Recombination Can Evolve in Large Finite Populations Given Selection on Sufficient Loci

Mark M. Iles,[1] Kevin Walters[2] and Chris Cannings

Mathematical Modelling and Genetic Epidemiology, Division of Genomic Medicine, University of Sheffield, Sheffield S10 2JF, United Kingdom

Manuscript received June 6, 2003
Accepted for publication August 25, 2003

ABSTRACT

It is well known that an allele causing increased recombination is expected to proliferate as a result of genetic drift in a finite population undergoing selection, without requiring other mechanisms. This is supported by recent simulations apparently demonstrating that, in small populations, drift is more important than epistasis in increasing recombination, with this effect disappearing in larger finite populations. However, recent experimental evidence finds a greater advantage for recombination in larger populations. These results are reconciled by demonstrating through simulation without epistasis that for m loci recombination has an appreciable selective advantage over a range of population sizes (a_m, b_m). b_m increases steadily with m while a_m remains fairly static. Thus, however large the finite population, if selection acts on sufficiently many loci, an allele that increases recombination is selected for. We show that as selection acts on our finite population, recombination increases the variance in expected log fitness, causing indirect selection on a recombination-modifying locus. This effect is enhanced in those populations with more loci because the variance in phenotypic fitnesses in relation to the possible range will be smaller. Thus fixation of a particular haplotype is less likely to occur, increasing the advantage of recombination.

THERE are many population-genetic hypotheses for the evolutionary advantage of recombination. However, the most plausible (Otto and Lenormand 2002) are based on the idea that the advantage to recombination is in breaking down deleterious linkage disequilibrium between advantageous alleles.

It has long been known that if loci interact multiplicatively with unchanging fitnesses, linkage disequilibrium (LD) will not arise in an infinite population if it does not already exist (Felsenstein 1965; Maynard Smith 1968). Thus, under such conditions, a modifier gene affecting the rate of recombination will not change in frequency. LD can be generated by (i) fluctuations in fitness, (ii) negative epistasis between loci, or (iii) finite population size.

Changing fitness: Several theories for the evolution of recombination are based around the generation of LD by fluctuations in epistatic fitness, whether the cause be simple environmental changes (Sturtevant and Mather 1938; Charlesworth 1976), parasite resistance (Jaenike 1978; Hamilton 1980), fluctuating selection for an intermediate phenotype (Maynard Smith 1980), or variation in fitness among patches (Williams 1975). It has been shown, however, that for such conditions to favor recombination, the environment must fluctuate rapidly between positive and negative epistasis (Barton 1995), a parasite must be highly virulent (Peters and Lively 1999), and in a heterogeneous environment, dispersal must be of a particular type (Maynard Smith 1976). Work on the effect of various types of heterogeneous environment on recombination is currently being carried out and it remains a viable model, for example, when LD is generated by migration (Lenormand and Otto 2000).

Epistasis: Increased rates of recombination may also evolve if there is epistatic interaction between loci combined with directional selection, fluctuating selection, or mutation-selection balance (Eshel and Feldman 1970; Feldman *et al.* 1980; Altenberg and Feldman 1987; Barton 1995). It has been shown that epistasis must be negative and weak to produce an increase in recombination (Barton 1995) but there is little experimental evidence for such epistasis (Rice 2002). An advantage to recombination over a wider range of values of epistasis is possible if spatial heterogeneity in selection exists (Lenormand and Otto 2000).

Finite population size: The final theory for the evolution of recombination is that LD may be generated by finite population size. The Fisher-Muller theory declares that the advantage to recombination is in bringing together beneficial mutations arising in different individuals (Fisher 1930; Muller 1932). Similarly, Muller (1964) proposed that, in a finite population without recombination, the best haplotypes would be lost either by chance

[1]*Corresponding author:* Mathematical Modelling and Genetic Epidemiology Group, Division of Genomic Medicine, University of Sheffield, Royal Hallamshire Hospital, Glossop Rd., Sheffield S10 2JF, United Kingdom. E-mail: m.m.iles@shef.ac.uk

[2]*Present address:* Cancer Research UK, Genetic Epidemiology Division, Cancer Genetics Bldg., St. James's University Hospital, Leeds LS9 7TF, United Kingdom.

or through recurrent deleterious mutations and without recombination could not be recreated—this is known as Muller's ratchet (Felsenstein 1974). Both of these describe populations in negative LD—the first because beneficial mutations are unlikely to appear together and the second because the fittest possible combinations of alleles are lost. In both cases, if the population were infinite all combinations of alleles would occur at their expected frequency, and thus these theories depend on the population being finite.

It was first shown by Hill and Robertson (1966) that even when a finite population is initially in linkage equilibrium, the change in frequency at one locus will be affected by the fitness of the alleles not only at that locus, but also at the surrounding loci even in the absence of fitness interactions. Specifically, negative disequilibrium between favorable alleles is generated by genetic drift (Hill and Robertson 1966; Felsenstein 1974). This interference between loci can be reduced by recombination. Felsenstein (1974) showed by simulation in populations of size $N = 100$ that favorable mutations increased more rapidly when recombination was present to break down negative linkage disequilibrium. Felsenstein and Yokoyama (1976) proceeded to demonstrate by simulation of up to 20 segregating loci in populations of size $N = 100$ and 200 that a modifier allele allowing recombination (as opposed to no recombination) will increase in frequency. Thus they were the first to show that recombination could evolve simply due to the short-term benefits to the individual of a finite population, without requiring group selection arguments. The idea that a finite population can give rise to LD through drift leads naturally to the (in our view erroneous) idea that since drift is increased in smaller populations, the biggest advantage to recombination will be in the smallest populations. This has led to rejection of stochastic models for the evolution of recombination as requiring "special circumstances, such as small or structured populations" (Crow 1999). It is demonstrated here that stochasticity can lead to increased recombination even in very large populations.

Recently Otto and Barton (2001) investigated the sizes of populations over which recombination would be expected to increase. By simulating populations of various sizes, both with and without epistasis, they demonstrated that, in small populations with no mutation and a constant homogeneous environment, recombination rates tend to increase, with this effect mostly explained by drift rather than by epistasis. Unfortunately, their results are limited to quite small population sizes. For larger population sizes ($> \sim 1000$), the advantage to drift is much decreased. This appears unsurprising as drift will decrease as population size increases. However, this suggests that the finite population effect is limited to small population sizes. Otto and Barton (2001) suggest that the effect could be extended to larger populations if they consisted of small subpopulations. Alternatively, populations may go through recurrent bottlenecks, leading to a small effective population size. Additionally, Otto and Barton's (2001) results may explain the increase in recombination that has been observed in laboratory experiments on very small populations when selecting on a particular trait (*e.g.*, Korol and Iliadi 1994).

In apparent contradiction to Otto and Barton's (2001) findings, a recent study on a homogeneous population has suggested that there is an increased advantage to recombination in larger populations (Colegrave 2002). This apparently supports the idea of clonal interference—that recombination is required to combine beneficial mutations that are newly arisen in different individuals as in the Fisher-Muller theory. As population size increases, so will the number of new beneficial mutations arising, and so the advantage to recombination increases. However, in this study, the populations were initially isogenic (no genetic variance) and so drift could act only on newly arisen mutations; thus if the population is small, few new beneficial mutations may arise, limiting the usefulness of recombination. In Otto and Barton's (2001) simulations there is variation in the population initially, so there may be an advantage to recombination even in small populations.

Thus those who look at the effect of drift on standing variation in populations see an advantage in smaller populations where drift is strongest and those who look at the effect of drift on newly arisen mutations see a greater advantage in larger populations where more new alleles will arise. Felsenstein has argued that the two are based on the same effect—the generation of negative LD through drift (Felsenstein 1988).

For tractability and ease of simulation, many previous models have focused on only two selected loci. However, in natural populations, selection is likely to be acting not on two, but on many loci at once. Here, the effect of increasing the number of selected loci on the evolution of recombination in finite populations is investigated by computer simulations.

We simulate a general model—a number of diallelic loci, at each of which one allele is fitter than the other(s). The population size and allele fitnesses remain constant over time and there are no mutations. In a randomly generated finite population, alleles at any pair of loci will on average appear together at their expected frequencies. However, given enough loci, it is impossible that all combinations of alleles would appear; thus initially there will be some LD in the population due to the lack of some combinations of alleles in the population. For instance, for 20 diallelic loci there are $2^{20} = 1{,}048{,}576$ possible haplotypes; so in a population of $<524{,}288$ diploid individuals, some haplotypes must be missing.

The effect of drift will depend not only on the population size, but also on haplotype frequencies and fitnesses. We predicted that the more loci that selection

acts upon, the smaller will be the observed variance in fitness compared to its potential range, and so the greater will be the advantage to recombination. Thus, as the number of loci increases, so recombination may be selected for at larger population sizes. An alternative way of thinking of this effect is that as the number of loci increases, so the average frequency of any particular haplotype will decrease. Thus, since stochasticity has a greater effect when numbers are low, the effect of drift will be greater as the number of loci increases.

This is the first time, to our knowledge, that the quantitative effect of the number of loci being selected on has been specifically and fully investigated in relation to the evolution of recombination in different population sizes.

METHODS

We followed a similar model to that of Otto and Barton (2001). Populations of $2N$ chromosomes were simulated. Each chromosome consisted of m loci. At each locus were two alleles, labeled 1 and 0. The fitness of the 1 allele was $(1 + s)^{\alpha}$ times the fitness of the 0 allele, where $s \geq 0$ and α is a function of m. The alleles at each locus were generated independently—allele 1 having a frequency of p. This is equivalent to randomly sampling haplotypes from an infinite population, thereby generating some initial LD due to the lack of some haplotypes. The larger the population, the smaller the effect sampling will have. Denoting the number of "1" alleles on a chromosome by k, the fitness of a chromosome was equal to $(1 + s)^{\alpha k}$. Using this function means that fitness is multiplicative within chromosomes, so there is no epistasis. The initial variance is then approximately $\alpha^2 s^2 mp(1 - p)$ for small s, and the possible range is $(1, (1 + s)^{\alpha m})$.

Here we describe our standard simulations. Alternative assumptions are explored in results. If $\alpha = \sqrt{10/m}$, then the fitness function is $(1 + s)^{\sqrt{10k}/\sqrt{m}}$, initial variance is $\sim 10 s^2 p(1 - p)$, and possible range is $(1, (1 + s)^{\sqrt{10m}})$. Thus, regardless of the number of loci under selection, the initial variance in fitness of the population would be approximately the same. We use this as our standard fitness function. Note that using this fitness function, the potential range in fitness is greater when there are more loci.

Individuals were assumed to be diploid, and the fitness, f_i, of the ith individual ($i = 1, \ldots, N$) was equal to the product of the fitnesses of its two chromosomes. Hence fitness was multiplicative over both loci and chromosomes. The probability of an individual in the current generation contributing a chromosome to an individual in the next generation was proportional to its fitness. Thus, the probability that a randomly selected chromosome in the next generation was from the ith individual in the current generation was $f_i / \sum_{i=1}^{N} f_i$. Each subsequent generation was produced by random mating with replacement according to these probabilities [the Wright-Fisher model (Fisher 1930; Wright 1931) with selection].

The probability of recombination between adjacent loci was constant within individuals. The recombination rate between any two adjacent loci in the ith individual is denoted θ_i. At the end of each chromosome there was an $m + 1$th locus that did not affect fitness but dictated the value taken by θ_i. $\theta_i = 0.001 + 0.0005\delta$, where δ = number of copies of allele 1 at locus $m + 1$ (so $\theta = 0.001$, 0.0015, or 0.002). Here even with 25 loci, the maximum recombination rate across the region is 0.05. Thus, if there were a selective advantage to recombination, allele 1 at the $m + 1$th locus would tend to increase in frequency and so recombination would increase. We measured the increase in recombination by calculating the change in frequency of allele 1 at locus $m + 1$ in the population after 50 generations, whereas Otto and Barton (2001) measured the increase in expected recombination rates. Note that the two measures are equivalent because, for the recombination fractions we use, our measure is simply Otto and Barton's (2001) measure scaled by a factor of $(2p' + 1)/2p'$ (where p' is the frequency of the allele increasing recombination). Because selection of the recombination modifier depends on a hitchhiking effect (Strobeck *et al.* 1976), a modifier in tight linkage is simulated. Selection for recombination will be diminished, but not be absent, for looser linkage (as demonstrated below in our results). Note that this model implies that the length of the chromosome considered is proportional (in map length) to the number of loci. Intrinsically, then, there is more recombination with more loci, which only strengthens the conclusion below.

The program was validated using simple examples (*e.g.*, $f_i = f_j \; \forall \; i, j$) and by altering the program slightly to reproduce Otto and Barton's (2001) results.

Initially the frequency of allele 1 (p) at loci 1, . . . , m was 0.1 and the frequency at locus $m + 1$ (p') was 0.5. Populations of size (N) 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 10^3, 10^4, and 10^5 individuals were generated. The number of loci simulated (m) was 2, 5, 10, or 25. Populations of size $N \leq 50$ were replicated 80,000 times, populations of size $100 \leq N \leq 500$ were replicated 40,000 times, populations of size $10^3 \leq N \leq 10^4$ were replicated 10,000 times, and populations of size $N = 10^5$ were replicated 5000 times.

We assume that, given the short timescale involved, mutation would be of relatively minor importance, so it is not included in our simulations. Nor are changes in fitness included as our interest here is simply in the effect on recombination of finite population size; fluctuating fitness, as already discussed, can lead to increased recombination in its own right.

Separating increasing number of loci from other effects: The problem with our approach here is that as

TABLE 1

Effect of increasing *m* on four quantities of interest for three fitness functions

Effect of increasing m	$(1 + s)^{\sqrt{10k}/\sqrt{m}}$	$(1 + s)^{10k/m}$	$(1 + s)^k$
Selection per locus	Decreased	Decreased	No effect
Mean(fitness)	Increased	No effect	Increased
Var(fitness)	No effect	Decreased	Increased
Var(fitness)/mean(fitness)	Increased	Decreased	No effect

the number of loci (m) is increased, a number of other factors also change. It is possible that these factors could be the cause of any differences in selection for recombination observed for different values of m. To investigate this explanation one could either try different fitness functions or examine the potential factors separately. In the three sections below we first look at various fitness functions and then investigate separately the effect of amount of selection per locus and the recombination fraction.

Fitness function: We examined two alternate fitness functions besides our original $(1 + s)^{\sqrt{10k}/\sqrt{m}}$. These were $(1 + s)^{10k/m}$ and $(1 + s)^k$. There is no best function when comparing selection for recombination with different numbers of loci. Each function has its own advantages and disadvantages. Ideally, as the number of loci increases, the fitness function chosen would be such that no other relevant quantity changed, but this is not possible. The quantities that can be controlled for here are the amount of selection per locus, the initial mean fitness, the initial variance in fitness, and the initial variance in fitness divided by initial mean fitness. The effect of increasing m on each of these quantities for the three fitness functions is summarized in Table 1. Although no one measure of fitness keeps all quantities constant in m, for each quantity there exists a fitness function that will hold it constant.

Selection per locus: Using our standard fitness function, the selection per locus will decrease with increasing m. If $s = 0.5$, the selection per locus for $m = 2$, 5, 10, and 25 is 2.476, 1.774, 1.5, and 1.29 (the ratio of the fitnesses of a "good" and "bad" allele, respectively). It may be that less selection per locus leads, in itself, to a change in any selective advantage to recombination at different population sizes (N), so this effect should be investigated. Using our standard fitness function, s was varied between 0.05 and 2, for $m = 2$ and $m = 4$ ($10 \leq N \leq 10^4$). Note that this issue is also addressed by changing the fitness function (above).

Recombination fraction: At the same time as m increases, the overall recombination fraction increases. It may be that increasing the recombination fraction leads to a change in the selective advantage to recombination at different values of N. We would expect increasing recombination to reduce this advantage as it has been demonstrated that selection of a modifier for recombination depends on hitchhiking (Strobeck *et al.* 1976), which requires tight linkage. Nonetheless, we investigated the effect of increasing recombination while holding m constant.

Recombination values were chosen such that the highest recombination rate was twice the baseline for those individuals homozygous for the high recombination allele and the intermediate recombination rate was midway between these. Minimum values of θ ranging between 0.001 and 0.4 were used, despite the highest value being biologically implausible. We looked at the effect of varying θ for $m = 2$ and $m = 5$ ($10 \leq N \leq 10^4$).

RESULTS

The results of our simulations over the various population sizes, chromosome sizes, and selective advantages are shown in Figures 1 and 2 (and Figures A–D in supplementary information at http://www.genetics.org/supplemental/). The trends in these are clear. Increasing the number of loci on the chromosome increased both the proportional change in recombination and the range of population sizes over which recombination appreciably increased. For instance, in Figure 1, where

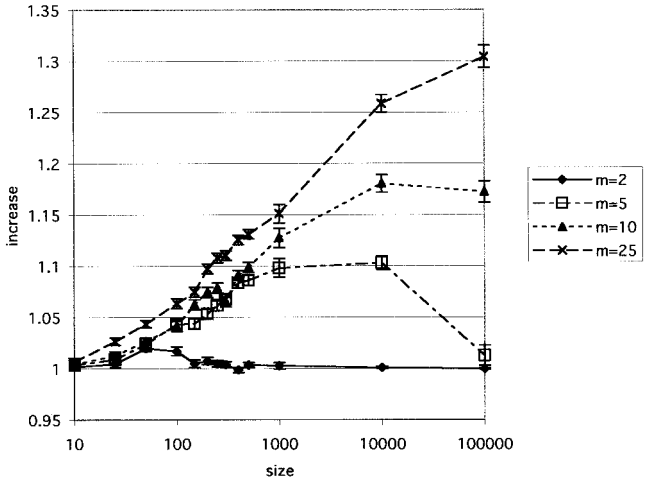

Figure 1.—The effect of number of loci on the evolution of recombination at different population sizes. $s = 0.5$; $m = 2, 5, 10, 25$; $N = 10$–10^5. The y-axis shows proportional change in frequency of allele 1 at the $m + 1$th locus after 50 generations. Bars indicate ±1 standard error of the mean.

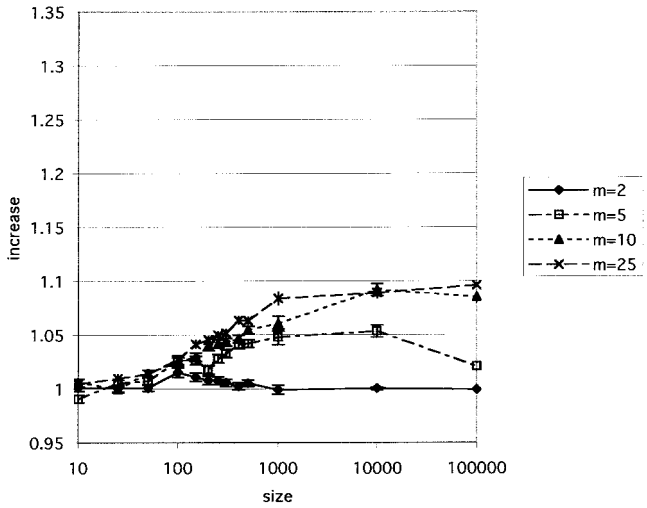

FIGURE 2.—The effect of the number of loci on the evolution of recombination at different population sizes. $s = 0.1$; $m = 2, 5, 10, 25$; $N = 10$–10^5. The *y*-axis shows proportional change in frequency of allele 1 at the $m + 1$th locus after 50 generations. Bars indicate ±1 standard error of the mean.

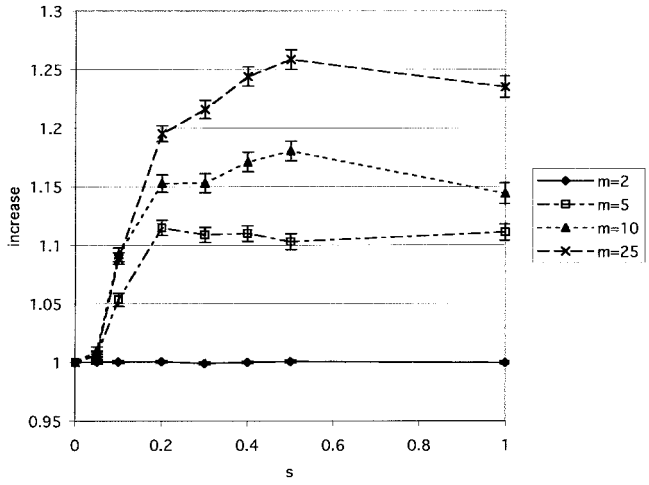

FIGURE 3.—The effect of selective advantage on the increase in recombination at $N = 10^4$. $s = 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0$; $m = 2, 5, 10, 25$. The *y*-axis shows proportional change in frequency of allele 1 at the $m + 1$th locus after 50 generations. Bars indicate ±1 standard error of the mean.

the selective advantage of the best allele at each locus was $1.5^{\sqrt{10}/\sqrt{m}}$, the maximum increase in recombination when the chromosomes consisted of two loci ($m = 2$) was 2%, for $m = 5$ it was 10%, for $m = 10$ it was 18%, and for $m = 25$ it was 30%. The advantage to recombination disappeared (the allele frequencies remain constant at the $m + 1$th locus) when $m = 2$ for $N \geq 150$ and when $m = 5$ for $N = 10^5$. But recombination was still strongly selected for (not much less so than the maximum) at a population size of 10^5 when $m = 10$ or $m = 25$. This pattern was repeated for lower selective advantages (Figure 2 and Figures A–D in supplementary information at http://genetics.org/supplemental/). At intermediate levels of selection ($s = 0.1$ in Figure 2 and $s = 0.4, 0.3, 0.2$, and 0.05 in Figures A–C in supplementary information) the pattern was still that the increase in recombination was generally greater and persisted for larger population sizes when m was larger. As the selective advantage decreased, the increase in recombination became less until, for $s = 1.05$ (Figure D in supplementary information), there seems to be only a slight advantage to recombination whatever the value of m.

Plotting the increase in recombination for $N = 10^4$ against selective advantage for the four chromosome sizes ($m = 2, 5, 10$, and 25; Figure 3), the pattern becomes clear. For $s \leq 0.5$, as selective advantage decreased, the advantage to recombination decreased. Increasing s from 0.5 to 1 showed some evidence of causing a decrease in recombination, perhaps because such strong selection leads to a greater loss of polymorphism. Logically, when $s = 0$ (*i.e.*, no selective advantage), there should be no increase in recombination. This explains why it is that at very small selective advantages there was little increase in recombination. It also suggests that for much larger values of m (the number of loci with alleles differing only slightly in fitness is more likely to number in the thousands) an even greater increase in recombination may be seen in large populations. Unfortunately, due to the computational time involved we were unable to simulate such large genomes.

Changing the fitness function: Clearly, the same pattern in the change in recombination is observed, regardless of the fitness function used (Figure 4). This suggests that the advantage to recombination from increasing m cannot be caused by any one of these quantities (selec-

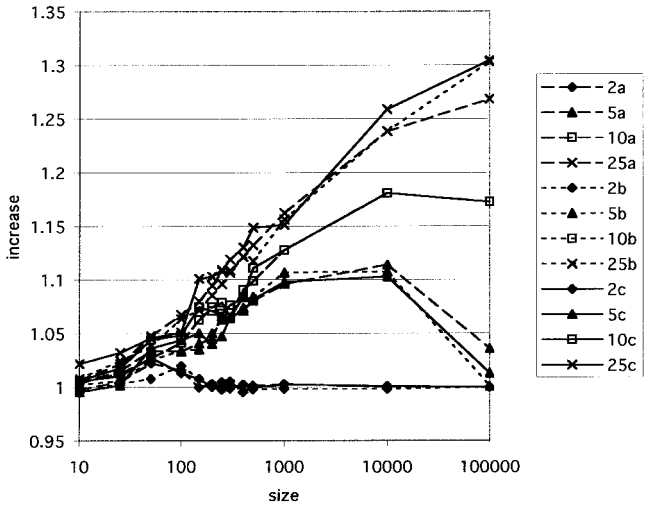

FIGURE 4.—Comparing the effect of the three fitness functions for various values of m directly. *a* (dashed lines), *b* (dotted lines), and *c* (solid lines) represent fitness functions $(1 + s)^{10k/m}$, $(1 + s)^k$, and $(1 + s)^{\sqrt{10}k/\sqrt{m}}$, respectively. $m = 2, 5, 10$, and 25 are represented by the diamonds, triangles, open squares, and crosses, respectively. $N = 10$–10^5, $s = 0.5$. Standard error bars are not included to improve legibility.

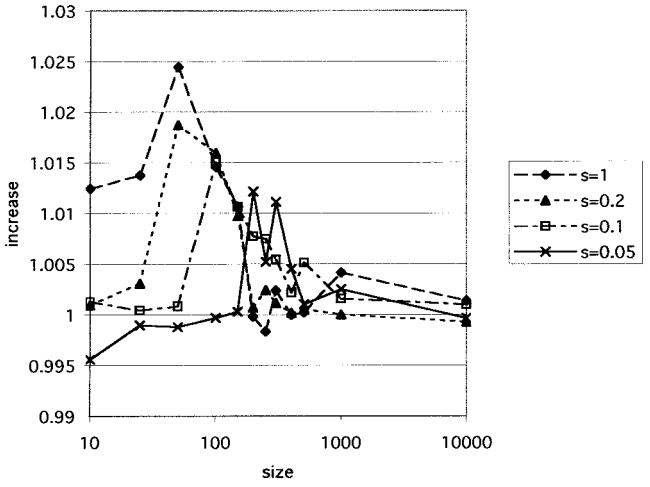

FIGURE 5.—Effect of varying selection for $m = 2$. Change in recombination for $N = 10$–10^4 is shown. $s = 0.05$–1. Standard error bars are not included to improve legibility.

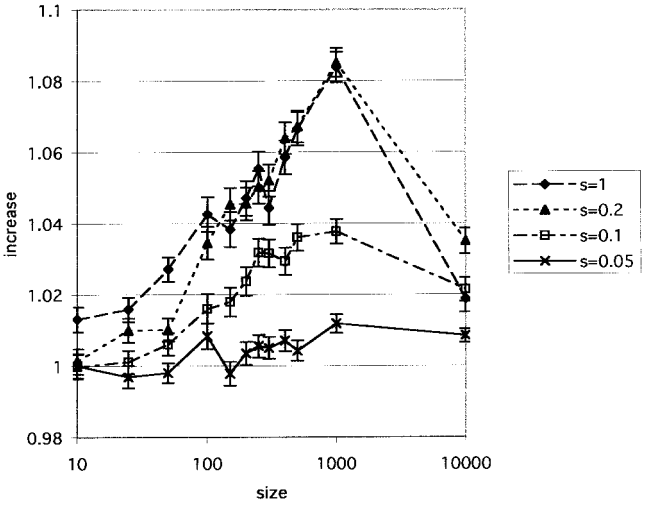

FIGURE 6.—Effect of varying selection for $m = 4$. Change in recombination for $N = 10$–10^4 is shown. $s = 0.05$–1. Bars indicate ±1 standard error of the mean.

tion per locus, mean fitness, variance in fitness, or variance in fitness divided by mean fitness). It may seem surprising that the three different measurements of fitness give such similar results. However, the different fitness functions are all of the form $(1 + s)^{\alpha k}$. Thus, changing the fitness function is equivalent to changing the value of s. Furthermore, for $m = 10$, the three functions are identical. When $s = 0.5$ and the fitness function is $(1 + s)^{\sqrt{10}k/\sqrt{m}}$, changing the fitness function to either $(1 + s)^{10k/m}$ or $(1 + s)^k$ is equivalent to changing s to a value >0.20 for $m = 2$, >0.33 for $m = 5$, and >0.30 for $m = 25$. It is clear from Figure 3 that changing the value of s in these ranges has little effect on the increase in frequency observed for the recombination-modifying locus. At smaller values of s we would be likely to see more of a change but, as seen in Figures 1 and 2 (and Figures A–D in supplementary information at http://genetics.org/supplemental), the effect is likely to be merely equivalent to a scaling down in the magnitude of the effect.

Selection per locus: As seen in Figure 5, when $m = 2$, there is some evidence that the population size at which recombination is most favored is larger for lower values of s, seen as a slight increase in the value of N giving the largest increase in recombination as s decreases. For $m = 4$, no such shift in the peak is observable (Figure 6). However, these shifts are nowhere near great enough to explain the differences in selection for recombination observed for different numbers of loci. When $m = 2$ and $s \approx 0.12$ the selection per locus is the same as for $m = 25$ and $s = 0.5$ (1.29 per locus). Clearly, reducing s does not shift the peak on the curve sufficiently: the peak for $m = 2$, $s = 0.1$ is at $N = 200$ (Figure 5), much lower than the location of the peak for $m = 25$, $s = 0.5$ at $N \geq 10^5$. Furthermore, decreasing s reduces the maximum increase in recombination (Figures 5 and 6), while the opposite is seen when m is increased (Figure 1).

Recombination fraction: For $m = 2$ a greater increase in the frequency of the recombination modifier is observed when the baseline θ (the recombination rate between adjacent pairs of loci) is increased from 0.001 to 0.01, but any further increase in θ leads to a decrease in selection for recombination (Figure 7). No change is seen in the range of N over which there is notable selection for recombination. Similar results are seen for $m = 5$ (Figure 8), but here as θ is decreased, the value of N at which recombination is most selected for is increased. This trend is the opposite of that observed when m is increased (and the total map length is in-

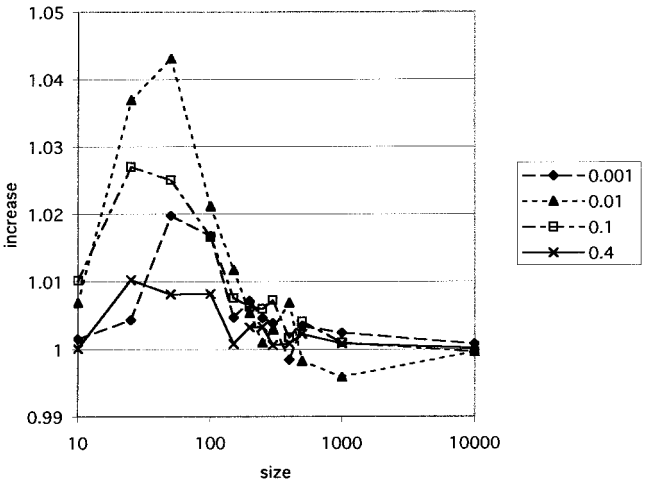

FIGURE 7.—Effect of varying recombination fraction, $m = 2$. Change in recombination for $N = 10$–10^4, $s = 0.5$ is shown. Minimum recombination fraction between consecutive loci = 0.001–0.4. Standard error bars are not included to improve legibility.

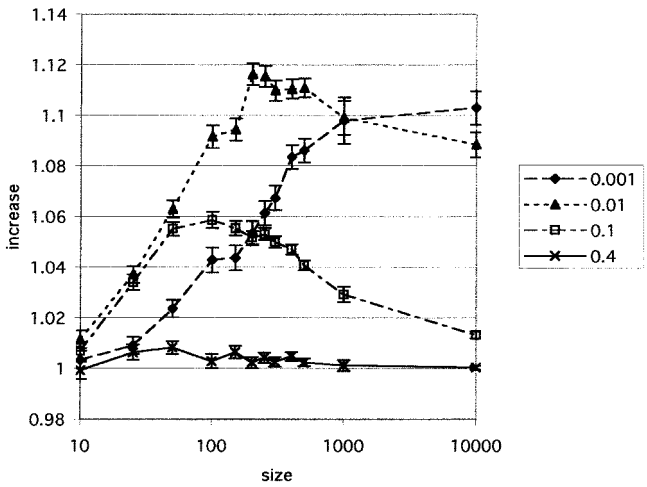

FIGURE 8.—Effect of varying recombination fraction, $m = 5$. Change in recombination for $N = 10$–10^4, $s = 0.5$ is shown. Minimum recombination fraction between consecutive loci = 0.001–0.4. Bars indicate ±1 standard error of the mean.

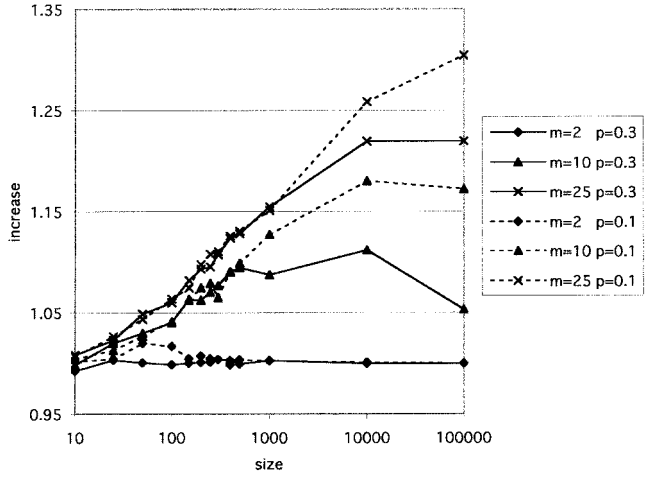

FIGURE 9.—The effect of increasing the frequency of the "fitter" allele, shown as in Figure 1, but with the frequency of the "best" allele now 0.3 (previously 0.1). $m = 5$ and standard error bars not included to improve legibility.

creased). This suggests that increasing the recombination rate cannot explain our original observation that selection for recombination occurs at larger N (Figure 1) with increasing m.

To see the effect of allele frequency we repeated our original simulations (as in Figure 1), but with p increased from 0.1 to 0.3 (Figure 9). The important finding here is that the same pattern is seen with $p = 0.3$ as with $p = 0.1$: the maximal advantage to recombination is seen at larger population sizes as m increases. Also worth noting is that the advantage to recombination decreases for all values of m as p increases. Thus, as the "best" allele becomes more frequent and the variance in fitness increases, the selective advantage to recombination is reduced and disappears at progressively smaller population sizes. The results were similar for $m = 5$, although these are not included in Figure 9 to improve legibility.

THE POPULATION-GENETIC MECHANISMS GENERATING INCREASED RECOMBINATION

The advantage to recombination is in breaking down nonrandom association between loci, thus reducing LD. Fisher's fundamental theorem states that the rate of change of fitness due to natural selection in sexual populations is proportional to the additive genetic variance in fitness. Following from this, BARTON (1995) demonstrated that the amount of selection acting on a modifier of recombination is proportional to the ratio of the variance in log fitness in the population if there were no LD compared to the observed variance in log fitness. If this ratio is >1, then the modifier for increased recombination is favored.

We simulated populations under our standard model, measuring variance in expected ln(fitness) when there is no LD (on the basis of the observed allele frequencies), divided by the observed variance. ln(fitness) is $(\sqrt{10}k/\sqrt{m})^{\ln(1+s)}$. This takes far longer to simulate than previously, especially as the number of loci increases. Thus it has not been possible to estimate values for $m = 25$.

Looking first at populations of size 10,000 with $m = 5$ and $m = 10$, we found that, in any particular generation, there was a strong correlation between the change in frequency of the recombination modifier and the ratio of ln(fitness) variance with and without LD (Figures E and F in supplementary information at http://www.genetics.org/supplemental/). This is in agreement with BARTON's (1995) results and suggests that the driving force favoring the recombination modifier is the effect of recombination on the variance in ln(fitness).

Averaging the variance ratio for each generation over all the simulations and then taking the geometric mean of these averages over the 50 generations indicates the overall strength of selection for the recombination modifier. A total of 20,000 replicates were produced for population sizes ≤50, 5000 replicates for a population of size 100,000, and 10,000 replicates for the remainder.

We found that, as m increases, the ratio of the variance in expected log fitness under no LD and that observed increases (Figure 10). This means that there is a greater selective advantage to recombination as m increases. This effect alone can explain why an increase is seen in the recombination-modifying allele for greater values of m, but it does not seem adequate to explain why the greatest increase in frequency of recombination is seen at larger population sizes (N) as m increases.

A possible explanation of this observation is found by studying the variance in fitness in the population (Figure 11). For very small population sizes, drift rapidly

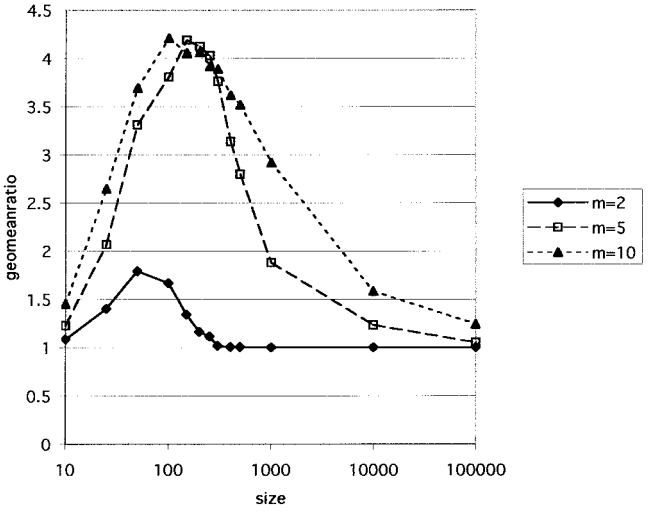

FIGURE 10.—Geometric mean of average variance in expected ln(fitness) when there is no LD divided by the observed variance. $m = 2, 5, 10$. $N = 10$–10^5, $s = 0.5$.

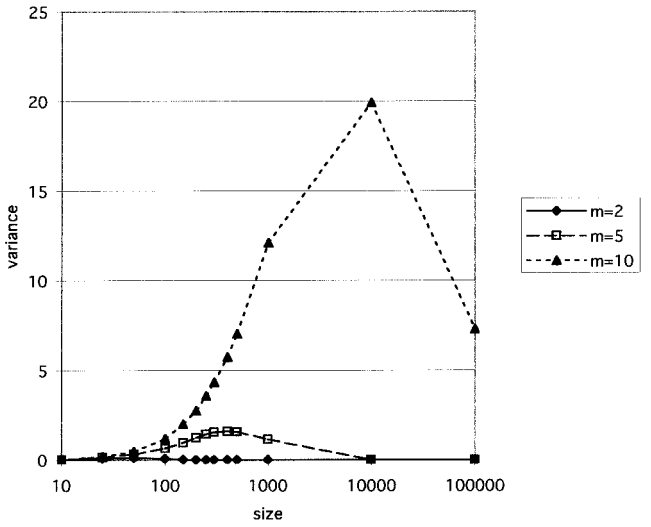

FIGURE 11.—Geometric mean of average variance. $m = 2, 5, 10$. $N = 10$–10^5, $s = 0.5$.

depletes variance. For slightly larger population sizes, there is more variance. If the population is sufficiently big, variance is reduced again. This is because in a large population the range of fitnesses will be greater and the fittest chromosomes will be much fitter—thus the population reaches fixation more quickly.

The maximum average variance occurs at larger values of N when m is greater. For $m = 10$ the maximum is at $N = 10^4$, for $m = 5$ the maximum is at $N = 400$, and for $m = 2$ the maximum is at $N = 50$. The key to understanding this difference in variance comes, we believe, from considering the time until fixation of the population. For a given value of m, if N is sufficiently small, drift has a large effect relative to selection and so fixation is reached quickly. Clearly, when the population is close to fixation the variance in fitness is low and there is little advantage to recombination. For the same value of m, if the population is sufficiently large, then the fittest haplotypes are likely to already be present in the population and so selection acts to push the population rapidly toward fixation. Here, again, variance is rapidly diminished as the population heads quickly toward fixation and there is little advantage to recombination. At intermediate values of N, however, the effect of drift is not great enough to outweigh that of selection and the population is unlikely to include the fittest haplotypes. Thus a number of recombination events are required until the fittest haplotype appears in the population. Fixation therefore occurs less quickly, variance is maintained in the population longer, and recombination is advantageous. Naturally, the more loci there are, the smaller the observed initial variance will be in relation to the potential variance as the fittest haplotypes are less likely to occur. So as m increases, variance will be maintained and recombination selected for at larger population sizes. There may, of course, be other explanations for the results we have observed and the problem requires further investigation.

DISCUSSION

We have confirmed previous findings (FELSENSTEIN and YOKOYAMA 1976; OTTO and BARTON 2001) that recombination may be positively selected simply by having a noninfinite population size. Here we have shown that this is possible for any population size (however large) provided there is selection on sufficient loci.

We have shown that, as the number of loci (m) increases, the population size for which recombination is maximally selected increases. Similarly, as m increases, the range of population sizes for which recombination is noticeably selected increases.

However, there is a confounding effect because as m increases, several other factors change. We have looked at the effect of altering these factors both directly and by controlling for them using alternative fitness functions. Our results suggest that these factors cannot account for the differences we observe in the evolution of recombination for different values of m.

In the final section of this article we looked directly at the population-genetic mechanisms affecting recombination. The most important of these is the ratio in variance in observed log fitness and the variance in log fitness under linkage equilibrium, which is proportional to the strength of selection for increased recombination. Our results show that this effect may explain to an extent the increased range of population sizes for which there is an appreciable increase in recombination, but may not fully explain the change in the population size at which the peak of the change in recombina-

tion occurs. If we plot the geometric mean of the variance in the population, we see that for greater values of m, this has a wider range, a higher peak, and a larger population size (N) at which this peak occurs.

A possible explanation of our observations is that, although the immediate advantage to recombination is simply in breaking down LD and therefore is dependent on the level and pattern of LD in the population, over the course of several generations the variance may also be an important factor. If the population is very small, then polymorphisms are rapidly lost, and so there is little time over which recombination is advantageous. If the population is too big, the fittest combinations of alleles will either already exist in the population or be created quickly by recombination and so fixation will occur quite quickly. Again this leaves little time over which recombination is advantageous. At intermediate population sizes, however, the population is of sufficient size that fixation will be slow and recombination can produce fitter haplotypes. As the least-fit haplotypes are lost from the population and these fitter haplotypes increase in frequency, so recombination can act again to produce still fitter haplotypes. For a given population size, the variance in fitness relative to the potential range will be smaller for larger m, so fixation will be reached more slowly and the advantage to recombination will be greater.

In summary we propose that, at intermediate population sizes, haplotypes made up of the best combinations of alleles are unlikely to be observed and can be created only through recombination. Thus such haplotypes are more likely to be carrying the higher-recombination modifier, leading to a hitchhiking effect as these fitter haplotypes increase in frequency.

Our model is limited by not investigating the effects of mutation and the changing fitness landscape. Given that both of these act randomly, it is likely that in the long term, as natural selection acts to improve the fitness of the haplotypes in the population, both these effects act on average to decrease the fitness of the haplotypes. Thus individuals are constantly improving in fitness, but never reaching the "optimal" fitness; hence, the continued advantage to recombination.

Otto and Barton (2001) suggest that their conclusions be tested experimentally by varying the size of the population being studied to see its effect on increasing recombination. If selection is on a single trait it may be that only a very few loci are experiencing selection pressures and Otto and Barton's predictions may hold, so that recombination rates may increase only for small population sizes. However, if the trait is complex or if several traits are selected for, it may support our predictions, because multiple loci are experiencing selection pressure, and thus recombination would increase even for larger population sizes. Thus we suggest that, as an addendum to Otto and Barton's (2001) suggestions for experimental tests of the drift theory, not only should a range of population sizes be investigated but also the number of loci being selected on. Small values of m can be investigated by selecting on traits that are known to be due to only a few loci, whereas selecting on complex traits or several traits at once would allow investigation of larger values of m. Investigators could also look at reducing the variance in fitness by bottlenecking without selection (*i.e.*, just randomly sampling from a population) and then seeing the effect of this on recombination when selection starts again (we predict that because the variance is smaller with smaller bottleneck sizes, the increase in recombination will be greater).

Note that the model used does not require any special conditions such as fluctuating fitness (in time or space) or epistasis to generate short-term selection for increased recombination. The only requirements, that multiple loci are involved in selection and that the population is not infinite, ought to be built into any realistic model. The main reason that they have not is probably that analytic calculations are much easier if populations are assumed to be infinite and/or selection acts only on two loci (see Otto and Barton 1997 for a model that begins to analyze the problem of drift). It may be that simply by considering multiple loci, recombination could be favored at any population size.

Thanks are given to Nick Barton, Thomas Lenormand, Alexey Kondrashov, and particularly Sarah Otto for their suggestions and insight on this and earlier versions of the manuscript.

LITERATURE CITED

Altenberg, L., and M. W. Feldman, 1987 Selection, generalized transmission and the evolution of modifier genes. I. The reduction principle. Genetics **117:** 559–572.

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

Charlesworth, B., 1976 Recombination modification in a fluctuating environment. Genetics **83:** 181–195.

Colegrave, N., 2002 Sex releases the speed limit on evolution. Nature **420:** 664–666.

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

Eshel, I., and M. W. Feldman, 1970 On the evolutionary effect of recombination. Theor. Popul. Biol. **1:** 88–100.

Feldman, M. W., F. B. Christiansen and L. D. Brooks, 1980 Evolution of recombination in a constant environment. Proc. Natl. Acad. Sci. USA **77:** 4838–4841.

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

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics **78:** 737–756.

Felsenstein, J., 1988 Sex and the evolution of recombination, pp.74–86 in *The Evolution of Sex*, edited by R. E. Michod and B. R. Levin. Sinauer Associates, Sunderland, MA.

Felsenstein, J., and S. Yokoyama, 1976 The evolutionary advantage of recombination II. Individual selection for recombination. Genetics **83:** 845–859.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Hamilton, W. D., 1980 Sex versus non-sex versus parasite. Oikos **35:** 282–290.

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

Jaenike, J., 1978 An hypothesis to account for the maintenance of sex within populations. Evol. Theor. **3:** 191–194.

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

Lenormand, T., and S. P. Otto, 2000 The evolution of recombination in a heterogeneous environment. Genetics **156:** 423–438.

Maynard Smith, J., 1968 Evolution in sexual and asexual populations. Am. Nat. **102:** 469–473.

Maynard Smith, J., 1976 A short term advantage for sex and recombination through sib-competition. J. Theor. Biol. **63:** 245–258.

Maynard Smith, J., 1980 Selection for recombination in a polygenic model. Genet. Res. **35:** 269–277.

Muller, H. J., 1932 Some genetic aspects of sex. Am. Nat. **66:** 118–138.

Muller, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. **1:** 2–9.

Otto, S. P., and N. H. Barton, 1997 The evolution of recombination: removing the limits to natural selection. Genetics **147:** 879–906.

Otto, S. P., and N. H. Barton, 2001 Selection for recombination in small populations. Evolution **55:** 1921–1931.

Otto, S. P., and T. Lenormand, 2002 Resolving the paradox of sex and recombination. Nat. Genet. **3** (4): 252–261.

Peters, A. D., and C. M. Lively, 1999 The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. Am. Nat. **154:** 393–405.

Rice, W. R., 2002 Experimental tests of the adaptive significance of sexual recombination. Nat. Rev. Genet. **3:** 241–251.

Strobeck, C. J., J. Maynard Smith and B. Charlesworth, 1976 The effects of hitchhiking on a gene for recombination. Genetics **82:** 547–558.

Sturtevant, A. H., and K. Mather, 1938 The interrelations of inversions, heterosis and recombination. Am. Nat. **72:** 447–452.

Williams, G. C., 1975 *Sex and Evolution.* Princeton University Press, Princeton, NJ.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

Communicating editor: S. Otto