THE EVOLUTIONARY ADVANTAGE OF RECOMBINATION. II. INDIVIDUAL SELECTION FOR RECOMBINATION

JOSEPH FELSENSTEIN AND SHOZO YOKOYAMA

Department of Genetics, University of Washington, Seattle, Washington 98195

Manuscript received June 10, 1974
Revised copy received March 11, 1976

ABSTRACT

Based on the FISHER-MULLER theory of the evolution of recombination, an argument can be constructed predicting that a recessive allele favoring recombination will be favored, if there are either favorable or deleterious mutants occurring at other loci. In this case there is no clear distinction between individual and group selection. Computer simulation of populations segregating for recessive or dominant recombination alleles showed selection favoring recombination, except in the case of a dominant recombination allele with deleterious background mutants. The relationship of this work to parallel investigations by WILLIAMS and by STROBECK, MAYNARD SMITH, and CHARLESWORTH is explored. All seem to rely on the same phenomenon. There seems no reason to assume that the evolution of recombination must have occurred by group selection.

THE preceding paper in this series (FELSENSTEIN 1974) reviewed theories of the intrinsic advantage of recombination, starting with the theory presented by FISHER (1930) and MULLER (1932). It pointed out that their conclusion that the presence of sexual recombination would speed the substitution of advantageous mutants follows from the general results of HILL and ROBERTSON (1966). The latter authors found that genetic drift continually produces random linkage disequilibrium. The average effect of these disequilibria is to retard the response to selection. Recombination speeds the response to selection by breaking down this random linkage disequilibrium. The HILL-ROBERTSON effect also predicts MULLER's (1964) "ratchet mechanism", an increase of mutational load in populations lacking recombination. This effect is also the result of random linkage disequilibrium. Computer simulation confirmed the reality of both of these evolutionary advantages associated with recombination.

All of the arguments in the previous paper were implicitly stated in terms of group selection, in that they described an advantage to a whole population which has recombination. MULLER (1932) did not discuss this point, but FISHER (1930) acknowledged explicitly that he was invoking group selection. While he was unwilling to credit group selection with much importance in evolution, he felt forced to make an exception for recombination, which "could be interpreted as evolved for the specific rather than the individual advantage."

A number of recent authors (NEI 1967, 1969; LEWONTIN 1971; FELDMAN 1972; KARLIN and MCGREGOR 1974) have presented convincing evidence that individual natural selection of modifiers of recombination between interacting overdominant polymorphisms should in general act to reduce recombination as much as possible. Their results also imply that there should be group selection against recombination in such cases. To explain the continued presence of recombination (unless it is a necessary byproduct of some "extrinsic" cellular process such as chromosome pairing or DNA repair), without invoking group selection, we must know whether individual selection can favor the presence of recombination.

In this paper, we will demonstrate that such selection can take place. We use a direct argument for some special cases, and computer simulation for others. We then examine three other studies which have come to the same conclusion: the work of WILLIAMS and MITTON (1973); WILLIAMS (1975) of MAYNARD SMITH (1976), and of STROBECK, MAYNARD SMITH and CHARLESWORTH (1976). We will argue that their models also rely on the effects of recombination on random linkage disequilibrium. In fact, there seem to exist only two conceptually different models for the evolution of recombination.

Intra-population Selection for Recombination

Neither FISHER nor MULLER ever presented any argument as to why individual selection, in the cases they considered, should *not* favor recombination. In some special cases, we can use their group selection argument to demonstrate that individual selection will favor recombination. This is possible because in certain cases there is no sharp distinction between individual selection and group selection. This is so when two conditions are met. First, the survival or extinction of a group must depend on the individual fitnesses of the members of the group, and not on any interaction between them. This condition is fully met by the FISHER-MULLER model, in which the advantage of recombination is that it allows the average dosage of favorable alleles per individual to increase more rapidly, and reduces the average dosage of deleterious alleles. The second condition is that if we divide the haploid genomes (or gametes) of the population into two subpopulations according to whether or not they possess the allele in question, that we be able to consider these either as subpopulations or as distinct populations. In other words, we must be able to divide the population into two subpopulations, between which there is no gene flow. If these conditions are met, then if we have an argument which predicts that a group which is fixed for the recombination allele will on average be favored over one which lacks it, this argument will also predict that the frequency of the subpopulation of genomes containing recombination alleles will tend to increase within a single population.

The second condition is met if there is a single locus at which one allele causes recombination and the other abolishes it, and the recombination allele is recessive. In the diploid stage of a haploid-diploid life cycle, no recombination must occur unless the individual is homozygous for the recombination allele. Then any

mutant arising in either subpopulation can never spread by recombination to the other. The FISHER-MULLER argument then predicts that favorable alleles at background loci will tend to increase more quickly in genomes containing the recombination allele. The Muller "rachet mechanism" argument predicts that deleterious alleles will correspondingly increase less rapidly in those genomes. We have not specified the degree of dominance of the mutant alleles at the background loci. As long as they are not over- or underdominant, the subpopulation which incorporates more favorable and fewer deleterious alleles will have a higher average fitness. If selection occurs during the haploid stage, this is equivalent to heterozygote fitness being the geometric mean of both homozygotes, and no difficulty arises. Since in the present case the genomes containing the recombination allele have a fitness advantage, individual selection will favor this allele over one which does not permit recombination.

But the argument cannot really be this simple. Suppose that, at the background loci, favorable mutants are occurring. Whichever of the two subpopulations is momentarily larger is the one more likely to accumulate the next favorable mutation. Thus there should be an inherent instability of the gene frequency of the recombination allele: whichever allele is more frequent should tend to increase to fixation. However, the other allele will presumably be re-introduced into the population by mutation. The frequency of the recombination allele should therefore oscillate rather wildly. If forward and backward mutation rates were equal, then in the absence of any natural selection for the recombination allele, its long-term expected frequency should be one-half. But there will be selection for the recombination allele. The recombining subpopulation will be somewhat more likely to incorporate the next favorable mutant than one would predict from its relative subpopulation size alone. So the long-term average frequency of the recombination allele will exceed one-half.

The above argument applies only to recessive recombination alleles with favorable mutants occurring at the background loci. When the background mutants are deleterious, the oscillations in the sizes of the subpopulations should be much smaller. The mean rate of accumulation of deleterious alleles per individual will not be very much different in a large subpopulation than in a small one. Nevertheless, there should be some tendency for large populations to have slightly fewer deleterious alleles per individual. This follows from the higher expected gene frequency of deleterious mutants in small populations, since natural selection is more effective in large populations.

When the recombination allele is dominant or partially dominant, qualitative predictions are harder to make. In this case, the two subpopulations will have gene flow at the background loci. Recombination in heterozygotes for the recombination locus will reduce the genetic difference between the subpopulations at the background loci. It will be possible for two favorable mutants to occur in different individuals in the nonrecombining subpopulations, to each cross over into the recombining subpopulation, to there be recombined into the same gamete, and then for this double mutant genome to be reincorporated into the nonrecombining subpopulation. In such a complex situation, recourse to intuition seems

unduly risky. When the two alleles at the recombination locus code for different nonzero levels of recombination, there is a similar difficulty. Whether or not the low-recombination allele is recessive, there will be recombination in the heterozygotes between the two populations. The genetic distinction between them will not be absolute. As in the case of a dominant recombination allele, the situation becomes too complicated to yield readily to intuition.

A further complication arises when the advantageous alleles at the background loci do not arise *de novo* by mutation. If they represent formerly deleterious alleles which have become favorable by an environmental change, recombination will have two contrasting effects. On the one hand, it will increase the probability of fixation of the newly favorable alleles. But these alleles start at gene frequencies which result from the previous equilibrium between their mutation and their elimination by natural selection. The MULLER "ratchet mechanism" predicts that they will be held to lower average gene frequencies in the recombining subpopulation. These lower initial frequencies will presumably lower their chances for fixation once the environment changes. We thus have two antagonistic effects of recombination. Qualitative arguments are clearly not useful in such a case, where we need to know which of two effects is quantitatively more important.

In all of these cases, we have a modifier locus linked to a set of linked background loci each of which is undergoing natural selection and mutation in a finite population. We are many years away from adequate theoretical tools to handle such cases, even very approximately. Our only recourse is simulation.

Computer Simulations

The object of the computer simulations was to discover whether there was any demonstrable selection for the recombination allele. It would have been impractical to try to simulate the long-term evolution of this allele. Instead, a rather unrealistic situation was chosen for its symmetry, to see whether selection would favor the recombination allele. In all cases, the recombination allele had frequency 0.5 in the initial population, all other loci being fixed for the "wild-type" allele. In one case, favorable mutants then occurred at the background loci, in the other case deleterious mutants occurred. As a result of genetic drift and the selection at the background loci, the recombination allele would ultimately either be lost or fixed. The results of the simulations were examined to see whether the probability of fixation differed significantly from one-half. Actual recombination levels must be a complex compromise between the selection described here and the reduction of recombination predicted by NEI (1967, 1969), LEWONTIN (1971), FELDMAN (1972), and KARLIN and MCGREGOR (1974). Our argument can make no prediction about selection among two nonzero levels of recombination. One of our alleles was therefore taken to cause complete linkage. To maximize the chance of seeing an effect, the recombination allele was assumed to cause free recombination.

It proved necessary to write different computer simulation programs for the cases of favorable and deleterious mutants at the background loci. We will first

describe the simulation for the case of favorable background mutants. The population consisted of N haploid genomes, each consisting of the recombination locus and twenty background loci. Two alternative alleles were possible at each locus. The next generation was chosen by the following procedure:

1. Two parents were chosen at random, sampling each with replacement from the N haploid adults. The probability of an individual being chosen was made proportional to its fitness. The fitness of an individual carrying k mutant alleles at the background loci was taken to be $(1 + s)^k$, where s was the selective advantage of a single mutant allele. Thus, the selection was, in effect, differential fertility, which was multiplicative between loci and occurred in the haploid stage of the life cycle.

2. If the two genomes chosen both carried the recombination allele, or if only one carried it and it was assumed to be dominant, a single offspring was produced by free recombination between all loci. Otherwise, the offspring was simply a copy of the first parent.

3. After N offspring had been chosen in this way, they were examined to see if any of the background loci had fixed for either allele. If a locus proved to be fixed for the favorable allele, it was reset to the wild-type allele in all N individuals, so as to be available for further mutation. Since the fitnesses were assumed multiplicative, this resetting of fixed loci to wild-type alleles would have no effect on the relative fitnesses of genotypes within the population.

4. The number of new mutants which were to occur at the background loci was drawn from a Poisson distribution with mean Nu, so that u is the rate of mutation *per genome*. Each mutant was placed in a different one of the available nonsegregating background loci, and in a randomly chosen individual. Since the background loci experienced either no recombination at all or free recombination, this re-use of loci was mathematically equivalent to having each mutant occur at a completely new locus.

5. These steps were repeated each generation until the recombination allele was fixed or lost. In a few replicates this did not happen within the appropriate number of generations, and in these the gene frequency at the recombination locus was recorded at that time. With the exception of these cases, the data recorded were the numbers of replicate runs fixing and losing the recombination allele.

If in even one replicate the number of background loci segregating exceeded 20, all results from that parameter combination were discarded. To discard only those replicates in which more than 20 background loci would have been needed might have biased the results. For example, it might be that those runs which are least likely to fix the recombination allele are most likely to exceed 20 segregating background loci.

In the case of deleterious mutants at the background loci, the simulation procedure was nearly the same. There was only one non-trivial difference. Instead of a maximum of 20 background loci segregating in the population, the maximum was 20 deleterious mutant alleles *per individual*. The genome at the background

loci was simply represented by a list of the deleterious mutants which the individual carried, up to a maximum of 20. Thus the number of loci which could be segregating simultaneously was 20 times the number of individuals. Each mutant was assigned a serial number, so that there was no need for the elaborate procedure of resetting fixed loci to wild-type alleles. As in the case of favorable mutants, recombination was either absent or free between all loci, so that the linkage map position of each locus had no biological significance.

Both programs were written in FORTRAN on the CDC 6400 computer at the University of Washington Computer Center. The word length of this machine is 60 bits. All computations using reals were single precision. The pseudorandom numbers were generated by the multiplicative-congruential method with multiplier 5^{19} modulo 2^{48}.

Results of Simulation—Favorable Mutants

Runs were made for all possible combinations of the following parameter values, each for both dominant and recessive gene action of the recombination allele:

$N = 100$ and 200;
$u = 0.1, 0.3, 0.6$, and 1.0;
and $s = 0.25, 0.50, 1$ and 3.

In all cases 100 replicates were run for each combination of parameter values. For one parameter combination (dominant gene action with $N = 200$, $u = 0.1$, and $s = 0.25$) three replicates were still segregating after 1000 generations each. In this case, since the mean gene frequency at the recombination locus in the segregating populations was 0.28, for the purposes of further analysis two of them were taken to have lost the recombination allele and one to have fixed it.

Table 1 shows the results. Parameter combinations in which more than 20 background loci would have been necessary in order to complete the simulation are indicated by the symbol ">20". In the recessive case, the overall numbers of times that the populations fixed and lost the recombination allele were 1354 fixed : 1146 lost. This is extremely significantly different from the expected numbers based on random fixation ($\chi^2 = 18.93$, d. f. $= 1$, $P < 0.00002$). So there is clear evidence of selection for the recombination allele. This is apparent from a glance at the results, which show a clear excess of cases in which more replicates have fixed than lost the recombination allele.

We can also ask whether effects of N, of u, and of s can be detected. Each must have some influence on the results, since setting either $N = 1$ or $u = 0$, or $s = 0$ would definitely abolish all selection for the recombination allele. Within the range of values of N, u, and s examined, we can ask whether the numbers fixed : lost differ. For N, the contingency table turns out to be:

$N =$	100	200
Fixed	684	670
Lost	616	530,

TABLE 1

Results of simulation of populations in which a recombination locus segregated, and twenty background loci were subject to the occurrence of favorable alleles by mutation

(a) Recessive

	$N = 100$			
u \ s	0.25	0.5	1	3
0.1	48:52	54:46	47:53	44:56
0.3	57:43	52:48	53:47	51:49
0.6	53:47	51:49	59:41	59:41
1	>20	>20	>20	56:44

	$N = 200$			
u \ s	0.25	0.5	1	3
0.1	56:44	53:47	54:46	43:57
0.3	59:41	55:45	49:51	58:42
0.6	>20	>20	61:39	58:42
1	>20	>20	63:37	61:39

(b) Dominant

	$N = 100$			
u \ s	0.25	0.5	1	3
0.1	53:47	45:55	61:39	56:44
0.3	52:48	54:46	64:36	76:24
0.6	59:41	63:37	78:22	86:14
1	>20	>20	86:14	92:8

	$N = 200$			
u \ s	0.25	0.5	1	3
0.1	47:53*	50:50	55:45	67:33
0.3	52:48	58:42	79:21	89:11
0.6	>20	77:23	96:4	94:6
1	>20	>20	>20	98:2

The details of the simulation are described in the text. Results are shown for two types of gene action of the recombination allele (Recessive and Dominant), two haploid population sizes (N), four selection coefficients of mutant alleles at the background loci (s), and four rates of mutation per genome at the background loci (u). The table shows the numbers out of 100 replicate populations fixing : losing the recombination allele. Parameter combinations for which more than twenty background loci would have been required in any replicate are denoted by ">20".

* In this case, 46 replicates fixed the recombination allele, 51 lost it, and three were still segregating at the recombination locus after 1000 generations of simulation. The mean gene frequency of the recombination allele among these three replicates at the end of simulation was 0.28, so for the purposes of statistical analysis two were taken to have lost the recombination allele and one to have fixed it.

giving $\chi^2 = 2.60$, d. f. $= 1$, $0.09 < P < 0.11$, so that there is no clear sign of an effect of varying N in this range of values. For u, the table is:

$u =$	0.1	0.3	0.6	1.0
Fixed	399	434	341	180
Lost	401	366	259	220,

giving $\chi^2 = 11.76$, d. f. $= 3$, $P < 0.01$, so that variation in u over this range has a significant effect on fixation probability of the recombination allele. A trend

toward more selection for recombination of higher values of u is evident. For s, the contingency table is:

$s =$	0.25	0.50	1	3
Fixed	273	265	386	430
Lost	227	235	314	370,

giving $\chi^2 = 0.63$, d. f. $= 3$, $0.87 < P < 0.90$, so that there is no sign of an effect of varying s over this range of values. These three tests may be slightly non-independent, since the exclusion of the cases where a replicate exceeded 20 segregating loci creates a partial confounding of the variables N, u, and s. Thus a run with $s = 0.5$ is more likely to be one with $u \leq 0.3$. In view of this partial confounding of variables and in view of the well known difficulties of interpreting higher-order interactions in multidimensional contingency tables, it would be risky to attempt to test the interaction of N, u, and s from these simulations. We may conclude that the data show selection for recessive recombination alleles, that variation in u over the range from 0.1 to 1 affects the strength of selection, and that there are strong theoretical arguments that sufficiently large reductions in N, u or s will abolish this selection.

For the case of a dominant recombination allele, the overall numbers of replicates Fixed : Lost are 1787 : 813. Tested against equal expected numbers, this gives $\chi^2 = 364.86$, d. f. $= 1$, and $P < 2 \times 10^{-81}$. This may be conservatively described as stupendously statistically significant. As in the recessive case, reduction of N to 1, or of u or s to 0, must abolish all selection for recombination. When we look for evidence of effects of variation in N, u, or s over the range of values simulated, we find significant effects in all three cases.

For N, the contingency table is:

$N =$	100	200
Fixed	925	862
Lost	475	338,

giving $\chi^2 = 9.98$, d. f. $= 1$, $P < 0.002$. For u, the table is:

$u =$	0.1	0.3	0.6	1
Fixed	434	524	553	276
Lost	366	276	147	24,

giving $\chi^2 = 191.87$, d. f. $= 3$, $P \ll 0.0001$. For s, the table is:

$s =$	0.25	0.50	1	3
Fixed	263	347	519	658
Lost	237	253	181	142,

giving $\chi^2 = 171.26$, d. f. $= 3$, $P \ll 0.0001$. So we not only expect to observe effects of varying N, u, and s when each is made small enough, but we actually do observe effects of varying each of them over the range of values simulated. The trend toward more selection for recombination with larger u or larger s is apparent, and larger N also seems to increase selection for recombination. Again, partly

because of confounding due to the omitted cases it would be risky to attempt to test for interactions between N, u, and s.

It should be pointed out that there is no evidence whatever that selection acts against the recombination allele in any of the cases run. Although some individual cases show fewer replicates fixed than lost, none of these cases deviates significantly in this direction from equal expected numbers.

A further comparison of interest can be made. We may inquired whether cases with the same values of Nu and Ns are comparable. We can make three such comparisons in each of the two parts of Table 1. The first two have $Nu = 60$, $Ns = 50$ or 100. The third compares $N = 100$, $u = 0.6$, $s = 3$ with $N = 200$, $u = 0.3$, $s = 1$. In that comparison we are equating cases with equal values of $(1 + s)^N$ rather than Ns. Of the six comparisons, five have a smaller probability of fixing the recombination allele when N is increased. The reality of this phenomenon is confirmed by taking the six 2×2 contingency tables, computing χ^2 for each, taking $\sqrt{\chi^2}$, and appending a sign depending on the direction of the deviation. Each of these values of $\pm\sqrt{\chi^2}$ should be drawn from a standard normal distribution under the null hypothesis. Instead, the sum of all six values is 6.76, which gives $P < 0.003$. So there is some evidence that we cannot equate cases with equal Nu and Ns.

Results of Simulation—Deleterious Mutants

Table 2 shows the results of running all combinations of

$$N = 50, 100;$$
$$u = 0.16, 0.32, 0.64, 1.28;$$
and $$s = -0.04, -0.08, -0.16, -0.32,$$

except for those combinations in which $u/|s| > 4$. This restriction was made in order to avoid replicates ever having more than 20 deleterious alleles at background loci in any individual. It was a successful restriction, in that no case had to be discarded for this reason. One replicate in one case (Dominant, $N = 50$, $u = 0.16$, $s = -0.08$) was still segregating after 300 generations. Since the frequency of the recombination allele was 0.76, in the statistical analysis it was taken to have fixed for this allele. For cases with $N = 100$, fifty replicates per case were run. When $N = 50$, one hundred replicates were run.

Overall, the case of a recessive recombination allele gives 857 Fixed : 643 Lost, so that $\chi^2 = 30.53$, d. f. $= 1$, $P < 3 \times 10^{-8}$. For a dominant recombination allele, we find 757 Fixed : 743 Lost, so that $\chi^2 = 0.13$, d. f. $= 1$, $P > 0.7$. We thus detect selection for a recessive recombination allele, but fail to detect it for a dominant recombination allele.

As for the effects of N, u, and s, we once again have every reason to expect that setting $N = 1$, $u = 0$, or $s = 0$ will abolish selection for the recombination allele. Therefore in the recessive case, each of N, u and s must have some influence if reduced far enough. Since we have failed to detect selection in the dominant case, we cannot argue that reducing N, u, or s would make any difference.

TABLE 2

Results of simulations of populations in which a recombination locus segregated, and background loci were subject to the occurrence of deleterious alleles by mutation

			$N = 50$		
u	s	−0.32	−0.16	−0.08	−0.04
			Recessive		
	1.28	46:54			
	0.64	52:48	58:42		
	0.32	49:51	64:36*	56:44	
	0.16	53:47	54:46	60:40	56:44
			$N = 100$		
u	s	−0.32	−0.16	−0.08	−0.04
	1.28	30:20			
	0.64	31:19	33:17*		
	0.32	31:19	35:15*	34:16*	
	0.16	30:20	25:25	30:20	30:20
			Dominant		
			$N = 50$		
u	s	−0.32	−0.16	−0.08	−0.04
	1.28	60:40			
	0.64	42:58	51:49		
	0.32	45:55	48:52	46:54	
	0.16	58:42	45:55	53:47†	52:48
			$N = 100$		
u	s	−0.32	−0.16	−0.08	−0.04
	1.28	31:19			
	0.64	26:24	24:26		
	0.32	29:21	23:27	23:27	
	0.16	30:20	27:23	22:28	22:28

The details of the simulation are described in the text. Analogous to Table 1, except that 50 replicates were run for cases in which N = 100. Individual cases in which the numbers of replicates fixing : losing the recombination allele deviated significantly from a 50:50 or 25:25 expectation, as determined by a two-tailed test with $\alpha = 0.05$ based on the binomial distribution, are denoted by asterisks.

* In this case, 52 replicates fixed the recombination allele, 47 lost it, and one was still segregating at the recombination locus after 300 generations. At that time the frequency of the recombination allele was 0.76 in that replicate, so that for the purposes of statistical analysis it was taken to have fixed the recombination allele.

We can compare runs having $N = 50$ with those having $N = 100$. For the case of a recessive recombination allele, the contingency table is

$N =$	50	100
Fixed	548	309
Lost	452	191,

giving $\chi^2 = 6.67$, d. f. $= 1$, $P < 0.01$. For the case of a dominant recombination allele, the table is

$N =$	50	100
Fixed	500	257
Lost	500	243,

giving $\chi^2 = 0.26$, d. f. $= 1$, $P > 0.6$. We can detect an effect of the variation in N in the recessive case, but not in the dominant case which is hardly surprising.

The triangular shape of the parts of Table 2 means that u and s are partly confounded. A conservative way of examining their effects is simply to see which individual cases show a significant departure from equal numbers of replicates fixed and lost. In Table 2, all cases showing a departure from 50 : 50 or 25 : 25 significant at $P = 0.05$ are marked with asterisks. Such cases seem to be confined to the recessive cases, and to tend to occur when $u/|s| \geq 2$ and when $|Ns| \cong 10$ (so that $Nu \geq 20$). In the previous paper of this series (FELSENSTEIN 1974) it was suggested that MULLER's ratchet mechanism would operate most strongly when Nu is large and $|Ns|$ is intermediate. The present results seem to bear this out.

As in the case of favorable mutants, we can check whether the results seem to depend only on Ns and Nu. In Table 2 there are six such comparisons possible within each of the two types of recombination genes—recessive and dominant. Each comparison consists of two parameter combinations with the same Nu and Ns. We can test each such comparison by computing a 2×2 heterogeneity chi-square. Since each has one degree of freedom, we take $\pm\sqrt{\chi^2}$, the sign being determined by the direction of deviation in the 2×2 table. Each of these values should have a standard normal distribution under the null hypothesis. We compare the sum of six such quantities with a normal distribution with $\sigma^2 = 6$. In the recessive case, $\Sigma(\pm\chi) = -5.71$, so that $P < 0.02$. Two of the six χ^2 heterogeneity tests are individually significant in this case. But in the case of dominance, $\Sigma(\pm\chi) = 0.007$, so that $P > 0.997$. This "significantly non-significant" value is somewhat disturbing. It could indicate inadequacy of the pseudorandom number generator, or be an effect of the discreteness of the multinomial distribution, or it could simply be a fluke. Once again, the significant deviation (here found only in the recessive case) is in the direction of more selection for recombination in the smaller population, for fixed Nu and Ns.

Relationships Between Theories

There are two previous investigations of within-population natural selection for recombination which seem to present theories which compete with ours. In fact, both turn out to be logically equivalent to the argument presented above. WILLIAMS and MITTON (1973; WILLIAMS 1975) presented a model of individual selection for recombination which envisaged habitats which were invaded by propagules. Within each habitat, a different randomly-chosen genotype is favored. Reproduction is asexual within the habitat. Within each habitat, the best available genotype takes over, completely eliminating all other clones. As each habitat becomes available for invasion, it is colonized by a number of asexually-produced propagules, and an (approximately) equal number of sexually-produced propagules. WILLIAMS (1975) analogizes the results to a lottery: the asexual propagules are like N copies of the same lottery ticket, while the sexual propagules are like N different lottery tickets. Clearly the organism with

the winning genotype is most likely to be sexual. Sexual reproduction is thus strongly favored by individual selection in this model.

MAYNARD SMITH (1976) has presented computer simulations of a population undergoing the sort of process WILLIAMS and MITTON envisaged. He has been able to confirm the validity of their argument. MICHAEL TURELLI (manuscript in preparation) has formulated a mathematical model of the WILLIAMS-MITTON situation, and has derived an approximate relationship between the advantage of the sexual subpopulation and the numbers and sizes of sibships entering the habitats.

The WILLIAMS and MITTON model may appear to be completely distinct from ours, but it is not. Note that their model implicitly assumes that all incoming asexual propagules are sibs. To avoid an asymmetry in their model, they must also assume that all incoming sexual propagules are also full sibs. Note also that the incoming propagules must include at least two asexuals and two sexuals for the lottery-ticket argument to work. The advantage of the sexuals is their greater *within-sibship* genetic variation. Their argument will continue to give the same qualitative result if the number of sibships among each of the two subpopulations of propagules (sexual and asexual) is increased. But the effect gradually lessens, and when the propagules entering a given habitat come from an infinite number of sibships, there is no net advantage of the sexual subpopulation. We can see this by considering the habitats from which the propagules came. Each such habitat will give rise to propagules whose genotypes reflect the temporary optimum in that habitat. By drawing an infinite number of sibships (whether sexual or asexual), we end up with parents whose genotype frequencies are a perfect reflection of habitat frequencies.

Suppose that there were three loci segregating in the species: each with two alleles. We can characterize a habitat by the genotype it favors. Suppose that the organisms are haploid, and the eight possible habitats *ABC*, . . . , *abc* occur in equal frequencies. Then the parents of the propagules will have genotype frequencies 0.125 *ABC*, 0.125 *ABc*, . . . , 0.125 *abc* within both asexual and sexual subpopulations of propagules. Since these genotype frequencies are in linkage equilibrium proportions, the random mating among the sexuals will leave the genotype frequencies among the sexual propagules unaltered. If there are an infinite number of propagules entering a habitat, both sexuals and asexuals will be in the same genotype frequencies, and there will be no net advantage to the sexuals: competition within the habitat will not change the relative proportion of sexuals.

We are thus led to the conclusion that there is selection favoring the sexuals only when the propagules come from a finite number of sibships and have more than one incoming propagule per sibship.

The requirement is essentially that the propagule population entering a habitat have passed through a bottleneck of a small number of parents. This guarantees that there will be linkage disequilibrium within most subpopulations of incoming propagules. The single generation of recombination reduces the amount of linkage disequilibrium among the sexual propagules. When genotype frequencies

are in linkage equilibrium proportions, all genotypes exist in each population. Extreme linkage disequilibrium is often associated with the absence of one or more genotype. The greater linkage disequilibrium among asexuals makes it less likely that they contain the particular genotype favored in that habitat. As is the FISHER-MULLER theory, the model of WILLIAMS and MITTON is dependent on the HILL-ROBERTSON effect: that random linkage disequilibrium has a net negative effect on response to directional selection.

Note that since the sexuals never mate with the asexuals, recombination is effectively recessive in the WILLIAMS-MITTON model. We can closely approximate their model by modifying our own model. We assume a number of populations connected by migration, instead of a single population. The substitutions at the background loci occur as a result of environmental change, with pre-existing deleterious mutants becoming favored. The selection coefficients s are very large, so that the best genotype in each population is virtually certain of taking over. The migrants entering any population tend to be sibs, which requires that there either be a finite number of populations, or that the migration of sibs is correlated. The resulting model is equivalent to the WILLIAMS-MITTON model. There is thus no hard and fast distinction between their model and ours.

STROBECK, MAYNARD SMITH and CHARLESWORTH (1976) have presented another model, involving the "hitch-hiking" effect (MAYNARD SMITH and HAIGH 1973). They assume that an overdominant locus is segregating for two alleles, A and a, and that a single favorable mutant, B, occurs at a linked locus. A third locus, with alleles C and c, controls recombination between A and B. It is itself linked to B with recombination fraction 0.01. STROBECK, MAYNARD SMITH and CHARLESWORTH assume a population of 1000 individuals. The B mutant is assumed to occur initially in complete coupling with one of the alleles at the A locus and one of the alleles at the C locus. Further change of gamete frequencies is taken to be deterministic, although they also did some fully stochastic simulations with no qualitative changes in the results. They find that C is, in fact, favored over c if it produces a higher level of recombination. Their paper should be consulted for further details. For our purposes, we need only make three points.

First, their model clearly relies on randomly-produced linkage disequilibrium (initially, only one of the alleles A, a is present in B-bearing gametes). The advantageous effect of recombination is to introduce the other allele at the A locus into the subpopulation of B gametes. Clearly, the net disadvantageous effect of hitch-hiking is an instance of the HILL-ROBERTSON effect.

Second, they find that selection for the recombination allele C is substantial only when the c allele brings about complete or nearly complete linkage. This is consistent with our arguments which apply only to cases in which the allele for less recombination causes complete linkage.

Third, their recombination allele did not modify its own linkage to B. This parallels the dominant case in our model. In that case, in heterozygotes at the recombination locus there was free recombination between that locus and the others. (That there was no such recombination in one homozygote at the recom-

bination locus is irrelevant for this particular comparison). We were able to detect natural selection for recombination even when the recombination locus was unlinked to the background loci. This calls into question the necessity for tight linkage between these loci which STROBECK, MAYNARD SMITH and CHARLESWORTH inferred. Of course, some linkage disequilibrium between the recombination locus and the background loci is necessary, else there could be no selection for recombination.

If we modify a diploid version of our model to have only two background loci, one with an overdominant polymorphism and the other subject to the occurrence of favorable mutants, and if we assume that the recombination allele is dominant, we closely approximate the model of STROBECK, MAYNARD SMITH and CHARLESWORTH. Clearly the model of WILLIAMS and MITTON, the model of STROBECK, MAYNARD SMITH and CHARLESWORTH and our own model are all special cases of a more general class of models, and all rely on the HILL-ROBERTSON effect. They must be considered to be one model rather than three.

There is, however, an alternative model which can be the basis of a model of individual selection for recombination. It was stated by MAYNARD SMITH (1971), who provided in this case his own strongest competition. This is a completely deterministic model in which linkage disequilibrium results from epistasis rather than from genetic drift. If coupling disequilibrium is favored in some generations and repulsion in others, recombination could be advantageous. MAYNARD SMITH suggested that the condition for recombination to be favorable was that there be a negative correlation between the type of linkage favored in successive generations. An essentially equivalent model was proposed by STURTEVANT and MATHER (1938), though without a quantitative treatment. CHARLESWORTH (1976) has recently presented a quantitative model of the modification of recombination in the presence of a fluctuating environment, confirming MAYNARD SMITH's intuition.

There thus appear to be two distinct classes of models of individual selection for recombination. These nicely parallel the two classes of models of the advantage of recombination to the population (FELSENSTEIN 1974).

We wish to thank JOHN MAYNARD SMITH, MICHAEL TURELLI, and the reviewers for this journal of a previous version of this paper for helpful comments. This research was supported by ERDA Contract AT (45-1) 2225 TA 5 with the University of Washington.

LITERATURE CITED

CHARLESWORTH, B., 1976 Recombination modification in a fluctuating environment. Submitted for publication.

FELDMAN, M. W., 1972 Selection for linkage modification: I. Random mating populations. Theoret. Pop. Biol. **3**: 324–346.

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

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford (Second edition, Dover Publications, New York, 1958).

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

KARLIN, S. and J. MCGREGOR, 1974 Towards a theory of the evolution of modifier genes. Theoret. Pop. Biol. **5**: 59–103.

LEWONTIN, R. C., 1971 The effect of genetic linkage on the mean fitness of a population. Proc. Natl. Acad. Sci. U.S. **68**: 984–986.

MAYNARD SMITH, J., 1971 What use is sex? J. Theoret. Biol. **30**: 319–335. ——, 1976 A short-term advantage for sex and recombination through sib-selection. J. Theoret. Biol., in press.

MAYNARD SMITH, J. and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23–35.

MULLER, H. J., 1932 Some genetic aspects of sex. Am. Naturalist **66**: 118–138. ——, 1964 The relation of recombination to mutational advance. Mutation Res. **1**: 2–9.

NEI, M., 1967 Modification of linkage intensity by natural selection. Genetics **57**: 625–641. ——, 1969 Linkage modification and sex difference in recombination. Genetics **63**: 681–699.

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

STURTEVANT, A. H. and K. MATHER, 1938 The interrelations of inversions, heterosis and recombination. Am Naturalist **72**: 447–452.

WILLIAMS, G. C. and J. B. MITTON, 1973 Why reproduce sexually? J. Theoret. Biol. **39**: 545–554.

WILLIAMS, G. C., 1975 *Sex and Evolution.* Monographs in population biology No. 8. Princeton University Press, Princeton, New Jersey.

Corresponding editor: R. C. LEWONTIN