Genet. Res., Camb. (1978), **31**, *pp.* 13–28
With 2 text-figures
Printed in Great Britain

Theoretical study on genetic variation in multigene families*

BY TOMOKO OHTA
National Institute of Genetics, Mishima, 411, Japan

(*Received* 16 *April* 1977)

SUMMARY

Genetic variation contained in a multigene family was theoretically investigated from the standpoint of population genetics. Unequal crossover is assumed to be responsible for the coincidental evolution of mutant genes in a chromosome. When the allowed latitude of the duplicated or deleted number of gene units at unequal crossover is $10 \sim 15\%$ of the total gene number in a chromosome, the arrangement of gene lineage in a chromosome is shown to be roughly random. The equilibrium properties of genetic variation or the probability of identity of two genes within a family (clonality) were studied under mutation, unequal crossover, interchromosomal crossover and sampling of gametes. The clonality of a multigene family within *a chromosome* is shown to be approximately

$$C_0 = \frac{\alpha}{\alpha + 2v + \frac{\beta}{3}\frac{4N_e v}{1+4N_e v}},$$

in which $\alpha = 2k/n^2$ with k = effective number of cycles of unequal crossover and with n = number of gene units in a family, v is the mutation rate per gene unit, β is the rate of interchromosomal crossover per family and N_e is the effective size of the population, all measured by the rate per generation. The clonality of a gene family between *two different chromosomes* becomes approximately $C_1 = C_0/(1+4N_e v)$. Some models of natural selection which lowers the clonality or increases genetic variation in a multigene family were investigated. It was shown that natural selection may be quite effective in increasing genetic variation in a gene family.

1. INTRODUCTION

The neo-Darwinism is mostly based on the theory of adaptive gene substitution in the population. Recent advances on multigene families such as immunoglobulin gene family (Hood, Campbell & Elgin, 1975, for review) suggest that the theory of individual gene substitution is not enough for understanding the mechanism of progressive evolution at the molecular level. The two remarkable characteristics, i.e. the contraction–expansion of the gene number in a family and the coincidental evolution should now be seriously considered for the theoretical study of the evolution of higher organisms.

I have attempted to clarify the nature of genetic variation contained in the

* Contribution No. 1172 from the National Institute of Genetics, Mishima, 411, Japan.

gene family by using the model of unequal crossing-over for the coincidental evolution (Ohta, 1977). I assumed that the gene number in a chromosome is constant and examined the effects of mutation, crossing-over and selection. I have further investigated the case where the gene number in a chromosome is a random variable by introducing a model of directional selection to increase the genetic information in a gene family (Ohta, 1978). In this report I shall present the more general and detailed analyses of the evolution and variation of multigene families, although concentrating on the situation where the number of gene units in a family is more or less constant.

2. CROSSOVER FIXATION TIME

The process of coincidental evolution is considered to be analogous to the mutant dynamics in population genetics, i.e. the number of a gene lineage contracts or expands within a chromosome just like the frequency of a mutant fluctuates in finite populations by random drift (Smith, 1974; Hood *et al.* 1975; Tartof, 1975). In particular, the model of unequal crossover of the coincidental evolution may be treated by the diffusion model of Kimura (1964) (Ohta, 1976). In this model, the time until fixation of a gene lineage in a family of a chromosome is called the crossover fixation time (Smith, 1974). As a first step for analysing the complex process of coincidental evolution, the single chromosomal family is considered. Even under this simplified situation, the additional assumption is necessary for an analytical treatment; I assumed that the duplication or deletion by unequal crossover occurs by one gene unit and that the duplication and deletion occurs alternately so that the number of units in a chromosome is kept constant, n. Two successive crossovers (duplication and deletion of one unit) is called one cycle of the process. Then it has been shown that one cycle of unequal crossover corresponds to $4N_e/n^2$ generations of random sampling drift in population genetics in which N_e is the effective population size (Ohta, 1976; Perelson and Bell, 1977).

In this section, I shall present some analyses for the cases in which the number of units of duplication or deletion is not restricted to one but is a random variable, by introducing the results of extensive Monte Carlo experiments.

Let n be the number of gene units in a chromosome but it now expands or contracts following a certain probability function by unequal crossover. Let us consider a following particular case due to Smith (1974). In this model we assign the allowed latitude (p) so that n changes within the interval $(n_0 - pn_0,\ n_0 + pn_0)$ in which n_0 is the initial number of units. The number of units after an unequal crossover is determined completely randomly in this interval by using a uniform random number but if it happens to be the same as the previous value it is discarded and the process is repeated. Then the mean change of n is zero but the mean of the absolute value of the change of n is average distance between two randomly chosen units from linearly arranged $(2pn_0 + 1)$ units, therefore the mean of the absolute value of the change of n is

$$m = \tfrac{2}{3}(np + 1). \tag{1}$$

https://doi.org/10.1017/S0016672300017778 Published online by Cambridge University Press

The symbol n now represents average value which is equal to n_0. As I have argued in my previous report (Ohta, 1976), if the arrangement of the gene lineages is random in the chromosome, one crossover should correspond to $m/2$ cycles of the previous model. In such a case, the crossover fixation time (t_1) can be approximately expressed by the following formula in terms of the number of *crossovers*:

$$t_1 = \frac{2n^2}{m} = \frac{3n^2}{np+1} \approx \frac{3n}{p}, \tag{2}$$

since t_1 is about n^2 cycles in my previous model (Ohta, 1976). If the gene lineages are more clustered than random arrangement, one would expect that the coincidental evolution is more rapid than the above estimate suggests due to the correlated change of gene frequencies. On the other hand, if the gene lineages are more dispersed than random arrangement, the coincidental evolution is expected to be slower than the prediction by the negative correlation.

I have carried out an extensive Monte Carlo experiment to examine the above relationship. Generally it is expected that if the allowed latitude is small, the gene lineages are more clustered and vice versa. This prediction is verified in the simulation studies as given below.

The experiment was done with a starting chromosome made up of linearly arranged $1 \sim n_0$ numbers, each number representing the gene unit. At the unequal crossover a uniform random number determines the number of units of the new chromosome in the allowed interval $(n_0 - pn_0,\ n_0 + pn_0)$. If the resulting number of units is the same as the previous value, it is discarded and the process is repeated. The two sister chromatids are arranged so that the new chromosome has a specified number of units. Another random number determines the point of crossover within the overlapping region of the chromatids. The unequal crossover was repeated until the entire chromosome was made up of a single gene lineage (number in this case).

Three levels of n and four levels of the allowed latitude (p) (total of 12 cases) have been done, with $15 \sim 48$ repetitions. Fig. 1 shows the result. The abscissa is the allowed latitude in percentage of the number of units and the ordinate is the number of crossovers until fixation of a single gene unit in the chromosome, both in logarithmic scale. In the figure the expected crossover fixation time by the formula (2) is given by the curve as functions of the allowed latitude and the observed values, by the white circles for the case of $n = 160$, by the crosses for the case of $n = 80$ and by the black dots for the case of $n = 40$. It can be predicted from the figure that the random arrangement of gene lineages may be obtained when the allowed latitude is $10 \sim 15\%$. This result is consistent with that of Smith (1974), who obtained empirically by Monte Carlo experiments that the crossover fixation time is roughly 20 times of the number of units, since t_1 is $20n$ when p is roughly 0·15 from formula (2).

In order to strengthen further the analogy between the unequal crossover and random genetic drift, the variance of the crossover fixation time has been studied. From the theory of Kimura & Ohta (1969*a*, *b*), the coefficient of variation of

the time until fixation of a neutral mutant in a finite population is a constant, 0·538. The same theory should apply in the present case (Ohta, 1976). Table 1 shows the observed coefficient of variation of the crossover fixation time in simulations. The figures in parentheses show the number of repetitions in each case. From the table it can be seen that the observed coefficient agrees with the

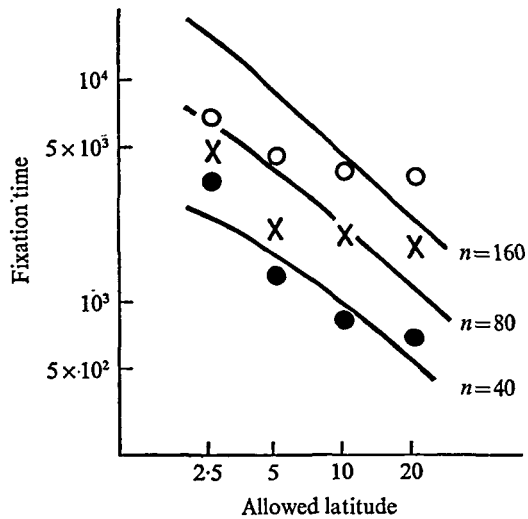

Fig. 1. The crossover fixation time in terms of the number of crossovers as functions of the allowed latitude. The curves show the theoretical fixation time (formula 2) and the observed values are shown by the white circles (n = 160), by the crosses (n = 80) and by the black dots (n = 40).

Table 1. *The observed coefficient of variation of the crossover fixation time by Monte Carlo experiments*

(The figures in parentheses are the number of repetitions of the experiments.)

	Allowed latitude (%)			
n	2·5	5	10	20
40	0·50 (22)	0·57 (29)	0·54 (48)	0·38 (20)
80	0·47 (20)	0·54 (34)	0·51 (28)	0·47 (17)
160	0·35 (17)	0·53 (15)	0·67 (16)	0·51 (24)

predicted value irrespective of the magnitude of the allowed latitude. This result implies that although the significant correlation may appear in the arrangement of gene lineages in the chromosome, the theory of random gene frequency drift roughly applies to the contraction–expansion of gene lineages in the chromosome by unequal crossover. This is a very convenient analogy since one can use the established theories in population genetics for analysing the process of unequal

crossover. For example, the genetic variation contained in a gene family may be treated as that in finite Mendelian populations (Ohta, 1977). Here the difficulty arises when many chromosomal lines are considered and the ordinary interchromosomal crossing-over becomes significant. In the next section, I shall present an analysis which takes interchromosomal exchanges into account.

3. GENETIC VARIATION IN MULTIGENE FAMILIES AT EQUILIBRIUM

(i) *Basic theory*

Let us consider a Mendelian population with the effective size N_e. In each generation, mutation and sister chromatids unequal crossover occur in the multigene family of each chromosome and the ordinary interchromosomal crossover and sampling of gametes follow. Mutation is the factor which increases genetic variation in the gene family whereas unequal crossover decreases it. The interchromosomal exchange and random sampling control the genetic variation among the chromosomes as well as that within the chromosome. We shall ask the question 'what is the nature of genetic variation when all these factors balance each other?'

Let n be the number of gene units of a family in a chromosome as before and let n be constant throughout. Let v be the mutation rate per gene unit per generation and we assume that all mutations are unique and detectable. Let k be the effective number of *cycles* of unequal crossover per family and let β be the rate of interchromosomal crossover per family per generation.

We measure the genetic variation by the clonality (C) defined by Smith (1974), which is equivalent to the homozygosity in population genetics. The clonality of a family within a chromosome is expressed as

$$C_i = \sum_l x_{i,l}^2, \tag{3}$$

in which the subscript i indicates the ith chromosome, $x_{i,l}$ is the frequency of the lth gene lineage in the ith chromosome and summation is over all lineages. By the similar way, the clonality of a family between the two chromosomes may be defined

$$C_{ij} = \sum_l x_{i,l} x_{j,l}, \tag{4}$$

in which the subscript ij indicates the clonality of a family between the ith and the jth chromosomes. Note C_{ij} is equivalent to the gene identity between the two populations in population genetics (cf. Nei, 1975).

Let us denote the expectation of C_i by C_0 and that of C_{ij} by C_1. We shall formulate the rate of change of the vector $\mathbf{C} = \begin{bmatrix} C_0 \\ C_1 \end{bmatrix}$ by mutation, sampling and crossover. First, the mutation decreases the value of C_0 and C_1 by the fraction, $2v$, each generation, therefore we have

$$\mathbf{C}' = \mathbf{M}\mathbf{C} = (1-2v)\,\mathbf{C}. \tag{5}$$

https://doi.org/10.1017/S0016672300017778 Published online by Cambridge University Press

By sampling, C_0 does not change whereas C_1 changes by inbreeding due to finite population size.

$$\mathbf{C}' = \mathbf{SC}, \tag{6}$$

where

$$\mathbf{S} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2N_e} & 1-\frac{1}{2N_e} \end{bmatrix}.$$

Next, by intrachromosomal crossover, C_1 does not change but C_0 increases by the amount $2k(1-C_0)/n^2$ (Ohta, 1976). Thus, we have

$$\mathbf{C}' = \mathbf{KC}+\mathbf{A}, \tag{7}$$

where

$$\mathbf{K} = \begin{bmatrix} 1-\alpha & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{A} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix},$$

with $\alpha = 2k/n^2$.

Finally, the change of C by interchromosomal crossover is evaluated. For the analysis I assume that the gene lineages are arranged randomly along the chromosome, so that the clonality does not depend on the distance between the units on the chromosome. This condition is satisfied when the allowed latitude is about 10–15 % of the total number of the gene family as shown in the previous section and may not be generally met. However, this assumption greatly facilitates the analysis. Now consider that one crossover takes place between the ith and the jth chromosome within the region of the family and that the crossover point

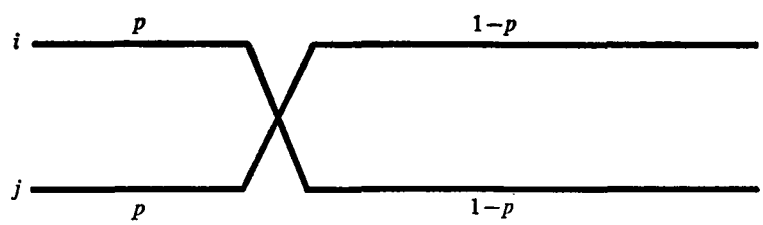

Fig. 2. Diagram illustrating the interchromosomal crossover.

divides the gene family into two parts, p and $1-p$ as in the Fig. 2. Then after the crossover, C_i and C_{ij} change following the formulas:

$$\left.\begin{aligned} C_i' &= p^2C_i+(1-p)^2\,C_j+2p(1-p)\,C_{ij}, \\ C_{ij}' &= \{p^2+(1-p)^2\}\,C_{ij}+p(1-p)\,C_i+p(1-p)\,C_j, \\ C_j' &= (1-p)^2\,C_i+p^2C_j+2p(1-p)\,C_{ij}. \end{aligned}\right\} \tag{8}$$

In order to take expectation of the formula (8), I assume the uniform distribution for p. One must further consider in taking the expectation that, by one crossover, C_i and C_j change according to the formula (8) (two chromosomes out of $2N_e$), whereas for C_{ij}, only one combination, ij, out of $2N_e(2N_e-1)/2$ possible combina-

tions follows this formula. In other words, the amount of change of C_1 should be multiplied by $1/(2N_e-1)$ as compared with that of C_0 in the formula (8) in taking the expectation.

Actually the rate of interchromosomal exchange per generation within the region of a multigene family is very small. Let β be this rate and we take the expectation of the formula (8). We have

$$\mathbf{C}' = \mathbf{RC}, \tag{9}$$

where

$$\mathbf{R} = \begin{bmatrix} 1-\frac{\beta}{3} & \frac{\beta}{3} \\ \frac{\beta}{3(2N_e-1)} & 1-\frac{\beta}{3(2N_e-1)} \end{bmatrix}.$$

The total change of $\mathbf{C}$ in one generation may be obtained by multiplying the matrices, $\mathbf{M}$, $\mathbf{S}$, $\mathbf{K}$ and $\mathbf{R}$ and, in general, the following equation holds:

$$\mathbf{C}' = \mathbf{GC}+\mathbf{T} \tag{10}$$

where $\mathbf{G}$ is the product of $\mathbf{M}$, $\mathbf{S}$, $\mathbf{K}$ and $\mathbf{R}$ and $\mathbf{T}$ is the product of $\mathbf{A}$ and some of $\mathbf{M}$, $\mathbf{S}$ and $\mathbf{R}$, depending on the model. If the sequence of these four events differs, the resulting equation is slightly different. For example, assume the sequence of events: sampling → mutation → sister-chromatids crossover → interchromosomal crossover (case 1). Then the equation becomes

$$\mathbf{C}' = \mathbf{G}_1\mathbf{C}+\mathbf{T}_1, \tag{11}$$

where

$$\mathbf{G}_1 = \mathbf{RKMS} = (1-2v) \times \begin{bmatrix} \left(1-\frac{\beta}{3}\right)(1-\alpha)+\frac{\beta}{6N_e} & \frac{\beta}{3}\left(1-\frac{1}{2N_e}\right) \\ \frac{\beta}{3(2N_e-1)}(1-\alpha)+\frac{1}{2N_e}\left(1-\frac{\beta}{3(2N_e-1)}\right)\left(1-\frac{\beta}{3(2N_e-1)}\right)\left(1-\frac{1}{2N_e}\right) \end{bmatrix}$$

and

$$\mathbf{T}_1 = \mathbf{RA} = \begin{bmatrix} \alpha\left(1-\frac{\beta}{3}\right) \\ \frac{\alpha\beta}{3(2N_e-1)} \end{bmatrix}.$$

On the other hand, if the sequence is sister-chromatids crossover → interchromosomal crossover → mutation → sampling (case 2), the equation becomes

$$\mathbf{C}' = \mathbf{G}_2\mathbf{C}+\mathbf{T}_2, \tag{12}$$

where

$$\mathbf{G}_2 = \mathbf{SMRK} = (1-2v)\begin{bmatrix} (1-\alpha)\left(1-\frac{\beta}{3}\right) & \frac{\beta}{3} \\ \frac{1}{2N_e}(1-\alpha) & \frac{\beta}{6N_e}+\left(1-\frac{1}{2N_e}\right)\left(1-\frac{\beta}{3(2N_e-1)}\right) \end{bmatrix}$$

https://doi.org/10.1017/S0016672300017778 Published online by Cambridge University Press

and

$$\mathbf{T}_2 = \mathbf{SMRA} = (1-2v)\begin{bmatrix} \alpha\left(1-\frac{\beta}{3}\right) \\ \alpha/2N_e \end{bmatrix}.$$

At equilibrium, $\mathbf{C}' = \mathbf{C}$ and one can readily get the equilibrium value of clonality. If we assume that all parameters, $\beta/3$, α, $2v$ and $1/2N_e \ll 1$, the solution is simplified. It becomes both in cases 1 and 2

$$\hat{C}_0 = \alpha \Big/ \left(\alpha + 2v + \frac{\beta}{3}\frac{4N_e v}{1+4N_e v}\right) \tag{13}$$

and

$$\hat{C}_1 = \hat{C}_0/(1+4N_e v).$$

It is interesting to note that $\hat{C}_1$ is obtained by multiplying $\hat{C}_0$ by a factor $1/(1+4N_e v)$, which is the expected homozygosity if one gene unit of the present model exists as a single and independent locus in the population (Kimura & Crow, 1964). Also, the equilibrium value of C_0 is reduced by the interchromosomal crossover by the product of its rate divided by three ($\beta/3$) and $4N_e v/(1+4N_e v)$, the latter being the expected heterozygosity at the single locus.

(ii) *Some models of natural selection*

In this section, two models of natural selection which lowers the clonality are presented. The first one has been briefly considered in my previous report (Ohta, 1977).

(a) *Model I*

Let s be the selective disadvantage at $C_i = 1$ such that the fitness of the family in the ith chromosome (W_i) is equal to

$$W_i = 1 - sC_i. \tag{14}$$

For simplicity's sake, the additive effect based on clonality is assumed. We shall investigate how the equilibrium property of the mean clonality, C_0, is influenced by this kind of selection. We first determine the amount of change of C_0 by selection. C_0 may be expressed as

$$C_0 = \sum_i C_i x_i, \tag{15}$$

in which x_i is the frequency of chromosomes with C_i in the population. After selection, x_i changes by the amount

$$\Delta x_i = \frac{-s(C_i - C_0)x_i}{\overline{W}}, \tag{16}$$

where $\overline{W}$ is the mean fitness of the population and is equal to $1 - sC_0$ (cf. Wright, 1969; Crow & Kimura, 1970). Therefore the amount of change of C_0 becomes

$$\Delta C_0 = \sum_i C_i(\Delta x_i) = \frac{-s\sigma^2_{C_i}}{\overline{W}}, \tag{17}$$

https://doi.org/10.1017/S0016672300017778 Published online by Cambridge University Press

where $\sigma^2_{C_i}$ is the variance of clonality among the chromosomes in the population. Note here that the above formulation is analogous to that of the fundamental theorem of natural selection due to Fisher (1930). Also note that the mean clonality decreases by an amount $s\sigma^2_{C_i}/\overline{W}$ each generation by this type of selection.

Thus, the variance of clonality among the chromosomes determines the effectiveness of natural selection. However, its evaluation seems to be difficult. Only under the very simple situations where the interchromosomal exchange is negligible, population size is large and no selection is involved, Stewart's (1976) formula for the variance of heterozygosity is applicable

$$\sigma^2_{C_i} = \frac{2\theta}{(1+\theta)^2(2+\theta)(3+\theta)}, \tag{18}$$

in which $\theta = n^2v/k = 2v/\alpha$. The more general solution for $\sigma^2_{C_i}$ awaits future investigation. By comparing the formula (17) with the formulas (6) ~ (9) for the selectively neutral cases in the previous section, it can be seen that the clonality may be greatly influenced by natural selection. The results of the simulation studies will be shown later.

(*b*) *Model II*

I shall next introduce a slightly modified model of natural selection. Let us suppose that the fitness of the family in a chromosome is expressed by the following formula,

$$\left.\begin{aligned} W &= 1-s(C_i-C_0) \quad \text{for} \quad C_i > C_0 \\ &= 1 \qquad\qquad\quad \text{for} \quad C_i \leqslant C_0. \end{aligned}\right\} \tag{19}$$

In other words, the family in the ith chromosome has selective disadvantage only when C_i is above the average value (C_0). This is analogous to the previous model of directional selection to increase genetic information with multigene family (Ohta, 1978). Let us again formulate the change of clonality due to selection. Let $f(C_i)$ be the distribution function of C_i in the population. Then it can be shown, based on the argument analogous to the fundamental theorem of natural selection as before that ΔC_0 becomes

$$\Delta C_0 = \frac{-s}{\overline{W}}\left[\int_{C_0}^{1}(C_i-C_0)^2 f(C_i)\,dC_i - \left\{\int_{C_0}^{1}(C_i-C_0)f(C_i)\,dC_i\right\}^2\right], \tag{20}$$

where

$$\overline{W} = 1-s\int_{C_0}^{1}(C_i-C_0)f(C_i)\,dC_i.$$

This formula is difficult to evaluate, however the very rough approximation procedure is to assume normal distribution for $f(C_i)$, even if it has no probability beyond $C = 1$. If $f(C_i)$ is normally distributed with mean C_0 and variance $\sigma^2_{C_i}$, the above formula becomes

$$\Delta C_0 = \frac{-s\sigma^2_{C_i}\left(1-\frac{1}{\pi}\right)}{2\overline{W}} = \frac{-s\sigma^2_{C_i}\left(1-\frac{1}{\pi}\right)}{2\left(1-\frac{s\sigma_{C_i}}{\sqrt{(2\pi)}}\right)}. \tag{21}$$

https://doi.org/10.1017/S0016672300017778 Published online by Cambridge University Press

By comparing this result with the formula (17) it can be seen that the effectiveness of natural selection in this model is $\frac{1}{2}(1-1/\pi)$ as compared with that of the previous model I. However, the genetic load (L, defined as the amount of selective death) becomes much smaller in model II than that in model I as long as $C_0 \gg \sigma_{C_i}$, as the following comparison shows:

$$\begin{aligned} L_I &= 1-\overline{W} = sC_0 \quad \text{in model I,} \\ L_{II} &= s\sigma_{C_i}/\sqrt{(2\pi)} \quad \text{in model II.} \end{aligned} \tag{22}$$

In general, the assumption of normal distribution for C_i may not be satisfied. However, such deviation from the normal distribution would not greatly affect the above conclusion that the genetic load is more dependent on the models than the selection response.

(iii) *Simulation studies*

I have carried out the Monte Carlo experiments to check the above theory. Each experiment started from a homogeneous population of chromosomes with $C_0 = 1$. The population size was either 12·5 or 25 ($2N_e = 25$ or 50). Each generation consisted of intrachromosomal crossover, interchromosomal crossover, mutation and sampling in this sequence (case 2). At the intrachromosomal crossover, a chromosome is randomly chosen from the population and another random number determines the point of crossover and a unit is duplicated. Next, another random number determines a unit deleted and a cycle of crossover is completed. As for the ordinary interchromosomal crossover, a pair of chromosomes was chosen from the population, they are exactly paired and the point of crossover is determined by another random number. Both intra- and interchromosomal crossovers are repeated for a specified number of times per one generation. Mutation occurs at the randomly chosen unit in the randomly chosen chromosome following a specified probability. Sampling was done by randomly choosing chromosomes $2N_e$ times. In each generation, the necessary quantities such as the mean and variance of clonality are calculated and printed out.

Table 2 shows the results of the simulations together with their predicted values. The figures are the averages of 101 $\sim$ 350th generations. It can be seen from the table that the agreement is satisfactory for the case of $n = 5$ but not so for the case of $n = 10$. This is because the correlation of gene lineages becomes significant as functions of the distance between the two units in the chromosome. In other words, the condition of random arrangements of genes in the chromosome is not satisfied in the Monte Carlo experiment in which duplication and deletion always occurs by one unit when $n = 10$. It is expected here that the genes are more clustered than random arrangement. In such cases, the effect of interchromosomal crossover is smaller than the theory predicts, since the more clustered gene lineages are more resistant to the interchromosomal crossing-over than randomly arranged ones. It is also possible that the equilibrium with respect to mutation, crossover and sampling has not yet been reached when $n = 10$ and $N_e = 50$.

Table 2. *The mean clonality (C_0) and the ratio of the two measures of clonality (C_1/C_0)*

(The theoretical and the observed values are compared. In all cases, $k = 1$ (one cycle of unequal crossover per chromosome per generation). The observed values are the average of about 101–350th generations. In the last two columns, the variance of clonality ($\sigma^2_{c_i}$) is given.)

				Mean clonality (C_0)			Ratio (C_1/C_0)		Variance of clonality ($\sigma^2_{c_i}$)	
N_e	v	n	β	Approximate solution (formula 13)	Exact solution (case 2)	Observed (simulation)	Exact solution (case 2)	Observed (simulation)	Approximate solution (formula 18)	Observed (simulation)
12·5	0·002	5	0·96	0·689	0·699	0·781	0·939	0·869	—	0·028
			0·48	0·812	0·809	0·837	0·926	0·881	—	0·023
			0·24	0·870	0·875	0·929	0·919	0·936	—	0·016
			0·0	0·952	0·952	0·951	0·913	0·881	0·015	0·014
12·5	0·004	5	0·96	0·566	0·550	0·748	0·886	0·827	—	0·035
			0·48	0·698	0·691	0·796	0·862	0·819	—	0·035
			0·24	0·789	0·786	0·810	0·851	0·715	—	0·041
			0·0	0·909	0·909	0·949	0·840	0·827	0·025	0·016
25	0·004	10	1·0	0·162	0·148	0·431	0·794	0·723	—	0·022
			0·48	0·266	0·261	0·684	0·753	0·858	—	0·038
			0·24	0·386	0·387	0·752	0·736	0·896	—	0·041
			0·0	0·714	0·713	0·677	0·719	0·661	0·050	0·038

Table 3. *The mean clonality (C_0), the ratio of the two measures of clonality (C_1/C_0) and the variance of clonality ($\sigma^2_{C_i}$) for the case with selection (model I, $s = 0{\cdot}5$)*

(The other parameters are the same as the selectively neutral case (Table 2).)

N_e	v	n	β	Mean clonality (C_0) observed (simulation)	Proportionate reduction in C_0 by selection		Ratio (C_1/C_0) observed (simulation)	Variance of clonality ($\sigma^2_{C_i}$) observed (simulation)
					Expected*	Observed†		
12·5	0·002	5	0·96	0·526	0·247	0·326	0·591	0·030
			0·48	0·583	0·284	0·303	0·629	0·032
			0·24	0·611	0·343	0·342	0·587	0 034
			0·0	0·716	0·333	0·247	0·594	0·039
12·5	0·004	5	0·96	0·492	0·199	0·342	0·554	0·026
			0·48	0·539	0·259	0·323	0·505	0·032
			0·24	0·601	0·288	0·258	0·574	0·034
			0·0	0·692	0·323	0·271	0·519	0·038
25	0·004	10	1·0	0·288	0·115	0·332	0·509	0·008
			0·48	0·309	0·202	0·548	0·447	0·010
			0·24	0·393	0·326	0·477	0·564	0·016
			0·0	0·417	0·462	0·384	0·425	0·016

* Calculated by the formula $\left(\frac{s\sigma^2_{C_i}}{C_0\bar{W}}\right)\left(\alpha+2v+\frac{\beta}{3}\frac{4N_ev}{1+4N_ev}+\frac{s\sigma^2_{C_i}}{C_0\bar{W}}\right)^{-1}$.

† Calculated from the observed C_0 values, $1-\frac{C_0\text{ (with selection, this table)}}{C_0\text{ (without selection, Table 2)}}$.

In the last two columns of the table the variance of $C_i(\sigma^2_{Ci})$ is given. The expected values are given only for the case of $\beta = 0$ (no interchromosomal crossover), since the approximate solution (formula 18) does not apply for the others. However, it can be seen from the table that the observed variance is not much smaller in magnitude than the above formula suggests relative to the mean clonality.

In the next set of the Monte Carlo experiments natural selection was incorporated. The model I in the previous section has been used. The selection and sampling were simultaneously carried out; a random number determines if a sampled chromosome survives or not such that its survival probability is $W_i = 1 - C_i$. However, selection was incorporated from the 51st generation, since there was not much genetic variation for selection to work before that. The sampling was repeated until the total of $2N_e$ chromosomes were sampled for the next generation. All other parameters except selection coefficient were the same as before. The results of the simulation are given in Table 3. The figures are again the averages of 101 ~ 350th generations. Both the mean clonality (C_0) and the ratio of the two measures of clonality (between and within chromosomes, C_1/C_0) significantly decrease by selection. In the table, the proportionate reduction of C_0 due to selection is also given. Its observed value is calculated by dividing the observed C_0 in Table 3 with the corresponding observed value in selectively neutral case (Table 2) and by subtracting this ratio from 1. On the other hand, from the analyses in the previous section, the expected decrease in C_0 per generation is roughly

$$C_0\left(\alpha + 2v + \frac{\beta}{3}\frac{4N_e v}{1+4N_e v}\right)$$

for selectively neutral case (the formula 13) and the decrease due to selection is $s\sigma^2_{Ci}/\overline{W}$ (the formula 17). Therefore the expected value of the proportionate reduction of C_0 due to selection becomes very roughly

$$\text{expected proportionate reduction} = \frac{s\sigma^2_{Ci}}{C_0\,\overline{W}\left(\alpha + 2v + \frac{\beta}{3}\frac{4N_e v}{1+4N_e v} + \frac{s\sigma^2_{Ci}}{C_0\,\overline{W}}\right)}.$$

Note here that the analytical solution for σ^2_{Ci} cannot be obtained at the moment and the observed values were used for $s\sigma^2_{Ci}/C_0\,\overline{W}$. Thus the estimation is circular and the agreement between the expected and the observed reduction in C_0 only shows that the selection is working just as expected. Close comparison indicates that the observed reduction in C_0 is slightly larger than the expected value particularly when the rate of interchromosomal exchange (β) is high. This is because the clonality of the family between the different chromosomes (C_1) is considerably reduced by selection and the interchromosomal crossover becomes more effective here in reducing C_0 than in the selectively neutral case.

4. DISCUSSION

In the present study two limitations of my previous model (Ohta, 1976, 1977), i.e. the duplication or deletion of *one* unit at unequal crossover and the neglection of the ordinary interchromosomal crossover have been taken into account. From the simulation studies with several levels of the allowed latitude, I have concluded that the allowed latitude of about 10–15% of the total number of units may result in the random arrangements of gene lineages. According to Wellauer *et al.* (1976) the ribosomal genes are more or less randomly arranged along the chromosome. Thus, it appears that the above range of the allowed latitude is an appropriate one.

The present model of assigning the number of units of the gene family at unequal crossover may be too simple, and actually the number of units should drift over a long period rather than staying within the fixed interval, unless natural selection works to keep it unchanged. The relationship between this kind of generation of random variation in gene number and natural selection to work upon it may be an interesting area of research in the future.

As to the problem of the interchromosomal crossover, the genetic variation contained in a gene family of a chromosome depends on the rate of interchromosomal exchange per chromosome per generation, β. Formula (13) tells that the interchromosomal crossover works through the product, $(\beta/3)(4N_ev/(1+4N_ev))$, which corresponds to the effect of producing new mutations in increasing genetic variation ($2v$). If the rate of interchromosomal crossover per nucleotide site is of the comparable magnitude to that of mutation in higher organisms (cf. Nei, 1968), β may be at least n times larger than v since v is the mutation rate per gene unit. Then the term, $(\beta/3)(4N_ev/(1+4N_ev))$ may be much larger than $2v$. Here, if the population size is very small such that $4N_ev/(1+4N_ev)$ is much less than one, the term becomes small even if β is large.

Formula (13) further shows that $C_1/C_0 = 1/(1+4N_ev)$. This implies that the probability of identity of a randomly chosen pair of gene units in the family in the different chromosomes is lower than that in the same chromosome by this fraction. Such a relationship could be experimentally examined. In fact, the results of Wellauer *et al.* (1976) suggest that the heterogeneity in repeat length is greater among individuals than within an individual in ribosomal gene family.

In the informational multigene families such as that of immunoglobulins, the genetic variation contained in the family is the crucial factor which determines the functional diversity. Although the possible importance of somatic mutation for antibody diversity complicates the interpretation, it is likely that natural selection may be very important here. In fact, the antibody combining sites of the variable part of immunoglobulins are shown to have more variations than the other part of the molecule (Capra & Edmundson, 1977, for review). This observation seems to fit well to the model of natural selection investigated in the previous section. Natural selection works through variation in fitness among chromosomes and, it is shown that the clonality is decreased or the genetic variation is

increased roughly by an amount $s\sigma^2_{C_i}$ each generation in model I. As compared with the effect of mutation, $2vC_0$, and that of interchromosomal crossover, $(\frac{1}{3}\beta)(4N_ev/(1+4N_ev))C_0$, the selection term, $s\sigma^2_{C_i}$, may have larger effect in increasing genetic variation of a family in a chromosome. Analytical treatment to determine $\sigma^2_{C_i}$ is an important future problem.

By comparing the models I and II of selection in the previous section, it can be seen that the genetic load is model dependent whereas the selection response is not so much influenced by the model. In other words, the optimum clonality chosen has a big effect on the amount of genetic load whereas it has a relatively minor effect on selection response. This is one aspect of the difference between the present model and the traditional one. Essentially the difference is caused by the interaction effects of mutant genes (such as that the fitness is determined by the clonality) and by the coincidental evolution in the chromosome.

As for the mutational load of multigene families, it may also be reduced compared to that based on the traditional formulation. In particular for the multiplicational multigene families such as genes for histone, rRNA or tRNA (Hood *et al.* 1975), the mutational load is considered to be much reduced. For example, consider the following simple case where the deleterious mutations are detected by natural selection only after they spread to a certain frequency, y, in the family by coincidental evolution. Since the coincidental evolution is controlled purely by chance, most of the deleterious mutations disappear by chance without selection. In fact the probability of a deleterious mutant to spread to the frequency, y, in the family is $1/(ny)$ and therefore the mutational load should also be reduced to $1/(ny)$ of the conventional prediction.

I thank Dr Motoo Kimura for his stimulating discussions and many helpful suggestions. Thanks are also due to a referee for many useful comments on the manuscript.

REFERENCES

Capra, J. D. & Edmundson, A. B. (1977). The antibody combining site. *Scientific American* **236** (1), 50–59.

Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.

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

Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975). The organization, expression, and evolution of antibody genes and other multigene families. *Annual Reviev of Genetics* **9**, 305–353.

Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Kimura, M. & Ohta, T. (1969*a*). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

Kimura, M. & Ohta, T. (1969*b*). The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **63**, 701–709.

Nei, M. (1968). Evolutionary change of linkage intensity. *Nature* **218**, 1160–1161.

Nei, M. (1975). *Molecular Population Genetics and Evolution*. Amsterdam, Oxford, New York: North-Holland/American Elsevier.

OHTA, T. (1976). A simple model for treating the evolution of multigene families. *Nature* **263**, 74–76.

OHTA, T. (1977). Genetic variation in multigene families *Nature* **267**, 515–517.

OHTA, T. (1978). A model of directional selection with multigene families. *Journal of Molecular Evolution* (submitted).

PERELSON, A. S. & BELL, G. I. (1977). Mathematical models for the evolution of multigene families by unequal crossing over. *Nature* **265**, 304–310.

SMITH, G. P. (1974). Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symposia on Quantitative Biology* **38**, 507–513.

STEWART, F. M. (1976). Variability in the amount of heterozygosity maintained by neutral mutations. *Theoretical Population Biology* **9**, 188–201.

TARTOF, K. D. (1975). Redundant genes. *Annual Review of Genetics* **9**, 355–385.

WELLAUER, P. K., REEDER, R. H., DAWID, I. B. & BROWN, D. D. (1976). The arrangement of length heterogeneity in repeating units of amplified and chromosomal ribosomal DNA from *Xenopus leavis*. *Journal of Molecular Biology* **105**, 487–505.

WRIGHT, S. (1969). *Evolution and the Genetics of Populations* Vol. 2. *The Theory of Gene Frequencies*. University of Chicago Press.