Genet. Res., Camb. (1984), **44**, *pp.* 199–217
Printed in Great Britain

Deleterious mutations as an evolutionary factor. 1. The advantage of recombination

By ALEXEY S. KONDRASHOV

Research Computer Center of the Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

(*Received 24 August 1983 and in revised form 13 January 1984*)

SUMMARY

A population with u deleterious mutations per genome per generation is considered in which only those individuals that carry less than a critical number of k mutations are viable. It has been shown previously that under such conditions sexual reproduction is advantageous. Here we consider selection at a locus that determines recombination frequency of the whole genome. The value $v = u/\sqrt{k}$ has been found to play the decisive role. When $v < 0{\cdot}35$ the direction of selection for recombination may be different for different cases, but the intensity of selection is always very small. The advantage of recombination becomes considerable when $v > 0{\cdot}5$, its growth under increasing v being approximately linear. If $v > 2$ no less than 95 % of the progeny are bound to die because of the selection against deleterious mutations. Since this seems to be too great a mutation load, we may assume $0{\cdot}5 < v < 2{\cdot}0$ for any sexual population if mutation really maintains crossing-over. Results on selection at a locus which controls mutability provide evidence that v is located within the specified interval if the physiological cost of a twofold reduction of the mutation rate is within the range 10–80 %. A number of consequences of this hypothesis about the mechanism of selection for sex and recombination are discussed.

1. INTRODUCTION

Mutations are the raw material of adaptive evolution. When selection acts only at one locus accumulation of advantageous mutations may be rapid. The rate of simultaneous evolution at many loci is limited by its 'cost' (Haldane, 1957), but under truncation selection this restriction is not essential (Kondrashov, 1983*b*).

The vast number of mutations, however, are deleterious and must be eliminated by selection. In every generation a few new mutations may appear in the genome (Mukai, *et al.*, 1972). Their elimination with selection acting independently at different loci would result in the death of a large proportion of the progeny. Under 'truncation' selection when only individuals having less than some critical (k) number of mutations survive, mutation load is substantially less (Kimura & Crow, 1979). The efficiency of such selection is high only in sexual populations, which may explain the maintenance of sex in nature (Kondrashov, 1982) independent of population number, environmental conditions (Maynard Smith, 1978) or inbreeding level (Shields, 1982).

Since sexual reproduction has no genetical effect without recombination, it seems natural that selection against deleterious mutations should maintain a sufficient recombination frequency (Crow, 1970, 1983; Feldman, Christiansen & Brooks, 1980). Recombination requires genes of two types: those coding for certain DNA metabolism enzymes, and some certain nucleotide sequences (recombination promotors) in which the process is initiated (Whitehouse, 1982).

The genes of the first type define the total recombination frequency, while those of the second type influence only their location. The exact recombination frequency at a given region of the genome may also depend on many different loci (Chinnici, 1971) as well as on the environment (Dishler, 1983). Genes of the first type are of special interest here because if selection does not create and maintain molecular mechanisms of recombination the latter will by no means exist.

We are going to investigate selection at such a locus, assuming that all other loci are situated in one (Section 2) or several (Section 3) chromosomes. The results allow us to suggest parameter values of the mutation process. In Section 4 selection on a locus controlling the mutability of the genome is considered.

2. RECOMBINATION IN A ONE-CHROMOSOME GENOME

The effect of the mutation process on recombination frequency was studied by Feldman *et al.* (1980). They consider three loci A, B and M, situated in that order in the same chromosome. The first two define the fitness of the individual and deleterious mutations may occur there. The third one, with alleles M and m, controls only the frequency of recombination between A and B, with m increasing it. Recombination frequency r between B and M does not depend on genotype. If the difference in fitness between individuals with one and two mutations is not less than that between those with 0 and 1 mutations (an analogue of truncation selection), allele m is advantageous under small r, and with the rise of r it may become maladapted (under some parameter values this happens with $r > 0{\cdot}5$). We are going to consider evolution of a locus which determines total recombination frequency and its effect on the population, under similar assumptions.

(i) *Model*

As before (Kondrashov, 1982), consider a large population with discrete generations having the life-cycle: mutation–mating–recombination–selection–mutation, assuming that selection works at the haploid stage. An individual genome contains a large number of loci subject to mutation, recombination and selection. Neglecting the loci situated near M, let us assume a coefficient of recombination r between M and all other loci of 0·5.

Mutations occur independently at all the loci, except M, with the rate u per genome per generation. Individuals carrying k or more mutations die. Later on, an important role will be played by $v = u\sqrt{k^{-1}}$. A definite number of crossovers (l) occur randomly in the zygote. With $l = 0$ linkage is complete, and with $l = \infty$ there is free recombination. Locus M alleles will be designated as M_l according to the number of crossovers induced by them; allele M_∞ is designated as m. In every

case the population will contain only two alleles: m and one of M_l and it is assumed that one of them is dominant. As before, the frequency of individuals carrying i mutations before mutation, recombination and selection will be designated as q_i, q_i', q_i'', respectively. For frequencies in the next generation capital letters will be used; the symbol $\sim$ defines stationary distributions. Let the frequency of allele M_1 in individuals carrying i mutations be p_i, p_i' and p_i'', respectively; and the frequency of individuals with i mutations among the ones having allele M_l be y_i, y_i' and y_i''. This means, e.g., that $y_i = p_i q_i p^{-1}$, where $p = \Sigma p_i q_i$ is a frequency of M_l in the population. Fitness of an individual with i mutations is determined by $s_i = 1-(i/k)^\alpha$, where $\alpha = 1, 2, \infty$ correspond to linear, intermediate and threshold selection (Kimura & Maruyama, 1966; Kondrashov, 1982). The proportion of viable progeny, i.e. the value of $\Sigma_{i<k} q_i'' s_i = \bar{w}$, will be called the average population fitness.

(ii) *Qualitative considerations*

Let a stationary population with threshold selection first contain only allelle m. Then the model is identical to that of the previous paper. Assume that a rare dominant allele M_o appears randomly in some generation after mutation. $p_i' = \epsilon$, or in other terms, $y_i' = \tilde{q}_i'$. What would its fate be? As there is no recombination in zygotes $M_o m$, $y_i'' = y_i' = \tilde{q}_i'$. On the other hand, as allele M_o is rare and its effect on the population is small, $q_i'' = \tilde{q}_i''$. Consequently, the frequency of M_o will rise in the next generation if

$$\sum_{i<k} \tilde{q}_i' > \sum_{i<k} \tilde{q}_i''. \tag{1}$$

As recombination does not change the average mutation number per individual but just increases its variance, so that $E\tilde{q}_i' = E\tilde{q}_i''$, $V\tilde{q}_i' < V\tilde{q}_i''$, where E and V denote the mean and variance of the corresponding distributions of number of mutations, respectively. Assuming all the distributions to be symmetrical, one may conclude that (1) is valid if most of the individuals escape the threshold, i.e. if $E\tilde{q}_i' < k-1$ (for illustration see Fig. 1*a*). Reasonably, because of symmetry, this implies $\bar{w} > 0{\cdot}5$; in this case allele M_o will spread. In the opposite case when $E\tilde{q}_i' > k-1$ (and $\bar{w} < 0{\cdot}5$) the increase of variance due to recombination raises the proportion of the survivors (Fig. 1*b*) so that allele M_o is maladapted right after its appearance. The same may be said about any allele M_l because restriction of recombination always results in $Vy_i'' < Vq_i''$.

Since $Vy_i'' < Vq_i''$, the same selection will lead to a smaller decrease of mutation number in individuals with M_l that in those with m. It follows that $EY_i > EQ_i$, i.e. in the next generation already, allele M_l will meet a loaded genetic background. The decrease of M_l fitness will be the greater the smaller r, i.e. the more M_l is connected with the consequences of its action.

Consider a rare dominant allele M_o with $r = 1$. In fact, $r \leqslant \frac{1}{2}$, however, this limiting case is instructive. If $r = 1$, the heterozygote $M_o m$ always transmits to its offspring an allele M_o obtained from one of its 'parents' together with all other loci from the other parent. Then after recombination M_o will always appear in a genome that has been subjected to free recombination in the previous generations. Therefore the frequency of M_o grows steadily under $\bar{w} > 0{\cdot}5$. This growth will lead

to a decrease of $\bar{w}$. M_0 fitness may then decrease, which would result in polymorphism. If $\bar{w} < 0{\cdot}5$ then allele M_0 is always ousted.

Under $r = 0{\cdot}5$, the rare dominant allele M_0 is located with probability $(0{\cdot}5)^n$ in a genome that has been congealed for n generations. As will be shown later, this leads to a decrease of M_0 fitness under given $\bar{w}$ as compared to the $r = 1$ case, though these cases are qualitatively identical.

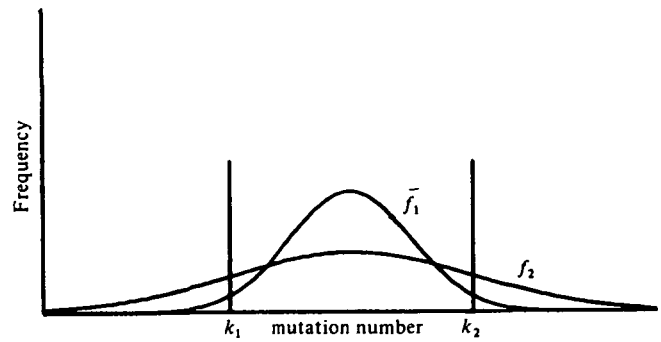

Fig. 1. Threshold selection. The critical numbers of mutations (k) are indicated by the vertical lines. The two distributions (f_1, f_2) have equal means and $Vf_{1i} < Vf_{2i}$. (a) $k = k_1$, i.e. the truncation point is to the left from the mean of the distributions. The growth of variance here leads to the increased proportion of the surviving individuals. (b) In the opposite case of $k = k_2$, $\sum_{i<k} f_{1i} > \sum_{i<k} f_{2i}$.

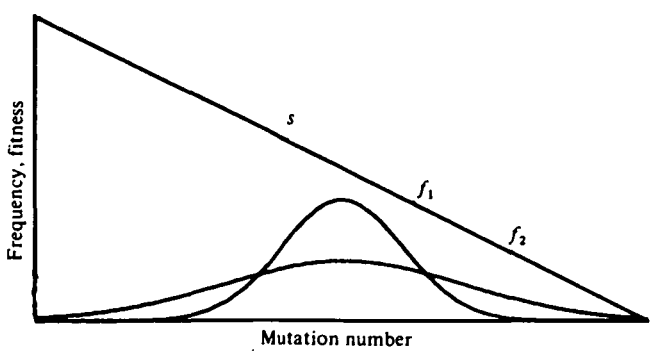

Fig. 2. Similar to Fig. 1 but selection is linear. Here $\sum s_i f_{1i} = \sum s_i f_{2i}$.

It is interesting to see whether allele m can be completely eliminated. If this happens then the population in fact will become asexual, and $\bar{w} = e^{-u}$ in the stationary state (Kimura & Maruyama, 1966; Kondrashov, 1982). After selection all individuals will carry exactly $k-1$ mutations. Let allele m appear with low frequency. Then in some zygotes recombination will occur. If the variance of mutation number in sexual progeny is large enough, then the probability of survival of these zygotes' offspring is about 0·5. Thus the frequency of allele m may grow if $e^{-u} < 0{\cdot}5$. Its fitness would further grow a little because it should appear against a better genetic background. Therefore, a fixation of allele M_0 may be expected to be under $u < -\ln(0{\cdot}5) \simeq 0{\cdot}69$.

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

The above refers to threshold selection. In the opposite case of linear selection, $Vy_i'' < Vq_i''$ will not result in a difference in survivability of M_o and of m individuals (Fig. 2). Hence, allele M_l will never enjoy a short-time advantage through reducing recombination. Since $EY_i > EQ_i$, as before, allele M_l may be expected to be eliminated under all conditions.

(iii) *Equations*

Under $p_i = 0$ the model repeats the one previously described, (Kondrashov, 1982). It is more complicated in the general case, but the approach to formulating the equations remains the same. The mutation process is described by:

$$q_i' = e^{-u} \sum_{j \leqslant i} q_j \frac{u^{i-j}}{(i-j)!}, \quad p_i' = (q_i')^{-1} e^{-u} \sum_{j \leqslant i} q_j p_j \frac{u^{i-j}}{(i-j)!}. \tag{2}$$

For the offspring from mating between individuals with g and h mutations with l crossovers in the zygote, the frequencies of individuals with i mutations will be $b_l(g, h, i)$. Obviously,

$$\left.\begin{aligned} b_\infty(g+h, i) &= \begin{cases} (0{\cdot}5)^i (0{\cdot}5)^{g+h-i} C_{g+h}^i, i \leqslant g+h, \\ 0; i > g+h, \end{cases} \\ b_0(g, h, i) &= \begin{cases} 0{\cdot}5, \quad i = g \quad \text{or} \quad i = h, \\ 0; \text{ other } i, \end{cases} \end{aligned}\right\} \tag{3}$$

Under $0 < l < \infty$, function b_i and its derivation are presented in the Appendix, under the assumption that mutations and crossovers are randomly distributed along the chromosome (see also Fig. 3). Although in nature the number of crossovers in zygote is a random variable, we assume this number to be determined by the genotype. After random mating and recombination, we have

$$\left.\begin{aligned} q_i'' &= \Sigma_g \Sigma_h q_g' q_h' [b_\infty \simeq a_1 + b_x(a_2+a_3) + b_l a_4], \\ p_i'' &= (q_i'')^{-1} \Sigma_g \Sigma_h q_g' q_h' [0{\cdot}5 b_x(a_2+a_3) + b_l a_4], \end{aligned}\right\} \tag{4}$$

where b_l indices are omitted, $x = \infty$ if m is dominant, $x = l$ if m is recessive, $a_1 = (1-p_g')(1-p_h')$; $a_2 = p_g'(1-p_h')$; $a_3 = (1-p_g')p_h'$; $a_4 = p_g' p_h'$. After selection

$$Q_i = q_i'' s_i \bar{w}^{-1}; \quad P_i = p_i''. \tag{5}$$

In the same way, more than two alleles of M, incomplete dominance, $r = 0$, and the case of $r = 1$ under dominant M_o may be considered. Calculations were made on an ES 10–40 computer.

(iv) *Results*

The quantitative investigation allows us to suppose that selection at locus M depends on the value of $\bar{w}$. Computation results are given in Fig. 4. Threshold selection, besides, allows an analytical estimation. Let mutations at different loci be distributed independently after recombination. Then q_i'' under large k may be

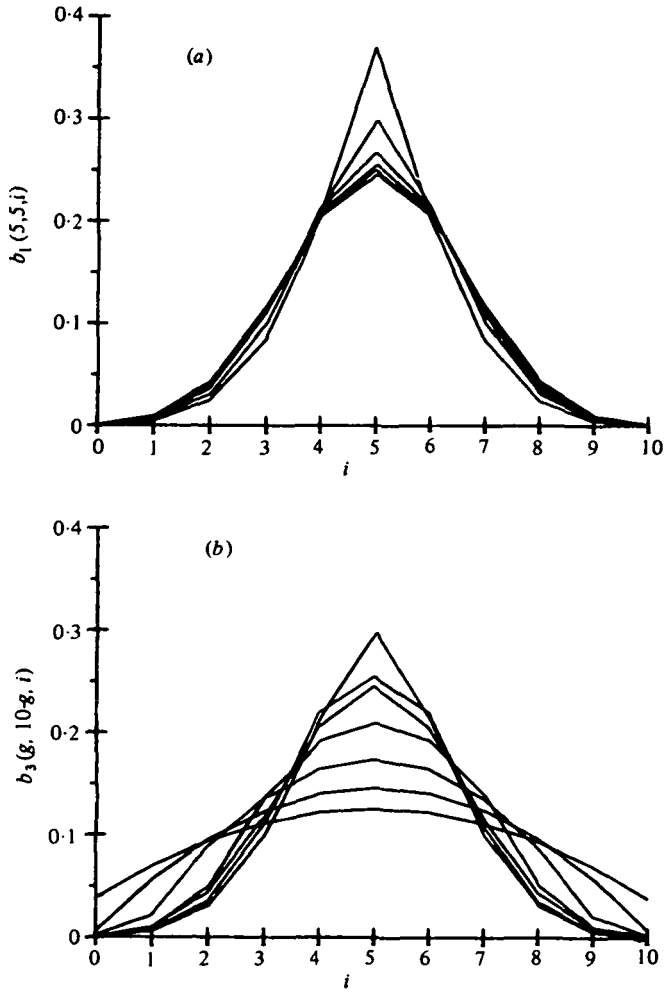

Fig. 3. Function $b_l(g, h, i)$. Compare it with $b_\infty(g+h, i)$ for $g+h = 10$. (*a*) $g = h = 5$. In the order of maximum increasing: b_∞; $l = 31, 15, 7, 3, 1$. (*b*) $l = 3, g+h = 10$. In the order of maximum increasing: $g = 10, 9, 8, 7$; b_∞; $g = 6, 5$.

regarded as normally distributed, with $Eq_i'' \simeq Vq_i'' \simeq k$. Since the selection differential of a stationary population is u, we have

$$F(\bar{w})\,\bar{w}^{-1} = \frac{u}{\sqrt{k}} = v,$$

which follows from equations (11) and (16) of Kimura & Crow (1978). Here $F(\bar{w})$ is the ordinate of a standard normal distribution at a point with the area under the plot to the right of it equal to $\bar{w}$. It is clear from Fig. 4 that our assumptions lead to but a small overestimation of $\bar{w}$.

Let v^* be the value of v under which fixation for allele m replaces polymorphism of locus M. Stationary frequencies of recessive (*a*) and dominant (*b*) alleles M under threshold selection are presented in Fig. 5, for $k = 20$, $r = 0{\cdot}5$. Evidently, $v^* \simeq 0{\cdot}35$

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

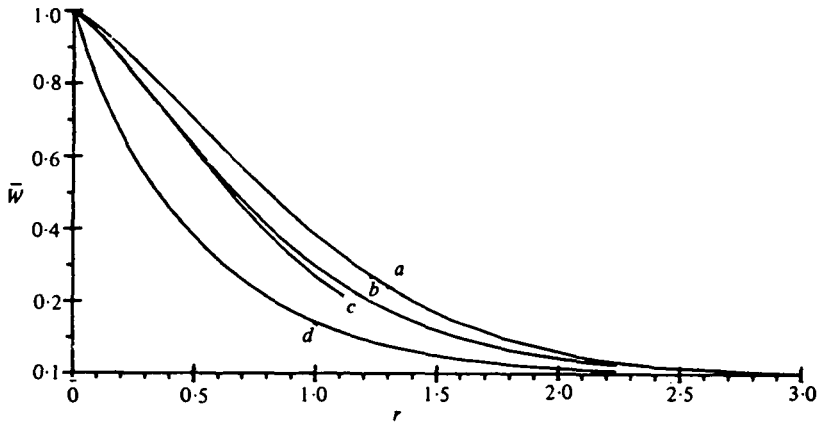

Fig. 4. Dependence of $\bar{w}$ of a stationary population with free recombination on v. (*a*) Analytical estimation; (*b*) threshold selection, $k = 20$; (*c*) threshold selection, $k = 80$; (*d*) intermediate selection, $k = 20$.

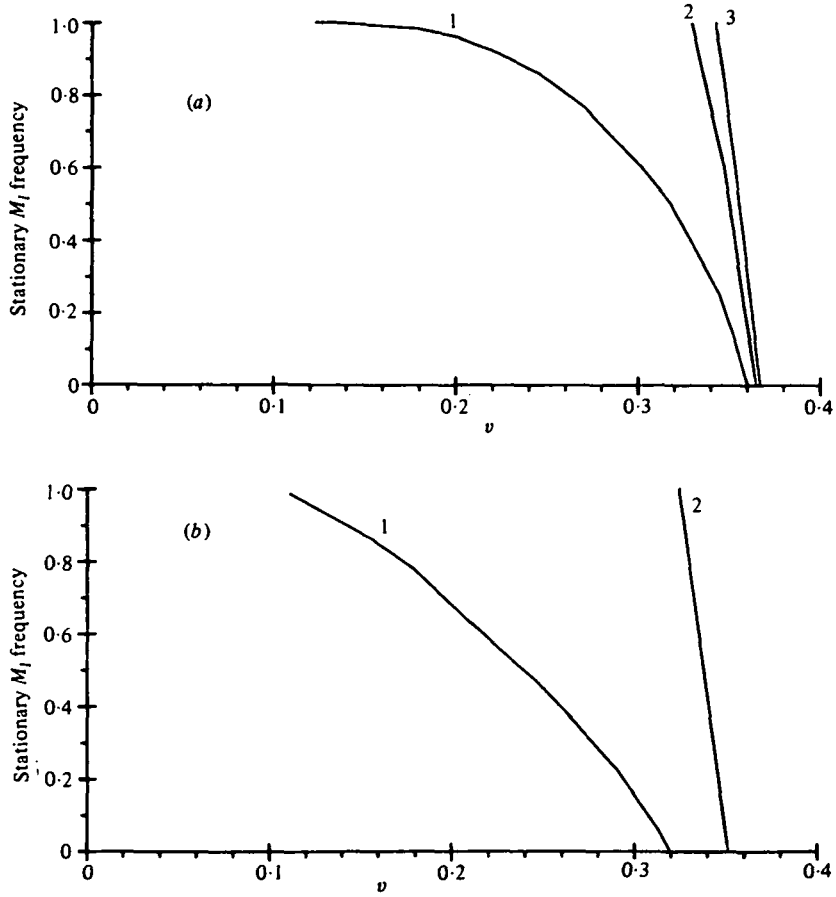

Fig. 5. Stationary frequencies of alleles M_0 (1), M_1 (2), and M_2 (3). (*a*) M_l is recessive; (*b*) M_1 is dominant.

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

in all cases. Under the same conditions, but with $k = 5$ and $k = 80$, the value of v^* remains practically unchanged. On the other hand, for $k = 5$, 20, 80 fixation for M_o replaces polymorphism with values of u close to 0·5 (data not presented), which is also in line with the qualitative conclusions.

The average fitness of a stationary population polymorphic for allele M_l is within the range of 0·73–0·77 under any value of v. With v gradually decreasing, the population fitness starts growing only after fixation of M_l (data not presented).

In a hypothetical case $r = 1$ (Section 2ii) $v^* = 0{\cdot}51$. For such a v the mean fitness of a population with free recombination is 0·61. The difference between the observed and the expected ($\bar{w} = 0{\cdot}5$) critical value of $\bar{w}$ may be due to disequilibria between mutations remaining after recombination, which lead to $Vq_i'' < Eq_i''$.

Polymorphism zones with $l \neq 0$ are narrow, because in this case even fixation of allele M_l would not result in any considerable decrease both of Vq_i'' and $\bar{w}$. For $u = 2$ and $k = 20$, the average fitness of a stationary population with fixed allele M_l is 0·135, 0·619, 0·639, 0·650, 0·656, 0·662 when $l = 0$, 1, 3, 7, 15 and ∞, respectively (results for $l = 0$ and $l = \infty$ are from Kondrashov, 1982, tables 1 and 2).

Under intermediate selection the results are similar to those under threshold selection but for $v^* \simeq 0{\cdot}1$. Under linear selection, as well as with $r = 0$, allele M_l is always ousted from the population (data not presented).

The intensity of selection at locus M is of an even greater interest than the stationary frequencies of the alleles. To determine the former one must describe the changes in distribution of allele M_l over the population. However, instead of p_i we shall take only p, the frequency of allele M_l. Numerous data show that under any initial p_i its shape rapidly tends to 'stationarity', and the latter changes slowly with the change of p. Therefore, soon after the beginning of the experiment, p_i in fact is determined by p.

To describe selection the 'gametes' may be assumed as having various fitnesses. Let W_0 and W_1 be the fitnesses of haploid individuals m and M_l. Then

$$P = \frac{W_1 p}{W_1 p + W_0 (1-p)}. \tag{6}$$

The relative fitness of allele m can be found through the change of its frequency

$$\frac{W_0}{W_1} = \frac{p(1-P)}{P(1-p)}. \tag{7}$$

Although selection works at the haploid stage, locus M operates in the 'zygotes', so that the viability of a haploid individual depends on what zygote it originates from. On the other hand, when $r = \frac{1}{2}$ the viability of the progeny of a zygote with specified genotype at locus M does not depend on which allele they have at this locus. Consequently, the fitnesses of zygotes here describe the selection completely. Since mating is random,

$$P = \frac{W_{11} p^2 + W_{01} p(1-p)}{W_{11} p^2 + 2 W_{01} p(1-p) + W_{00} (1-p)^2},$$

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

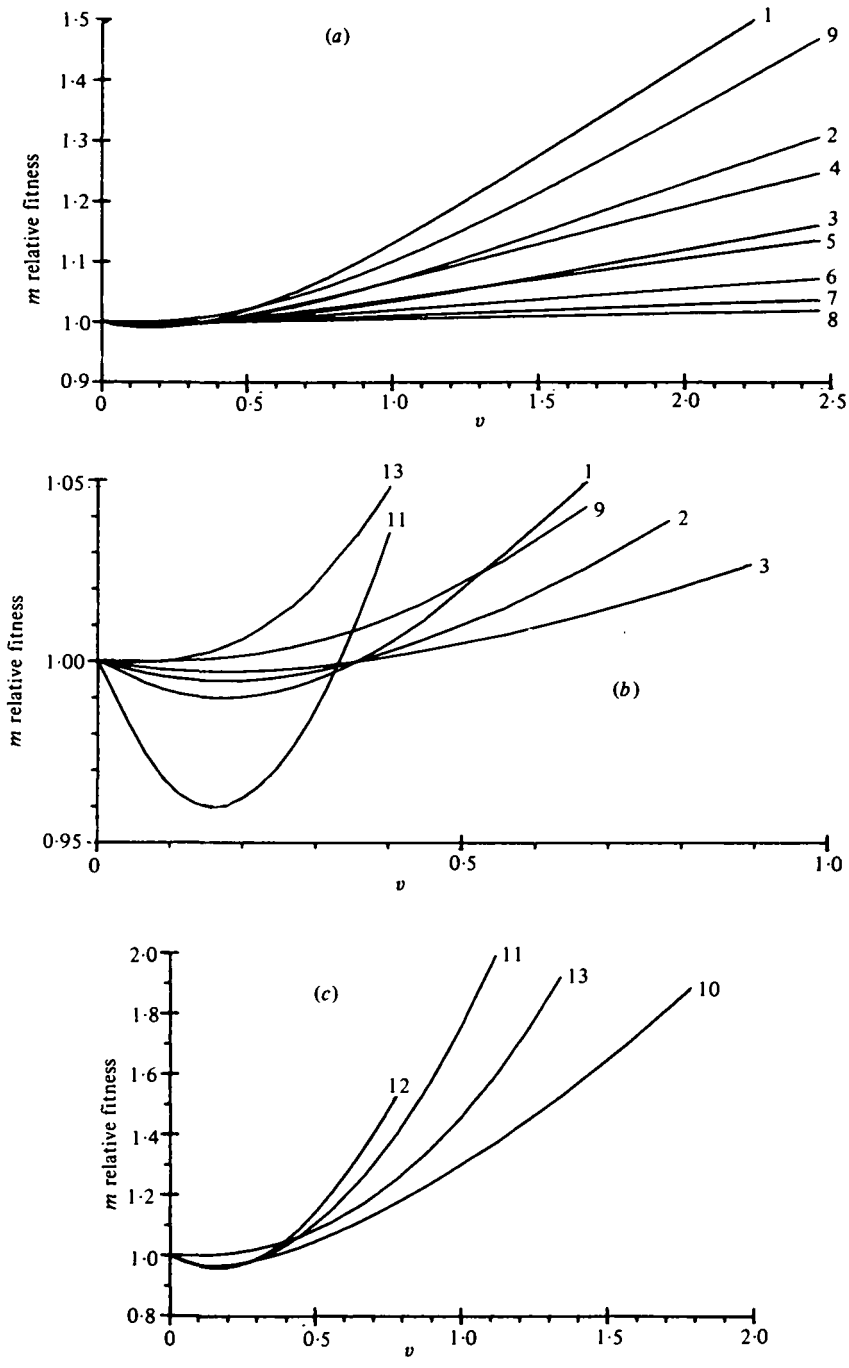

Fig. 6. Relative fitness of m with rare dominant M_l. Threshold selection: $k = 20$, $l = 1$ (1); $l = 3$ (2); $l = 7$ (3); $k = 5$, $l = 1$ (4); $l = 3$ (5); $l = 7$ (6); $l = 15$ (7); $l = 31$ (8). Linear selection; $k = 20$; $l = 1$ (9). Threshold selection: $l = 0$, $k = 5$ (10); $k = 20$ (11); $k = 80$ (12). Intermediate selection. $l = 0$, $k = 20$ (13). In (*a*), (*b*) and (*c*) scales are different.

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

where W_{11}, W_{01} and W_{00} are fitnesses of zygotes $M_l M_l$, $M_l m$ and mm. They reflect the viability of progeny and are described as

$$W_{11} = p^{-2}\Sigma_g \Sigma_h q'_g q'_h \Sigma_i s_i b_l a_4;$$

$$W_{01} = (2p(1-p))^{-1} \Sigma_g \Sigma_h q'_g q'_h \Sigma_i s_i b_x (a_2 + a_3);$$

$$W_{00} = (1-p)^{-2} \Sigma_g \Sigma_h q'_g q'_h \Sigma_i s_i b_\infty a_1.$$

The experimental data show that fitnesses of zygotes depend little both on the frequencies of the alleles at locus M and on which of the alleles is dominant. The values of W_1 and W_0, on the contrary, depend greatly on these factors. This fact made us use the zygote fitnesses to describe selection. In the case of a rare dominant allele M_l, presented in Fig. 6, the relative fitness of allele m, W_{00}/W_{01} equals W_0/W_1.

It is shown in Fig. 6 that, with $v < 0{\cdot}35$, when the direction of selection at M depends on s_i, its intensity for $l \neq 0$ is very small. With large v, allele m enjoys an advantage which is little dependent on s_i and grows approximately linearly with increasing v. Under a given v it grows with k and is inversely proportional to $l+1$. Relative fitnesses of $M_l m$ and mm, W_{01}/W_{11} and W_{00}/W_{11} decrease but slightly when the frequency of allele m increases (data not presented).

3. SELECTION FOR RECOMBINATION IN SUBDIVIDED GENOME

Recombination in eucaryotes is known to be due to two mechanisms: independent segregation of non-homologous chromosomes, and crossing-over. It is reasonable to suppose that consideration of a single chromosome leads to an overestimation of selection at locus M. A complete description of the distribution of mutations in a population of individuals with L chromosomes each will require C^L_{L+k-1} values, which makes the above approach impossible. Therefore, we performed a Monte-Carlo simulation with a population of size 2000.

(i) *Model*

The program simulated a population of haploid individuals with discrete generations. Their life-cycle remained the following: mutation–mating–recombination–selection–mutation. The computer memory stored genotypes of all the individuals of the given generation. Every genome was described by a set of $L+1$ values. The first L values were the numbers of mutations in the respective chromosomes of an individual, while the $(L+1)$st number denoted an allele of locus M in a separate chromosome that determined the crossover frequency. Recombination in this model was set by two numbers: L and the number of crossovers in each chromosome, l. The allele causing l was designated as μ_l to emphasize its difference from allele M_l. Allele μ_∞ retained the designation m because it leads to free recombination as above.

All chromosomes were assumed to carry equal proportions of the genome, so that after mutation the number of mutations in every chromosome increased by i with

probability $e^{-u/L}(u/L)^i(i!)^{-1}$. Then individuals mated randomly, each individual mating only once. If there was no crossing over in a particular zygote, an offspring received one chromosome from each pair of its parents' homologous chromosomes. With l crossovers per chromosome, offspring had i mutations in the respective chromosome with probability $b_l(g, h, i)$, assuming that one of the parents' homologous chromosomes carried g mutations and the other h. All matings had equal fertility, which resulted in from 6 to 50 offspring in different experiments.

Table 1. *Selective coefficient of allele m when alleles m and μ_0 are present*

L	$k=5$	$k=20$	$k=80$
	$u=0{\cdot}5$		
1	−0·04	−0·08	−0·03
2	−0·02	−0·03	−0·02
4	−0·01	−0·01	−0·01
8	−0·01	0	0
16	0	0	0
	$u=0{\cdot}2$		
1	+0·16	+0·05	−0·05
2	+0·08	+0·02	−0·02
4	+0·04	+0·01	−0·01
8	+0·02	0	0
16	+0·01	0	0
	$u=8{\cdot}0$		
1	+2·17	+2·04	+0·64
2	—	+1·05	+0·35
4	—	+0·47	+0·17
8	—	+0·25	+0·08
16	—	+0·09	+0·04

When progeny formation was complete, all individuals with k or more mutations were eliminated. Under non-threshold selection individuals died with probability $1-s_j$, where j was the mutation number in the genome. Then some individuals were randomly eliminated, so that the rest should not exceed 2000. Fertility was adjusted to minimize the number of excess individuals and, on the other hand, so as not to let the population die out. The model was run on an ES 10–40 computer. Pseudo-random numbers were obtained using the URAND standard program.

Each experiment lasted 25 generations, in each generation average numbers of the survived progeny of zygotes $\mu_l\mu_l$, $\mu_l m$ and mm were taken as their fitnesses. The average fitnesses for generations 11–25 were recorded, when presenting the result of the experiment. With measurable intensity of selection W_{01} proved to be close to W_{11} if μ_l was dominant, and close to W_{00} if μ_l was recessive. In the first case, therefore, we adopted the value of the $[W_{00}/\frac{1}{2}(W_{01}+W_{11})]-1$ and in the second the value of $[\frac{1}{2}(W_{01}+W_{00})/W_{11}]-1$ as selection coefficient of allele m.

(ii) *Results*

The data of Tables 1 and 2 were obtained by averaging out the results of at least 10 experiments. Their 95% confidence intervals were calculated employing

Student's t-distribution. Its width was about 0·02 with $u = 0{\cdot}5$, $u = 2{\cdot}0$ and $u = 8{\cdot}0$ when $k = 80$. Under $u = 8$, $k = 20$ it was 0·4, 0·15, 0·1, 0·1, when $L = 2, 4, 8, 16$, respectively. This difference is due to the fact that the variance of the experimental results increased when $u = 8$, $k = 20$ because of the reduction of the 'effective number' of the populations.

Table 2. *Selection coefficient of allele m under* $u = 8$, $k = 20$ *when alleles m and* μ_l *are present*

L	$l = 0$	$l = 1$	$l = 3$	$l = 7$
1	2·04	0·40	0·22	0·14
2	1·05	0·15	0·11	0·07
4	0·47	0·11	0·05	0·04
8	0·25	0·07	0·06	0·06

Table 1 shows the selection coefficient of allele m when m and μ_0 are present in a population under threshold selection. The initial frequency of the recessive allele μ_0 was 80%. The results for $L = 1$, presented for comparison, are calculated with the method of the above section. It is clear that, under the same conditions, the directions of selection in this and the previous models coincide (Fig. 6 and Table 1). However, the difference between the fitness of alleles μ_0 and m with some L is always slightly more than that of alleles M_l and m with $l = L-1$. This may be due to the difference between the models, and may also be explained by the assumption that recombination due to mixing L congealed chromosome of the same length is not so effective, as that which is due to $L-1$ crossovers occurring at random points of a single chromosome containing the whole genome. We found it impossible to test this assumption by considering the average fitness of population fixed for allele μ_0, because of the rapid operation of Muller's ratchet, i.e. the random loss of mutation-free chromosomes. However, with $L = 1$, $l \neq 0$ and threshold selection, the advantage of allele μ_l (Table 2) approximates the results of the previous section for alleles M_l. This may mean that, under identical conditions, these models produce similar results. Neither the differences in allele frequencies, nor the fact that allele m was recessive in the experiments presented in Fig. 6 and dominant in those of Table 1 are likely to explain the difference, as numerous results show that the role of these factors is not essential. On the other hand, it was assumed for the derivation of b_l that mutations were distributed along chromosomes randomly, which could have resulted in a slight overestimation of recombination frequency. Whatever the causes, the differences between the results of Fig. 6 and Table 1 are only quantitative.

The results of Table 2 show that the advantage of m over μ_0 is far greater than that over μ_1. For $L > 1$, allele μ_l is subjected to selection which is similar to that of the case of $L = 1$ for allele $M_{L(l+1)-1}$. Under intermediate and linear selection with $v < 0{\cdot}35$, selection at locus M is practically absent. When $v > 0{\cdot}35$ it was slightly weaker than under threshold selection, which is in line with the results of the above section. For example, with intermediate selection, $k = 80$, $L = 4$, the advantage of m over μ_0 was $0{\cdot}006 \pm 0{\cdot}008$; $0{\cdot}023 \pm 0{\cdot}006$; $0{\cdot}045 \pm 0{\cdot}01$; $0{\cdot}18 \pm 0{\cdot}08$ under $u = 1{\cdot}8$, $3{\cdot}0$, $4{\cdot}5$ and $8{\cdot}0$, respectively, i.e. under v close to 0·2, 0·33, 0·5 and 1·0.

The zones of parameter values for which selection should establish polymorphism of alleles μ_l under all l with $L > 1$ seem to have been as narrow as previously with $l \neq 0$; at least an alteration of the direction of selection at locus M was never observed when the allele frequencies were changed.

4. EVOLUTION OF THE PARAMETERS OF THE MUTATION PROCESS

The significance of v for the evolution of recombination suggests a consideration of the evolution of v.

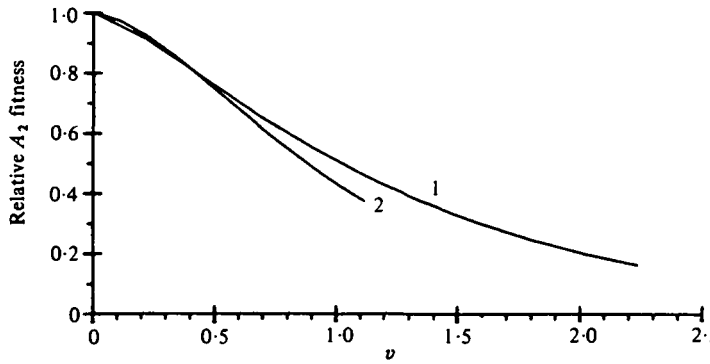

Fig. 7. Relative fitness of the rare allele A_2. $k = 5$ (1); $k = 20$ (2).

(i) *Model*

Let recombination be free in all zygotes; instead of the locus M, consider a locus A with alleles A_1 and A_2 controlling total mutability of the genome. In haploid individuals carrying allele A_1, there occur u mutations per genome on average, and in the ones carrying allele A_2 there are $2u$ mutations. In other ways, the model repeats that of Section 2. Allele A_2 frequencies among individuals with i mutations will be p_i, p_i', p_i'', respectively. Mutation and recombination are described by:

$$q_i' = e^{-u} \sum_{j \leqslant i} q_j (1-p_j)\, u^{(i-j)} (i-j)!^{-1} + e^{-2u} \sum_{j \leqslant i} q_j p_j (2u)^{(i-j)} (i-j)!^{-1}, \qquad (2')$$

$$p_i' = (q_i')^{-1} e^{-2u} \sum_{j \leqslant i} q_j p_j (2u)^{(i-j)} (i-j)!^{-1},$$

$$q_i'' = \Sigma_g \Sigma_h q_g' q_h' b_\infty (g+h, i), \qquad (4')$$

$$p_i'' = (q_i'')^{-1} \Sigma_g \Sigma_h q_g' q_h' b_\infty (g+h, i) \,.\, 0{\cdot}5 (p_g' + p_h').$$

Selection is described by the equation (5).

(ii) *Results*

Relative fitnesses of rate allele A_2 in relation to A_1 under threshold selection with $k = 5$, $k = 20$ were calculated by (7) and are presented in Fig. 7.

They are practically uninfluenced by the type of selection and allele frequencies

https://doi.org/10.1017/S0016672300026392 Published online by Cambridge University Press

(data not presented) and depend mainly on the value of v. In contrast, in asexual populations the relative fitness of clones with mutabilities $2u$ and u is always e^{-u} (Kimura & Maruyama, 1966; Kondrashov, 1982).

5. DISCUSSION

In small populations, selection against deleterious mutations is known to provide an advantage for recombination due to Muller's ratchet (Maynard Smith, 1978). Our results confirm the view that under synergistic interactions among mutations, recombination is advantageous in large populations as well (Crow, 1970, 1983; Feldman *et al.* 1981; Kondrashov, 1982). When $v < v^*$ an intermediate frequency of recombination is established and when $v > v^*$ free recombination has an advantage. One might conclude that a great variability of recombination in nature (Maynard Smith, 1978) means that actually $v < v^*$. We would rather accept a different view that selection against deleterious mutations is essential for the evolution of recombination only when $v > v^*$. There are three arguments for this:

(1) Polymorphism at locus M can be maintained only in a narrow spectrum of conditions: either with threshold selection or with a small v and intermediate selection.

(2) A zone of polymorphism is only wide with allele M_0. Such a situation is artificial, as even two chromosomes are sufficient to make allele μ_0 behave not like M_0 but more likely M_1 (Section 3). As the polymorphism zone of locus M for $L \neq 0$ is narrow, selection in a multichromosome genome would either maximize crossover frequency or lead to its extinction.

(3) Under any conditions (except $L = 1$) selection at locus M under $v < v^*$ is weak (Fig. 5, Table 1). Since crossing over has physiological disadvantages (Tucic, Ayala & Marinkovic, 1981), to exist at all it should be maintained by a considerable selection pressure, present only when $v > v^*$.

Thus, we presume crossing over frequency to be a compromize between its genetic advantage and physiological harm, which does not contradict its great variability. There is evidence (Charlesworth & Charlesworth, 1976) that after a break in crossing over for one generation the progeny appear with an enhanced fitness. Our results (Section 2ii) account for a stable existence of crossing over in such a population.

Selection for a further increase of recombination frequency is much weaker when allele μ_1 and not μ_0 is present (Table 2). This is in line with the fact that different eucaryotes have on average 1 cross over per chromosome regardless of the number of chromosomes and the size of the genome (Perkins & Barry, 1977).

According to the Karlin–McGregor principle (1974) the evolution of modifier genes goes in such a way as to increase the average fitness of population. Evidently, a stable existence of alleles M_l or μ_l maintained by population factors contradicts this condition. However, if in nature $v > v^*$, then the Karlin–McGregor principle is in fact valid in this case.

Our results show that mutation can provide a considerable advantage for recombination when $v > 0{\cdot}5$ (or at least $v > 0{\cdot}3$). Can these values of v be realistic?

Experimental data on the values of u and k are not abundant (Ohta, 1976).

Indirect evidence that follows from molecular experiments (see Kunkel & Loeb, 1981), from measurement of single loci mutability (Mukai & Cockerham, 1977; Neel & Rothman, 1978; Neel, Mohrenveiser & Meisler, 1980) and from estimation of the evolution rate of nucleotide sequences (Li, Gojobori & Nei, 1980; Miyata, 1982) can only provide the upper limit of u, as it remains obscure what part of the genome is functioning. If we assume the rate of neutral substitutions, as well as the mutation rate equal to 1×10^{-8} per base per year (Li *et al.* 1980; Miyata, 1982), then, with a generation of man lasting 20 years and his genome having 5×10^9 base pairs, one can conclude that $u \leqslant 1000$. Recent data also indicate a very high molecular variability (Kreitman, 1983). This suggests that k is large and selection against individual mutations is very weak (Langley, Montgomery & Quattlebaum, 1982).

Before discussing the result of direct measurements (Mukai *et al.* 1972), note that in our model the coefficient of selection against a single deleterious mutation is $s = u/k$, assuming that the population is stationary, and the number of mutations per individual in close to k (Kimura & Maruyama, 1966, equation 1.9). Then $v^2 = us$. Mukai measured the rate of mutations in chromosome II that reduce preadult viability of *Drosophila melanogaster* (u'_{II}) and their individual effect on viability in the homozygous state (s'_{hom}). By his method, $u'_{\mathrm{II}} s'_{\mathrm{hom}}$ and less precisely, u'_{II} and s'_{hom}, separately, could be estimated. He found that $u'_{\mathrm{II}} s'_{\mathrm{hom}} \simeq 0{\cdot}004$. In nature these mutations are probably selected against in heterozygous state (Crow, 1979) where those effects (s') are about a half of s'_{hom} (Mukai, 1980). On the other hand, as the single second chromosome comprises about one-sixth of a diploid genome, the rate of such mutations per genome (u') is approximately $6 \,.\, u'_{\mathrm{II}}$. Hence, $s'u' \simeq 0{\cdot}012$, which would suggest $v \simeq 0{\cdot}1$.

There is some evidence that $u's'$ was underestimated because of incomplete dominance of Cy chromosome (Mukai, 1980) and because of residual selection in the course of the experiment (Simmons *et al.* 1980). Note that the additive viability variance in nature is much greater than that predicted by Mukai *et al.* (1972) (see Mukai, Kusakabe & Tachida, 1983). But the central point is that Mukai measured only preadult viability neglecting the other components of fitness: longevity, mating success and fertility, which may have a much greater effect (Simmons & Crow, 1977). Then u' should be less than u, as there are mutations that affect fitness without changing viability; and s' should be less than s, as mutations that reduce viability may reduce the other components of fitness as well (Simmons *et al.* 1980). A most reliable estimate, $s \simeq 0{\cdot}02 - 0{\cdot}03$, follows from the data on inbreeding depression (Crow, 1979). With such an s, u should be more than 10 to satisfy $v > 0{\cdot}5$, and more than 4 to satisfy $v > 0{\cdot}3$. The data available do not seem to contradict these values of u, taking into account, particularly, that an excess of slightly deleterious mutations is possible (Mukai *et al.* 1972). Our models assume the effects of all mutations to be equal. Thus, we presume that considerable theoretical and experimental work is needed to find whether the parameters of the mutation process correspond to the condition advantageous for recombination. The same refers to the type of selection against deleterious mutations (Crow & Kimura, 1979).

Data of Fig. 4 show that when $v = 2$, even under a most effective threshold

selection and free recombination 95% of the progeny surviving random elimination and other forms of selection will have to die in the process of selection against deleterious mutations; note that when $v = 3$ mutation load is 99·6%. This conclusion appears to be robust and does not depend on the mode of selection for recombination. It is very unlikely that a population can overcome such a mutation load. Therefore we presume that if mutation does maintain recombination, the $0{\cdot}5 < v < 2{\cdot}0$. It is clear that such an 'invariance' of v may appear only due to the action of appropriate selection.

It is shown in Fig. 7 that selection for a decrease of mutability increases rapidly with the growth of v: when $v = 0{\cdot}25$ the relative fitness of A_2 is 90%, and when $v = 2{\cdot}0$ it is only 20%. Reasonably, if the difference in fitness of individuals carrying alleles A_2 and A_1 is more than 10%, then with $v = 0{\cdot}25$ allele A_1 is eliminated and v grows to 0·5. Though this consideration does not claim to be strict it allows us to conclude that if the physiological cost of reducing the mutability twice is within the range of 10–80%, then $0{\cdot}5 < v < 2{\cdot}0$. Unfortunately, experimental estimates of this cost are unknown to us, though the molecular mechanisms of mutability reduction are known to be complicated (Krutyakov *et al.* 1983).

Our results suggest, in contrast with the conclusion of Leigh (1973), that selection for reducing the mutability is stronger in an asexual than in sexual population. This is in accord with the fact that procaryotes in whose populations sexual reproduction does not play an important role (Selander & Levin, 1980) have a smaller value of u than eucaryotes (Maynard Smith, 1978). This may give an advantage in adaptation to eucaryotes. However, useful mutations are hardly likely to play an essential role in evolution of mutability (Leigh, 1973).

If the size of the genome increases, v must stay in the range of $v < 2{\cdot}0$ (3·0). It is clear that there are two ways to reach this that are not mutually exclusive. Mutation rate per base pair may decrease. The fact that mutability of single loci of man is lower than that of *Drosophila* in spite of the higher body temperature and higher number of DNA replications per generation (Neel *et al.* 1980) seems to support such a possibility. On the other hand, if a further decrease of mutation rate per nucleotide is connected with the high cost, the only possibility that remains is the growth of k, as increasing resistance of onthogenesis to deleterious mutations may be necessary for progressive evolution (Schmalhausen, 1949, 1982).

There are three other consequences of our hypothesis concerning selection for recombination. It is possible that with a very large k selection for recombination with the same v is considerably stronger than that obtained here. Still, we assume that with $v < 2$ and multichromosome genome this selection is always rather weak. This limits the possible intensity of selection for linkage disequilibrium, as recombination load should be considerably smaller than the advantage of crossing over (Tucic *et al.* 1981): the lack of linkage disequilibria in nature is hardly surprising.

If the number of chromosomes in a haploid set is very large, then with any admissible value of v the advantage of crossing over is small. It may lead to its elimination, which in the long run decreases the adaptability of the population. As the increase in the number of chromosomes is a rare event, it may be counterbalanced by group selection.

If $v > 0{\cdot}5$ the mutation load is 50 % or more (Fig. 4). As the mutation component of mortality is likely to be less than 0·5 (Crow & Denniston, 1981), it follows that in a sexual population females must produce 8–10 offspring minimum.

We think that the suggested hypothesis is a reasonable answer to the question: Why does the genome not congeal? (Turner, 1967). However, more experimental data is needed to confirm it with all its consequences.

The author is obliged to A. D. Bazykin, M. V. Mina and A. S. Rozanov for helpful discussions, to A. B. Kirillov for mathematic support, to Natasha Kondrashova for the translation, and to the referee for many helpful suggestions.

APPENDIX

By Alexander B. Kirillov

Let $l = 2n-1$; this means that the number of crossovers is odd. Consideration of the $l = 2n$ case is analogous. We thus have $2n-1$ independent uniformly distributed random variables $x_1, \ldots, x_{2n-1}$ (i.u.d.r.v.), which cut the interval (0, 1) into $2n$ parts. Enumerate them from the left to the right: 1, 2, ..., $2n$. Let the location of the ith mutation in the first parent be y_i and that of the jth mutation in the second parent be z_j. It follows from the assumption that they represent $g+h$, i.u.d.r.v., independent of $x_1, \ldots, x_{2n-1}$. Let us now find $\bar{b}_{2n-1}(g, h, k, r)$, the probability that an offspring receives k mutations from the first parent and r mutations from the second parent, which is the probability with which k out of g values of y_i are to be in the odd numbered parts of the cut interval (0, 1), and r out of h values of z_j are to be in the even numbered intervals. It is sufficient to compare the reciprocal locations of points x_m, y_i and z_j in (0, 1). There are $(2n-1+g+h)!$ of them altogether, and they all have equal probability due to their symmetry. Next by uniting locations of the points differing in their numbers only, e.g. $x_3y_2z_1x_2y_1z_2$ with $x_1y_1z_1x_3y_2z_2$, into single elementary events, we obtain $p(2n-1, g, h) = (2n-1+g+h)!\ [(2n-1)!g!h!]^{-1}$ equally probable locations of $2n-1$ indistinguishable points of x type, g points of y type, and h indistinguishable points of z type. Consider different locations of k points of y type and $h-r$ points of z type in odd intervals, where an interval is a line between the two closest points of x type. We face the analogous problem about the number of different locations of $n-1$ indistinguishable points x, of k indistinguishable points y, and of $h-r$ indistinguishable points z. The number of these locations is $p(n-1, k, h-r)$. Analogously, the remaining even intervals contain $p(n-1, g-k, r)$ locations. Therefore

$$\bar{b}_{2n-1}(g, h, k, r) = \frac{p(n-1, k, h-r) \cdot p(n-1, g-k, r)}{p(2n-1, g, h)}.$$

The function to be found, i.e. $b_l(g, h, i)$ $(l = 2n-1)$ is calculated using $\bar{b}_{2n-1}(g, h, k, r)$. Thus, $b_l(g, h, i) = \sum_{k+r=i} \bar{b}_l(g, h, k, r)$.

REFERENCES

CHARLESWORTH, B. & CHARLESWORTH, D. (1976). An experiment on recombination load in *Drosophila melanogaster*. *Genetical Research* **23**, 167–183.

CHINNICI, J. P. (1971). Modification of recombination frequency in *Drosophila*. II. The polygenic control of crossingover. *Genetics* **69**, 75–96.

CROW, J. F. (1970). Genetic loads and the cost of natural selection. In *Mathematical Topics in Population Genetics* (ed. K. Kojima). Berlin. Springer-Verlag.

CROW, J. F. (1979). Minor viability mutants in *Drosophila*. *Genetics* **92** (suppl.), 165–172.

CROW, J. F. (1983). Disease and evolution. *Evolution* **37**, 863–865.

CROW, J. F. & KIMURA, M. (1979). Efficiency of truncation selection. *Proceedings of the National Academy of Sciences, U.S.A.* **76**, 396–399.

CROW, J. F. & DENNISTON, C. (1981). The mutation component of genetic damage. *Science* **212**, 888–893.

DISHLER, V. YA. (1983). Induced recombinogenesis in higher plants. *Zinatne, Riga* (in Russian).

FELDMAN, M. W., CHRISTIANSEN, F. B. & BROOKS, L. D. (1980). Evolution of recombination in a constant environment. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 4838–4841.

HALDANE, J. B. S. (1957). The cost of natural selection. *Journal of Genetics* **55**, 511–526.

KARLIN, S. & MCGREGOR, J. (1974). Towards the theory of evolution of the modifier gene. *Theoretical Population Biology* **5**, 59–103.

KIMURA, M. & MARUYAMA, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

KIMURA, M. & CROW, J. P. (1978). Effect of overall phenotypic selection on genetic change at individual loci. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 6168–6171.

KONDRASHOV, A. S. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genetical Research* **40**, 325–332.

KONDRASHOV, A. S. (1984). Rate of evolution in changing environment. *Journal of Theoretical Biology* **107**, 249–260.

KREITMAN, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–416.

KRUTYAKOV, V. M., BELYAKOVA, N. V., KLEINER, N. E., LEGINA, O. K., SHEVELEV, I. V. (1983). Eucaryotic 3-5 exonuclease correcting DNA polymerase errors. *Dokladi Akademyi Nauk SSSR* **272**, 1491–1494.

KUNKEL, T. A. & LOEB, L. A. (1981). Fidelity of mammalian DNA polymerases. *Science* **213**, 765–767.

LANGLEY, C. H., MONTGOMERY, E. & QUATTLEBAUM, W. F. (1982). Restriction map variation in the Adh region of *Drosophila*. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 5631–5635.

LEIGH, E. G. (1973). The evolution of mutation rates. *Genetics* **73**, (suppl.), 1–18.

LI, W.-H., GOJOBORI, T. & NEI, M. (1981). Pseudogenes as a paradigme of neutral evolution. *Nature* **292**, 8237–8239.

LOUCAS, M., VEGRINI, Y. & KRIMBAS, C. B. (1981). The genetics of *Drosophila subobscura* populations. XVII. Further genic heterogeneity within electromorphs by urea denaturation and the effect of the increased genic variability on linkage disequilibrium studies. *Genetics* **97**, 429–441.

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

MIYATA, T. (1982). Evolutionary changes and functional constraints in DNA sequences. In *Molecular Evolution, Protein Polymorphism and Neutral Theory* (ed. M. Kimura), pp. 233–268. Berlin: Springer-Verlag.

MUKAI, T. (1980). The genetic structure of natural populations of *Drosophila melanogaster*. XIV. Effects of the incomplete dominance of the IN(2LR) SM1 (Cy) chromosome on the estimates of various genetic parameters. *Genetics* **94**, 169–184.

MUKAI, T., CHIGUSA, S. T., METTLER, L. E. & CROW, J. P. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

MUKAI, T. & COCKERMAN, C. C. (1977). Spontaneous mutation rates of enzyme loci in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences U.S.A.* **74**, 2514–2517.

MUKAI, T., KUSAKABE, S. & TACHIDA, H. (1983). Indirect negative evidence for frequency-dependent selection with respect to viability in *Drosophila melanogaster*. *Japanese Journal of Genetics* **58**, 297–314.

NEEL, J. V. & ROTHMAN, E. D. (1978). Indirect estimates of mutation rates in tribal *Amerindians*. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 5585–5588.

NEEL, J. V., MOHRENVEISER, H. W. & MEISLER, M. H. (1980). Rate of spontaneous mutations at human loci encoding protein structure. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 6037–6041.

OHTA, T. (1976). Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theoretical Population Biology* **10**, 254–275.

PERKINS, D. P. & BARRY, E. G. (1977). The cytogenetics of *Neurospora*. *Advances in Genetics* **19**, 133–285.

SCHMALHAUSEN, I. I. (1949). *Factors of Evolution: The Theory of Stabilizing Selection*. Philadelphia.

SCHMALHAUSEN, I. I. (1982). Organisms as a whole in an individual and historical development. *Moscow, Nauka* (in Russian).

SELANDER, R. K. & LEVIN, B. R. (1980). Genetic diversity and structure in *Escherichia coli* populations. *Science* **210**, 545–547.

SHIELDS, W. M. (1982). *Philopatry, Inbreeding and the Evolution of Sex*. State University of New York Press, Albany.

SIMMONS, M. J. & CROW, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics* **11**, 49–78.

TUCIC, N., AYALA, F. J. & MARINKOVIC, D. (1981). Correlation between recombination frequency and fitness in *Drosophila melanogaster*. *Genetica* **56**, 61–71.

TURNER, J. R. G. (1967). Why does the genome not congeal? *Evolution* **21**, 645–656.

WHITEHOUSE, H. L. K. *Genetic Recombination: Understanding the Mechanisms*. New York: John Wiley.