The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect)*

By TOMOKO OHTA and MOTOO KIMURA

National Institute of Genetics, Mishima, Japan

(*Received* 11 *March* 1975)

SUMMARY

A diffusion model was developed to investigate the effect of a mutant substitution by natural selection on heterozygosity at a linked neutral locus. Using this theory, we made extensive numerical analyses to compute the expected total heterozygosity (i.e. the sum of the fraction of heterozygotes over all generations until fixation or loss) at the neutral locus. It was shown that the hitch-hiking effect is generally unimportant as a mechanism for reducing heterozygosity. The effect becomes significant only when the recombination fraction between the selected and the neutral marker loci is smaller than the selection coefficient. In order to check the validity of the mathematical theory, Monte Carlo experiments were performed, and the results were in agreement. It has been suggested that linkage is important only in transient small populations such as at the time of speciation.

1. INTRODUCTION

The first persons to investigate the 'hitch-hiking' effect, whereby the path of a weakly selected mutant on its way to eventual loss or fixation is influenced by linkage to a strongly selected locus, were Kojima & Schaffer (1967). More recently, Maynard Smith & Haigh (1974) studied the effect of a selected allele at one locus on the heterozygosity of a neutral allele at a linked locus and concluded that in populations of 10^6 or more the hitch-hiking effect is more important than random drift. Their treatment was deterministic and we believe that it overemphasizes the linkage effect.

In this paper we treat the problem stochastically. Using a diffusion model we shall formulate a relevant equation which enables us to compute the amount of decrease in heterozygosity at a neutral locus in a finite population as influenced by the hitch-hiking effect of a definitely advantageous mutant. To corroborate the theoretical treatment, we also performed Monte Carlo experiments, and the results will be presented in this paper.

2. BASIC THEORY

We consider a random mating diploid population of the effective size N_e, and assume a pair of selectively neutral alleles A and a in the first locus. In the second locus we assume that an advantageous mutant B is in the process of replacing its allele b.

We propose to investigate how much the sum of the heterozygosity at the neutral locus over all generations (from the initial appearance of the mutant until its complete fixation or loss by random drift) is influenced by selected gene substitution at the second locus. Since the frequency of heterozygotes is a quadratic function of the gene frequencies under random mating, it is necessary to obtain the second moments as well as the first moments of the gene frequencies at the neutral locus. A complete treatment which takes stochastic elements into account at both loci is very difficult, so we make the simplifying assumption that, although the change at the neutral locus is subject to random drift, the change at the selected locus by natural selection can be treated deterministically with enough accuracy. Such an assumption is realistic if the selective advantage of allele B (over b) is large and if its frequency is not extremely close to either 0 or 1. Thus, if we denote by y_t the frequency of allele B in the tth generation, then y_t satisfies the differential equation;

$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = sy_t(1-y_t), \tag{1}$$

so that we have

$$y_t = \frac{1}{1 + \left(\dfrac{1-y_0}{y_0}\right)\mathrm{e}^{-st}}, \tag{2}$$

where s is the selective advantage of B.

In order to investigate the effect of the selected locus on the neutral one, we conceptually divide the population into two parts: a part consisting of chromosomes carrying the advantageous mutant B and another part carrying the disadvantageous allele b. We shall denote by x_1 the frequency of allele A at the first locus among chromosomes carrying B, and by x_2 the frequency of A among chromosomes carrying b. Then, as we shall soon show, for a function f which is an arbitrary polynomial in x_1 and x_2, we have the equation

$$\frac{\partial E(f)}{\partial t} = E\left\{\frac{x_1(1-x_1)}{4N_e y_t}\frac{\partial^2 f}{\partial x_1^2} + c(1-y_t)(x_2-x_1)\frac{\partial f}{\partial x_1}\right.$$
$$\left. + \frac{x_2(1-x_2)}{4N_e(1-y_t)}\frac{\partial^2 f}{\partial x_2^2} + cy_t(x_1-x_2)\frac{\partial f}{\partial x_2}\right\}, \tag{3}$$

where c is the recombination fraction between the two loci, and E stands for the operator for taking the expectation with respect to gene frequency distribution. The above equation is a special case of a more general equation

$$\frac{\partial E(f)}{\partial t} = E\{L(f)\}, \tag{4}$$

where L is the differential operator of the Kolmogorov backward equation that, in the case of n random variables $(x_1, x_2, \ldots, x_n)$, takes the form

$$L = \frac{1}{2} \sum_{i=1}^{n} V_{\delta x_i} \frac{\partial^2}{\partial x_i^2} + \sum_{i>j} W_{\delta x_i \delta x_j} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} M_{\delta x_i} \frac{\partial}{\partial x_i}$$

in which M, V and W designate respectively, the mean, the variance and the covariance of the rate of change of the random variables that appear as subscripts. The present method as developed by Ohta & Kimura (1969, 1971a) has proved to be particularly useful in evaluating the moments of the distribution for two linked loci. For details concerning the derivation and application of equation (4), readers are invited to consult Kimura & Ohta (1971, pp. 183–190). Then, derivation of equation (3) as a special case of equation (4) is straightforward: in generation t, there are $2N_e y_t$ chromosomes carrying B in the population, and the variance of change in x_1 by random sampling is $x_1(1-x_1)/(2N_e y_t)$, which leads to the first term in the right hand side of equation (3). Similarly the third term can be derived by noting that there are $2N_e(1-y_t)$ chromosomes carrying b. The second and the fourth terms represent the effects of crossing-over on the mean changes in x_1 and x_2. Finally we note that equation (3) is valid even if the right-hand side depends on t through y_t, since equation (4) holds for non time-homogeneous processes.

Let us calculate the moments of gene frequencies at the neutral locus using equation (3), starting with the first moment. Letting $f = x_1$, equation (3) yields

$$\frac{\mathrm{d}E(x_1)}{\mathrm{d}t} = c(1-y_t)\, E(x_2-x_1). \tag{5a}$$

Similarly for $f = x_2$ we get

$$\frac{\mathrm{d}E(x_2)}{\mathrm{d}t} = c y_t\, E(x_1-x_2). \tag{5b}$$

Solving this set of equations, we obtain the following formulae for the first moments of x_1 and x_2 in the tth generation:

$$E(x_{1,t}) = x_{1,0} - c(x_{1,0} - x_{2,0}) \int_0^t \frac{(1-y_0)\, \mathrm{e}^{-(s+c)\tau}}{y_0 + (1-y_0)\, \mathrm{e}^{-s\tau}} \,\mathrm{d}\tau, \tag{6a}$$

$$E(x_{2,t}) = x_{2,0} + c(x_{1,0} - x_{2,0}) \int_0^t \frac{y_0\, \mathrm{e}^{-c\tau}}{y_0 + (1-y_0)\, \mathrm{e}^{-s\tau}} \,\mathrm{d}\tau. \tag{6b}$$

In these equations $E(x_{1,t})$ and $E(x_{2,t})$ stand for the expected value of x_1 and x_2 in the tth generation, and $x_{1,0}$ and $x_{2,0}$ are the initial values. For example, if we consider a situation in which a single neutral mutant appears in the population in which gene substitution by natural selection is taking place, we put either $x_{1,0} = 1/(2N_e y_0)$ and $x_{2,0} = 0$, or $x_{1,0} = 0$ and $x_{2,0} = 1/\{2N_e(1-y_0)\}$, depending on whether the initial mutant is linked to the advantageous allele or the disadvantageous allele. The above results are essentially equivalent to those obtained by Maynard Smith & Haigh (1974), who treated the problem deterministically by using a discrete generation model. We now proceed to obtain the second moments of x_1 and x_2, as they are essential to compute the level of heterozygosity as influenced by 'hitch-hiking'.

Letting successively $f = x_1^2$, $f = x_1 x_2$ and $f = x_2^2$ in our basic equation (3), we obtain

$$\frac{\mathrm{d}E(x_1^2)}{\mathrm{d}t} = E\left\{\frac{x_1(1-x_1)}{2N_e y_t} + 2c(1-y_t)\,x_1(x_2-x_1)\right\}, \tag{7a}$$

$$\frac{\mathrm{d}E(x_1 x_2)}{\mathrm{d}t} = E\{c(1-y_t)\,(x_2-x_1)\,x_2 + cy_t(x_1-x_2)\,x_1\}, \tag{7b}$$

$$\frac{\mathrm{d}E(x_2^2)}{\mathrm{d}t} = E\left\{\frac{x_2(1-x_2)}{2N_e(1-y_t)} + 2cy_t\,x_2(x_1-x_2)\right\}. \tag{7c}$$

From the solutions to (6) and (7) we can write the total heterozygosity as

$$\int_0^\infty E\{2x_t(1-x_t)\}\,\mathrm{d}t,$$

where $x_t = x_{1,t}y_t + x_{2,t}(1-y_t)$.

Since the analytical solution of this set of equations appears to be difficult, we resorted to numerical solution in conjunction with (5a) and (5b) by a computer. This can be done readily by replacing the differential quotients in the left-hand side of (7a)–(7c) and (5a) and (5b) by corresponding finite differences, taking one generation as the unit step. Then the total heterozygosity can be computed as the sum of $2x(1-x)$ with $x = yx_1 + (1-y)x_2$ over all generations. Detailed analysis based on this method under various combinations of parameters will be presented in the next section.

For the special case of complete linkage ($c = 0$), the equations can readily be integrated, and we obtain

$$E(x_{1,t}^2) = x_{1,0} + (x_{1,0}^2 - x_{1,0})\,\mathrm{e}^{-\lambda_1(t)}, \tag{8a}$$

$$E(x_{1,t}x_{2,t}) = x_{1,0}x_{2,0} \tag{8b}$$

and

$$E(x_{2,t}^2) = x_{2,0} + (x_{2,0}^2 - x_{2,0})\,\mathrm{e}^{-\lambda_2(t)}, \tag{8c}$$

where

$$\lambda_1(t) = \frac{1}{2N_e}\left\{t + \frac{1}{s}\left(\frac{1-y_0}{y_0}\right)(1-\mathrm{e}^{-st})\right\}$$

and

$$\lambda_2(t) = \frac{1}{2N_e}\left\{t + \frac{1}{s}\left(\frac{y_0}{1-y_0}\right)(\mathrm{e}^{st}-1)\right\}.$$

These formulae can also be derived by noting that with complete linkage ($c = 0$), the population can be regarded as split into two independent subpopulations, one consisting of chromosomes carrying B and another carrying b. The former subpopulation consists of $2N_e y_\tau$ chromosomes in generation τ so that the rate of decrease of heterozygosity in this generation is $1/(2N_e y_\tau)$. Thus the expected heterozygosity in the tth generation is

$$E\{2x_1(1-x_1)\} = 2x_{1,0}(1-x_{1,0}) \exp\left\{-\int_0^t \frac{\mathrm{d}\tau}{2N_e y_\tau}\right\},$$

which leads to the same formula as (8a) if we put

$$\frac{1}{y_\tau} = 1 + \left(\frac{1-y_0}{y_0}\right)\mathrm{e}^{-s\tau}.$$

in the integrand, and note that $E(x_1) = x_{1,0}$ under no recombination. Formulae (8c) as well as (8b) can also be derived in a similar way.

The total heterozygosity over all generations, is

$$\int_0^\infty E\{2x(1-x)\}\,dt = 2x_{1,0}(1-x_{1,0})\int_0^\infty y_t^2\,e^{-\lambda_1(t)}\,dt$$
$$+ 2(x_{1,0}+x_{2,0}-2x_{1,0}x_{2,0})\left(\frac{1-y_0}{s}\right)$$
$$+ 2x_{2,0}(1-x_{2,0})\int_0^\infty (1-y_t)^2\,e^{-\lambda_2(t)}\,dt. \qquad (8d)$$

The second term on the right follows from (1), since

$$\int_0^\infty y(1-y)\,dt = \frac{1}{s}\int_{y_0}^1 dy = \frac{1-y_0}{s}. \qquad (8e)$$

3. NUMERICAL ANALYSIS

We shall investigate the total heterozygosity – that is, the sum of the fraction of heterozygotes over all generations – contributed by a single neutral mutant as influenced by the hitch-hiking effect. Clearly, the total heterozygosity will be increased by hitch-hiking if the initial mutant happens to be linked with advantageous allele B (as compared with the case of no hitch-hiking), while it will be decreased if it happens to be linked with disadvantageous allele b. Note that the average value of the total heterozygosity contributed by a single neutral mutant is approximately 2 (Kimura & Crow, 1963; Kimura, 1969; Maruyama, 1971). More generally, if p is the initial frequency of a neutral allele, the expected value of the total heterozygosity without the hitch-hiking effect in a random mating population is

$$H_T = \sum_{n=0}^\infty 2p(1-p)\left(1-\frac{1}{2N_e}\right)^n = 4N_e p(1-p), \qquad (9)$$

since heterozygosity decreases each generation at the rate of $1/(2N_e)$ on the average due to random sampling of gametes. Therefore we can use the above value, i.e. 2, as a standard, and we shall compare the corresponding values under hitch-hiking obtained numerically by using the method presented in the previous section.

We designate the total heterozygosity as $H_{T(+)}$ when the initial mutant is linked with the advantageous allele (to be called positive hitch-hiking) and as $H_{T(-)}$ when the initial mutant is linked with a disadvantageous allele (negative hitch-hiking).

In calculating the total heterozygosity, the moments of gene frequency at the neutral locus were computed throughout until the 10^4-th generation. This was considered to be sufficient for our purpose since in practice the mutant allele is completely lost or fixed during this period with the population size, i.e. 100 or 200 assumed in this analysis. (Note that in the diffusion models, the distribution of gene frequencies is determined by the products $N_e c$ and $N_e s$ rather than the individual parameters N_e, c and s separately.) We also adopt the following approximation procedure; after the frequency of the advantageous allele becomes so close to 1 that the

product of the frequency of the disadvantageous allele and the population size is equal to or less than unity $(N_e(1 - y_t) \leqslant 1)$, the neutral mutant is treated as if it were independent of hitch-hiking. We have also tried several values of $N_e(1 - y_t)$, such as 5 or 10, after which this procedure was started, and found that the effects of such an approximation are negligible.
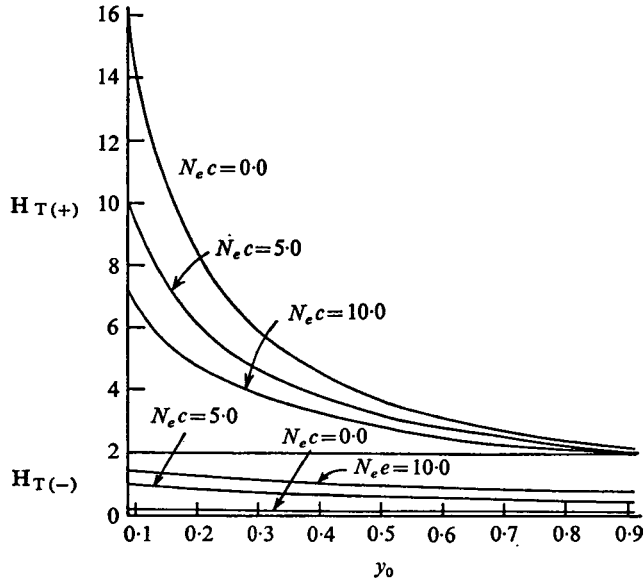

Fig. 1. Relationship between the total heterozygosity of a neutral mutant and the initial frequency of the advantageous allele at the linked locus (y_0). $H_{T(+)}$ is the value for positive hitch-hiking and $H_{T(-)}$ is that for negative hitch-hiking. Parameters are $N_e s = 20$ and three levels of $N_e c$ as shown in the figure.

We shall first examine how the initial frequency of the advantageous allele influences the total heterozygosity at the neutral locus. Fig. 1 shows the relationship between the total heterozygosity and the initial frequency of the advantageous allele at the second locus (y_0) for various values of $N_e c$. In this figure $N_e s$ is assumed to be 20. Curves above the flat line ($H_T = 2$) represent the cases of positive hitch-hiking, that is, when the neutral mutant occurs on the chromosome carrying the advantageous allele ($H_{T(+)}$), and curves below the line represent the cases of negative hitch-hiking ($H_{T(-)}$). As seen from the figure, positive hitch-hiking is more effective with smaller y_0 and it may be quite pronounced for extremely small y_0. The negative hitch-hiking effect is not greatly influenced by y_0, yet it is more effective with larger y_0. Fig. 2 shows the relationship between the total heterozygosity and y_0 for various values of $N_e s$ under complete linkage ($N_e c = 0$). The way by which the total heterozygosity depends on y_0 for the case $N_e s = \infty$ may be interpreted as follows: the increase of the advantageous allele by selection is so rapid as compared with its change by random drift that the heterozygosity of a neutral mutant is determined by the effective initial frequency which is equal to the frequency of the mutant *among*

chromosomes carrying the advantageous allele at the time of occurrence. Hence $H_{T(+)}$ is almost inversely proportional to y_0. By taking the value of $N_e s$ larger, we have actually found that the total heterozygosity approaches the limiting value $2/y_0$. Therefore this curve represents the maximum effect of positive hitch-hiking.
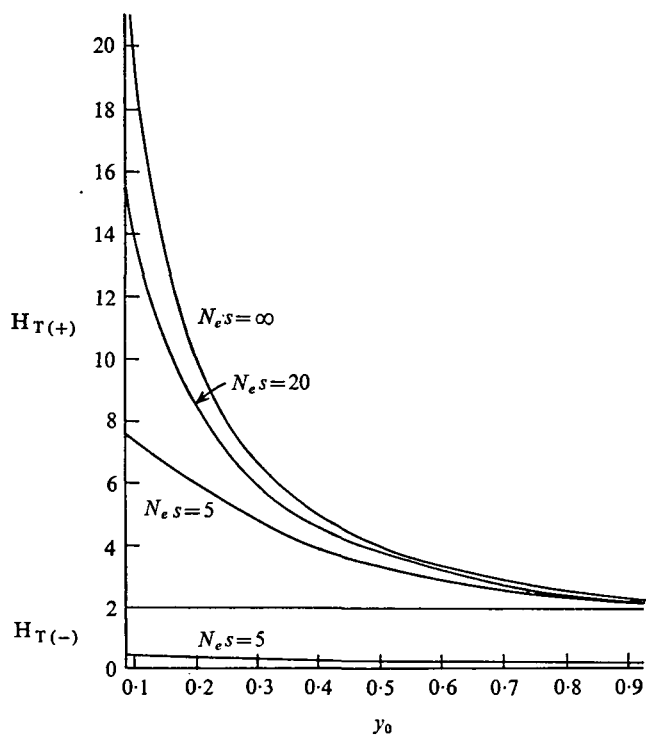

Fig. 2. Relationship between the total heterozygosity of a neutral mutant and y_0 for various values of $N_e s$ as shown in the figure. $N_e c$ is assumed to be 0.

Next, we shall investigate the dependence of the total heterozygosity on $N_e c$, the product of the recombination fraction and the population size. Fig. 3 illustrates this for various values of $N_e s$ including $N_e s = \infty$, assuming $y_0 = 0 \cdot 1$. From the figure, we can see that the hitch-hiking effect rapidly decreases as $N_e c$ increases.

So far we have examined the positive and the negative hitch-hiking effects separately. However, the real interest lies in the average behaviour of the plus and minus effects and therefore we shall evaluate the average of the total heterozygosity by assuming that a neutral mutant has equal chance of occurring on each chromosome irrespective of whether the chromosome carries an advantageous or disadvantageous allele at the second locus.

Fig. 4 shows the relationship between the average total heterozygosity and y_0 for some values of $N_e s$ and $N_e c$. It is interesting to note that the average total heterozygosity is always smaller than two and that it is not greatly different from two even when the individual positive and negative effects are large. It might be thought from

22-2

Fig. 4 that the total heterozygosity will continue to decrease as y_0 approaches $1/2N$. However, this is an artifact of our deterministic assumption. At the limit of a single mutant ($y_0 = 1/2N$) the value becomes very close to 2. This is easily understood as follows. The probability is $1/2N$ that the mutant will occur in a favoured chromosome in which case the total heterozygosity is approximately 4 (Kimura, 1969), and $1 - 1/2N$ that it will occur in one of the remaining chromosomes in which case
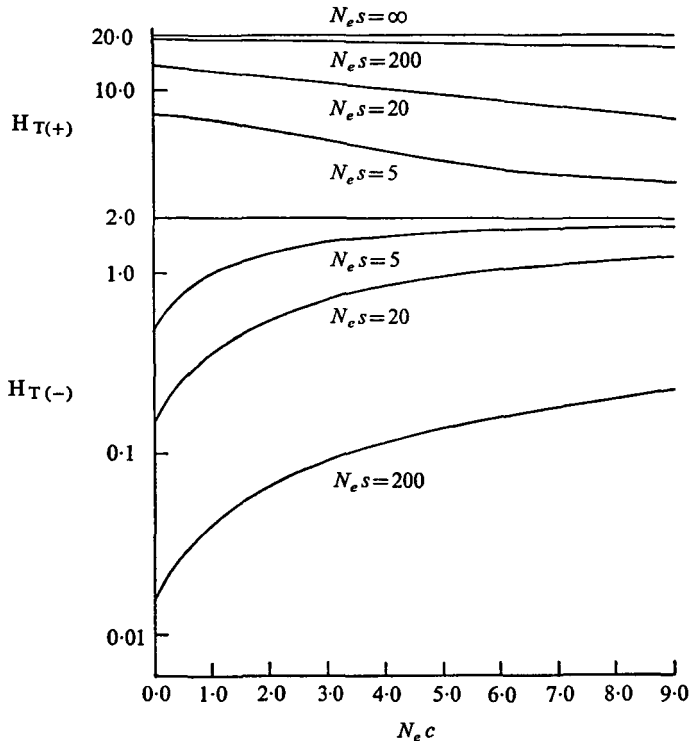

Fig. 3. Total heterozygosity of a neutral mutant with hitch-hiking as a function of $N_e c$. Parameters are $y_0 = 0.1$ and various levels of $N_e s$ as shown in the figure.

the total heterozygosity is roughly 2; so the average is about 2. So the curves in Fig. 5 eventually start to curve upward as y_0 becomes smaller, the exact shape depending on the extent to which there is stochastic variation in the selected locus.

Fig. 5 illustrates the average total heterozygosity as a function of $N_e s$ for some values of $N_e c$, assuming $y_0 = 0.1$. It is again interesting to find that the average total heterozygosity is not greatly influenced by hitch-hiking. In particular, for a given $N_e c$ there is a certain value of $N_e s$ at which the hitch-hiking effect becomes maximum with at most 43 % reduction of average heterozygosity when $y_0 = 0.1$. For smaller y_0, the maximum reduction may get larger, yet one can conclude that the hitch-hiking effect is quite small even for very tight linkage, since only under quite restricted conditions does the amount of reduction in the average heterozygosity become pronounced. Also an interesting result that has emerged from the

above analysis is that for $N_e c = 0$ and at the limit $N_e s \to \infty$, the plus and minus effects cancel each other and the average total heterozygosity approaches 2. This may be understood by noting that for such a case the total heterozygosity is contributed
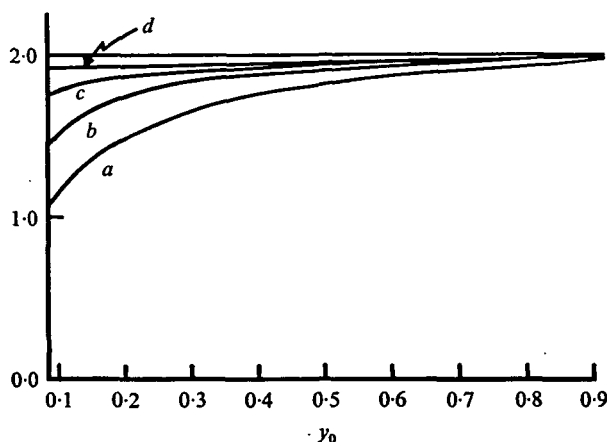

Fig. 4. Total heterozygosity of a neutral mutant as the average of $H_{T(+)}$ and $H_{T(-)}$. Parameters are (a) $N_e s = 5$, $N_e c = 0$; (b) $N_e s = 20$, $N_e c = 0$; (c) $N_e s = 20$, $N_e c = 5$; (d) $N_e s = 5$, $N_e c = 5$.
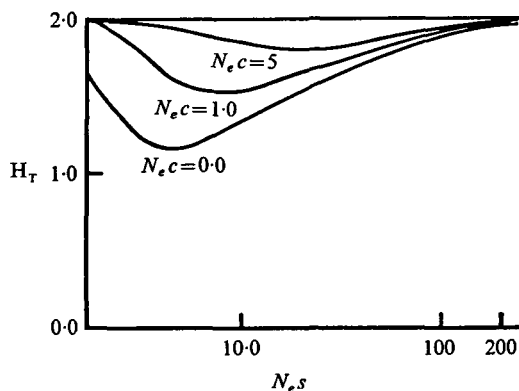

Fig. 5. Total heterozygosity of a mutant (as the average of $H_{T(+)}$ and $H_{T(-)}$) as a function of $N_e s$. The value of y_0 is assumed to be $0 \cdot 1$.

solely from the positively hitch-hiking mutant and that the heterozygosity due to that mutant is inversely proportional to y_0 as explained before, while the relative frequency of the neutral mutant occurring on a chromosome carrying the advantageous mutant is y_0.

We shall now investigate the effect of hitch-hiking by asking a slightly different question: To what extent is the existing heterozygosity at the neutral locus reduced by the spreading of an advantageous allele at a linked locus? This is nearer to the approach used by Maynard Smith & Haigh (1974). We again calculate the total heterozygosity by generating moments and compare the results with what is expected for an independent neutral allele. The latter can be obtained from equation

(9). The solution for $s = 0$ is $H(p) = 4Np(1-p)$, which is $2Nh_0$, where h_0 is the initial heterozygosity. To simplify the treatment, we assume that the advantageous allele, after its appearance by mutation in a chromosome carrying A, rapidly increased its frequency to reach y_0 and thereafter can be treated deterministically. We further assume that the initial increase to reach y_0 is so rapid that the recombination between the two loci concerned has not taken place during that period. Then we can use the same procedure, by setting $x_{1,0} = 1\cdot 0$ and $x_{2,0} = (x_0 - y_0)/(1\cdot 0 - y_0)$ as the initial condition for generating the moments, where $x_0 \, (\geq y_0)$ is the initial frequency of the neutral allele A, and y_0 is that of B. Note that the initial frequency
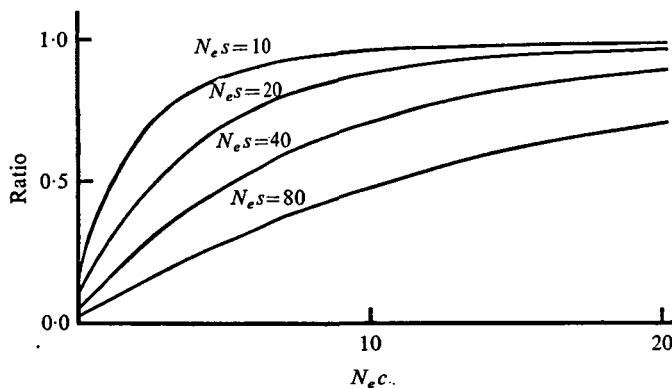

Fig. 6. Total heterozygosity at a neutral locus with hitch-hiking expressed as a fraction of what is expected at an independent locus, i.e. $4N_e x_0(1-x_0)$, where x_0 is the initial frequency of the neutral allele. Four curves represent the ratio of total heterozygosities with and without hitch-hiking, as functions of $N_e c$ for four levels of $N_e s$ as shown in the figure.

of AB is y_0, while aB does not exist at the start. In studying the hitch-hiking effects, we have found that the symmetric case ($x_0 = 0\cdot 5$) may be regarded as representative since the plus and minus effects tend to cancel each other even if we assume an asymmetric situation (such as $x_0 = 0\cdot 1$ vs $x_0 = 0\cdot 9$ or $x_0 = 0\cdot 2$ vs $x_0 = 0\cdot 8$). Therefore only results from the symmetric case will be presented here. The value of y_0 has been chosen as $0\cdot 1$ in the numerical study; however, as long as y_0 is small enough, the following results will not be changed much.

Fig. 6 shows the expected total heterozygosity as a fraction of what is expected at an independent neutral locus. The abscissa represents $N_e c$ and the ordinate the ratio of the expected total heterozygosities with and without hitch-hiking. Four curves correspond to the cases of $N_e s = 10, 20, 40$ and 80 respectively. The figure suggests that the hitch-hiking effect is negligible if $c > s$, and that the effect is mostly determined by the ratio s/c. For example, when $s/c = 2$, the ratio is about $0\cdot 87$ when $N_e c = 5$, $0\cdot 88$ when $N_e c = 10$ and $0\cdot 89$ when $N_e c = 20$. In general the probability is rather low that an advantageous mutant occurs at a locus so tightly linked to a particular neutral locus that $c < s$ is satisfied. Therefore, the hitch-hiking effect is likely to be generally unimportant in reducing the pre-existing heterozygosity.

Table 1. *The total heterozygosity due to a single neutral mutant*

(Observed values are from Monte Carlo experiments assuming $N_e = 100$.)

s	c	y_0	Positive hitch-hiking		Negative hitch-hiking	
			Observed	Expected	Observed	Expected
0·05	0·0	0·1	5·74	7·33	0·757	0·478
		0·2	4·83	5·94	0·459	0·389
		0·5	3·38	3·39	0·291	0·287
	0·1	0·1	3·05	2·97	1·724	1·862
		0·2	2·49	2·74	1·896	1·779
		0·5	2·53	2·34	1·613	1·606
0·1	0·0	0·1	8·84	10·91	0·320	0·258
		0·2	6·07	7·37	0·215	0·210
		0·5	3·87	3·62	0·142	0·160
	0·1	0·1	4·41	4·24	1·782	1·681
		0·2	3·93	3·52	1·451	1·551
		0·5	2·80	2·61	1·604	1·319

4. MONTE CARLO EXPERIMENTS

In order to check the validity of our mathematical treatments, we performed extensive Monte Carlo experiments. In simulating the process, we started each experiment by setting the initial condition in terms of gametic frequencies, and whenever the neutral mutants become either lost or fixed in the population, we started again from the same initial condition. Each experiment was continued 10^5 generations and the total heterozygosity, the frequency of loss or fixation of the mutant, etc. were counted during the experiment. Thus, we obtain the total heterozygosity per neutral mutant.

Each generation of the experiments consisted of crossing over, selection and sampling of gametes. Crossing-over and selection were carried out deterministically using the following formulae:

$$\Delta g_1 = \{(w_1 - \overline{w})\,g_1 - cD\}/\overline{w}, \quad \Delta g_2 = \{(w_2 - \overline{w})\,g_2 + cD\}/\overline{w},$$

$$\Delta g_3 = \{(w_3 - \overline{w})\,g_3 + cD\}/\overline{w}, \quad \Delta g_4 = \{(w_4 - \overline{w})\,g_4 - cD\}/\overline{w},$$

where g_i's stand for gametic frequencies, w_i's are the fitnesses in terms of selective values, $\overline{w}$ is the average population fitness and D is the coefficient of linkage disequilibrium (cf. Kimura, 1956; Lewontin & Kojima, 1960). Sampling of zygotes was performed following a simple scheme which we have used previously (Ohta & Kimura, 1971a).

Table 1 shows the results of simulation experiments for the total heterozygosity per mutant together with the theoretical values obtained by numerical analysis for comparison. N_e was assumed to be 100 and the observed values are the averages of $3023 \sim 20\,472$ trials, which means that a neutral mutant was supplied $3023 \sim 20\,472$ times in a single experiment. As seen from the table, the agreement between the observed and the expected total heterozygosity is satisfactory.

5. DISCUSSION

Our approach is essentially different from that of Maynard Smith & Haigh (1974), who treated the problem deterministically and who concentrated on the ratio of the initial to the final heterozygosities at the neutral marker locus when the gene substitution by natural selection takes place at a linked locus. Rather than looking just at the end-points of the process, we have evaluated the total heterozygosity at the neutral marker locus during the whole process. We believe that our treatment is more suitable for judging the effect of hitch-hiking on heterozygosity.

We have shown that the hitch-hiking effect is generally unimportant as a mechanism for reducing heterozygosity. It may be significant only when the recombination fraction is smaller than the selection coefficient. Such a situation must be quite rare at least in sexually reproducing organisms. In this regard, the phenomenon known as periodic selection in micro-organisms (Atwood, Schneider & Ryan, 1951; Koch, 1974) may represent a special case of large selection pressure and extremely small recombination fraction. It is usually observed in chemostat experiments, but the significance of this phenomenon in nature is not clear even in micro-organisms. A large reduction in heterozygosity depends on the mutant arising at a time when the favourable gene has a relatively low frequency, but with absolute numbers large enough to behave essentially deterministically. At other times the effect is small and if several favourable mutants are being selected simultaneously the average effect cannot be very large. Another situation where hitch-hiking may be important is with group selection, such as may have occurred among ancient human groups. Here the fates of mutant genes are associated with the expansion or extinction of the groups, since all genes within a group are effectively linked.

Together with our previous studies on linkage disequilibrium (Ohta & Kimura, 1971b; Ohta, 1973), our analyses indicate that linkage is important only in small and transient populations such as those at the time of speciation, and not in large and stable populations. In small populations, linkage disequilibrium due to random drift may have significant effects on the behaviour of surrounding genes. However, the average effects are rather small even if the individual effect may get large. In this respect, linkage only makes chance effects somewhat larger in transient populations. Lewontin (1974) seems to be overemphasizing the role of linkage, especially since, except where inversions are involved, linkage disequilibrium is rarely found (Mukai, Watanabe & Yamaguchi, 1974). For large and stable populations, the concept of quasi-linkage equilibrium (Kimura, 1965; Nagylaki, 1974) together with the single locus theory is sufficient to treat most problems realistically.

Maynard Smith & Haigh (1974) have argued that the adaptive gene substitutions at many loci, simultaneously occurring, may reduce the heterozygosity at other loci drastically if the population size is large. They suggested that this may explain why the observed average heterozygosity per locus is not much different among various species. Unfortunately for the neutral theory we have to deny their conclusion and state our belief that the total size of the species is the most important parameter that determines the amount of random drift. Then we have to search for

another explanation for the relative uniformity of average heterozygosity among various species. It is possible, as proposed by one of us (Ohta, 1974), that the very slight negative selection based on functional constraints of the protein molecule becomes effective in very large populations leading to mutation-selection balance at many loci and this prevents the level of heterozygosity from increasing indefinitely as the population size increases. On the other hand, random drift prevails in relatively small populations.

REFERENCES

ATWOOD, K. C., SCHNEIDER, L. K. & RYAN, F. J. (1951). Selective mechanisms in bacteria. *Cold Spring Harbor Symposia on Quantitative Biology* 16, 345–355.

KIMURA, M. (1956). A model of a genetic system which leads to closer linkage by natural selection. *Evolution* 10, 278–287.

KIMURA, M. (1965). Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52, 875–890.

KIMURA, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61, 893–903.

KIMURA, M. & CROW, J. F. (1963). On the maximum avoidance of inbreeding. *Genetical Research* 4, 399–415.

KIMURA, M. & OHTA, T. (1971). *Theoretical Aspects of Population Genetics*. Princeton University Press.

KOCH, A. L. (1974). The pertinence of the periodic selection phenomenon to prokaryote evolution. *Genetics* 77, 127–142.

KOJIMA, K. & SCHAFFER, H. E. (1967). Survival process of linked mutant genes. *Evolution* 21, 518–531.

LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York, London: Columbia University Press.

LEWONTIN, R. C. & KOJIMA, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* 14, 458–472.

MARUYAMA, T. (1971). An invariant property of a structured population. *Genetical Research* 18, 81–84.

MAYNARD SMITH, J. & HAIGH, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* 23, 23–35.

MUKAI, T., WATANABE, T. K. & YAMAGUCHI, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics* 77, 771–793.

NAGYLAKI, T. (1974). Quasilinkage equilibrium and the evolution of two-locus systems. *Proceedings of the National Academy of Sciences, U.S.A.* 71, 526–530.

OHTA, T. (1973). Effect of linkage on behaviour of mutant genes in finite populations. *Theoretical Population Biology* 4, 145–162.

OHTA, T. (1974). Mutational pressure as the main cause of molecular evolution and polymorphisms. *Nature* 252, 351–354.

OHTA, T. & KIMURA, M. (1969). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* 63, 229–238.

OHTA, T. & KIMURA, M. (1971a). Linkage disequilibrium between two segregating nucleotide sites under steady flux of mutations in a finite population. *Genetics* 68, 571–580.

OHTA, T. & KIMURA, M. (1971b). Behavior of neutral mutants influenced by associated overdominant loci in finite populations. *Genetics* 69, 247–260.