genes which descend from females. Both have a 50% chance of being passed to a male and a 50% chance of being passed to a female. Consequently, sex differences in the strength of selection on modifiers should depend solely on sex differences in the effects of recombination on linkage disequilibria. Most linkage disequilibria are attributable to effects on the variance (Charlesworth 1990; Barton 1995). Therefore, sex differences in selection on recombination modifiers should be determined by the sex difference in the effect of recombination on the variance. In other words, $\frac{d_m}{\Re_m} - \frac{d_f}{\Re_f} \cong 0$ (2.5).

Under this assumption, Equation 2.4 can be used to estimate the sex difference in selection on recombination modifiers. Assuming $hs_m << 1$ gives an estimate for $\Re^b$ of $\frac{d\Re_m}{2d_m}$. Combining Equation (2.4) and Equation (2.5), and using this estimate for $\Re^b$ gives: $d_m - d_f \cong -d(hs_m - hs_f)$ (2.6).

These expectations are tested for one set of parameters in table 2.1. The table shows sex differences in selection on recombination , $\frac{d_f}{d_m}$, and the sex difference in the effect of the modifier on the variance in mutation number, $\frac{(\frac{\delta V}{\delta r})_f}{(\frac{\delta V}{\delta r})_m}$. The two values are almost identical, showing that sex differences in selection on recombination are caused by sex differences in the effect of recombination on the variance. The absolute sex difference in the gradient of selection of recombination modifiers ($d_m - d_f$) is also compared with the value derived from the approximation $d_m - d_f \cong -d(hs_m - hs_f)$. The approximation is extremely good for the values shown.

Table 2.1 Sex difference in selection on recombination modifiers
The fitness of males carrying n mutations is $w(n) = \exp(-0.002n - 0.0004n^2)$. There is no selection or mutation in females. U_m is the mutation rate per male individual per generation.
(a) Properties of the unmodified population; the mean fitness of males, $\bar{w}_m$, the mean number of mutations per individual before selection, $\bar{n}$, and the reduction in variance due to linkage disequilibrium before selection, V-$\bar{n}$.
(b) Properties of the modified population given modification in one sex (the male); the selection gradient on a modifier which alters male recombination, d_m, the difference in mean , $(\frac{\delta \bar{n}}{\delta r})_m$, and variance, $(\frac{\delta V}{\delta r})_m$, in mutation number between individuals heterozygous for a modifier of male recombination and individuals from the unmodified population.
(c) Comparison of selection on modifiers of recombination in males with modifiers of recombination in females; the sex difference in selection on recombination modifiers, $d_m - d_f$, the prediction for the sex difference in selection on recombination given by Equation 2.6, $-d(hs_m - hs_f)$, and the ratio of the three quantities in (b) for a modifier of female recombination and a modifier of male recombination.

(a)

U_m	$\bar{w}_m$	$\bar{n}$	V-$\bar{n}$
2.0	0.334	50.4	-0.84
1.5	0.439	43.3	-0.63
1.0	0.575	34.9	-0.42
0.5	0.753	24.0	-0.20
0.1	0.940	9.87	-0.04

(b)

U_m	d_m	$(\frac{\delta \bar{n}}{\delta r})_m$	$(\frac{\delta V}{\delta r})_m$
2.0	0.000152	-0.0046	0.249
1.5	0.000080	-0.0031	0.190
1.0	0.000029	-0.0018	0.129
0.5	0.000002	-0.0007	0.065
0.1	-0.000002	-0.00007	0.012

(c)

U_m	$d_m - d_f$ $\times 10^4$	$-d(hs_m - hs_f)$ $\times 10^4$	$\frac{d_f}{d_m}$	$\frac{(\frac{\delta \bar{n}}{\delta r})_f}{(\frac{\delta \bar{n}}{\delta r})_m}$	$\frac{(\frac{\delta V}{\delta r})_f}{(\frac{\delta V}{\delta r})_m}$
2.0	-0.1258	-0.1254	1.083	1.082	1.083
1.5	-0.0581	-0.0579	1.072	1.071	1.072
1.0	-0.0179	-0.0176	1.060	1.060	1.060
0.5	-0.0009	-0.0009	1.043	1.043	1.043
0.1	0.0003	0.0003	1.021	1.020	1.021

Table 2.1a can be compared with Table 1.1a, which shows equivalent values when selection and mutation occur in both sexes. Mean fitness in males and mean mutation number are very similar. The level of linkage disequilibrium is reduced by about a half because of the recombination that occurs in females. Similarly, Table 2.1b can be compared with Table 1.1b. Values are smaller, both because the modifier only acts in males, and because there is less linkage disequilibrium.

Additionally, I have investigated, for a wide range of model parameters, whether the sex difference in selection on recombination modifiers can be explained by the sex difference in the effect of recombination on the variance in mutation number. I assumed that selection and mutation were restricted to males and varied a) male and female recombination rates independently, b) the strength of epistasis, c) the mutation rate, d) the fitness effect of each mutation. In general, I find that $\frac{d_m}{\Re_m} - \frac{d_f}{\Re_f} \cong 0$. The only exception occurs when modifiers of recombination are approximately neutral (see Section 1.6 for the conditions in which this occurs). When $d_m + d_f = 0$, I find that $d_f > 0 > d_m$. However, d_f and d_m are generally both very small in magnitude. The only circumstances in which $d_m + d_f = 0$ and d_f and d_m are significant occurs when selection is very strongly synergistic, the individual fitness effect of each mutation is large, and the mutation rate is quite low. Biologically these conditions are entirely implausible, because strong synergism has not been observed empirically and seems less likely to evolve if the mutation rate is low (Chapter 4). Given that these conditions are unrealistic, there is no need to discuss them in any detail. It is sufficient to state that this selection for a sex difference in recombination arises from sex differences in the effect of recombination on the skew in mutation number.

2.4.3 How large are predicted sex differences in recombination?

Suppose that, in the beginning, r_m and r_f are identical. If $d_m < d_f$, then this may lead to r_f's evolving to be larger than r_m. However, this change in recombination rates does not affect the selection for sex differences in recombination. The results in 2.4.2 imply that the sign of $d_m - d_f$ is largely independent of the values of r_m and r_f, depending instead on $hs_m - hs_f$. Therefore, after higher recombination in females has evolved, it will remain true that $d_m < d_f$ and selection to increase r_f at the expense of r_m will continue. For this reason, selection for sex differences in recombination might be expected to lead to large sex differences in recombination.

Two factors which may constrain sex differences in recombination are given in Sections 2.2.2 and 2.2.3 above. The first factor is drift. I argued that most modifiers of recombination are likely to have a similar effect in males and females. However, empirical evidence on this matter is lacking. The effect of drift on sex differences is therefore difficult to assess.

The second factor is the influence of the diminishing returns of chiasmata. In this section I provide a simple illustration of how the diminishing returns of chiasmata might constrain sex differences in recombination frequency. I assume that the recombination frequency has reached an evolutionarily stable rate, henceforth described as the ESS. At the ESS, altering recombination has two consequences, with opposite effects on the fitness of a modifier of recombination. First, recombination alters linkage disequilibrium. At the ESS, modifiers increasing recombination increase in frequency as a result of the reduction which they bring about in linkage disequilibria. Secondly, recombination imposes a mutational load. The mutational load always causes the frequency of modifiers increasing recombination to decrease. In Figure 2.3, the ESS recombination rate is plotted against the mutation rate in males. Mutation and selection are restricted to males. Two examples are given.

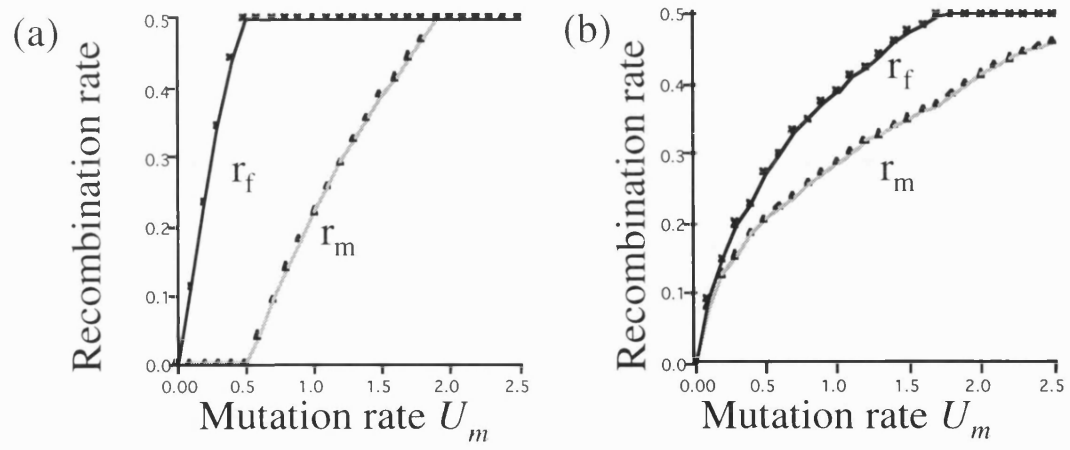
In the first example, the mutational load of recombination is proportional to the recombination rate *r*. In this case, reallocating recombination from males to females does not alter the mutation rate, provided average recombination is kept the same. Figure 2.3a shows that the ESS recombination rate is higher for high mutation rates. This reflects stronger selection to increase recombination as a result of greater linkage disequilibrium between mutations. When the average recombination rate at the ESS is less than or equal to 0.25, $r_m = 0$, so that all recombination is performed in females. When the average recombination rate at the ESS is greater than 0.25, $r_f = 0.5$. Therefore, the sex difference in recombination is always as large as is possible given the average recombination rate.

This example corresponds to a biological scenario in which each chiasma imposes a mutational load, but there is never more than one chiasma per chromosome, so that there are no diminishing returns. This applies to the smaller chromosomes of the mouse, which almost without exception have only one chiasma in each meiosis (Lawrie, Tease, and Hulten 1995). There is nevertheless a sex difference in recombination in these chromosomes; as stated in Section 2.1, chiasmata in males are more likely to occur close to the telomeres (Lawrie, Tease, and Hulten 1995). It may

be that these chiasmata are "localised" (Hultén 1990) and do not serve a recombinational function; their function would instead be to ensure the orderly disjunction of the meiotic chromosomes during meiosis I. The results shown in Figure 2.3a imply that sex differences in selection on recombination modifiers should lead to an evolutionary endpoint in which the localisation of chiasmata is restricted to one sex.

In the second example, the mutational load is proportional to r^2. This corresponds to a biological scenario in which there is often more than one crossover per chromosome. In this case diminishing returns of chiasmata ensure that the mutational cost of recombination is larger if there is a sex difference in recombination. The figure (2.3b) shows that the sex difference in recombination is reduced from the previous example; female recombination is never more than 50% higher than male recombination.

Figure 2.3 Sex differences in recombination at ESS
The ESS recombination rates in males and females, r_m and r_f, are plotted against the mutation rate in males U_m. $U_f = 0$. In males, there is quadratic epistasis, with fitness function $w(n) = 1 - \left(\frac{n}{10}\right)^2$, where *n* is the number of mutations an individual contains. There is negligible selection in females. In both examples, individuals produce gametes with an additional mutational load which depends on their recombination rate. In (a) recombination imposes a mutational load of $0.01r$. In (b) recombination imposes a mutational load of $0.02r^2$. The method by which the ESS was found is described in the appendix to this chapter, Section 2.7.



These considerations lead me to make two predictions. First, the degree of localisation of chiasmata should show great dimorphism between the sexes. Secondly, sex differences in the number of non-localised chiasmata should be smaller in organisms with many crossovers per chromosome. Unfortunately, these predictions are difficult to test because of the variable quality of meiotic preparations. This variability makes it

impossible to give a definition of localisation which is consistent across the sexes and across species. Furthermore, whilst the realisation of these predictions would show that sex differences in chiasma frequency were constrained by the diminishing returns of chiasmata, it would not prove that there was significant selection for sex differences in crossing over.

2.5 Do epistatic interactions between deleterious mutations cause significant selection for sex differences in recombination?

Having presented my model results, I wish to assess their biological importance. In Section 2.4 I found that so long as (a) epistasis is similar in males and females and (b) the mean number of mutations in maternal and paternal gametes is similar $[(M-P)^2 << D_P + D_M]$, then $d_m - d_f \cong -d(hs_m - hs_f)$, d_f and d_m are the selection gradients of male and female modifiers of recombination, d is the equivalent of a modifier of recombination in both sexes ($d = d_m + d_f$), and hs_f and hs_m are the effective selection on deleterious mutations in males and females. This means that selection on recombination is strongest in the sex in which selection is weaker. In this section, I ask whether this sex difference in selection is large enough to cause detectable sex differences in recombination rates.

Evidence concerning the size of d is discussed in Chapter 1. To summarise, selection is likely to act to increase recombination provided that the mutation rate is high (i.e. $U > 1$) and interactions between most pairs of mutations are synergistic. Knowledge of the relevant empirical parameters is quite limited, especially in organisms other than *Drosophila*, and there are no reliable estimates of the strength of selection on real modifiers of recombination. However all available evidence suggests that selection on real modifiers of recombination is weak.

Because d is small, it is implausible that there is significant selection for sex differences in recombination if $hs_m - hs_f$ is also small. There are no direct estimates of $hs_m - hs_f$ for any species. Therefore, I shall split my discussion of the size of $hs_m - hs_f$ into two parts. In the first part, I discuss the magnitude of the average effective selection on each mutation hs. $hs = \frac{hs_m + hs_f}{2}$, so the average effective selection on each mutation provides a maximum magnitude for $hs_m - hs_f$: $| hs_m - hs_f | < 2hs$. Secondly, I discuss the current empirical and theoretical knowledge on sex differences in selection.

It is important to note that the sex difference in selection on recombination modifiers is not determined by the average value of $hs_m - hs_f$, but rather by the average value of $hs_m - hs_f$ for the mutations which cause selection on recombination. Mutations with a very large or very small effect on fitness do not cause substantial selection on recombination. Mutations with a very large effect are removed by selection before significant linkage disequilibria develop and are therefore not significantly affected by recombination. A modifier may cause significant changes in the linkage disequilibria between mutations of very small effect. But, unless the modifier is very tightly linked, it will recombine away from the mutations before the change in linkage disequilibria caused by modified recombination has a significant effect on the modifier's fitness.

2.5.1 Empirical estimates of *hs*

2.5.1.1 *Estimates from mutation accumulation*

The estimate for the strength of selection on each mutation incorporated in the standard selection parameters is obtained from estimating the effect of each mutation on homozygous viability during mutation accumulation. This approach suffers from two serious limitations:

(1) Mutation accumulation does not provide enough information to estimate U and hs independently. Additional information is necessary to obtain an estimate of either (Mukai et al. 1972; Keightley 1994). After mutation accumulation it is possible to measure the average reduction in viability of accumulation lines and the increase in the variance between lines. The estimates of U and hs that are obtained depend on the underlying distribution of mutation effects that is assumed. Assuming all mutations have similar effects on viability gives a maximum estimate of hs, and a minimum estimate of U. However, the experiment is consistent with a higher U, and a lower hs, depending on the distribution of viability effects of deleterious mutations. Mutations have a wide variety of phenotypic effects, so this distribution is difficult to assess *a priori*. The distribution may vary widely according to the type of mutation in question. For example, point mutations may typically have smaller effects on fitness than transposable element insertions or other insertions and deletions. As a result, the distribution is not amenable to empirical measurement. In practice, mutation accumulation experiments do not detect mutations with very small fitness effects, because laboratory experiments do not provide enough time for a significant number of these mutations to build up.

(2) The estimates obtained are for the effect of each mutation on homozygous viability in a particular laboratory environment. It is not obvious how this estimate for the effect

of mutations on one particular fitness trait relates to the total fitness effect of each mutation in natural populations. First, the effect of each mutation on a fitness trait is highly dependent on the competitiveness of the environment in which it is measured (Kondrashov and Houle 1994; Shabalina, Yamplonsky and Kondrashov 1997). Secondly, the proportion of the variance in fitness accounted for by individual traits (e.g. viability) is unknown. Many fitness traits, such as male competitive ability, are hard to measure, but may account for an important component of the overall variation in fitness. For example, Simmons et al. (1978) found that recessive lethals induced by radiation reduced total heterozygous fitness by an average 11%, whereas Temin (1966) found that similarly induced lethals reduced heterozygous viability by 1-3%. Thirdly, extrapolating to heterozygous fitness effects requires assumptions about dominance. Dominance estimates vary widely (Crow and Simmons 1983; Houle et al. 1997), adding to the degree of uncertainty surrounding the estimates.

2.5.1.2 *Comparison of standing and mutational variance*

An alternative approach is to compare the genetic variation that exists in a natural population with the variation that arises through mutation each generation. Under the assumption that the variation in the natural population is maintained by mutation/selection balance, the comparison allows the fitness effect of each mutation to be estimated.

This approach was used to estimate the fitness effect of recessive lethals when in heterozygous state by Crow and Simmons (1983). Recessive lethals have an unambiguous phenotype, so there are good estimates of both their frequency in natural populations and their rate of origination. Crow and Simmons found that the standing frequency is approximately 50 times the rate of origination. Assuming that the mutations are removed because of their fitness effects as heterozygotes gives an estimate of the effective selection against a heterozygous recessive lethal of 0.02.

This estimate is probably an underestimate of the heterozygous fitness effect of the average newly-arisen recessive lethal. Not all of the recessive lethals are removed because of their fitness effects as heterozygotes. In particular, a proportion are maintained because they are overdominant (the heterozygote has a fitness advantage over the wild type allele). These alleles are maintained in the population by their heterozygous advantage, and therefore may occur at high frequency despite a low rate of origination. These alleles may affect Crow and Simmons's estimate if they make up a significant proportion of the naturally occurring lethals. In this case, the average selective coefficient against newly-arising lethals may be higher than Simmons and Crow suggest.

A well known example of an overdominant locus is the sickle-cell polymorphism in humans, which is thought to be maintained by heterozygous advantage. The evidence for this is summarised by Cavalli-Sforza and Bodmer (1991). The homozygote has sickle-cell anemia, which causes a large reduction in fitness. The frequency of the sickle cell allele is very high (>20%) in some geographical areas, reflecting strong selection in favour of heterozygotes. Heterozygotes have been shown to be less severely affected by malarial disease than individuals without the sickle cell allele. Because the frequency of the heterozygote is higher in malarial areas, it is thought that malaria is the selective agent responsible for the maintenance of the polymorphism. In malarial areas, the frequency of this single overdominant allele is of the same order of magnitude as the total frequency of recessive lethals on the *Drosophila X* chromosome in natural populations (around 25%, Crow and Temin 1964).

The proportion of recessive lethals that are overdominant is unknown. Attempts to assess the fitness effects of recessive lethals from natural populations directly have come to contradictory conclusions (reviewed by Crow and Simmons 1983); some studies find that recessive lethals are beneficial on average when heterozygous, whereas others have found that they are deleterious on average. This disagreement is not surprising. In the absence of inbreeding, recessive lethals can be maintained at significant frequency in the population by slight overdominance (Falconer and Mackay 1996). Therefore the fitness effects of overdominant alleles may be difficult to detect. Conversely, if none of the lethals were overdominant, the average fitness effect when heterozygous would still be too small to be easily detectable experimentally.

Crow and Simmons (1983) did obtain an estimate of the fitness effects of one type of deleterious mutation. However, if a substantial proportion of the recessive lethals observed in natural populations are overdominant, then the average heterozygous fitness effect of newly-arising deleterious recessives may be substantially higher than their 2% estimate.

Comparison of the variance increase in a fitness trait per generation during mutation accumulation (V_M) with the standing mutational variance (V_G) can in principle be used to obtain an estimate for the average fitness effect of all classes of deleterious mutation. The comparison gives the persistence time in generations of each mutation affecting the trait. Most deleterious mutations seem to affect all fitness traits; the value of one fitness trait measured for a mutation accumulation line is highly correlated with the value of another fitness trait measured for the same line (Houle et al. 1994). Therefore, the

persistence times for mutations affecting different fitness traits should be similar, and each should give an estimate of the inverse of the total effective selection on each deleterious mutation.

Unfortunately, the estimates of V_G and V_M that have been made are not readily comparable. After making a number of assumptions, for example about the dominance of mutations, Houle et al. (1996) used existing estimates of V_G and V_M to arrive at estimates of V_G/ V_M of 48 for viability, 26 for fecundity and 57 for longevity. These values should be regarded as overestimates for $1/hs$ for the following reasons:

(1) Mutation accumulation experiments underestimate V_M as a result of unavoidable viability selection during the course of mutation accumulation.

(2) Some of the standing variation that exists may not be maintained by mutation/selection balance.

Therefore, the available data is consistent with a large proportion of the mutations that arise having a significant effect on fitness, with hs being of the order of 0.04.

2.5.2 Empirical evidence on sex differences in selection

Trivers (1988) suggested that, in most species, the genetic variance in fitness of males is higher than the genetic variance in fitness of females. Typically, males undergo sexual selection, being chosen by females or competing for access to them. Sexual selection can in principle lead to stronger selection in males than in females for two complimentary reasons:

(1) *Capture of genetic variance by sexually selected traits*

Sexually selected traits have been shown to be condition-dependent in expression; individuals in good condition produce better ornaments than individuals in bad condition. Rowe and Houle (1996) have suggested that the evolution of condition dependence has led to the capture of genetic variance by sexually selected traits. Their argument is quite straightforward; sexually selected traits have evolved to become condition-dependent. The condition of each male is dependent on his genetic quality. Therefore, his genetic quality is reflected in the sexually selected trait.

(2) *A larger opportunity for selection*

One male can inseminate many females. Therefore, males with good genes may achieve a very high level of fitness, impossible for females to achieve (Bateman 1948).

The overall variation in fitness of males that is observed is often higher than the overall variation in females. However, the difference is constrained by a number of factors. In some species, such as *Drosophila melanogaster*, males invest a considerable amount of resources into each mating (Partridge and Farquhar 1981), restricting the number of females with which each male can mate. In other species, such as the red deer *Cervus elaphus*, variation in male reproductive success in a single season is much higher than variation in female reproductive success (Clutton-Brock, Guinness, and Albon 1982). However, the variation in lifetime reproductive success differs less dramatically because male reproductive success is much more strongly related to age than female reproductive success. Males are only able to compete in ruts with a high level of success in four years of their life, whereas successful hinds may produce offspring in ten or more seasons (Clutton-Brock, Guinness, and Albon 1982).

In practice, it is not clear whether either of these two factors necessarily results in the genetic variance in fitness being higher in males than in females:

(1) Females contribute directly to the success of their offspring. Variation in condition of females is also reflected in their reproductive success. Therefore female fitness may be as condition-dependent as male fitness.

(2) The only necessary connection between the opportunity for selection and the realised strength of selection is that the former gives a maximum bound on the latter (Crow 1989). Males may have a higher variance in fitness simply because the stochastic variation in fitness is higher. The stochastic variation is likely to be highest in the sex in which the investment in each offspring is smallest. This is simply illustrated by a hypothetical example in which there is no selection in either sex, and each female has the same number of offspring, mating with a single male chosen at random. In the example there is no selection in either sex. There is no variance in fitness in females, but the variance equals the mean in males.

Because of these difficulties, there are currently no good theoretical or empirical estimates of the sex difference in the genetic variance in fitness. I list three possible avenues for future work to rectify this:

(1) *Measurement of the condition dependence of fitness*
It is possible to experimentally manipulate condition, for example by altering the amount of food available to individuals. The effect of an equivalent manipulation on the fitness of males and females could be compared. Assuming that condition and

genetic quality have similar effects on fitness, this should give an indication of the sex difference in the effect of genetic quality on fitness.

(2) *Direct estimates of the genetic variance in fitness*
A more direct approach is to assess the fitness effects on males and females of genotypes with known effects on total fitness. Estimates of the fitness of individual chromosomes can be obtained using population cage experiments.

For example, Fowler et al. (1997) competed twelve wild type chromosomes as heterozygotes against two balancer chromosomes in population cages. The chromosomes each had a recessive lethal, but were otherwise chosen at random from a cage adapted population. The chromosomes varied in fitness substantially, and the fitness of each chromosome on the genetic background of one balancer was correlated with its fitness on the genetic background of the other. The variance in fitness of the chromosomes was estimated as 0.06. Of course, the proportion of the variance in fitness accounted for by the recessive lethal each chromosome contained is unknown. If it is assumed that the recessive lethals are responsible for little of the genetic variation observed, then the total variance in fitness of *Drosophila* individuals is estimated as 0.46 (Fowler et al. 1997). This experiment is significant because it indicates that normal chromosomes in heterozygous state show considerable genetic variance in fitness. This observation is consistent with a high mutational load, and with individual mutations having a large effect on fitness.

It is possible to assess the variance in fitness of each of the chromosomes for individual fitness traits. For example, the same study found that, in an uncompetitive environment, the viability of each chromosome was uncorrelated with the fitness of that chromosome (although the statistical power may have been rather limited, since no between-replicate correlation was observed for viability). Male fitness traits, such as mating ability, are inherently more difficult to measure than female fitness traits, such as fecundity. Nevertheless, obtaining an estimate for the proportion of the genetic variance in fitness accounted for by easily measurable fitness traits would give an upper bound on the sex difference in the strength of selection.

(3) *Evolutionary arguments*
The degree of condition dependence in fitness is subject to evolution. Therefore it may be possible to use evolutionary arguments to relate the condition dependence in fitness to observable trade-offs in the life history strategies of males and females.

2.6 Conclusions

Two hypotheses can explain the observation of Burt et al. (1991) that crossing over rates are higher in males and in females in an approximately equal number of species. The first hypothesis is drift. This hypothesis can only be falsified by demonstrating that there is a (phlyogenetically controlled) pattern to the direction of sex differences.

The second hypothesis is that sex differences in crossing over are determined by sex differences in the direction of selection. I argue that it is plausible, in theory, that sex differences in the direction of selection could exert significant selection for sex differences in crossing over. More crossing over is expected in the sex in which selection is weaker. The necessary conditions are fairly strong synergistic interactions between deleterious mutations and a high frequency of mutations with a large individual effect on fitness. Empirical evidence on the strength of synergism in obligately sexual organisms is extremely limited (see Chapter 4). The more extensive evidence on the mutation rate is consistent with the necessary conditions, but is currently inconclusive.

It is plausible that the sex difference in the direction of selection differs between animal species, since each species has a different sexual ecology. However, at present there are no satisfactory hypotheses which predict which species should have a higher genetic variance in fitness in females and which should have a higher genetic variance in fitness in males. Establishing such a prediction is a priority for future work.

2.7 Appendix. ESS finding.

Finding an evolutionary stable recombination rate (ESS) entails finding r_m and r_f such that $d_f = 0$ and $d_m = 0$. This involves searching the two dimensional space (r_m, r_f). Finding an ESS in a two dimensional space is difficult, even if the surface is benign and the ESS is unique. In fact, the surface is challenging because the strength of selection on recombination falls off very quickly as the recombination rate increases. This makes methods based on the magnitude of selection on recombination difficult to apply.

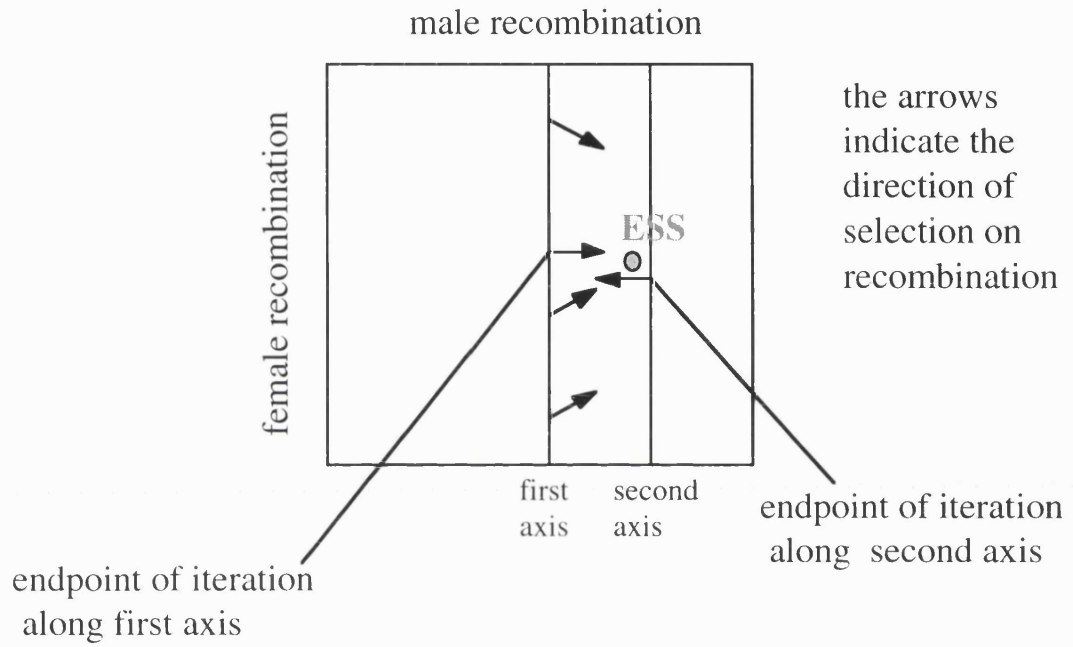
A convenient method is suggested by analogy with interval halving in one dimension. Interval halving requires the function g to be weakly monotonic, i.e. for all x and y, if $x<y$ then $g(x>0) \Rightarrow g(y>0)$. In this case, if $g(x)>0$ and $g(y)<0$ and there is a value z such that $g(z)=0$, then $x>z>y$. Interval halving starts from a value x_0. If $g(x_0)>0$ then $x_1 = x_0 - I_0$ otherwise, $x_1 = x_0 + I_0$. This process continues iteratively, with $I_{t+1} = \frac{I_t}{2}$ -hence the name "interval halving". Provided that there is an ESS z, with $x_0 - 2I_0 < z < x_0 + 2I_0$, then this ESS will be approached by the method.

Two dimensional interval halving proceeds by performing one dimensional interval halving along a number of perpendicular axes. These axes are themselves chosen iteratively using an interval halving algorithm. One dimensional interval halving is used to find the point on the axis at which the vector $g(x,y)$ is perpendicular to the axis. For example, if the axis x=0.25 is chosen, then y is varied by one dimensional interval halving to find the value y^* at which $g(0.25,y^*)=(a,0)$, where a is any number. The choice of the next axis is dependent on the value of a. If $a>0$ then the axis $x = 0.25 + J_0$ is chosen, otherwise the axis $x = 0.25 - J_0$ is chosen. As for one dimensional interval halving, $J_{t+1} = \frac{J_t}{2}$. The value of J_0 is chosen so that the whole of the two dimensional space of interest is covered. The algorithm is illustrated pictorially in Figure 2.4.

Since selection to increase recombination becomes weaker as linkage is loosened (section 1.6), it is probable that both d_m and d_f are weakly monotonic, i.e. if $d_m(r_m, r_f)$ is negative, then it is probable that $d_m(r_m', r_f')$ will be negative provided $r_m' > r_m$ and $r_f' > r_f$. This monotonicity makes the interval halving method appropriate. The interval halving algorithm is performed for the vector (d_m, d_f). In order to check that the endpoint reached by interval halving is an ESS, interval halving

is performed assuming at least two different starting axes. Additionally, the value of d_m, d_f at the endpoint is checked to ascertain whether the endpoint is an ESS. In the two examples in the text (Figure 2.3a,b), the algorithms converged on the same point, regardless of the starting axis. These points are ESSs.

Figure 2.4 Illustration of interval halving algorithm



Chapter 3

Can deleterious mutations explain variation in mammalian chiasma frequency?

Abstract

In this chapter I investigate whether variation in recombination rates can be explained by differences in mutation rates. In general, it is difficult to obtain empirical evidence on how differences in mutation rates affect natural populations. This is because mutation is a ubiquitous process; it affects all populations all of the time. However, one of the few things that we do know about mutation rates is that they are high in males of long-lived mammalian species, as a result of the large number of cell divisions in the male germ line. I use this knowledge to investigate the evolution of the recombination rate in mammals as a consequence of recurrent deleterious mutations. I predict that female recombination rates are more variable than male recombination rates. This prediction is supported by the limited empirical data.

3.0 Introduction

This chapter begins with some theoretical preliminaries (Section 3.1). I explain how the effect of recombination depends on the mutation rate. I then give a detailed account of the mammalian data concerning recombination rates (Section 3.2) and mutation rates (Section 3.3). On the basis of the mutation rate data, I propose a relationship between age at maturity and the mutation rate. Assuming this relationship, I go on to examine whether deleterious mutations can explain the observed variation in recombination rates in the mammalian data (Section 3.4).

3.1 Why should selection on recombination depend on mutation rates?

The strength of selection on recombination modifiers depends on the amount by which modified recombination alters the variance in mutation number. In chapter 2, I defined the potential for recombination to increase the variance in mutation number, $\Re^a$. Recombination with rate r increases the variance in mutation number of gametes, compared to unrecombined gametes by $\frac{1}{2}r\Re^a$ (Section 2.3). The superscript a implies that the quantity is measured after selection. In this section, I will discuss how mutation rates affect the size of $\Re^b$, the potential for recombination to increase the variance in mutation number, measured before selection. The terms of $\Re^b$ are more easily related to the mutation rate than are the terms of $\Re^a$, and, while $\Re^b - \Re^a$ is crucial to sex differences in selection, it is a small proportion of $\Re^a$ (Section 2.4.1).

$\Re^b = -D_M^b - D_P^b - (\overline{M}^b - \overline{P}^b)^2$ depends on two types of parameters, the linkage disequilibria of gametes before selection, D_M^b and D_P^b and the difference in mutation number between maternal and paternal gametes, $\overline{M}^b - \overline{P}^b$. Synergistic selection creates negative linkage disequilibria between mutations which contribute positively to $\Re^b$. More disequilibria build up if more mutations are under selection, so D_M^b and D_P^b are larger for high mutation rates than for low mutation rates.

$\overline{M}$ and $\overline{P}$ are the mean number of mutations in maternal and paternal gametes, respectively. Before selection, males and females are identical. The difference between maternal and paternal gametes arises in the course of a single generation. Selection removes $\frac{\overline{n}hs_f}{2}$ mutations per female gamete, and $\frac{\overline{n}hs_m}{2}$ mutations per male gamete, where $\overline{n}$ is the mean number of mutations in diploid individuals, measured

before selection and hs_f and hs_m are the effective selection on each mutation in males and females. Mutation adds $\frac{U_m}{2}$ mutations per male gamete and $\frac{U_f}{2}$ mutations per female gamete. This gives:

$$(\overline{M}^b - \overline{P}^b)^2 = \frac{1}{4}(U_m - \bar{n}hs_m - U_f + \bar{n}hs_f)^2 \qquad (3.1)$$

At mutation/selection balance, $U_m + U_f = \bar{n}hs_m + \bar{n}hs_f$. Therefore, the magnitude of each of the terms of $(\overline{M}^b - \overline{P}^b)^2$ is dependent on the mutation rate. The actual magnitude of $(\overline{M}^b - \overline{P}^b)^2$ is dependent on sex differences in selection, $hs_m - hs_f$, and mutation, $U_m - U_f$.

With these preliminaries completed, it is now possible to examine how changes in mutation rates affect selection on recombination. For example, if, as in chapter 2, selection and mutation are both restricted to males ($U_f = 0$ and $hs_f = 0$), then $\Re^b = -D_M^b - D_P^b$. Under synergistic selection $\Re^b$ will be positive and an increasing function of U_m. If, on the other hand, selection is restricted to females but mutation is restricted to males ($hs_m = 0$ and $U_f = 0$), then $(\overline{M}^b - \overline{P}^b)^2 = \frac{1}{4}(U_m + \bar{n}hs_f)^2$. At mutation/selection balance, $\bar{n}hs_f = U_m$. In this case, $\Re^b = -D_M^b - D_P^b - {U_M}^2$. Under standard selection parameters, the linkage disequilibria is approximately linearly dependent on the mutation rate ($V - \bar{n} = D_M^b + D_P^b$ in Table 1.1). The sign of $\Re^b$ will depend on the strength of epistasis and the mutation rate. If the mutation rate is sufficiently low, the largest contribution to $\Re^b$ will be from the linkage disequilibria. Under synergistic epistasis, $\Re^b$ will be positive. However, if the mutation rate is sufficiently high, the largest contribution will be from $(\overline{M}^b - \overline{P}^b)^2$. Therefore, if the mutation rate is high enough, $\Re^b$ will be negative, whatever the sign of epistasis.

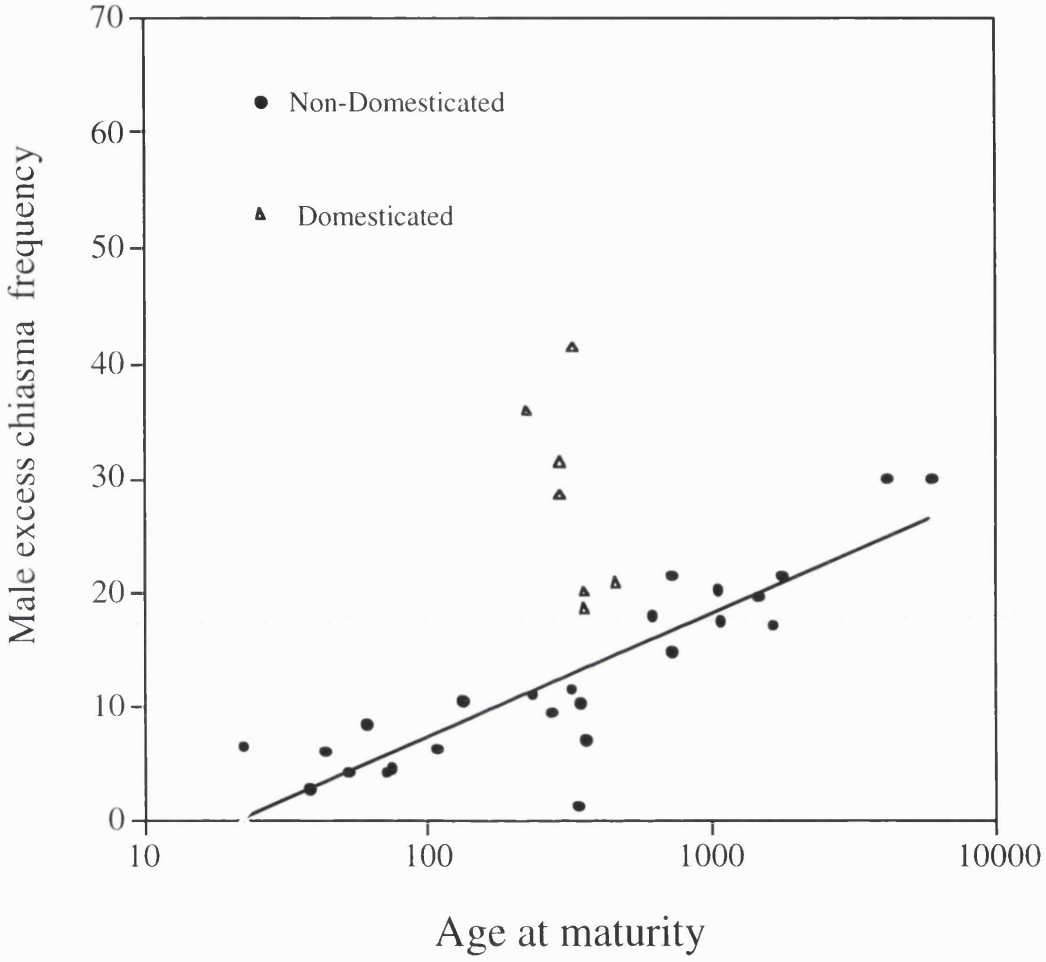
Having described how the strength of selection on recombination might depend on the mutation rate, it would be desirable to test whether an association is supported by the empirical data. The best available dataset on recombination rates was compiled by Burt and Bell (1987) for mammalian chiasma frequency. Fortunately, our knowledge of sex specific mutation rates in mammals is better than in any other taxonomic group. I summarise the two datasets, and then describe how I will use them to test the theoretical results.

3.2 Mammalian recombination data

3.2.1 Data from males

Burt and Bell assembled male chiasma and life history data from 32 species, 7 of them domesticated, in order to study recombination rates. They examined the relationship between excess chiasma frequency in each species, that is the number of chiasmata per meiosis minus the haploid number of chromosomes, and adult body weight, age at reproductive maturity and litter size. The justification for using excess chiasma frequency as a measure of recombination is that at least one chiasma is necessary to ensure orderly segregation in mammals, so that the first chiasma is not a reliable indicator of selection for recombination. Burt and Bell found that, for non-domesticated animals, long-lived species tended to have more chiasmata. The dependence is strong; age at reproductive maturity accounts for 75% of the variance in excess chiasma frequency observed between species. The other life history traits they examined could explain neither the remaining variation in excess chiasma frequency nor its correlation with age at reproductive maturity. Additionally, Burt and Bell suggest that excess chiasma frequency is a good measure of selection on recombination, while chromosome number is not. Excess chiasma frequency is independent of chromosome number, while chromosome number is independent of age at reproductive maturity. In fact the assumption that excess chiasma frequency is a good measure of selection on recombination is questionable. A potential problem with the measure is discussed in Section 5.6. The male chiasma data is summarised in figure 3.1.

Figure 3.1 Mammalian male chiasma data (Burt and Bell 1987)
Excess chiasma frequency (number of chiasmata per meiosis minus haploid number of chromosomes) is plotted against age at reproductive maturity in days (log scale). Solid line is least squares regression of chiasma frequency and log(age at reproductive maturity).



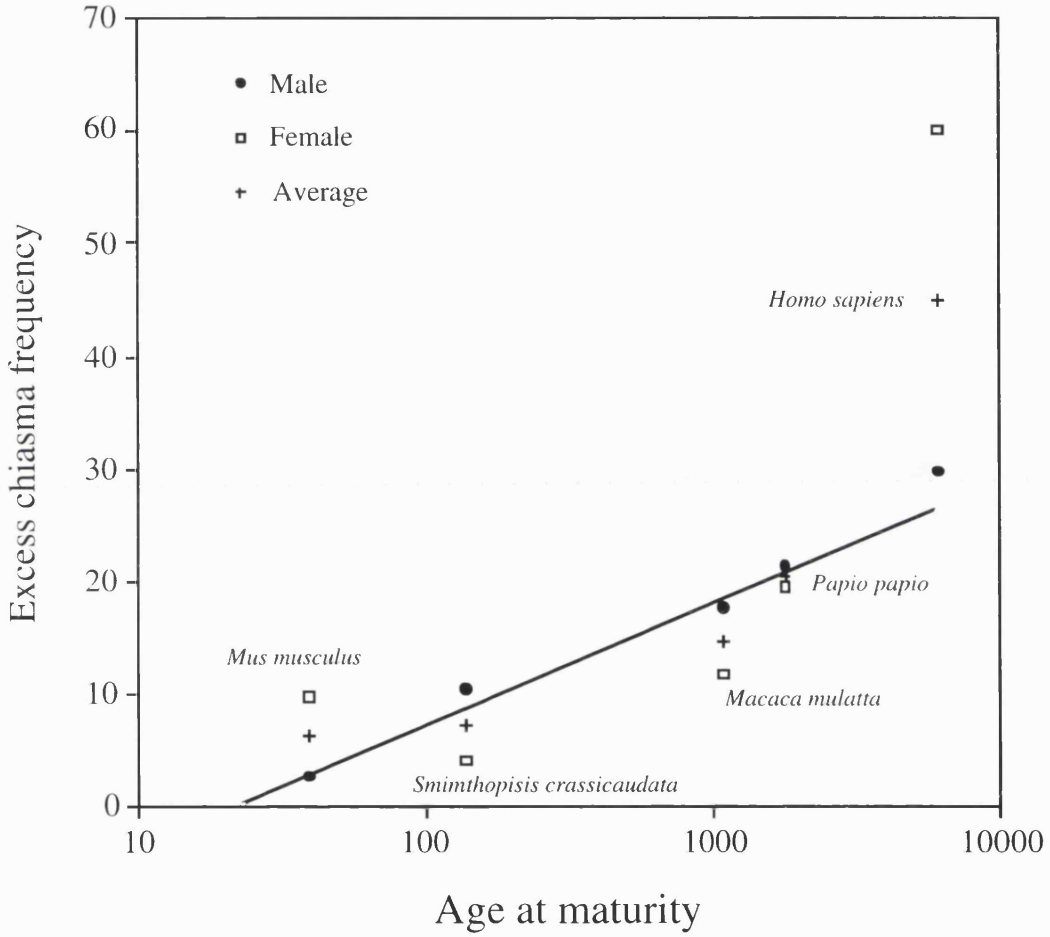
Burt and Bell also found that several of the domesticated species have much higher numbers of excess chiasmata than other species with a similar age at reproductive maturity. The higher rate of recombination could, in principle, be caused by artificial selection or by the recent change in the selective environment associated with domestication. Artificial selection often involves directional selection on specific quantitative traits, such as milk yield. Therefore, the increase in recombination might be comparable to the increase in recombination caused by selection on geotaxis observed by Korol and Iliyadi (1994) and discussed in Section 1.1.3. Moreover, artificial selection often involves selection for individuals with particular combinations of new mutations. This might also cause selection for increased recombination (Hill and Robertson 1968; Otto and Barton 1997). Dogs and cats both have highly elevated levels of recombination, while horses and cows do not. It is not obvious which traits have undergone artificial directional selection in the cat, but it

seems likely that there has been arbitrary selection for new varieties of appearance. Therefore, selection for particular combinations of mutations might explain much of the increase that is observed.

3.2.2 Sex differences in recombination in mammals

Unfortunately, data on crossing over in the female germ line of mammals is quite limited. We have chiasma data or comprehensive genetic maps for five species (Burt Bell and Harvey 1991). For humans and the house mouse, there are extensive genetic maps, which give a reliable indication of the sex difference in recombination (Dib et al. 1996; Dietrich et al. 1996). For the mouse, we also have chiasma data, which agrees qualitatively with the genetic map (section 2.1). Both species have more crossovers in females. Chiasma data from a marsupial mouse, *Smimthopsis crassicaudata* appears to shows the reverse pattern, namely more recombination in males. This conclusion is corroborated by a limited amount of genetic data (Bennett, Hayman and Hope 1986). Chiasma data from two other species reveals that macaque, *Macaca mulatta*, has a higher chiasma frequency in female, while a baboon, *Papio papio,* has approximately the same frequency in males and females (Burt, Bell and Harvey 1991).

Figure 3.2 Chiasma frequencies for two sexes, plotted against age at reproductive maturity, in days (log scale)
Excess chiasma frequency is shown for the five species for which data from females is available. For the human, female chiasma frequency is estimated from a recent version of the human genetic map (Dib et al. 1996). The remainder is chiasma data, compiled by Burt, Bell and Harvey (1991). The least squares regression is taken from figure 3.1.



The sex difference in excess chiasma frequency is plotted against age at reproductive maturity for these five species in figure 3.2. The regression of male chiasma frequency and log age at reproductive maturity, taken from figure 3.1, is also shown. The figure allows comparison of sex differences in recombination with overall variation in recombination. Additionally, the average recombination rate for each species is plotted. This can be compared with the expected recombination rate derived from the regression of male data. Obviously, the data is limited, but the graph shows that sex differences in recombination are at least as large in magnitude as differences in male recombination between species with a similar age at reproductive maturity (compare Figures 3.1 and 3.2). Figure 3.2 also shows that in the species for which recombination is higher in females, average recombination is

above the regression line, whereas in species where recombination is lower in females it is below the line. In *Papio papio*, which shows little sex difference, the average recombination rate is approximately on the regression line.

3.3 Mammalian mutation rates

3.3.1 A physiological perspective

The potential for mutations caused by incorrect DNA replication is much higher in male mammals than female mammals because of the larger number of cell divisions that the male germ line undergoes. Female germ cells undergo a total of about 28 cell divisions during embryonic development in rats, and about 33 divisions in humans (Li et al. 1996) and then are held at prophase I of meiosis until reproductive maturity. The meiotic division is completed shortly prior to fertilisation. The number of germ line divisions is presumably similar in the female germ line of all mammals. In contrast, sperm are produced by repeated stem cell division throughout reproductive maturity. Stem cell divisions organised according to a "spermatogenic cycle" (Johnson and Everitt 1995). Each cell divides once in each cycle. The length of this cycle shows slight variation between mammalian species. For example, in humans the cycle lasts 16 days, while in the rat it lasts 12 days (Johnson and Everitt 1995). Clearly, the number of stem cell divisions in the male germ line will depend on the length of time between sexual maturity and mating. The average length of this interval is hard to measure, but is likely to be correlated with age at reproductive maturity. Consequently, in mammals, the male point mutation rate will depend on longevity. For example, the sperm cells of a twenty year old man will have undergone approximately 200 divisions, while the sperm cells of a 5 month old male mouse will have undergone approximately 56 divisions (Li et al. 1996).

3.3.2 Empirical evidence

3.3.2.1 *Direct estimation of the mutation rate*

It is possible, though laborious, to estimate the mutation rate directly by observing mutational events. For example, Crow and Temin (1964) estimated the recessive lethal mutation rate of the X chromosome in female *Drosophila melanogaster* at 0.0013, while Abrahamson, Meyer, and Jongh (1981) estimated the equivalent male mutation rate at 0.0026. However, other authors do not find evidence for similar sex difference in mutation rate in *Drosophila* (Charlesworth 1993). Such measurements are less practical in mammals. A direct estimate of mutation rates in mice found a higher rate in males, but the estimates have high confidence limits (Russell and Russell 1992).

3.3.2.2 *Inferences from disease mutations*

A less laborious method of estimating sex differences in mutation rate is to infer the parental origin of newly arising disease mutations. Haldane (1947) used a population genetic argument to infer the sex ratio of mutations for the X linked genetic disease Hemophilia from the proportion of affected males who had carrier mothers. Haldane found that most mothers were carriers, and concluded that the majority of mutations had arisen in the male germ line in previous generations.

It is now possible to use linkage mapping to ascertain directly the parental origin of most dominant disease mutations (Crow 1997), as well as a large proportion of recessive X linked mutations (Grimm et al. 1994). This facilitates reliable estimation of sex differences in origination rate for the disease mutation in question. For many classes of disease mutation, such as point mutations, the mutations seem to be overwhelmingly paternal (Crow 1997). However, insertions and deletions often show a female bias (Grimm et al. 1994).

Genetic diseases which originate in males also show a strong age effect; children born to older fathers are much more likely to be affected than those born to younger fathers (Risch et al 1987). This provides direct evidence that the male mutation rate depends on age.

Disease loci provide the most direct, and therefore the most reliable, estimates of sex difference in mutation rate in mammals currently available. However, disease mutations may be unrepresentative of all deleterious mutations, since the mutation rate at known disease loci may be unusually high, and the mutations in question may have an unusually large effect on phenotype.

3.3.2.3 *Inferences from molecular evolution*

An alternative approach is to study the relative rates of *X* chromosome, *Y* chromosome and autosome evolution (Miyata et al. 1987). Autosomes spend an equal number of generations, on average, in males and females. In contrast, *Y* chromosomes spend every generation in males, while X chromosomes are inherited by females two thirds of the time and by males one third of the time. The rate of evolution at a neutral locus is equal to the mutation rate (Kimura 1983). Therefore, the rate of evolution at neutral loci on the *X*, *Y* and autosomes should depend on the mutation rate in males and females. Miyata et al. (1987) suggested that the relative rates of divergence between species of X, Y and autosomal loci should give an estimate of the sex difference in mutation rate. In the study of divergence of homologous loci between primates, they found that the Y chromosome evolved

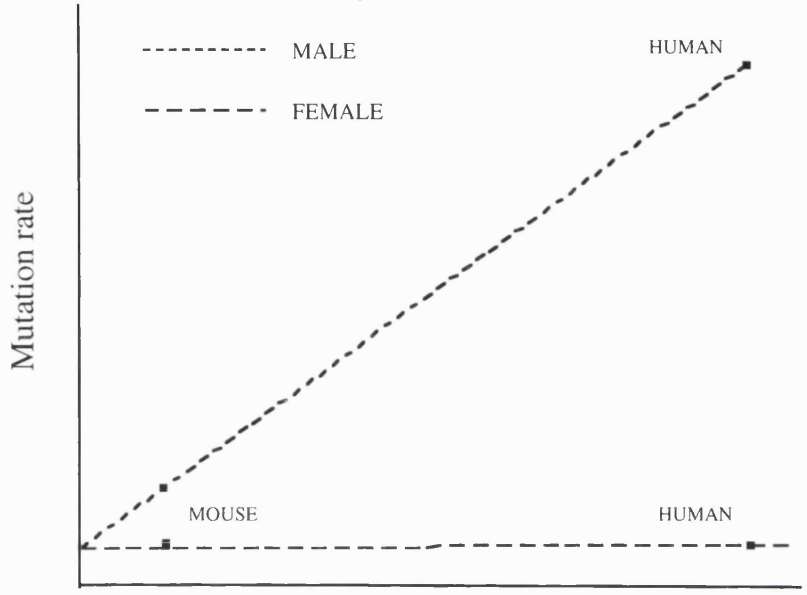
approximately twice as fast as the X chromosome, suggesting that most mutations occurred in males. They termed this phenomenon "male-driven evolution". Similar data has been obtained for other primate loci by Shimmin et al. (1993) and for rodents by Chang et al. (1994). The data suggests that, in primates, the male point mutation rate is many times higher than the female point mutation rate. In rodents, the difference is more modest, with the male mutation rate only twice as high.

Miyata et al.'s (1987) method makes the assumption that there are no systematic differences in mutation rate or fixation probability between the sex chromosomes and the autosomes. In fact, this assumption is questionable; the X chromosome is hemizygous in males, so partially recessive mutations on the X are likely to undergo stronger selection than similar mutations on the autosomes (Chang et al. 1994). Additionally, there may be selection for a lower mutation rate on the X chromosome than on the autosomes because of stronger selection against newly arising deleterious mutations (McVean and Hurst 1997). Any systematic differences between the X or Y chromosomes and the autosomes will bias the estimates for the sex difference in mutation rate that are obtained. A study of molecular evolution in birds, performed by Ellgren and Fridolfsson (1997), has shown that while such differences may exist, they do not account for the rate differences that are observed. In birds, males are homogametic, having two Z chromosomes. Females have a W and a Z. Ellgren and Fridolfsson found that the W chromosome evolved much more slowly than the Z chromosome. Therefore, it is sex and not gamety which appears to be the strongest force determining the different rates of evolution of sex chromosomes and autosomes.

3.3.3 Conclusions on mammalian mutation rates

The sexual physiology of mammals suggests that point mutation rates will be higher in males than females and, moreover, will be higher in males with greater longevity. This is supported by the empirical evidence on the mutation rate. On the basis of this evidence, I propose a relationship between age at reproductive maturity and mutation rate, shown in Figure 3.3. I use this proposed relationship between age at maturity and mutation rate to investigate selection on recombination.

Figure 3.3 Proposed relationship between age at reproductive maturity and male and female mutation rate in mammals
U_f is independent of age at reproductive maturity. For short lived mammals, such as mouse and rat, $U_f = 2U_m$. For long lived mammals, U_m may be as much as ten times greater than U_f.



3.4 How does selection on recombination depend on mutation rates?

In this section, I use my genes-in-a-pot model to generate values for selection on recombination based on the empirical data on mutation rates. I then ask whether these results can explain the mammalian chiasma data. This analysis will be split into two subsections. The first subsection investigates the effect of mutation rate on selection on overall recombination. The second investigates the relationship between mutation rate and sex differences in recombination.

3.4.1 The higher recombination rate in long- lived mammals

Section 3.2.1 described how long lived mammals have a higher level of recombination than short lived ones. Section 3.3 described the evidence that longer lived mammals have a higher mutation rate, because of the greater number of cell divisions in the male germ line. My aim is to determine whether epistatic interactions between the deleterious fraction of those mutations can account for the higher level of recombination. This is a reasonable hypothesis because Table 1.1 shows that selection to increase recombination is an increasing function of the mutation rate.

However, where there are sex differences in the mutation rate, the term $(\overline{M}^b - \overline{P}^b)^2$ may be significant and needs to be included in calculating $\Re^b$ (section 3.1). I use my

model to investigate how the gradient of selection on recombination, *d*, depends on sex differences in mutation rate. *d* was calculated first for a two-to-one ratio of male and female mutation rates, and secondly with mutation restricted to males. These values are compared with the selection on recombination with no mutation bias, but the same average mutation rate. Simulation results are shown in Table 3.1.

Table 3.1 Effect of sex differences in mutation rate on selection on recombination
The table shows the gradient of selection on a modifier, *d*, and the difference in mean, $\frac{\delta\bar{n}}{\delta r}$ and variance, $\frac{\delta V}{\delta r}$, between an individual heterozygous for a modifier and an individual from the unmodified population. Standard selection parameters are assumed: $w(n) = \exp(-0.002n - 0.0004n^2)$. Three examples are shown. The first is taken from Table 1.1, with $U_m = U_f = U$. In the second, the male mutation rate is twice the female mutation rate. In the third, mutation is restricted to males. The average mutation rate in each row is the same for all three examples. The properties of the unmodified population, ($\bar{w}, \bar{n}$,V-$\bar{n}$), are very similar to those given in Table 1.1 in each case, and are therefore not shown.

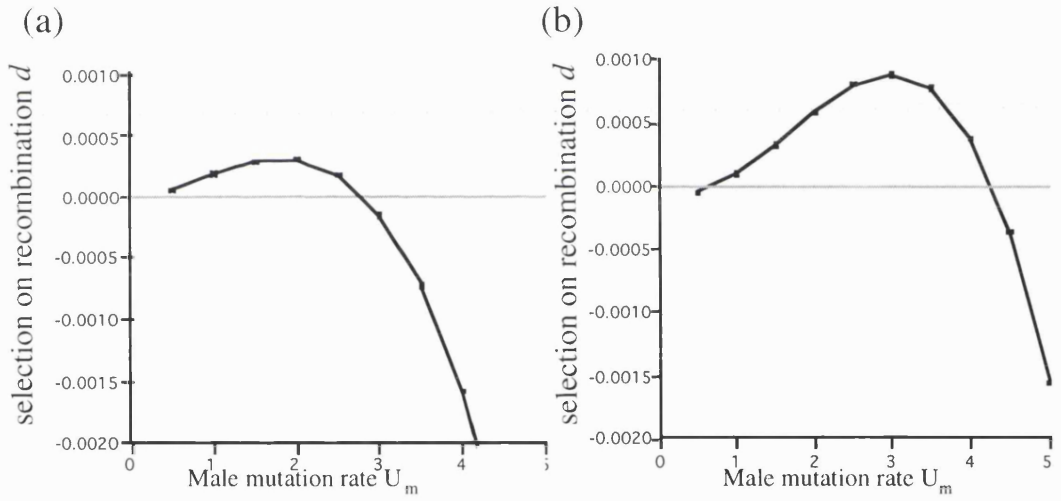
Sexes identical				$U_m = 2U_f$				$U_f = 0$			
U	d $*10^2$	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$	U_m	d $*10^2$	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$	U_m	d $*10^2$	$\frac{\delta\bar{n}}{\delta r}$	$\frac{\delta V}{\delta r}$
2.0	0.181	-0.0337	0.912	2.67	0.131	-0.0244	0.661	4	-0.268	-0.0497	-1.35
1.5	0.101	-0.0231	0.704	2.00	0.080	-0.0184	0.560	3	-0.085	-0.0194	-0.594
1.0	0.042	-0.0133	0.484	1.33	0.036	-0.0116	0.419	2	-0.009	0.0029	-0.107
0.5	0.007	-0.0051	0.249	0.67	0.006	-0.0048	0.232	1	0.003	-0.0020	0.096
0.1	-0.001	-0.0005	0.047	0.13	-0.001	-0.0005	0.046	0.2	-0.001	-0.0004	0.041

The table shows that for low mutation rates, sex differences in mutation make little overall difference to selection on recombination, *d*. However, for higher mutation rates sex differences rapidly become more important. When the male mutation rate is twice that of females, *d* is reduced by about 1/3, compared to the case with no bias, as a result of the change in size of $(\bar{M}^b - \bar{P}^b)^2$. This level of bias is consistent with empirical data from *Drosophila* (Section 3.3.2.1). When mutation is restricted to males, $(\bar{M}^b - \bar{P}^b)^2$ is larger still, and selection to reduce recombination occurs for $U_m > 1$.

I now investigate how selection on recombination is likely to depend on age at reproductive maturity. As proposed in Figure 3.3, I assume that female recombination rate is constant, and that male mutation rate is a function of age at

maturity. Therefore, I plot the strength of selection on recombination, *d*, as a function of male mutation rate. Values are shown for standard selection parameters, with $U_f = 0.5$, and selection identical in males and females (Figure 3.4a). The figure shows that selection to increase recombination is maximised for a male to female mutation ratio of four to one, and selection for reduced recombination occurs when the male to female mutation ratio is greater than six to one.

Figure 3.4 Strength of selection on recombination as a function of male mutation rate
The gradient of selection on recombination, *d*, is plotted against the male mutation rate U_m. $U_f = 0.5$, and is therefore less than or equal to the male mutation rate for all values plotted. Selection is identical in males and females, and there is free recombination. In (a) standard selection parameters are assumed: $w(n) = \exp(-0.002n - 0.0004n^2)$; in (b) stronger epistasis is assumed: $w(n) = \exp(-0.000006n^3)$.



Assuming that standard selection parameters apply to all mammals, the model suggests that reduced recombination should be favoured in long lived mammals. This is clearly not the case (section 3.2 above). Either selection on deleterious mutations is unimportant in determining recombination rates, or some of the assumptions used to calculate values in figure 3.4a are inappropriate. The assumptions are based on the findings of Mukai (Mukai 1969; Mukai et al. 1972). Specifically, I assume:

(1) Epistasis is constant for mammalian species, and has the same strength as in *Drosophila*.

(2) Each mutation has a similar effect on fitness, with effective selection on each mutation of the order of a few percent.

(3) The female mutation rate is lower than the male mutation rate and is independent of age at reproductive maturity.

(4) The female mutation rate is of the same order of magnitude as the average mutation rate in *Drosophila* measured by Mukai et al. (1972). $U_f = 0.5$ for all mammals.

(5) Selection is identical in males and females.

Each of these assumptions may be violated for mammalian data.

3.4.1.1 *The level of epistasis*

In chapter 4, I will argue that the level of epistasis is dependent on the mutation rate. Stronger synergistic epistasis is likely to evolve in species which have a higher mutation rate. *Drosophila* has many fewer genes than either human or mouse (Bird 1995) and, consequently, is likely to have a lower deleterious mutation rate. Therefore, the level of epistasis may differ between *Drosophila* and the mouse, and between the mouse and humans. Under stronger epistasis, more linkage disequilibria build up. Therefore, the effect of stronger epistasis is to increase the mutation rate at which selection to increase recombination is maximised (Figure 3.4b).

3.4.1.2 *The fitness effect of each mutation*

Differences in the fitness effect of each mutation will affect the relative size of $D_M^b + D_P^b$ and $(\overline{M}^b - \overline{P}^b)^2$. All mutations contribute to $(\overline{M}^b - \overline{P}^b)^2$, regardless of the size of their effect. However, because greater linkage disequilibria can build up between mutations of small effect, these mutations may make a larger contribution to $D_M^b + D_P^b$. Mutation accumulation experiments do not provide evidence on the distribution of mutations, nor on the number of mutations with a small effect on fitness (Section 2.5.1). Consequently, using standard selection parameters may underestimate the true linkage disequilibrium, and, thus, the selection for increased recombination. Of course, mutations of small effect do not cause significant selection for sex differences in recombination (Chapter 2).

3.4.1.3 *Sex differences in mutation rate*

The assumption that the total female deleterious mutation rate is lower than the male mutation rate may be unrealistic if the number of insertions and deletions is large. Little is known about this class of mutations, because it is extremely hard to infer the exact nature of the mutations which underlie the observed divergence in sequence between lineages. Additionally, insertions and deletions may undergo much stronger selection than point mutations since they often disrupt open reading frames. D. Tautz (personal communication) performed an informal study of sequence divergence of a *Drosophila* pseudogene, and found that the number of insertion and deletion events was similar in magnitude to the number of point mutations. Human disease studies have found a female bias in insertions and deletions. If insertions and deletions form a substantial proportion of the total deleterious mutation rate, then $(\overline{M}^b - \overline{P}^b)^2$ will be reduced, relative to the linkage disequilibria.

3.4.1.4 *The mutation rate in females*

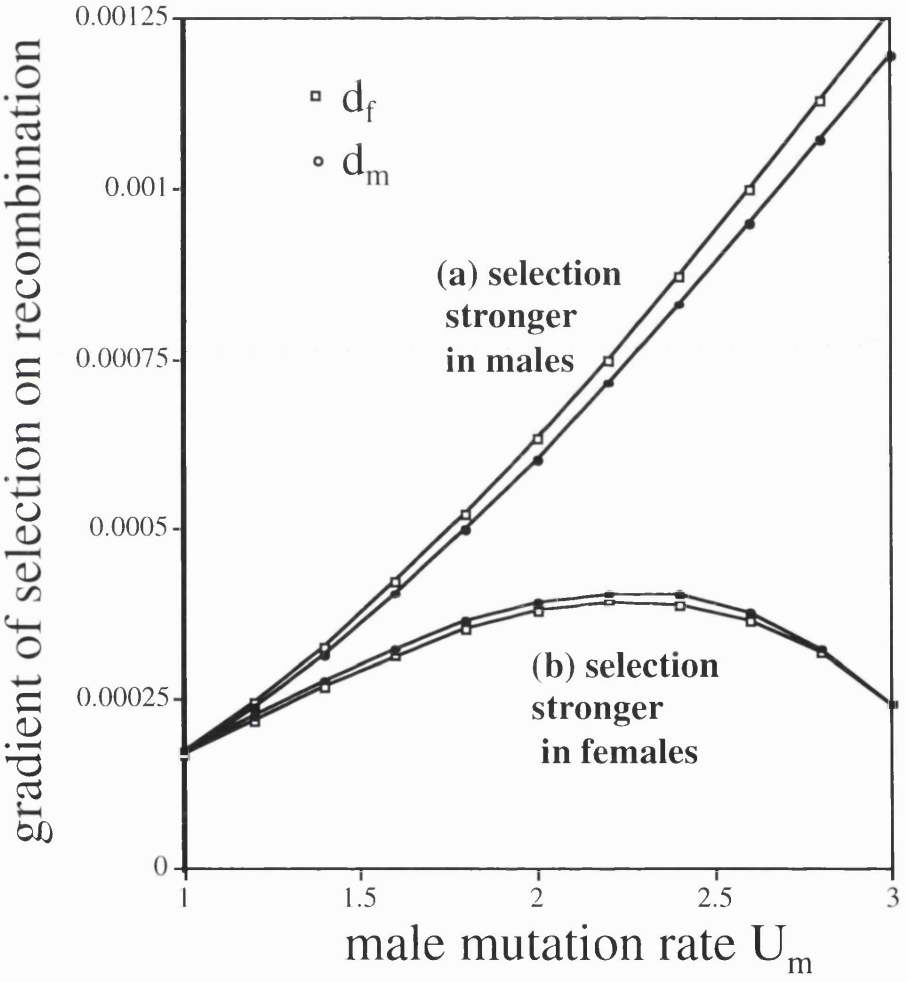
In Figure 3.4, it is assumed that $U_f = 0.5$. This value is loosely based on Mukai et al.'s (1972) data, which suggests a minimum average deleterious mutation rate of about 0.5 in *Drosophila*. In fact, the value of U_f is unknown. Mammals have many more genes than *Drosophila* (Bird 1995), so the female mutation rate may be higher than the average *Drosophila* mutation rate. Estimates the of deleterious mutation rate provided by mutation accumulation vary greatly; one recent experiment implies that the mutation rate is two orders of magnitude lower than Mukai et al. suggest (Keightley and Caballero 1997), another implies that Mukai et al's estimates may be quite accurate (Shabalina, Yamplonsky and Kondrashov 1997). The ratio of male to female mutation rates at which selection on recombination is maximised depends on the value of U_f. If $U_f < 0.5$, then $(\overline{M}^b - \overline{P}^b)^2$ will be reduced, relative to the linkage disequilibria. As a result, the value of $\frac{U_m}{U_f}$ at which selection to increase recombination is maximised will be higher. However, if the mutation rate is low, then selection on recombination is likely to be weak for all mammals. Unfortunately, we do not have reliable estimates for the strength of selection on realistic recombination modifiers (Section 1.4). Nevertheless, this analysis allows me to conclude that male-driven evolution can only explain the high recombination rates of long lived mammals under the assumption that the mutation rate is neither very high, nor very low.

3.4.1.5 The effect of sex differences in selection is described in the next section.

3.4.2 Sex differences in selection

Sex differences in selection between species may shed light on the overall question of selection on recombination. Section 3.1 demonstrated that the magnitude of $(\overline{M}^b - \overline{P}^b)^2 = \frac{1}{4}(U_m - U_f + \bar{n}hs_f - \bar{n}hs_m)^2$ is dependent on sex differences in selection, $hs_m - hs_f$, and mutation, $U_m - U_f$. In mammals, $U_m > U_f$, so that $(\overline{M}^b - \overline{P}^b)^2$ will be larger if $hs_m > hs_f$ than if $hs_m < hs_f$. Consequently, the effect of differences in $(\overline{M}^b - \overline{P}^b)^2$ on recombination rates can be investigated by determining whether sex differences in selection can explain variation in average recombination rates in mammalian species. The dependence of d on sex differences in selection is illustrated in Figure 3.5. The figure shows that sex differences in selection have two consequences for selection on recombination. First, the gradient of selection on recombination, d, is higher when selection is stronger in males than when selection is stronger in females. Secondly, where selection is stronger in males, $d_m < d_f$, and, where selection is stronger in females, $d_m > d_f$.

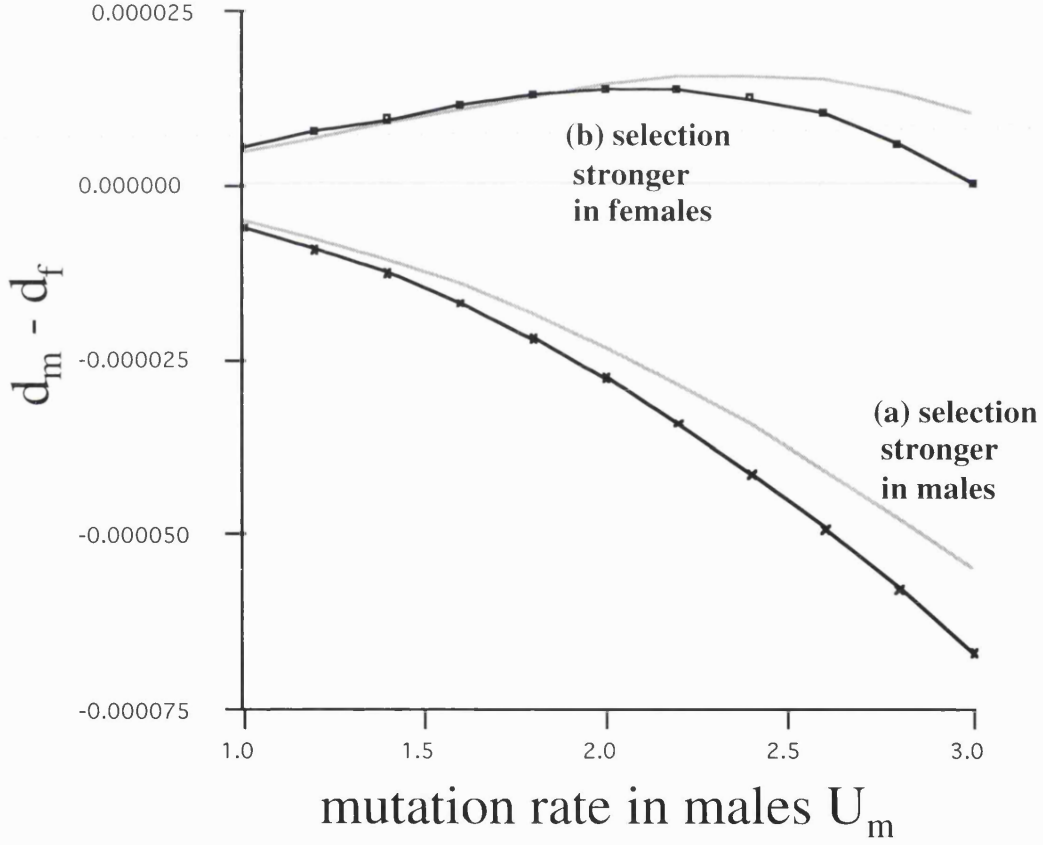
Figure 3.5 Dependence of sex specific recombination on sex differences in selection
The gradient of selection on modifiers of male recombination, d_m, (circles) and female recombination, d_f, (squares) is shown for two different examples. In both examples, $U_f = 1$ and U_m is varied between 1 and 3. In the first example, (a), selection is stronger in males. The fitnesses of males and females with *n* mutations are $w_m(n) = \exp(-0.000006n^3)$ and $w_f(n) = \exp(-0.000003n^3)$ respectively. In the second example, (b), the fitnesses are reversed so that selection is stronger in females; $w_m(n) = \exp(-0.000003n^3)$, $w_f(n) = \exp(-0.000006n^3)$.



Recall that, in chapter 2, I demonstrated that $d_m - d_f \cong -d(hs_m - hs_f)$ when $(\overline{M}^b - \overline{P}^b)^2 = 0$. It turns out that this relationship is robust when $(\overline{M}^b - \overline{P}^b)^2$ is quite large. I found that, when $(\overline{M}^b - \overline{P}^b)^2 = 0$, differences between d_m and d_f arise because selection removes some of the mutations responsible for linkage disequilibrium. However, here, when $(\overline{M}^b - \overline{P}^b)^2$ is significant, selection also

removes the mutations responsible for the difference in mutation number between maternal and paternal haplonts. Therefore, $(\overline{M}^b - \overline{P}^b)^2$ has an effect on sex differences in recombination similar to its effect on overall recombination. Figure 3.6 shows the sex difference in recombination $d_m - d_f$ and the predicted value $-d(hs_m - hs_f)$ for the same parameters as in Figure 3.5. $-d(hs_m - hs_f)$ is a good estimate of $d_m - d_f$ when the mutation rate is low, but becomes worse as the mutation rate increases. The predicted strength and direction of selection on recombination becomes unreliable when $(\overline{M}^b - \overline{P}^b)^2$ is comparable to the linkage disequilibria (Figure 3.6b, when $U_m = 3$).

Figure 3.6 Sex difference in selection on recombination, $d_m - d_f$, (black line) compared to estimate given by equation 2.6, $-d(hs_m - hs_f)$ (grey line). Parameters are as described in figure 3.5.



It would be desirable to test whether the data bears out the prediction that average recombination is higher where selection is stronger in males ($hs_m > hs_f$). Unfortunately, there are no direct estimates of $hs_m - hs_f$ (Section 2.5.2). However, we can infer sex differences in selection from the sex difference in recombination, because where selection is stronger in males, we expect $d_m < d_f$. Figure 3.2 shows that, in the two species, humans and mice, where recombination is higher in females,

indicating $hs_m > hs_f$, average recombination is above the regression line. In the two species, where recombination in males is higher than recombination in females, *Smimthopsis crassicaudata* and *Macaca mulatta*, indicating $hs_m < hs_f$ average recombination is below the regression line. In *Papio papio* there is little sex difference, and the average recombination is on the regression line.

3.5 Conclusions

The purpose of this chapter has been to demonstrate that differences in recombination rates can be determined by differences in mutation rates. I demonstrated the theoretical basis for this relationship and examined whether it could explain the chiasma data from mammalian males. I found that it can, but only if some of the assumptions embodied in the standard selection parameters are violated. Having examined the dependency of recombination on mutation rates, I used the dependency to examine how sex differences in selection effect sex differences in recombination. Clearly, the empirical data is limited, yet Figure 3.2 does suggest higher variation in female recombination rates than in male recombination rates. My explanation for this pattern suggests that both sex differences in recombination rates and variation in average recombination rates are determined by sex differences in selection. If this pattern is confirmed, it would demonstrate that subtle differences in the sexual ecology of species determines a key component of our genetic system.

Chapter 4

The evolution of epistasis

Abstract

The most reliable measurements of epistasis have found no evidence for synergism. However, these measurements are for asexual or facultatively sexual organisms with low mutation rates. I show that synergistic epistasis evolves in response to high mutation rates and high recombination rates. Therefore it remains plausible that there is strong synergism in mammals.

4.0 Introduction

The previous chapters have described the evolution of recombination rates caused by synergistic fitness interactions between recurrent deleterious mutations. A premise of the work is that the fitness interactions between most combinations of deleterious mutations are significantly synergistic. This chapter discusses the plausibility of this premise. First, I will briefly recapitulate the reasons why epistasis is important. Secondly, I will describe the existing empirical evidence concerning the level of epistasis. Thirdly, I will suggest that there is a general tendency for synergism to evolve in obligately sexual species with high mutation rates, but that this tendency is greatly reduced in facultatively sexual species or in species with low mutation rates. The best available empirical data comes from facultatively sexual species with low mutation rates. My argument implies that the absence of consistent synergism found by these experiments does not show that synergistic interactions have been unimportant in moulding recombination rates in mammals.

4.1 The importance of epistasis

Epistasis is important in species with high mutation rates for two complementary reasons. First, synergistic epistasis may cause selection for increased recombination, as discussed in chapters 1-3. In the presence of consistently synergistic interactions, the level of recombination may be determined by the mutation rate (Sections 1.5, 3.4.1). Secondly, synergistic interactions may reduce the mutational load, provided the recombination rate is high (discussed in Section 1.1.1). Therefore, species with a high mutation rate may evolve a high recombination rate, which will reduce an otherwise ruinous mutational load.

If synergistic epistasis is ubiquitous then the variation in recombination between species may be explained by variation in their mutation rate. This is the mutational deterministic hypothesis for the evolution of sex (Kondrashov 1988). Here I will suggest a less general, but nevertheless important role for synergism. The level of epistasis may itself depend on the mutation rate and the recombination rate. In particular, I will argue that consistent synergism may only arise in obligately sexual species with high mutation rates. However, in these species, synergism may be important in increasing recombination and reducing the mutational load.

4.2 Empirical evidence for epistasis

4.2.1 Experiments where the number of mutations is known

The most direct approach to estimating the strength of epistasis is to construct genotypes with known numbers of mutations, and then to measure their fitness. This approach was used recently by Elena and Lenski (1997) and by de Visser et al. (1997). Elena and Lenski constructed individuals with between one and three mutations by inserting transposons with a different antibiotic resistance marker. The mutants were collected under permissive conditions, which were designed to ensure that the genotypes had not undergone significant natural selection before being recovered. The fitness was then measured under stringent conditions using a competitive fitness assay. They found that there was no overall tendency for epistasis: three deleterious mutations reduced log mean fitness almost exactly three times as much as one deleterious mutation. Additionally, Elena and Lenski recombined pairs of individuals with one deleterious mutation each in order to produce individuals with two deleterious mutations. They found that 14 of the 27 pairs of mutations examined showed detectable interaction in the fitness effects of the mutations. The interaction was often large. In seven cases the interaction was synergistic, In seven cases antagonistic (which would imply diminishing returns epistasis). Therefore, the variance in epistasis was considerably higher than average epistasis.

De Visser et al. (1997b) performed a similar experiment in the filamentous fungus *Aspergillus niger*, and obtained comparable results. Instead of using randomly induced mutations, de Visser et al. created strains with different combinations of eight markers. Not all combinations with eight markers were recovered; indeed it was harder to recover strains with many mutations than strains with few. Ignoring unrecovered strains with low fitness and many mutations is likely to bias the results towards diminishing returns epistasis. De Visser et al. minimised the bias towards antagonism by examining the fitness of the largest subset of the markers for which all of the combinations were recovered. Using this approach, they found no evidence for significant epistasis on average, but significant evidence for epistasis between individual pairs of mutations. They also suggested that the effect of pairwise interactions was reduced when many mutations were involved. This result suggests that, when fitness is determined by the action of many mutations of small effect, the contribution of pairwise epistasis to variability in fitness may be slight.

4.2.2 Experiments where the number of mutations is unknown

4.2.2.1 *Assumption of a constant mutation rate*

Mukai (1969) inferred synergistic epistasis from the observation that viability decreased more slowly in the early stages of mutation accumulation than in later stages. In order to make this inference, it must be assumed that the mutation rate is constant over time. In fact, the relaxation of selection on fitness implies relaxed selection on the mutation rate. Therefore, the mutation rate might have increased over the course of the experiment. Therefore, little can be concluded from their observations.

4.2.2.2 *Assumption of mutation/selection balance*

Charlesworth and Charlesworth (1973) compared the viability and fecundity of individuals containing chromosomes extracted from males and those containing chromosomes extracted from females. Female chromosomes undergo recombination, but male chromosomes do not. They found that the fitness of female chromosomes was lower than male chromosomes. However, this observation does not test between synergistic or antagonistic epistasis, since recombination may cause an immediate reduction in fitness at mutation/selection balance in either case (Barton 1995). Additionally, many other factors may cause a difference in fitness of male and female chromosomes: a sex difference in mutation rate (Section 3.3.2.1), the mutagenicity of crossing over (Section 2.2.2.1) and transposition (Charlesworth and Barton 1996) would all have this effect. In fact, the difference in fecundity that Charlesworth and Charlesworth observed is larger than seems possible as a result of the effects of recombination under mutation/selection balance (Charlesworth and Barton 1996).

De Visser et al. (1996) crossed two mutagenised strains of *Neurospora crassa* and compared the fitness of the unrecombined parental strains with the fitness of the offspring. Their experiment suffers from similar problems of interpretation to that of Charlesworth and Charlesworth (1973). It is not known whether the offspring have a higher or a lower variance in mutation number than the parents. Therefore a lower log mean fitness of offspring is compatible with both synergism and diminishing returns epistasis (West, Peters and Barton in press).

4.2.2.3 *Known differences between strains*

A more promising approach is to cross individuals from two strains which are known to have very different numbers of mutations. The mean log fitness of offspring is then compared with the mean log fitness of the parents. The variance in mutation number of the offspring is lower than the variance in mutation number between the two parents. Under synergistic epistasis, the offspring should have a higher mean log fitness than the parents. Under diminishing returns epistasis, the offspring should have a lower mean log fitness. An advantage of this approach is that the change in variance caused by recombining the two strains will be much larger than the change in variance caused by recombining strains with a similar number of mutations (West, Peters and Barton in press). As a result, fitness differences between parents and offspring should be large enough to be detectable experimentally, and are less likely to be confounded by other factors.

4.2.2.4 *Measurement of skew*

Finally, De Visser et al. (1997a) measured the skew in fitness of the offspring of a single cross. The distribution of mutation number of the offspring of a cross between two individuals is symmetrical. Consequently, the distribution of log fitness of the offspring should have a negative skew under synergistic epistasis, and a positive skew under diminishing returns epistasis. Measuring this skew gives the sign of epistasis. This approach suffers from the problem that the variance in fitness of the offspring of a particular cross is small, being comparable in size to the variance in mutation number at mutation/selection balance. As a result, the predicted skew is small (West, Peters and Barton in press). Additionally, the approach is prone to confounding factors, such as skew in the error with which fitness is measured (West, Peters and Barton in press).

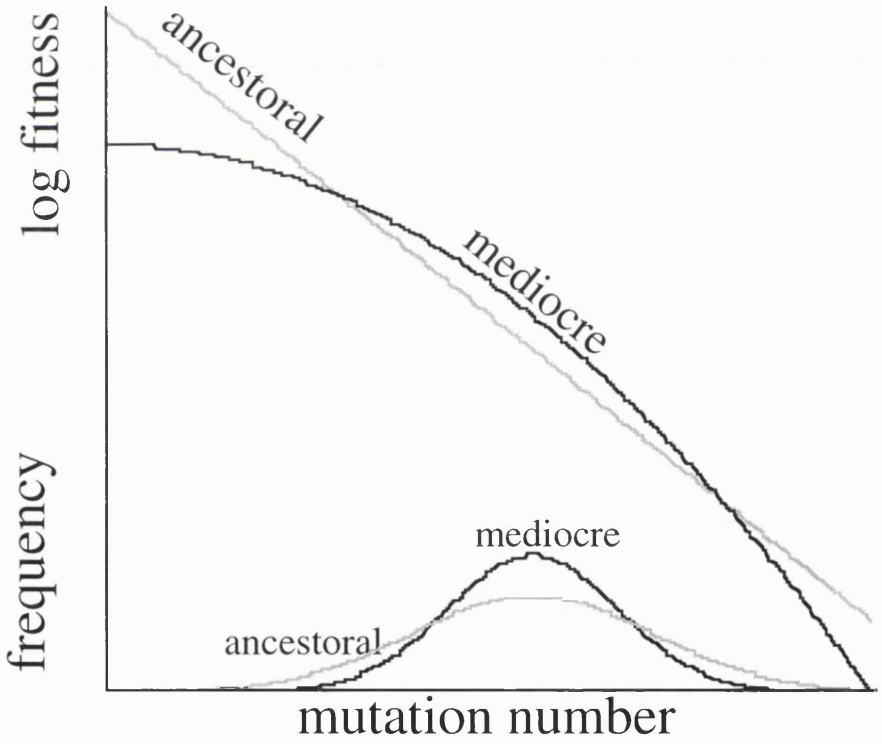
4.3 The evolution of epistasis

Here I will give a preliminary description of how epistasis might evolve. Evolution occurs through the fixation of what can be called "mediocrity genes". I define a mediocrity gene, give a biological example, and then discuss how selection on mediocrity depends on the mutation rate and the level of recombination. I then discuss the co-evolution of mediocrity and the level of recombination.

4.3.1 Definition of mediocrity

Consider a population at mutation/selection balance. The mean number of mutations per individual is $\bar{n}$. I define a "mediocrity gene" as a gene which increases the mean fitness of individuals who have $\bar{n}$ mutations, but decreases the mean fitness of both individuals with many fewer than $\bar{n}$ mutations and individuals with many more. The fixation of mediocrity genes results in synergistic epistasis. This is illustrated in Figure 4.1. In the absence of mediocrity genes, there is assumed to be no epistasis. Once mediocrity genes have fixed, the fitness function becomes synergistic; i.e. log mean fitness decreases increasingly quickly as mutation number increases.

Figure 4.1 The effect of mediocrity genes
The figure shows fitness as a function of the number of deleterious mutations (above) and the equilibrium frequency of individuals with each number of mutations (below), for an ancestral population and for a similar population after the fixation of mediocrity genes.



4.3.2 A biological example

One type of mediocrity gene alters reproductive capacity. All sexual organisms have their reproductive capacity limited in some way, either through senescence (Hamilton 1966) or through a fixed number of germ cells (Hodgkin and Barnes 1991). Limited capacity evolves because additional capacity is seldom used (Hamilton 1966) and is costly to create (Chapman et al. 1995). An organism with a given genotype has an

optimal capacity which reflects a trade-off between increased capacity and fecundity in adverse conditions. The optimal capacity for an individual with many mutations is likely to be lower than the optimal capacity for an individual with few. Therefore, the trade-off between high capacity and high fecundity in adverse conditions is likely to lead to a trade-off between fitness on a good genetic background and fitness on a poor genetic background.

A gene which specifies the optimal capacity for an individual with $\bar{n}$ deleterious mutations is necessarily a mediocrity gene. An individual which has the gene, but which has no deleterious mutations, will have lower than optimal capacity for the genotype, and, as a result, may not significantly outperform genotypes with many more mutations. An individual with the gene and many more than $\bar{n}$ mutations will attempt to create a higher than optimal reproductive capacity. This genotype may not be capable of surviving until reproductive maturity. In this case, the genotype will have zero fitness.

This example shows that mediocrity genes may occur as a result of commonplace and observable trade-offs.

4.3.3 Selection on mediocrity genes

At mutation/selection balance, the mutation number per individual is approximately normally distributed (Charlesworth 1990). In the absence of epistasis, the variance in mutation number is equal to the mean. Therefore, most individuals have between $\bar{n}-2\sqrt{\bar{n}}$ and $\bar{n}+2\sqrt{\bar{n}}$ mutations. Individuals with few mutations contribute disproportionately to the next generation. Nevertheless, most of the next generation will be produced by individuals with between $\bar{n}-2\sqrt{\bar{n}}$ and $\bar{n}+2\sqrt{\bar{n}}$ mutations.

Consider a mediocrity gene which increases the fitness of individuals with between $\frac{1}{2}\bar{n}$ and $\frac{3}{2}\bar{n}$ mutations but decreases the fitness of other individuals. Provided $\bar{n}$ is large, a gene with this effect will increase in frequency in the short term. However, the gene also reduces the additive genetic variance in fitness by reducing the number of offspring with less than $\frac{1}{2}\bar{n}$ mutations or more than $\frac{3}{2}\bar{n}$ mutations. The reduction in additive genetic variance may cause a reduction in frequency of the gene in future generations. Recall from Section 1.6 that a gene which reduces recombination under synergistic epistasis has a similar effect, increasing mean fitness, but reducing the additive genetic variance in fitness. Thus, the conditions for the spread of a mediocrity

gene are analogous to the conditions for the spread of a modifier reducing recombination. Reduced recombination is more likely to evolve if the recombination rate between the modifier and its background is loose (Barton 1995). So mediocrity genes are more likely to spread when the recombination rate is high. In fact, mediocrity genes cannot spread in an entirely asexual population. In an asexual population, the only genotypes which contribute to the gene pool in the long run are those which have very few mutations. Mediocrity genes reduce the fitness of these individuals.

Mediocrity genes are more likely to spread if $\bar{n}$ is large. Increasing $\bar{n}$ reduces the standard deviation of mutation number compared to the mean. Therefore, mediocrity genes are more likely to spread if the mutation rate is high, or if each mutation has a small fitness effect.

4.3.4 Mediocrity is self-reinforcing

The spread of mediocrity genes causes synergistic epistasis. Once synergistic epistasis has evolved, selection causes negative linkage disequilibria to build up (Section 1.1.1). The build up of disequilibria may make the fixation of further mediocrity genes more likely for two reasons. First, the reduction of the variance in mutation number ensures that a mediocrity gene increases the fitness of a larger proportion of the population (Figure 4.1). This increases the advantage of mediocrity genes. Secondly, synergism may cause the evolution of increased recombination (Chapter 1). This will dissipate some of the linkage disequilibria, but will reduce the selection against mediocrity genes that results from their effect on the additive genetic variance (Section 4.3.3). Therefore, mediocrity and recombination may co-evolve. The fixation of mediocrity genes leads to selection for higher recombination rates, which in turn causes the fixation of further mediocrity genes.

4.4 Conclusions

The level of epistasis may depend on the mutation rate. Species with low mutation rates tend to have low recombination rates, and, in these species, little or no mediocrity is likely to evolve. Above a certain threshold, which depends on both the mutation and recombination rates, selection for mediocrity may become significant. Organisms which lie above this threshold evolve higher recombination rates because of selection on deleterious mutations. Higher recombination rates in turn cause stronger selection for synergistic epistasis. Thus, there may be a "mediocrity threshold", above which synergism becomes important. This threshold may coincide with the evolution of obligate sexuality.

The only convincing estimates of epistasis to date were made by Elena and Lenski (1997) and de Visser et al. (1997). Both authors used a direct approach, constructing genotypes with known numbers of mutations and measuring their fitness. However, the approach may not be feasible for obligately sexual organisms, because the measurement of fitness of individual genotypes is more difficult.

Of the indirect approaches, the best is the comparison of the fitness of progeny from crosses between genotypes with very different numbers of mutations with the fitness of their parents. However, if these experiments are performed in organisms with low mutation rates, then they would not provide a good indication of the level of epistasis in organisms with high mutation rates. An alternative approach is to measure the capacity for mediocrity to evolve. The persistence of mediocrity genes is dependent on a trade-off between fitness on a genetic background with many mutations and fitness on a genetic background with few mutations. This trade-off may exist in every organism, regardless of its natural mutation rate. The existence of the trade-off may be testable as follows:

Two selection lines are created. The first is maintained on a mutagenised genetic background, the second is maintained on an unmutagenised background. It is necessary to renew the genetic backgrounds frequently in order to ensure that the first selected line becomes adapted to mutagenised backgrounds in general, rather than to mutations in a specific mutagenised genotypes. After several generations of selection, the fitness of the two lines are compared on mutatagenised and unmutagenised backgrounds. Provided the trade-off exists and there is suitable genetic variation, the fitness of the first line should be higher than the second on a mutagenised background, but lower on an unmutagenised background. It is necessary to devise an experiment which ensures that, while the lines undergo selection on a mutagenised or unmutagenised background, they do not recombine with that background. This is possible, at least in principle. Rice (1996) cycled chromosomes from females of a non responding stock out of the population every generation. A similar design could ensure that mutagenised and unmutagenised chromosomes were cycled out of the population each generation.

Chapter 5

Chiasma interference and the distribution of crossovers

Abstract

It has been suggested that recombination is often an incidental result of selection for crossing over (Bernstein and Michod 1987; Otto and Barton 1997; Takahashi et al. 1997). This would imply that chiasmata have an independent selective function. I describe the phenomenon of chiasma interference and examine whether it is compatible with the alternative explanations for selection on crossing over. I find that, despite possible physiological roles for crossovers, selection on recombination must play an important role in determining the distribution of chiasmata.

5.1 The problem: explaining chiasma distribution

The distribution of crossovers in higher organisms has two contrasting features which need to be reconciled. The first is that crossovers occur throughout the chromosome and within each meiosis the location of crossovers is apparently random. The second is that crossovers are not random with respect to each other; they rarely occur close together. This phenomenon is called chiasma interference.

The level of chiasma interference differs considerably between species. It occurs in *Drosophila melanogaster,* is weaker in *Neurospora crassa* (Foss et al. 1993), and is not observed at all in *Aspergillus nidulans* (Stadler 1996). In species with good meiotic cytology, such as humans (Laurie, Hultén, and Jones 1981) and grasshoppers (Fox 1973), interference appears to be even stronger. For example, for human chromosome 1, chiasma counts from the testes of four individuals revealed that a majority had exactly four chiasmata, with a variance in number less than a tenth of the mean for each individual (Laurie, Hultén, and Jones 1981). This chapter discusses the mechanical basis of chiasma interference, and asks whether what one could call the "organised randomness" of chiasma formation is consistent with the different theories for the maintenance of crossing over.

5.2 Models of chiasma interference

In this section I discuss attempts to formulate mechanical models which reconcile the random location of chiasmata with chiasma interference. I concentrate on one family of models, which involve chiasma "precursors". In these models, precursors have three important features. First, they are all formed before chiasma formation itself begins. Secondly, they are randomly distributed with respect to each other; there is no precursor interference. Thirdly, the frequency of precursor formation may differ between loci. It has been postulated that chiasma precursors are cytologically observable (Carpenter 1987), although this remains unproved (John 1990).

It is helpful to contrast two proposed precursor models, that of Foss et al. (1993) and that of King and Mortimer (1990). In both models, the effect of a precursor being converted into a chiasma is to prevent the surrounding precursors from becoming chiasmata. This blocking causes chiasma interference. In the model of Foss et al, precursor choice is determined by a counting mechanism. A precursor is chosen and converted into a chiasma. Counting starts from this chiasma, and proceeds in both directions along the chromosome. The adjacent n precursors are counted and blocked, and the n+1th precursor is turned into a chiasma. Counting then proceeds analogously from this chiasma until the end of the chromosome is reached. In the model of King and Mortimer, blocking is caused by a polymer which extends along the chromosome

when a chiasma is formed, preventing the precursors it reaches from becoming chiasmata.

I have attempted to test precursor models of chiasma formation using cytological data from humans, mice and grasshoppers. Meiotic chromosomes are photographed, the number of chiasmata on the chromosome is counted, and their position on the chromosome measured. My objective was to test whether precursor models could explain the observed distribution of chiasmata and the level of interference. I simulated a linear chromosome in which precursors formed randomly. Precursors were converted into chiasmata assuming a specific interference mechanism. I compared the observed distribution of chiasmata for a sample of computer-generated chromosomes with the actual distribution of chiasmata in a sample of real chromosomes. The distribution of precursors and the parameters of the interference mechanism were adjusted using a hill climbing algorithm in order to ensure that the distribution of chiasmata on the computer generated chromosomes was as similar as possible to distribution on the real chromosomes. I compared the fit of model and data for a variety of interference mechanisms. My findings can be briefly summarised as follows:

(1) Human male chiasma data is not consistent with counting models of chiasma interference. Counting models imply that the level of interference in chiasma frequency between loci is dependent on their genetic distance. In fact, the level of interference between loci is more accurately predicted by their cytological distance.

(2) The level of interference that is observed in both humans and grasshoppers is much stronger than suggested by King and Mortimer's simple model of polymer formation. Their model can be modified by allowing the rate at which the polymer grows to change over time. In fact, the best fit to either human or grasshopper data is provided by a model in which the polymer extends along a fixed distance instantaneously. This suggests that realistic polymerisation models cannot explain the strength of interference that is observed.

(3) Precursor models in which precursors are chosen at random cannot explain the observed distribution of crossovers. For some human and grasshopper chromsomes, there is a chiasmata close to the telomere in more than 50% of meioses. For this to be consistent with random choice of precursors, the number of precursors at the end of the chromosome would have to be unrealistically large. I investigated sequential models, in which chiasma formation starts at one end of the chromosome and proceeds to the other. These models cannot explain the distribution of chiasmata on smaller mouse

chromosomes. These chromosomes always have one chiasma, which is randomly distributed on the chromosome.

I conclude that there is no simple mechanism involving chiasma precursors which can explain the pattern of interference of all chromosomes. Unfortunately, interpretation of my analysis is complicated by limitations inherent in the data. The measurement of chiasma position is subject to a number of possible measurement errors. Potentially, some of these errors may bias any analysis. The most serious type of error involves systematic stretching of the chromosomes in response to the position of chiasmata. In a careful study, Fox (1973) was able to show that, within an individual meiosis, the overall length of a chromosome was unaffected by the number of chiasmata it contained. Nevertheless, the possibility of systematic measurement error cannot be ruled out.

5.3 Recombination and the evolution of interference

Despite the absence of a good mechanical model for chiasma interference, interference clearly does occur. This section discusses possible interference regimes and their consequences for recombination rate. I then go on to ask how selection on recombination might cause interference to evolve.

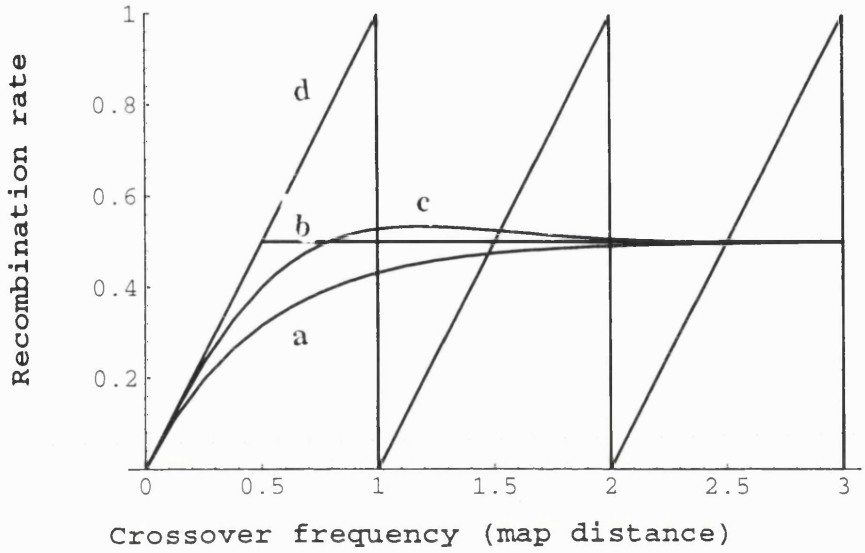
5.3.1 Possible interference regimes

At meiosis I, the maternal and paternal chromosomes are both present in two copies, giving a total of four chromatids. Each crossover recombines two of the four nonhomologous chromatids (Weinstein 1936). Which of the strands recombine is apparently random. Consequently, there is little or no chromatid interference (Zhao, McPeek, and Speed 1995), so that the strands that are involved in one crossover are no more or less likely to be involved in adjacent crossovers as a result. As a consequence, the maximal recombination rate between two loci is 0.5.

The pattern of strong chiasma interference and no chromatid interference that is observed in higher organisms represents an efficient way of achieving a high level of recombination per crossover. However, high levels of efficiency could be achieved in other ways. Four possible interference regimes are shown in Figure 5.1. M is the map distance between two loci, which is measured in Morgans. Two loci are separated by one Morgan if there is an average of one recombination event between them. r is the recombination rate between the two loci. Four different interference schemes are shown. When there is no interference of any sort, the recombination rate between two genes separated by a map distance M is equal to $0.5(1-e^{-M})$ (Figure 5.1a). When there

is maximal chiasma interference but no chromatid interference, the recombination rate is equal to M for $M \leq 0.5$, and equal to 0.5 thereafter (Figure 5.1b).

Figure 5.1 Recombination rate of a pair of genes plotted against their map distance. a) no interference, b) maximal chiasma interference and no chromatid interference, c) maximal chromatid interference and no chiasma interference, d) maximal chiasma interference and maximal chromatid interference.

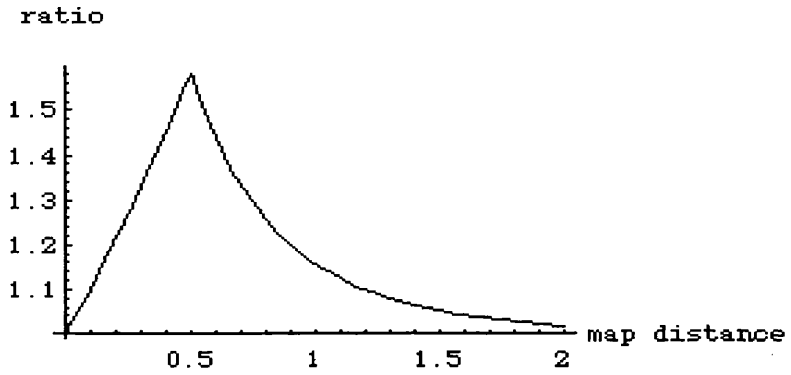


When there is maximal chromatid interference and no chiasma interference the strands which cross over alternate but crossovers themselves are randomly placed. The recombination rate is equal to $e^{-M}(0.5(\frac{M}{1!}+\frac{M^3}{3!}+...)+\frac{M^2}{2!}+\frac{M^4}{4!}+...)$ (Figure 5.1c). In this case, the recombination rate exceeds 0.5 for intermediate map lengths but asymptotically approaches 0.5. The recombination rate is always higher than the recombination frequency assuming no interference, but is lower than the recombination frequency assuming maximal chiasma interference for short map distances. Figure 5.1d shows the effect of maximal chiasma interference and maximal chromatid interference combined. The recombination rate approaches 1 as M approaches 1, but becomes M-1 for $1<M<2$.

These examples illustrate that there is no single interference mechanism which globally maximises the amount of recombination achieved by a given level of crossing over. For chromosomes which are less than one Morgan in length, maximal recombination is achieved by the combination of maximal chiasma interference and maximal chromatid interference. However, this combination is very inefficient in recombining loci more than one Morgan apart. Maximal chiasma interference in combination with no chromatid

interference represents a good compromise, since it recombines genes at quite a high frequency at all map distances. Nevertheless, the increase in recombination achieved, compared to no interference, is limited, being less than 60% for all map lengths. The increase is small for very short map lengths, and increases to a maximum for loci separated by 0.5 Morgans, before declining steeply thereafter (Figure 5.2).

Figure 5.2 The ratio of recombination rates at each map distance. The ratio is the recombination rate assuming maximal chiasma interference divided by the recombination rate assuming no interference.



5.3.2 The evolution of interference

Interference may evolve because of its effect on recombination rates. Goldstein, Bergman, and Feldman (1993) investigated this possibility using a four locus model with three selected loci and a modifier of interference arranged in a linear sequence. In their models, the modifier did not alter the recombination rate between adjacent selected loci but did modify the level of interference. In each of the cases they investigated, they found that the direction of selection on the modifier could be predicted from its overall effect on recombination. Therefore, a good starting point for the analysis of the effects of interference is its effect on pairwise recombination rates. We have seen that chiasma interference significantly increased the recombination rate between loci separated by a distance of 0.2 to 0.8 Morgans. Therefore, if there is significant selection on the recombination rates of these loci, it may cause selection on the level of chiasma interference.

The observed distribution of chiasmata is consistent with selection to increase recombination. Chiasma interference may have evolved to increase the level of recombination achieved by a given number of crossovers. This possibility is particularly likely if crossing over is costly (see Section 2.2.2.1). Selection to increase recombination may have led to the occurrence of crossovers throughout the chromosome, since linkage disequilibria will build up between any loci which are not

recombined. Therefore, selection to increase recombination can, in principle, explain the qualitative features of chiasma distribution.

5.4 Repair theories and the evolution of interference

Crossing over entails the formation and repair of a double strand break at the crossover site. Repair theories, such as the double strand break repair theory (Bernstein and Michod 1987), and the excision of transposable element theory (Takahashi et al. 1997) claim that the functional significance of crossing over is due to its local action at the site of crossing over, rather than its recombination of genes. The former theory claims that crossovers occur at the site of double strand breaks. The latter claims that crossovers remove heterologous inserts, such as transposable elements.

The scattered distribution of crossovers on the chromosome is clearly consistent with repair theories, since both transposable elements and double strand breaks can, in principle at least, occur throughout the chromosome. The non-random distribution of crossovers with respect to each other is more troublesome. Double strand breaks that occur through random mutational events should be randomly distributed with respect to each other. Similarly, transposable elements may insert themselves close together as well as far apart. So the potential sites of transposon excision or repair should be quite random with respect to each other. The randomness of the distribution of sites at which repair is necessary can only be consistent with the strong interference found in many organisms if one of the following is true:

(1) Most or all potential sites of excision or repair become crossovers, but there are few potential sites. In this case, most crossovers occur at sites where repair is unnecessary.

(2) There are very many potential sites of excision or repair. However, most sites are prevented from becoming crossovers by chiasma interference.

(3) Potential sites of excision or repair become crossovers with a higher probability than other sites, but are otherwise similar.

None of these possibilities is entirely satisfactory. (1) implies that crossing over must have another function, such as ensuring the segregation of chromosomes. It would predict that, if there are many sites at which crossing over is necessary, interference should be substantially reduced. This has not been experimentally demonstrated. (2) and (3) would imply that interference is maladaptive, because interference prevents crossing over from occurring at sites where it is postulated to be beneficial. In

conclusion, repair theories are not inconsistent with the distribution of crossovers, but they do not predict the evolution of chiasma interference, and, indeed, sit uneasily with it.

5.5 Chromosome disjunction and the evolution of crossover distribution

The potential importance of segregational factors in influencing crossover distribution was recently highlighted in an evolutionary context by Otto and Barton (1997). Meiosis involves the pairing of homologous chromosomes, their condensation, their alignment on a plane, and subsequent separation and disjunction. Chiasmata play an important role in this mechanism in all chiasmate organisms by holding the homologues together until they part.

Chiasma interference can obviously help in this process. First, it reduces the total number of crossovers required to ensure one crossover per chromosome. Secondly, it may help to ensure the structural stability of chromosomes throughout meiosis. The scattered distribution of chiasmata on the chromosome is less obviously beneficial; Koehler et al. (1996) reviewed evidence which showed that crossovers in particular regions of the chromosome reduced the probability of correct disjunction in meiosis I, and that crossovers in other regions reduced the probability of nondisjunction in meiosis II. So if there is an optimal distribution of chiasmata which minimises the total probability of nondisjunction in meiosis I and meiosis II , why is it not repeated in every meiosis? Two possibilities present themselves:

(1) The formation of chiasmata is intimately connected with the pairing of chromosomes. Pairing is a stochastic process, and involves random interactions between the homologues. The scattered distribution of chiasmata may then be a necessary consequence of the randomness of this process. Arguing against this, localisation of chiasmata in one sex, at least on a cytological scale, has evolved in many organisms (John 1990). Otto and Barton (1997) conclude their discussion of physiological hypotheses with the statement that the need for proper segregation acts as "a selective force constraining the position as well as the number of chiasmata". This is undoubtedly true, but on the simplest view at least, the scattered distribution of chiasmata seems to occur in spite of the need to ensure structural stability, and not because of it.

(2) The persistence of recombination hotspots is paradoxical, because, according to the current molecular model of recombination, any allele which promotes crossing over at its own locus tend to drive itself out of the population through unequal gene conversion at the locus, unless viability selection to maintain the allele is extremely strong

(Boulton, Myers, and Redfield 1997). Hotspots can persist if selection is strong relative to the recombination frequency at the locus. If the viability effects of the gene are recessive, however, then a polymorphism of the hotspot locus and the repressor locus can be maintained, entailing a large fitness cost (Boulton, Myers, and Redfield 1997). The problem is more acute in obligately sexual diploid organisms (Boulton, Myers, and Redfield 1997). The problem can be resolved through spreading recombination sufficiently thinly across a sufficient number of loci, so that viability selection on each locus is strong enough to maintain the high crossing-over allele at each one. If polymorphisms exist at a significant number of sites, then modifiers which increase the scatter of chiasmata may be favoured. This is speculation, however, since the natural resolution of the paradox is unknown.

5.6 Conclusions

Otto and Barton (1997) advocated a synthetic hypothesis for the evolution of recombination rates. According to their hypothesis, recombination rates are selected to increase in periods of rapid evolutionary change, but are otherwise mostly influenced by segregational factors. Is such a synthetic hypothesis necessary to explain the qualitative features of crossover distribution?

This chapter has argued that local action theories, such as the double strand break repair hypothesis, cannot explain chiasma interference. Segregational theories cannot, on the simplest view, explain the scattered distribution of crossovers. Recombinational theories can explain the scattered distribution of crossovers, and can also explain the evolution of chiasma interference. However, selection to increase the efficiency with which crossing over recombines genes does not seem to be a strong enough force to create a complex adaptation such as chiasma interference. Overall, a synthetic approach does seem most appealing.

Several synthetic hypotheses are possible. A good candidate is the following: that the principal role of chiasma interference is in helping to ensure the correct disjunction of chromosomes, but that selection to increase recombination acts to keep the distribution of crossovers scattered. This hypothesis is attractive because selection to maintain some recombination between all pairs of loci will be stronger than selection to increase average recombination rates. Because of this, selection to ensure that crossovers are scattered may be important for organisms with low genomic mutation rates, even in times of evolutionary stasis.

Elaboration of this hypothesis can provide a testable paradigm within which to measure the amount of crossing over which is not purely segregational in function. For example,

the hypothesis that localised chiasmata have a segregational function while other chiasmata have a recombinational function can be tested by comparing the nondisjunction rate of meioses in which chiasmata are localised with the nondisjunction rate of those in which they are not. Under this hypothesis, the rate of nondisjunction should be lower when chiasmata are localised.

Pragmatically, we can ask whether there are simple measures of recombination rate which avoid problems caused by our uncertainty. One reasonable measure, used by Burt and Bell (1987) and described in chapter 3, is excess chiasma frequency. An assumption underlying this measure of recombination is that excess chiasmata do not play a segregational role. This is unproven. The rate of nondisjunction of chromosomes in the meiosis of human females rises sharply with age, and this age-related nondisjunction is more likely in chromosomes with fewer chiasmata (Koehler et al. 1996). This age effect could, in principle, explain the correlation of chiasma frequency with age at reproductive maturity. This possibility could be refuted by demonstrating that female chiasma frequency was more variable than male chiasma frequency. Another problem is that excess chiasma frequency is clearly an inadequate measure of recombination rate in species where crossover localisation provides most of the variation in recombination rates. This highlights the importance of developing an understanding of selection on recombination distribution caused by the segregational needs of meiosis.

SUMMARY

The exchange of genetic material through recombination underlies the evolutionary process. However, the factors which determine the evolution of the rate of exchange are poorly understood. In mammals and other obligately sexual animals, the rate of exchange is itself determined by the level of crossing over. Variation in the level of crossing over is heritable, and varies between species and sexes (Section 1.1.2). Understanding the nature of this variation will give a crucial insight into a fundamental genetic mechanism.

Various theories have been proposed to account for the evolution of recombination (Michod and Levin 1988). This thesis has concentrated on one, the "deterministic mutation hypothesis" (Kondrashov 1988). Under this hypothesis, recombination has evolved to moderate the fitness effects of deleterious mutations. This is an attractive hypothesis because the ubiquitous nature of deleterious mutations necessitates fundamental mechanisms for their control. However, this ubiquity provides the single greatest challenge to investigating the effects of deleterious mutation and the selection on recombination that they cause.

Previous approaches have not been able to give an accurate estimate of the strength of selection on recombination, or how it might vary between species or sexes. Because of the mathematical difficulties of modelling the behaviour of chromosomes, these approaches have been limited to considering either few loci, or, alternatively, multiple loci under the assumption that selection is weak relative to recombination. In reality, the strength of selection on mutations is unknown (Section 2.5.1). Additionally, the weak selection approximation does not apply to real linkage maps. Loci which are close together on the same chromosome have low rates of recombination between them. For these combinations of loci, selection is strong, relative to recombination.

I have developed a modelling approach which overcomes many of the limitations of previous methods (Chapter 1). My main innovation is to make a simplifying assumption; I do not consider genes in linear arrays on chromosomes. Instead of treating genes as beads on a string, I treat them as beads in a pot. This approach has two key advantages. First, the weak selection approximation becomes unnecessary (Section 1.3). Secondly, maternal and paternal alleles can be considered separately, allowing me to investigate the effects of sex differences (Chapter 3).

An important aspect of the variation in recombination is the variation between the sexes. For example, in humans, the female genetic map is 60% longer than the male

genetic map (Dib et al. 1996). Trivers (1988) speculated that sex differences in recombination are likely to arise because of sex differences in selection. I use my genes-in-a-pot method to show that this is plausible; in fact, higher recombination should evolve in the sex in which selection is weaker (Chapter 2). I show that selection for sex differences in recombination requires selection to be strong. Therefore, the weak selection approximation is unsuitable for studying this problem.

If recombination has evolved in order to moderate the effect of deleterious mutations - the deterministic mutation hypothesis- then the rate of recombination should depend on the mutation rate. I investigate this possibility in Chapter 3. In mammals, most mutations occur in males, and male mutation rates increase with longevity. As a result, longevity has two counterbalancing effects on recombination. On the one hand, increasing mutation rate increases selection for recombination. On the other hand, increasing the sex difference in the mutation rate decreases selection for recombination (Section 3.1). I show that under standard selection parameters, obtained from experiments in *Drosophila* (Section 1.5), recombination would not be an increasing function of longevity. However, while selection parameters are not yet available for mammals, basic biological considerations suggest that selection for recombination is likely to increase with longevity in this taxon (Section 3.4.1).

A relationship between male recombination rates and longevity has been demonstrated by Burt and Bell (1987). However this relationship alone does not imply causation because many other factors may covary with longevity. I show that female recombination rates should show less dependence on longevity, being determined instead by sex differences in selection (Section 3.4.2). This pattern is dependent on male-biased mutation rates, and could not be predicted by other theories. The limited data available from species where both male and female recombination rates are known follows the pattern which I predict (Section 3.4.2).

In the first three chapters of the thesis, I assume that there are consistently synergistic interactions between deleterious mutations. However, this assumption is not supported by existing empirical evidence (Section 4.2). In Chapter 4, I show that synergistic epistasis is only likely to evolve in obligately sexual species with high mutation rates. Existing empirical data are largely drawn from facultatively sexual species, with low mutation rates, and, consequently, do not give a good indication of the level of epistasis in mammals. The evolution of epistasis requires a genetic trade-off between fitness on a genetic background with many mutations and fitness on a background with fewer. The plausibility of synergism could be investigated by measuring this trade-off.

Some authors have suggested that recombination is merely a byproduct of selection for crossing over, and not itself subject to direct selection. In Chapter 5 I ask whether the distribution of chiasmata is compatible with these alternative explanations. I find that, despite possible physiological roles for crossovers, selection on recombination must play an important role in determining the distribution of chiasmata.

The purpose of this thesis has been to ask whether selection on deleterious mutations has moulded recombination rates in animals. By developing a novel method for simulating recombination modifiers, I have been able to make testable predictions about the effects of deleterious mutations on the evolution of recombination rates. I show that the empirical data from mammals conforms my predictions. However, existing empirical data is sparse. More data on recombination rates in females would not only provide a stronger test of my prediction, but would also shed light on the nature of selection in different mammalian species.

LITERATURE CITED

Abrahamson, S., H.U. Meyer, and C.D. Jongh (1981) The shapes of the radiation dose-mutation response curves in *Drosophila.* In: Measurement of risks, pp 477-92. Edited by G. C. Berg and H. D. Maille. Plenum Press, New York.

Barton, N.H. (1995) A general model for the evolution of recombination. *Genetical Research, Cambridge* 65, 123-144.

Barton, N.H., and M. Turelli (1991) Natural and sexual selection on many loci. *Genetics* 127, 229-255.

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

Bernstein, H.F.A. Hopf, and R.E. Michod (1988) Is meiotic recombination an adaption for repairing DNA, producing genetic variation, or both? In: The evolution of sex, pp Edited by R. E. Michod and B. R. Levin. Sinauer Associates inc, Sunderland, Massachusetts.

Bernstein, H.F.A., and R.E. Michod (1987) The molecular basis of the evolution of sex. *Advances in Genetics* 24, 323-370.

Bird, A.P. (1995) Gene number, noise reduction and biological complexity. *Trends in Genetics* 11, 94-100.

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

Brooks, L.D. (1988) The evolution of recombination rates. In: The evolution of sex, pp. 87-105. Edited by R. E. Michod and B. R. Levin. Sinauer Associates Inc., Sunderland, Massachusetts.

Bulmer, M.G. (1980) The mathematical theory of quantitative genetics. Oxford University Press; Oxford.

Burt, A., and G. Bell (1987) Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* 326, 803-805.

Burt, A., G. Bell, and P.H. Harvey (1991) Sex differences in recombination. *Journal of Evolutionary Biology* 4, 259-277.

Carpenter, A.T.C. (1987) Gene conversion, recombination nodules, and the initiation of meiotic synapsis. *Bioessays* 6, 232-6.

Cavalli-Sforza, L.L., and J. Lederberg (1955) Isolation of pre-adaptive mutants in bacteria by sib selection. *Genetics* 41, 367-381.

Cavalli-Sforza, L.L., and W.F. Bodmer (1991) The Genetics of Human Populations.

Chang, B.H.J., L.S. Shimmin, S.K. Shyue, D. Hewett-Emmett, and W.H. Li (1994) Weak male-driven molecular evolution in rodents. *Proceedings of the National Academy of Sciences U.S.A.* 91, 827-831.

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

Charlesworth, B. (1990) Mutation-selection balance and the evolutionary advantage of recombination. *Genetical Research, Cambridge* 55, 199-221.

Charlesworth, B. (1993) More mutations in males. *Current Biology* 3, 466-467.

Charlesworth, B., and N.H. Barton (1996) Recombination load associated with selection for increased recombination. *Genetical Research, Cambridge* 67, 27-41.

Charlesworth, B., and D. Charlesworth (1975) An experiment on recombination load in *Drosophila melanogaster*. *Genetical Research, Cambridge* 25, 267-274.

Chinnici, J.P. (1971) Modification of recombination frequency in Drosophila. I. Selection for increased and decreased crossing over. *Genetics* 69, 71-83.

Chinnici, J.P. (1971) Modification of recombination frequency in Drosophila II. The polygenic control of crossing over. *Genetics* 69, 85-96.

Clutton-Brock, T.H., F.E. Guinness, and S.D. Albon (1982) Red Deer. Behaviour and ecology of two sexes. University of Chicago Press, Chicago.

Crow, J.F. (1970) Genetical loads and the cost of natural selection. In: Mathematical topics in population genetics, pp 128-177. Edited by K. I. Kojima. Springer-Verlag, New York.

Crow, J.F. (1989) Fitness variation in natural populations. In Evolution and animal breeding, pp Edited by W. G. Hill and T. F. C. Mackay. C A B International, Oxford.

Crow, J.F. (1997) The high spontaneous mutation rate: Is it a health risk. *Proceedings of the National Acadamy of Sciences* 94, 8380-8386.

Crow, J.F., and M.J. Simmons (1983) The mutation load in *Drosophila.* In The Genetics and Biology of Drosophila, pp 1-35. Edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic Press, London, New York.

Crow, J.F., and R.G. Temin (1964) Evidence for the partial dominance of recessive lethal genes in natural populations of *Drosophila. The Americal Naturalist* 98, 21-33.

de Visser, J.A.G.M., R.F. Hoekstra, and H.V.D. Ende (1996) The effect of sex and deleterious mutations on fitness in Chlamydomonas. *Proceedings of the Royal Society Series B* 263, 193-200.

de Visser J.A.G.M., R.F. Hoekstra, and H.V.D. Ende (1997a) An experimental test for synergistic epistasis and its application in *Chlamydomonas. Genetics* 145, 815-819.

de Visser, J.A.G.M., R.F. Hoekstra, and H.V.D. Ende (1997b) Test of interactions between genetic markers that affect fitness in *Aspergillus Niger. Evolution* 51, 1499-1505.

Dib, C., S. Faure, C. Fizames, D. Samson N. Drouot, A. Vignal, P. Millasseau et al. (14 authors) (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380, 152-154.

Dietrich, W.F., J. Miller, R. Steen, M.A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, et al. (21 authors) (1996) A comprehensive genetic map of the mouse genome. *Nature* 380, 149-152.

Drake, J.W. (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences, USA* 88, 7160-7164.

Dunn, L.C., and D. Bennett (1967) Sex differences in recombination of linked genes in animals. *Genetical Research, Cambridge* 9, 211-220.

Elena, S.F., and R.E. Lenski (1997) Test of synergistic interactions among deleterious mutations in bacteria. *Nature* 390, 395-397.

Ellgren, H., and A. Fridolfsson (1997) Male-driven evolution of DNA sequences in birds. *Nature Genetics* 17, 182-184.

Fain, P.R., E.N. Kort, C. Yousry, M.R. James, and M. Litt (1996) A high resolution CEPH crossover mapping panel and integrated map of chromosome 11. *Human Molecular Genetics* 5, 1631-1636.

Falconer, D.S., and T.F.C. Mackay (1996) Introduction to quantitative genetics. fourth edition. Longman, Harlow, Essex.

Feldman, M.W., F.B. Christiansen, and L.D. Brooks (1980) Evolution of recombination in a constant environment. *Proceedings of the National Academy of Sciences USA* 83, 4838-4841.

Feldman, M.W., S.P. Otto, and F.B. Christiansen (1996) Population genetic perspectives on the evolution of recombination. *Annual review of Genetics* 30, 261-295.

Fields, C., M.D. Adams, O. White, and J.C. Venter (1994) How many genes in the human genome? *Nature Genetics* 7, 345-346.

Foss, E., R. Lande, F.W. Stahl, and C.M. Steinberg (1993) Chiasma interference as a function of genetic distance. *Genetics* 133, 681-691.

Fowler, K., C. Semple, N.H. Barton, and L. Partridge (1997) Genetic variation for total fitness in *Drosophila melanogaster. Proceedings of the Royal Society Series B* 264, 191-199.

Fox, D.P. (1973) The control of chiasma distribution in the Locust, *Schistocerca gregeria*(Forskål). *Chromosoma(Berl.)* 43, 289-328.

Goldstein, D.B., A. Bergman, and M.W. Feldman (1993) The evolution of interference: Reduction of recombination among three loci. *Theoretical Population Biology* 44, 246-259.

Grimm, T., G. Meng, S. Liechti-Gallati, T. Becttecken, C.R. Müller, and B. Müller (1994) On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. *Journal of Medical Genetics* 31, 183-186.

Haldane, J.B.S. (1922) Sex ratio and unisex sterility in hybrid animals. *Journal of Genetics* 12, 101-109.

Haldane, J.B.S. (1947) The mutation rate of the gene for haemophilia and its segregation ratios in males and females. *Annals of Eugenics* 13, 262-271.

Halliday, T.R., and S.J. Arnold (1987) Multiple mating by females: a perspective from quantitative genetics. *Animal Behaviour* 35, 939-941.

Hamilton, W.D. (1966) The moulding of senescence by natural selection. *Journal of Theoretical Biology* 12, 12-45.

Hodgkin, J., and T.M. Barnes (1991) More is not better: brood size and population growth in a self-fertilizing nematode. *Proceedings of the Royal Society Series B* 246, 19-24.

Houle, D., B. Morikawa, and M. Lynch (1996) Comparing mutational variabilities. *Genetics* 143, 1467-1483.

Houle, D., K.A. Hughes, D.K. Hoffmaster, J. Ihara, S. Assimacopoulos, D. Canada, and B. Charlesworth (1994) The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. *Genetics* 138, 773-785.

Houle, D., K.A. Hughes, S. Assimacopoulos, and B. Charlesworth (1997) The effects of spontaneous mutation on quantititative traits. II. Dominance of mutations with effects on life-history traits. *Genetical Research, Cambridge* 70, 27-34.

Hultén, M. (1990) The topology of meiotic chiasmata prevents terminalization. *Annals of Human Genetics* 54, 307-314.

Hultén, M., J.M. Luciani, V. Kirton, and M. Devictor-Vuillet (1978) The use and limitations of chiasma scoring with reference to human genetic mapping. *Cytogenetic Cell Genetics* 22, 37-58.

John, B. (1990) Meiosis. Cambridge University Press, Cambridge.

Johnson, M.H., and B.J. Everitt (1995) Essential Reproduction. Fourth edition . Blackwell Sciences, Oxford.

Keightley, P.D. (1994) The distribution of mutation effects on viability in *Drosophila melanogaster. Genetics* 138, 1315-1322.

Keightley, P.D. (1996) Nature of deleterious mutation load in Drosophila. *Genetics* 144, 1993-1999.

Keightley, P.D., and A. Caballero (1997) Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans. Proceedings of the American National Academy of SciencesUSA* 147, 3823-3827.

Kimura, M. (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kimura, M., and T. Maruyama (1966) The mutation load with epistatic gene interactions in fitness. *Genetics* 54, 1337-1351.

King, J.S., and R.K. Mortimer (1990) A polymerization model of chiasma interference and corresponding computer simulation. *Genetics* 126, 1127-1138.

Koehler, K.E., R.S. Hawley, S. Sherman, and T. Hassold (1996) Recombination and nondisjunction in humans and flies. *Human Molecular Genetics* 5, 1495-1504.

Kondrashov, A.S. (1984) Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. *Genetical Research, Cambridge* 44, 199-217.

Kondrashov, A.S. (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature* 336, 435-440.

Kondrashov, A.S., and D. Houle (1994) Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster. Proceedings of the Royal Society of London Series B* 258, 221-227.

Kondrashov, A.S., and J.F. Crow (1993) A molecular approach to estimating the human mutation rate. *Human mutation* 2, 229-234.

Korol, A.B., and K.G. Iliadi (1994) Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila. Heredity* 72, 64-68.

Laurie, D.A., M. Hultén, and G.H. Jones (1981) Chiasma frequency and distribution in a sample of human males: Chromosomes 1, 2, and 9. *Cytogenetics and Cell Genetics* 31, 153-166.

Lawrie, N.M., C. Tease, and M.A. Hulten (1995) Chiasma frequency, distribution and interference maps of mouse autosomes. *Chromosoma* 104, 308-314.

Lindsley, D.L., and G.G. Zimm (1992) *The genome of Drosophila.* Academic Press, San Diego California.

Macgregor, H.C., and H. Horner (1980) Heteromorphism for chromosome 1, a requirement for normal development in crested newts. *Chromosoma (Berl.)* 76, 111-122.

McVean, G.T., and L.D. Hurst (1997) Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* 386, 388-392.

Magni, G.E., and R.C.V. Borstel (1962) Different rates of spontaneous mutation during mitosis and meiosis in yeast. *Genetics* 47, 1097-1108.

Michod, R.E., and B.R. Levin (eds.) (1988) The evolution of sex. Sinauer Associates Inc; Sunderland, Massachussets.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu, and T. Yasunaga (1987) Male-driven molecular evolution: A model and nucleotide sequence analysis. *Cold Spring Harbor Symposia on Quantitative Biology* 52, 863-867.

Mukai, T. (1969) The genetic structure of natural populations of *Drosophila melanogaster.* VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* 61, 749-61.

Mukai, T., S.I. Chigusa, L.E. Mettler, and J.F. Crow (1972) Mutation rate and dominance of genes affecting viability in *Drosophila Melanogaster. Genetics* 72, 335-55.

Muller, H.J. (1958) Our load of mutations. *Americal Journal of Human Genetics* 2, 111-176.

Nei, M. (1967) Modification of linkage intensity by natural selection. *Genetics* 57, 625-41.

Otto, S.P., and M.W. Feldman (1997) Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology* 51, 134-147.

Otto, S.P., and N.H. Barton (1997) The evolution of recombination:removing the limits to natural selection. *Genetics* 147, 897-906.

Partridge, L., and M. Farquhar (1981) Sexual-activity reduces lifespan of male fruitflies. *Nature* 294, 580-582.

Rice, W.R. (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38, 735-742.

Rice, W.R. (1992) Sexually antagonistic genes: experimental evidence. *Science* 256, 1436-1439.

Rice, W.R. (1996) Sexually antagonistic male adaption triggered by experimental arrest of female evolution. *Nature* 381, 232-234.

Risch, N., E.W. Reich, M.M. Wishnick, and J.G. McCarthy (1987) Spontaneous mutation and paternal age in humans. *American Journal of Human Genetics* 41, 218-248.

Rowe, L., and D. Houle (1996) The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London Series B* 263, 1415-1421.

Russell, L.B., and W.L. Russell (1992) Frequency and nature of specific locus mutations induced in female mice by radiations and chemicals: a review. *Mutational Research* 292, 107-127.

Keightley, P.D. (1994) The distribution of mutation effects on viability in *Drosophila melanogaster. Genetics* 138, 1315-1322.

Shabalina, S.A., L.Y. Yamplonsky, and A.S. Kondrashov (1997) Rapid decline of fitness in panmictic populations of Drosophila melanogaster maintained under relaxed natural selection. *Proceedings of the American National Academy of SciencesUSA* 94, 13034-13039.

Shaw, D.D. (1972) Genetic and environmental components of chiasma control. II. The response to selection in Schistocerca. *Chromosoma* 37, 381-301.

Shimmin, L.C., B.H.J. Chang, and W.H. Li (1993) Male driven evolution of DNA sequences. *Nature* 362, 745-747.

Simmons, M.J., E.W. Sheldon, and J.F. Crow (1978) Heterozygous effects on fitness of EMS-treated chromosomes in *Drosophila melanogaster. Genetics* 88, 575-590.

Stadler, D.R. (1996) Meiotic recombination in filamentous fungi. *Journal of Genetics* 75, 265-280.

Stamenkovic-Radek, M., L. Partridge, and M. Andjelkovic (1992) A genetic correlation between the sexes for mating speed in *Drosophila melanogaster. Animal Behaviour* 43, 389-396.

Takahashi, N.K., K. Sakagami, K. Kusano, K. Yamamoto, H. Oshikura, and I. Kobayashi (1997) Genetic recombination through double strand break repair: shift

from two-progeny mode to one-progeny mode by heterologous inserts. *Genetics* 146, 9-26.

Tease, C., and G.H. Jones (1978) Analysis of exchanges in differentially stained meiotic chromosomes of *Locusta migatoria* after BrdU substitution and FPG staining I. Crossover exchanges in monochiasmate bivalents. *Chromosoma* 69, 163-78.

Temin, R.G. (1966) Homozygous viability and fertility loads in *Drosophila melanogaster*. *Genetics* 53, 27-46.

Trivers, R. (1988) Sex differences in rates of recombination and sexual selection. In: The Evolution of Sex: An Examination of Current Ideas, pp 270-286. Edited by R. E. Michod. Sinauer Associates, Sunderland, Massachusetts.

True, J.R., J.M. Mercer, and C.C. Laurie (1995) Differences in crossover frequency and distribution among three sibling species of Drosophila. *Genetics* 142, 507-523.

Turelli, M., and N.H. Barton (1994) Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics* 138, 913-941.

Weinstein, A. (1936) The theory of multiple-strand crossing over. Genetics 21, 155-199.

West, S.A., A. Peters, N.H. Barton Testing for epistasis between deleterious mutations. *Genetics* in press.

Zhao, H., M.S. McPeek, and T.P. Speed (1995) Statistical analysis of Chromatid Interference. *Genetics* 139, 1057-1065.

Appendix

A threshold size for microsatellite expansion

Disclaimer
The following letter, which will shortly be published in Molecular Biology and Evolution, was produced in collaboration with Dr Owen Rose. The letter was produced as the result of creative discussion between me and Dr Rose. Dr Rose suggested the idea of using total genomic sequence data to study microsatellite evolution. The sequence was searched using software written to my specification by Erik Corry. Interpretation and presentation of the results was performed jointly by myself and Dr Rose.

When is a DNA repeat sequence a microsatellite? Microsatellites are tandem arrays of short (1-5 b.p.) repeats, characterised by rapid expansion and contraction through a process of 'dynamic mutation' (Sutherland and Richards 1995). Despite their importance in modern genetic analyses (Bruford and Wayne 1993; Dib et al 1996), and association with several human genetic diseases (Mandel 1993), little is known about the initial conditions necessary for dynamic mutation to occur - the evolutionary origin of a microsatellite. Here we use data on the frequency of large arrays in the genome of *Saccharomyces cerevisiae* to demonstrate the existence of a minimum threshold size necessary for a repeat sequence to undergo dynamic mutation. The existence of this threshold provides an important insight into the evolutionary properties of repeat arrays in the genome.

The major component of the dynamic mutation process is thought to be polymerase slippage during replication (Strand et al 1993), resulting in an increase or decrease of array size by one repeat unit. Estimates of slippage mutation rates range from 10^{-5} to 10^{-2}, values which are orders of magnitude greater than point nucleotide substitution rates (Edwards et al 1992; Mahtani and Willard 1993). Large *in vivo* studies of pedigrees show a 2:1 bias in favour of gain of repeats (Weber and Wong 1993; Banchs et al 1994), and coupled with high mutation rates this suggests that microsatellites show a tendency to rapidly increase in size over time. Phylogenetic analysis of two putative primate microsatellite loci further suggests that dynamic mutation may only affect arrays with a certain minimum number of repeat units (Messier, Li and Stewart 1996). Thus an understanding of the rapid evolutionary dynamics of microsatellites is developing, but both *in vivo* and phylogenetic studies are limited by the large amount of data which must be accumulated in order to observe a few mutation events.

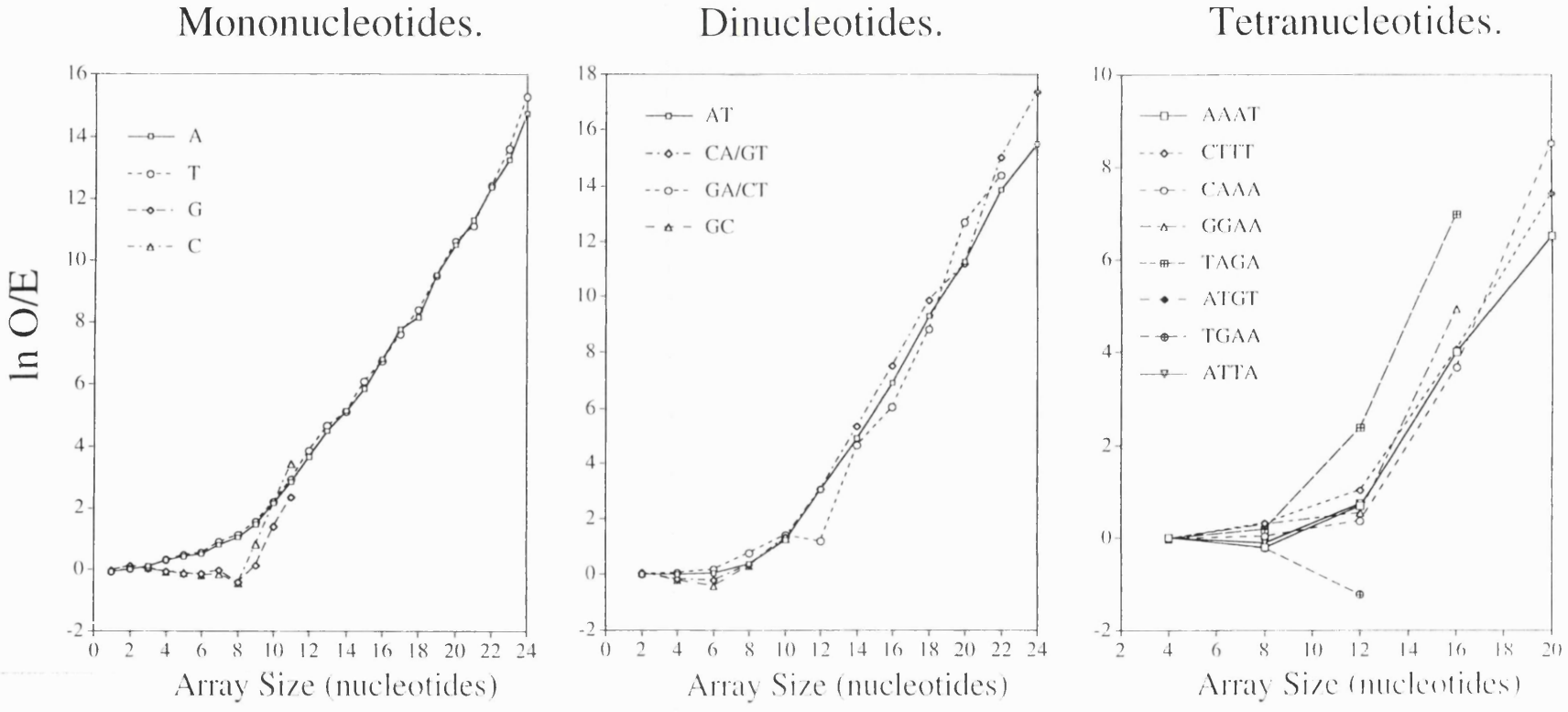
Large-scale genome sequencing projects allow a novel approach to the analysis of slippage mutation and the evolutionary origins of microsatellites. We present a study of the size distribution of short tandem repeats in the genome of the yeast (*Saccharomyces cerevisiae)*, to illustrate the mutational processes acting on these arrays. We assume a null hypothesis in which there is no dynamic mutation and repeat units (ie CA, AT, AGAT etc) are randomly distributed within the genome.This null hypothsesis is consistent with a simple model of genome evolution in which nucleotides evolve by random transitions and transversions. Under this null hypothsesis there are many more short repeat arrays than long ones. Therefore, if rapid dynamic mutation does occur, it will tend to create an excess of long repeat arrays compared to the null hypothesis, even if dynamic mutation is unbiased or slightly downwardly biased. For any particular type of array, the size at which this transition to overrepresentation occurs may reflect the beginning of dynamic mutation and the 'birth' of a microsatellite. We show evidence

for such a minimum size threshold for slippage mutation, and suggest that for a variety of different microsatellites this threshold is consistently determined by the number of nucleotides in the array, rather than the number of repeats.

The 12.06 Mb of *S. cerevisiae* DNA represents the first complete eukaryotic genome to be sequenced. This provides an ideal system for studying microsatellite size distribution because the use of an entire genome precludes any ascertainment bias arising through sampling of sequences from a composite database. Using software written by Erik Corry, we have performed a genome-wide search for mononucleotide, dinucleotide and tetranucleotide repeat arrays. For each type of repeat, observed numbers of arrays in each size class were compared with those expected under the assumption of a random distribution of repeat units (figure 6.1). Trinucleotides were excluded from the analysis, as all array sizes of trinucleotide repeats coincide with the periodicity of codons within open reading frames. Consequently the pattern of trinucleotide array sizes is likely to be affected by selection on amino acid sequences. To calculate our expected values for a particular array, we count the total number of repeat units of that type in the yeast genome and use this to calculate the expected numbers of arrays at each size, assuming that the repeat units are randomly distributed within the genome. For example, the expected number of CACACA arrays is determined by the probability of finding exactly three consecutive CA's by chance. As long as chromosomes are reasonably large, this is very closely approximated by $\frac{(N-CA)^2}{N^2}\frac{CA^3}{N^3}N$, where CA is the number of CAs in the genome and N is the total number of nucleotides.

The data show that for each type of microsatellite, large arrays tend to be greatly overrepresented (ln O/E >> 0). With the exception of the mononucleotides A and T, all repeat types show a reasonably clear transition to overrepresentation. Below about eight nucleotides there are significant deviations between observed and expected values for most repeat types, however the arrays are both under and overrepresented - there is no obvious pattern. At sizes of greater than eight nucleotides almost all arrays are overrepresented with observed numbers typically orders of magnitude greater than expected. Under the assumption that overrepresentation of repeat arrays is caused by upwardly biased slippage mutation, there is evidence for a minimum threshold size for this process at about eight nucleotides. This concurs with the phylogenetic analysis of the primate η-Globin pseudogene by Messier et al. They found that a substitution event which created a tetranucleotide repeat of length two, and another which converted a dinucleotide repeat of length three to length five, were followed by subsequent array expansion (Messier, Li and Stewart 1996).

Figure 6.1 Ratio of observed to expected frequencies of microsatellite repeats in the yeast genome. Expected values are calculated assuming that repeat units are randomly distributed at their observed genomic frequencies. Tetranucleotide data comprise both homologous forms of each repeat, thus AAAT = AAAT+TTTA. Some G/C rich tetranucleotides do not exist in arrays of greater than two repeats, however this is probably due to the general scarcity of G's and C's in the genome. For the sake of clarity, only the eight commonest tetranucleotides are shown.



In yeast, the mononucleotides A and T are consistently more numerous than expected at small array sizes. We found a similar situation in the genome of the bacterium *Mycoplasma genitalium,* which has very few long arrays but contains a great overabundance of short repeats of A's and T's. In both yeast and *Mycoplasma* these short sequences are so frequent and widely distributed that most must lie within coding regions. Consequently they are unlikely to have arisen through slippage mutation as this would frequently disrupt open reading frames.

Obviously selective constraints affect the distribution of bases in the genome, but it is unlikely that such constraints would act in the same way on each different type and size class of microsatellite. It is also possible that localised variation in base composition, such as particular regions of AT or GC richness could cause an overrepresentation of microsatellites containing these nucleotides. However, this would affect all array sizes and could not account for the threshold-linked change in array size abundance that we observe. The genome of *S. cerevisiae* is extremely compact for a eukaryotic organism, comprising approximately 70% coding DNA, with an average of one gene every 2kb. Despite this, it contains approximately 4000 expanded (>8 nucleotide) mono, di and tetranucleotide repeats which together contain about 0.38% of the genomic DNA.

Despite a detailed knowledge of DNA replication, the mechanism of slippage mutation remains elusive. Our results suggest that it is a threshold linked process, affecting repeats of approximately eight nucleotides or more. Above this threshold, repeat sequences become subject to the dynamic mutation characteristic of microsatellites. It has been shown that rates of dynamic mutation can be drastically altered by mutations affecting DNA mismatch repair (Strand 1993). The microsatellite expansion threshold that we observe may therefore represent a structural constraint on the accuracy of this repair process.

BANCHS, I., A. BOSCH, I. GUIMERA, C. LAZARO, A. PUIG, and X. ESTIVILL. 1994. New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell-lines. Human Mutation **4**: 232-232.

BRUFORD, M. and R. K. WAYNE.1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics and Development **3**: 939-943.

DIB, C., S. FAURE, C. FIZAMES, et al (14 co-authors). 1996. A comprehensive genetic-map of the human genome based on 5,264 microsatellites. Nature **380**: 152-154.

EDWARDS, A., H. A. HAMMOND, L. JIN, C. T. CASKEY, and R. CHAKRABORTY. 1992. Genetic-variation at 5 trimeric and tetrameric tandem repeat loci in 4 human-population groups. Genomics 12: 241-253.

MANDEL, J. L. 1994. Trinucleotide diseases on the rise. Nature Genetics **7**: 453-455.

MAHTANI, M. M.and H. F. WILLARD. 1993. A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. Human Molecular Genetics **2**: 431-437.

MESSIER, W., S. H. LI, and C. B. STEWART. 1996. The birth of microsatellites. Nature **381**: 483.

STRAND, M., T. A. PROLLA, R. M. LISKAY, and T. D. PETES. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365**: 274-276.

SUTHERLAND, G. R., and R. I. RICHARDS. 1995. Simple tandem DNA repeats and human genetic disease. Proc Natl. Acad. Sci. USA **92**: 3636-3641.

WEBER, J. L., and C. WONG. 1993. Mutation of Human short tandem repeats. Human Molecular Genetics **2**: 1123-1128.