Selection for recombination in a polygenic model

By J. MAYNARD SMITH

*School of Biological Sciences, University of Sussex,
Brighton BN1 9QG, U.K.*

SUMMARY

Normalizing selection with a moving optimum can result in strong selection for increased recombination. This conclusion is based on computer simulation of an infinite, random-mating diploid population with additive, polygenic determination of the selected phenotype. There can also be selection for increased recombination with a fluctuating optimum, and with directional selection. These effects which arise in an infinite population appear to be large compared to the additional effects arising in a finite population because of the random generation of linkage disequilibrium.

1. INTRODUCTION

Three different situations have been described in which there can be selection for increased recombination:

(i) Sib competition (Williams & Mitton, 1973; Williams, 1975; Maynard Smith, 1976).

(ii) Randomly generated linkage disequilibrium in finite populations (Hill & Robertson, 1966; Felsenstein & Yokoyama, 1976; Strobeck, Maynard Smith & Charlesworth, 1976).

(iii) A fluctuating environment in which selection favours sometimes positive and sometimes negative linkage disequilibrium (Maynard Smith, 1971; Charlesworth, 1976).

These situations were reviewed by Maynard Smith (1977, 1978). The second of them seemed the only one universal enough to counterbalance the selection for reduced recombination which is bound to occur in populations in selective equilibrium in a uniform environment (given that there are epistatic fitness interactions). Sib competition requires a rather special kind of population structure. A fluctuating environment requiring repeated changes in sign of correlations between selectively relevant features seems highly implausible.

The present paper points out that selection for a polygenically determined trait can rather easily lead to changes in the sign of the linkage disequilibrium which is favoured, and hence can select quite strongly for increased recombination. In Section 2 this will be demonstrated for polygenic inheritance in a deterministic, infinite-population model. In Section 3 some of the same phenomena

are shown to occur in a simple 3-locus model. In Section 4 finite population effects are introduced into the polygenic model, in order to compare the magnitude of these effects with those occurring in an infinite population.

2. THE DETERMINISTIC MODEL

The model is of an infinite diploid random-mating population. A phenotypic trait, X, is determined by genes at six loci linearly arranged on a chromosome. Each locus has two alleles, 0 and 1, having equal and additive effects on X. Hence X is an integer in the range 0 to 12. The fitness of an individual, W, depends only on the value of X. A Gaussian fitness model is used,

$$W = \exp\left(-(X-Z)^2/2S^2\right),$$

where Z is the optimum phenotype (which may be constant or changing) and S measures the intensity of selection (a large value of S specifying weak solution). If Z is chosen in the range 0–12 then selection is in part normalizing, whereas if $Z \geqslant 12$ selection is monotonically directional, in the sense that fitness increases with X over the whole phenotypic range.

Genetic recombination is determined by alleles at a seventh locus, terminally situated on the same chromosome. There are two alleles at the locus, CH (high recombination) and CL (low recombination). In CH/CH homozygotes, all loci recombine freely. In genotypes CL/CL and CH/CL, recombination is determined by a parameter r. The probabilities of 0, 1 and 2 crossovers in the whole chromosome are taken as $(1-r)^2$, $2r(1-r)$ and r^2 respectively; double crossovers within a region are excluded. Within each class, crossovers in the six regions are equally likely. The probability c of a crossover between neighbouring genes is therefore $\frac{1}{6} \times 2r(1-r) + \frac{1}{3}r^2 = r/3$. Appreciable computing time is saved by ignoring triple and higher order recombinants.

Each simulation started by specifying initial gene frequencies, and calculating gamete frequencies in linkage equilibrium. Selection was specified by choosing values of Z (constant or variable) and S (constant). Gamete and gene frequencies in subsequent generations were then calculated deterministically according to the laws of genetics.

In the absence of mutation, either normalizing or directional selection ultimately exhausts all genetic variance at the six fitness loci (after which the recombination locus is selectively neutral). In most cases, simulations were not continued long enough for this to matter (typically, 1 minute of computer time gave 16 generations). However, in a few cases a mutation process was incorporated. In the infinite-population model considered in this section, mutation is supposed to occur with equal frequency at the six fitness loci, in both directions, in such a way as not to generate linkage disequilibrium. Thus suppose the frequency of gamete type i before mutation is p_i, and that m is the per site mutation rate. Then the frequency of gamete type i is scaled by $(1-m)^6$; an increment of $p_i m(1-m)^5$ is added to each gamete type derived from i by a single mutation, of $p_i m^2(1-m)^4$ to

each type derived by two mutations, and so on. Correspondingly, p_i is increased by mutation from other gamete types. In Section 4 a more realistic mutation process, involving unique random mutations, is considered.

In each generation the total phenotypic variance, V, was calculated, and also the 'genic variance' (Bulmer, 1971), V_g, which is the phenotypic variance which would exist for the given gene frequencies if there was no linkage disequilibrium. For the present model,

$$V_g = \sum_{i=1}^{6} 2p_i q_i.$$

If (as is always the case in these simulations) $V_g > V$, then there is repulsion linkage disequilibrium; that is, there is an excess of $+ - + -$ chromosomes.

Some results from this model are given in Tables 1 and 2. In runs 1 and 2, Table 1, there was a constant optimum phenotype, $Z = 6$, and the initial gene frequencies were such that the population mean was also at $Z = 6$. Thus there

Table 1. *Normalizing selection and monotonic directional selection in an infinite population*

		Normalizing		Directional	
Run ...		1	2	3	4
Per genome mutation rate	M	0·1	0	0	0
Distance between neighbouring genes	c	0·03	0·03	0·003	0·03
Optimal phenotype					
Initial	Z_o	6	6	12	12
Increase/generation	δZ	0	0	0·5	0·5
Range	S	1	1	7·07	7·07
Initial frequencies of		3 at 0	3 at 0·1	0·01, 0·05, 0·10	
+alleles		3 at 1·0	3 at 0·9	0·20, 0·25, 0·29	
Frequency of CH allele	0	0·50	0·50	0·5	0·5
in generation	10	0·498	0·479	0·508	0·505
	16	0·488	0·474	0·548	0·521
Variances in	V	0·297	0·060	1·953	1·986
generation 10	V_g	0·540	0·212	2·245	2·245

was pure normalizing selection. In run 1 the population started genetically homozygous and variability was generated by mutation during the run, whereas in run 2 the population was initially variable and there was no mutation. In both cases, selection generated a high degree of repulsion disequilibrium, and reduced the frequency of the CH allele. This is in accord with the results reported by Maynard Smith (1979) for a similar model.

Runs 3 and 4, Table 1, simulate monotonic directional selection. An idea of the intensity of selection is given by the fact that, initially, $X = 12$ has a fitness of 1·0 and $X = 2$ of $1/e = 0·42$. The initial population mean was 1·8. In both runs a small degree of repulsion disequilibrium is generated, but there is now selection for increased recombination, particularly in later generations (in fact,

the frequency of CH fell slightly for the first five generations in run 4). Both effects are stronger in run 3, in which linkage in CL/CL and CL/CH genotypes was tighter.

The reason for the increase in frequency of CH under directional selection appears to be as follows. Initially, selection generates repulsion linkages. Hence, when there is strong selection in later generations for 111111 gametes, these are produced mainly as recombinants, and hence carry the allele CH. This interpretation is confirmed in Section 3.

Table 2. *Combined normalizing selection and directional selection in an infinite population*

Run ...		5	6	7	8	9
Mutation rate	M	0	0	0	0	0
Distance between neighbouring genes	c	0·03	0·03	0·03	0·003	0·10
Optimal phenotype						
Initial	Z	2	2	2	2	2
Increase/generation	δZ	0·125	0·25	0·5	0·5	0·5
Range	S	1	1	1	1	1
Initial frequencies of + alleles		0·01, 0·05, 0·10, 0·20, 0·25, 0·29				
Frequency of CH allele in generation	0	0·5	0·5	0·5	0·5	0·5
	10	0·442	0·460	0·590	0·754	0·501
	16	0·397	0·474	0·739	0·972	0·525
Variances in generation 10	V	0·448	0·405	0·502	0·552	0·507
	V_g	1·817	2·035	2·046	2·019	2·027

Stronger selection for the CH allele can arise from a combination of normalizing and directional selection, because higher levels of repulsion disequilibrium are built up. This is shown in runs 5–9, Table 2. Whether selection is for increased or decreased recombination depends on the relative strengths of normalizing selection (reducing CH) and directional selection due to a shifting optimum (increasing CH). Consider first run 7, in which the initial optimum of 2 units was increased by 0·5 in each generation, and the strength of normalizing selection was such that individuals 2 units away from the optimum had a fitness of 0·42. (The initial population mean was 1·8.) With these parameter values there was strong selection favouring CH. In contrast, run 5, differing only in that the optimum was increased by 0·125 units per generation, gave selection favouring lower recombination. In run 6, in which the optimum shifted 0·25 units per generation, selection first favoured CL and later favoured CH.

The other factor influencing the strength of selection is the tightness of linkage in CL/CL and CL/CH genotypes. Runs 7, 8 and 9, which differ only in the recombination fraction c, show that, as expected, the tighter the linkage the greater the effect. However, it is significant that there is still appreciable selection for allele CH even when in its absence the chromosome is 60 centimorgans long (run 9).

There can also be selection for higher recombination with a fluctuating optimum.

No attempt was made to investigate the range of parameters in any detail, but one example will be given to illustrate the possibility. The optimum fluctuated between $X = 3$ and $X = 9$, with a total period of 8 generations per cycle. Mutation rate, $M = 0.1$; selective range, $S = 1.0$; distance between neighbouring genes, $c = 0.003$. Selection was continued for 40 generations, after which the frequency of the CH allele had risen from 0.50 to 0.799.

3. A THREE-LOCUS MODEL

It was suggested in the last section that recombination increased under mono-tonic directional selection because selection first generates repulsion disequilibrium. This effect can be illustrated in a simple three-locus model with constant fitnesses. The phenotype is determined additively by two loci, with two alleles at each, and ranges from 0 to 4. A third locus determines recombination, so that in CH/CH homozygotes there is free recombination between all three loci, and in CH/CL and CL/CL there is no recombination.

The system is started in linkage equilibrium. If fitnesses between loci are multiplicative, then the population will remain in linkage equilibrium indefinitely (Felstenstein, 1965). If individuals with the same phenotype have the same fitness, this requires that fitnesses within loci are also combined multiplicatively, as shown in Table 3.

Table 3

	Locus 2		
	0/0	0/1	1/1
Locus 1 0/0	1	V	W
0/1	V	V^2	VW
1/1	W	VW	W^2

Multiplicative fitness scheme. If $\dfrac{0\ 1}{0\ 1}$ and $\dfrac{0\ 0}{1\ 1}$ have the same fitness, then $W = V^2$.

Table 4 shows some results of simulating this model. As predicted theoretically, if fitnesses are multiplicative there is no linkage disequilibrium and hence no selection for or against recombination. Normalizing selection produces negative and disruptive selection positive linkage disequilibrium, but both select for lower recombination. Directional selection in which fitnesses rise less steeply with phenotype than required by the multiplicative assumption generates negative linkage disequilibrium, but selects for increased recombination.

4. THE SIGNIFICANCE OF FINITE POPULATION EFFECTS

An exact investigation of the significance of finite population effects would require Monte Carlo simulation. However, approximate methods have been used, for two reasons. First, it became possible to use the same basic programme with minor modifications. Second, and more relevant, it became possible to study

separately two different effects of finiteness. These effects are the linkage dis-equilibrium arising from the random loss of particular gamete types, and the disequilibrium arising from unique mutational events. They will be discussed in turn.

Table 4. *Results from 3-locus model*

	Fitnesses of phenotypes 0–4	Frequency of CH allele		Linkage disequilibrium (00) (11)−(01) (10)	
		gen 0	gen 10	gen 0	gen 10
Normalizing	0, 0·5, 1, 0·5, 0	0·5	0·344	0	− 0·221
Disruptive	1, 0·75, 0·5, 0·75, 1	0·5	0·485	0	+ 0·108
Directional					
(a) Additive	1, 2, 3, 4, 5	0·5	0·519	0	− 0·023
(b) Multiplicative	1, 2, 4, 8, 16	0·5	0·5	0	0

(i) *Random loss of gametes*

In a finite population, even if all alleles are present, some gamete types are likely to be absent. Thus in the present simulations the initial frequencies of the + alleles at the six fitness loci were usually 0·01, 0·05, 0·1, 0·2, 0·25 and 0·29. Hence the expected frequency of 111111 gametes was of the order 10^{-6}. Such gametes would be absent in many large but finite populations, and could then only arise by mutation or, more probably, by recombination.

The following procedure was adopted in every generation to simulate this effect in a population of equivalent size E (the meaning of 'equivalent size' is defined by the procedure, which approximately simulates random gamete loss in an actual population whose size is of order E). For any gamete type with frequency $< 1/E$ (i.e. less than two copies in the population), with probability $\frac{1}{2}$ the frequency was set to zero, and with probability $\frac{1}{2}$ the frequency was doubled. The frequency of the remaining gametes was then normalized to add up to 1·0.

In order to measure the effect of this procedure on selection for recombination, it was desirable to compare a population of equivalent size E with an infinite population in which there was no selection for or against recombination. This was done by simulating a population with multiplicative fitnesses. The fitness was therefore taken as $1 \cdot 2^X$, where X is the phenotype. Simulation of an infinite population with this selective regime gave no linkage disequilibrium, and no selection for or against recombination; this can be regarded as evidence that the programme is in fact doing what it is intended to do.

The first four rows of Table 5 show simulations with random gamete loss, and various values of population size E and recombination fraction c between neighbouring alleles. As expected, there is selection for increased recombination. Population size is relatively unimportant in determining the magnitude of the effect. The important parameter is the recombination fraction c. This confirms the finding of Strobeck *et al.* (1976) that hitch-hiking of this kind can select for

increased recombination, but that the effect is small unless recombination in the absence of the selected high-recombination allele is low.

(ii) *Unique mutational events*

If some alleles are missing they can arise only by mutation. In Section 2, simulating an infinite population, mutation was supposed to occur in such a way as to preserve linkage equilibrium. It is more realistic to treat mutations as unique events. To simulate a population of equivalent size $E/2$ (i.e. with E gametes) and a per genome mutation rate M, the following procedure was adopted in every generation. $R = ME$ gametes were selected at random (allowing for their frequencies). In each, a fitness locus was chosen at random and a frequency $1/E$ transferred from the selected gamete type to the gamete type derived from it by changing the allele at the selected locus. If the initial frequency F of the selected gamete was less than $1/E$, then the whole frequency F was transferred. In these simulations, no process of random gamete loss was incorporated.

Table 5. *Multiplicative fitnesses in a finite population*

M	Gamete loss	E	c	0	Frequency of CH in generation 16	32
0	Yes	5000	0·003	0·5	0·500	0·745
		5000	0·03	0·5	0·503	0·548
		50000	0·003	0·5	0·517	0·707
		50000	0·03	0·5	0·509	0·565
0·02	No	5000	0·003	0·5	0·477	0·472
				0·5	0·511	0·623
		5000	0·03	0·5	0·509	0·547
				0·5	0·486	0·507
		50000	0·003	0·5	0·501	0·538
				0·5	0·500	0·557
		50000	0·03	0·5	0·502	0·487
				0·5	0·501	0·515

(Fitness = $1 \cdot 2^X$, where X is the phenotype. Initial gene frequencies in the runs with random gamete loss were 0·01, 0·05, 0·10, 0·20, 0·25 and 0·29, and with mutation were 0, 0, 0, 0·2, 0·3 and 0·4.)

The results are shown in the last 8 rows of Table 5. There is appreciable difference between replicates, particularly for the smaller values of E. However, there is an increase in the frequency of the CH allele in most runs, although the effect is somewhat less than in the simulations with random gamete loss. Again, the value of c is more important than E or Ec in determining the extent of the increase.

In most generations of most runs shown in Table 5 there was a slight degree of repulsion disequilibrium ($V < V_g$), but this was a poor guide to the extent of selection for recombination.

In comparing the magnitude of the changes caused by deterministic (Tables 1

and 2) and stochastic (Table 5) effects, it should be noted that the stochastic runs were continued for 32 generations and the deterministic ones for only 16. A more valid comparison may be of the extent of the change in the mean phenotype during a run. In deterministic runs 7–9 the mean phenotype changed from 1·8 in generation 0 to approximately 8 in 14 generations and 9 in 16 generations. In the stochastic runs in Table 5 it changed from 1·8 in generation 0 to approximately 8 in generation 16 and 11 in generation 32. Thus the intensity of directional selection was approximately the same in the two cases.

(iii) *Combined normalizing and directional selection in a finite population*

As a final check on the relative importance of deterministic and stochastic effects, run 7 (Table 2) was repeated, but with unique mutation ($M = 0·02$) and random gamete loss ($E = 50000$) incorporated. In each of two replicates, the increase in frequency of CH was slightly lower than in run 7. The lower value was presumably a matter of chance, but these runs do confirm the view that stochastic effects are less important than deterministic ones.

5. CONCLUSIONS

The main conclusion of this paper is that normalizing selection with a moving optimum can result in strong selection for increased recombination. The mechanism is as follows. Normalizing selection generates repulsion linkage disequilibrium, and, if continued with a fixed optimum, selects for reduced recombination. However, if the selected optimum changes with time, previously deficient gamete types are favoured by selection. Since these can be produced by recombination, this selects for high recombination alleles linked to the fitness loci.

The effect is a large one. It operates in an infinite random-mating population. It does not depend on linkage in the absence of the high recombination allele being particularly tight. Although the strongest selection for high recombination occurs when there is an optimum phenotype which is changing continuously in the same direction, it is also possible to select for increased recombination with a fluctuating optimum. If selection is monotonic and directional, in the sense that fitness increases monotonically with phenotypic value, there will still be selection for increased recombination if fitness increases less than multiplicatively (e.g. additively). With multiplicative fitnesses (i.e. fitness $= C^X$, where X is the phenotypic value), if the population is initially in linkage equilibrium, it remains so, and recombination alleles are selectively neutral.

An attempt was made to compare the magnitude of this infinite-population selection for increased recombination with selection arising because a population is finite. The latter type of selection arises because, in a finite population, some gamete types are missing altogether, and because new mutations arise singly in unique gamete types. The simulations suggest that the finite-population effects are small compared to those which occur also in infinite populations. In particular, the finite population effects are substantial only when the recombination fraction

is very low in the absence of the allele for high recombination, whereas this restriction does not hold for the infinite population effects. However, this conclusion should be treated with caution, for two reasons. First, unrealistically high values of selection and mutation rates have been assumed. Second, the method of simulating finite effects was only approximate. This latter reservation may not be very important, because the results were not sensitive to the equivalent population size E. Thus provided that mutations were treated as unique events, and that E was small enough to ensure that rare gamete types were lost stochastically, the actual value of E was unimportant.

How important is the infinite-population effect described here in maintaining recombination? Contrary to my earlier opinion (Maynard Smith, 1977, 1978), I now think that this effect may be the major one. Selection for an optimum which changes in time, either undirectionally, or with changes of direction which are not so frequent as to prevent the population from tracking the optimum, may well be a widespread phenomenon. The crucial point is that there is conflict between selection for a reduction in recombination arising from normalizing selection, and for an increase because the optimum shifts. Thus my simulation results resemble the analytical conclusion of Slatkin & Lande (1976) that the direction of selection on a modifier of variance depends on whether fluctuations in the optimum phenotype exceed a certain threshold. Recombination is favoured only if the directional component of selection is large enough compared to the stabilizing one.

REFERENCES

BULMER, M. G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.

CHARLESWORTH, B. (1976). Recombination modification in a fluctuating environment. *Genetics* **83**, 181–195.

FELSENSTEIN, J. (1965). The effects of linkage on directional selection. *Genetics* **52**, 349–363

FELSENSTEIN, J. & YOKOYAMA, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**, 845–859.

HILL, W. G. & ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

MAYNARD SMITH, J. (1971). What use is sex? *Journal of Theoretical Biology* **30**, 319–335.

MAYNARD SMITH, J. (1976). A short term advantage for sex and recombination through sib-competition. *Journal of of Theoretical Biology* **63**, 245–256.

MAYNARD SMITH, J. (1977). Why the genome does not congeal. *Nature* **268**, 693–696.

MAYNARD SMITH, J. (1978). *The Evolution of Sex*. Cambridge University Press.

MAYNARD SMITH, J. (1979). The effects of normalising and disruptive selection on genes for recombination. *Genetical Research*, **33**, 121–128.

SLATKIN, M. & LANDE, R. (1976). Niche width in a fluctuating environment-density independent model. *American Naturalist* **110**, 31–55.

STROBECK, C., MAYNARD SMITH, J. & CHARLESWORTH, B. (1976). The effects of hitch-hiking on a gene for recombination. *Genetics* **82**, 547–558.

WILLIAMS, G. C. (1975). *Sex and Evolution*. Princeton University Press.

WILLIAMS, G. C. & MITTON, J. B. (1973). Why reproduce sexually? *Journal of Theoretical Biology* **39**, 545–554.