Genet. Res., Camb. (1980), **35**, *pp.* 65–83
With 1 *text-figure*
Printed in Great Britain

Non-random association between electromorphs and inversion chromosomes in finite populations

BY MASATOSHI NEI AND WEN-HSIUNG LI

Center for Demographic and Population Genetics, University of Texas at Houston, Texas 77025

(*Received* 13 *July* 1979)

SUMMARY

With the aim of knowing the probable magnitude of non-random association between inversion chromosomes and electromorphs, both deterministic and stochastic studies are conducted on the evolutionary change of non-random association, which is defined as the difference in the frequency of a given allele between inversion and non-inversion chromosomes. In these studies inversion chromosomes are assumed to be subject to selection but electromorphs are selectively neutral, and recombination is allowed to occur between inversion and non-inversion chromosomes with a low frequency. The deterministic study has shown that in a variety of selective schemes for inversion chromosomes the non-random association decays at a rate equal to the recombination value in every generation. Thus, if the recombination value is of the order of $10^{-5} \sim 10^{-4}$, it would take a long time for the non-random association to disappear. Furthermore, the stochastic study has indicated that random genetic drift generates non-random association of inversions and electromorphs in finite populations and the standard error of non-random association often becomes larger than the mean. In addition to these problems the time required for the electromorph frequencies in the inversion and non-inversion chromosomes to become equal in a finite population and the probability that the population of inversion chromosomes remains monomorphic for the allele which existed in the initial inversion introduced are studied. Considering all these quantities, it is concluded that data on the non-random association between electromorphs and inversions are not very informative for the study of the maintenance of protein polymorphism. It is also indicated that in the study of association between electromorphs and inversion chromosomes non-random association or Yule's coefficient of association has a better property than the usual linkage disequilibrium measure or correlation coefficient. Implications of this study on some experimental observations are discussed.

1. INTRODUCTION

A number of authors (e.g. Prakash & Lewontin, 1968; Kojima, Gillespie & Tobari, 1970; Mukai, Mettler & Chigusa, 1971; Mukai, Watanabe & Yamaguchi, 1974; Langley, Tobari & Kojima, 1974) observed non-random association between electromorphs (alleles detected by electrophoresis) and inversion chromosomes in Drosophila. The most conspicuous is Prakash and Lewontin's observation that

0016-6723/80/2828-7720 $01.00

gene arrangement in the *ST* phylad in chromosome III, which are shared by *Drosophila pseudoobscura* and *D. persimilis*, always carries allele 1.04 at the Pt-10 locus, whereas the gene arrangements in the *SC* phylad in *D. pseudoobscura* mostly carries allele 1.06. Many authors (particularly Prakash and Lewontin) regarded this as evidence for the co-adaptation of enzyme loci in the inversion chromosome. Nei (1975), however, argued that this type of observation can be explained either by the initial linkage disequilibria when the inversion was formed or by random genetic drift. Nei & Li (1975) conducted a mathematical study of this problem, examining the probability that an inversion chromosome remains monomorphic for the same allele in two populations or species under the neutral mutation hypothesis. From this study, they concluded that Prakash and Lewontin's observation can be accommodated with the neutral mutation hypothesis.

In this study, however, Nei and Li assumed that there is no recombination between the *ST* and other gene arrangements. This assumption was based on Dobzhansky & Epling's (1948) results about the recombination frequency between inversion and non-inversion chromosomes (*ST*, *CH*, *TL*, etc.) in *D. pseudoobscura*. These authors showed that in the presence of inversion the recombination value is reduced drastically even outside the inverted segment. Inside the inverted segment they observed no recombinants among 21439 individuals tested. Because of this observation Dobzhansky and Epling speculated that crossing over in the inverted segment would act as lethals in zygotes. Nevertheless, this does not rule out the possibility of very rare recombination. Indeed, Levine (1956) observed recombination in a number of inversion heterozygotes in this organism. Furthermore, in other species of Drosophila there are a number of reports about the occurrence of recombination inside the inverted segment. Therefore, the effect of recombination should be examined carefully.

Recently Ishii & Charlesworth (1977) studied the rate of decay of non-random association between neutral electromorphs and gene arrangements in the presence of recombination by using a deterministic model. Their model, however, seems to be quite restricted, since the inversion polymorphism has been assumed to be maintained by a pair of epistatic genes located inside the inverted segment. Furthermore, in the study of the dynamics of neutral alleles the effect of random genetic drift cannot be neglected, since this is the major force of changing the frequency of neutral alleles.

The purpose of this study is first to develop a general deterministic model of the dynamics of non-random association between neutral alleles and gene arrangements in the presence of recombination and then study the effect of random genetic drift. Our main concern will be the magnitude of the non-random association in a given generation after a new inversion is formed. We shall also study the time required for the allele frequencies in the inversion and non-inversion chromosomes to become equal in a finite population and the probability of an inversion chromosome being monomorphic for a given allele, taking into account the effect of recombination. The implication of these studies on the experimental observations will then be discussed. In this paper the effect of mutation will be neglected.

2. DETERMINISTIC TREATMENT

We consider an inversion chromosome, of which the frequency is P in a large random-mating population, and the frequency of non-inversion chromosomes is $Q = 1-P$. We also consider a pair of neutral alleles (A_1 and A_2) at a given locus in these chromosomes. The locus may be located either inside or outside the inverted segment. There are four possible types of chromosomes, i.e. inversion chromosomes with allele A_1 (designated as IA_1) and $A_2(IA_2)$ and non-inversion chromosomes with $A_1(NA_1)$ and $A_2(NA_2)$, of which the frequencies are denoted by X_1, X_2, X_3 and X_4, respectively. We denote the gene frequency of A_1 in the group of inversion chromosomes by $x = X_1/P$ and that in the group of non-inversion chromosomes by $y = X_3/Q$. The frequencies of A_2 in the inversion and non-inversion groups are obviously $1-x$ and $1-y$, respectively.

The degree of non-random association between alleles and gene arrangements may be measured by a quantity similar to the linkage disequilibrium for two loci with two alleles, i.e.

$$\begin{aligned} D &= X_1X_4 - X_2X_3 \\ &= PQ(x-y). \end{aligned} \tag{1}$$

For our purpose, however, a more convenient measure is

$$d = x-y. \tag{2}$$

It is clear that if d is 0, D is also 0. We call d non-random association. Another measure of non-random association is Yule's coefficient of association. In our case it becomes

$$\begin{aligned} A_s &= (X_1X_4 - X_2X_3)/(X_1X_4 + X_2X_3) \\ &= (x-y)/(x+y-2xy). \end{aligned} \tag{3}$$

A_s takes a value between -1 to 1 and is again independent of P and Q. The usual correlation coefficient, $r = (X_1X_4 - X_2X_3)/[(X_1+X_2)(X_1+X_3)(X_3+X_4)(X_2+X_4)]^{\frac{1}{2}}$, also takes a value between -1 to 1, but it depends on P and Q as well as on x and y. In the theoretical study, however, A_s and r are not easy to work with particularly when x and y are random variables. For this purpose d is much better, and we will be mainly concerned with d in this paper.

Any type of inversion chromosome seems to occur very rarely by mutation, and thus all of the same type of gene arrangement in a population are apparently descendants of a single mutational inversion. Therefore, if this ancestral inversion happens to have allele A_1 (or A_2) at the locus under consideration, all descendant inversions will have A_1 (or A_2) unless mutation and recombination occurs. It is clear that when a new inversion has occurred, x is either 1 or 0, and consequently d is either $1-y$ or $-y$. (Note that A_s is correspondingly 1 or -1, but r is not necessarily so.) In course of time, however, d will gradually decline because of new mutation at the A locus and recombination. To understand this process, we must know the dynamics of inversion chromosome as well as the neutral alleles.

Under random mating there arise ten different genotypes, and the genotype frequencies are given by the expansion of $(X_1+X_2+X_3+X_4)^2$ (see Table 1). We

assume that natural selection occurs with respect to inversion chromosomes but alleles A_1 and A_2 are selectively neutral. There are many ways to formulate selection for inversion chromosomes. As mentioned earlier, Ishii & Charlesworth (1977) assumed that two epistatic genes inside the inverted segment maintain the inversion polymorphism. As shown by Haldane (1957), such epistatic genes generate a balanced polymorphism if there is cumulative overdominance. However, it is more likely that selection for inversion is controlled by many genes both inside and outside the inverted segment. Furthermore, inversion chromosomes are not always maintained as balanced polymorphisms. The frequencies of many inversion chromosomes in natural populations are apparently changing slowly (Dobzhansky, 1970), and comparison of gene arrangements in different species indicates that replacement of gene arrangements must occur quite frequently in the evolutionary process. Therefore, we need a model which is valid both for the cases of balanced polymorphism and changing frequency. Thus, in this paper we assume that the fitnesses of inversion genotypes are controlled by the entire set of genes in the chromosome whether there is epistasis or not. We denote the fitnesses of inversion homozygotes (II), heterozygotes (IN), and non-inversion homozygotes (NN) by W_{II}, W_{IN}, and W_{NN}, respectively (Table 1). In general, these fitnesses would vary from generation to generation, since the gene contents in inversion and non-inversion chromosomes will not remain the same because of mutation, selection acting on individual genes, and recombination (Nei, Kojima & Schaffer, 1967).

After selection, recombination is assumed to occur. Recombination is important in the sense that the allele in inversion chromosome is exchanged with the allele in non-inversion chromosome. Let r be the rate of such exchange per generation. It is noted that the exchange of alleles between two types of gene arrangements occurs only in heterozygotes IA_1/NA_2 and IA_2/NA_1. Recombination in these genotypes gives rise to the four types of chromosomes with the probabilities given in Table 1.

Therefore, the chromosome frequencies in the next generation are given by

$$\begin{aligned}\overline{W}X_1' &= X_1[(X_1+X_2)W_{II}+(X_3+X_4)W_{IN}]-rDW_{IN}\\ &= X_1W_I-rDW_{IN}, \qquad (4a)\\ \overline{W}X_2' &= X_2W_I+rDW_{IN}, \qquad (4b)\\ \overline{W}X_3' &= X_3W_N+rDW_{IN}, \qquad (4c)\\ \overline{W}X_4' &= X_4W_N-rDW_{IN}, \qquad (4d)\end{aligned}$$

where $W_I=(X_1+X_2)W_{II}+(X_3+X_4)W_{IN}$, $W_N=(X_1+X_2)W_{IN}+(X_3+X_4)W_{NN}$, $\overline{W}=(X_1+X_2)W_I+(X_3+X_4)W_N$, and the prime (′) indicates the quantities in the next generation. It is also noted that

$$P'=PW_I/\overline{W}, \quad Q'=QW_N/\overline{W}.$$

Therefore, after a straightforward algebra, we have

$$\begin{aligned}D' &= X_1'X_4'-X_2'X_3'\\ &= (x-y)P'Q'-r(x-y)PQW_{IN}/\overline{W}. \qquad (5)\end{aligned}$$

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

Thus, the non-random association in the next generation is

$$d' = D'/(P'Q') = d\left(1 - r\frac{W_{IN}\overline{W}}{W_I W_N}\right). \quad (6)$$

Table 1. *Genotype frequencies and the gametes produced*

Genotype	Frequency		Fitness	Gametes produced IA_1	IA_2	NA_1	NA_2
IA_1/IA_1	X_1^2	P^2	W_{II}	1	—	—	—
IA_1/IA_2	$2X_1X_2$			1/2	1/2	—	—
IA_2/IA_2	X_2^2			—	1	—	—
IA_1/NA_1	$2X_1X_3$	$2PQ$	W_{IN}	1/2	—	1/2	—
IA_1/NA_2	$2X_1X_4$			$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$
IA_2/NA_1	$2X_2X_3$			$r/2$	$(1-r)/2$	$(1-r)/2$	$r/2$
IA_2/NA_2	$2X_2X_4$			—	—	1/2	1/2
NA_1/NA_1	X_3^2	Q^2	W_{NN}	—	—	1	—
NA_1/NA_2	$2X_3X_4$			—	—	1/2	1/2
NA_2/NA_2	X_4^2			—	—	—	1

Therefore, the change in d is affected by selection on the inversion chromosome as well as the recombination rate. In practice, however, $Z \equiv W_{IN}\overline{W}/(W_I W_N)$ seems to be equal or close to 1 in most cases. For example, in the case of multiplicative-genic selection, we may write $W_{II} = 1$, $W_{IN} = 1-s$, and $W_{NN} = (1-s)^2$. In this case, $W_I = 1-sQ$, $W_N = (1-s)(1-sQ)$, and $\overline{W} = (1-sQ)^2$, so that $Z = 1$. It should be noted that s may vary from generation to generation. As long as the multiplicative-genic selection applies, $Z = 1$ holds irrespective of the value of s. Nei *et al.* (1967) have shown that W_{II}, W_{IN} and W_{NN} take the form of time-dependent multiplicative-genic selection under certain circumstances. In this case $Z = 1$ for all generations, though W_{II}, W_{IN}, and W_{NN} change every generation.

When selection is not multiplicative-genic, Z is not strictly 1. However, it is close to 1 in most cases. For example, in the case of overdominant selection we can write $W_{II} = 1-s$, $W_{IN} = 1$, and $W_{NN} = 1-t$. Therefore, $W_I = 1-sP$, $W_N = 1-tQ$, $\overline{W} = 1-sP^2-tQ^2$, and $Z = (1-sP^2-tQ^2)/[(1-sP)(1-tQ)]$. Suppose $s = 0{\cdot}1$, $t = 0{\cdot}2$, $P = 0{\cdot}1$, and $Q = 0{\cdot}9$. Then, $Z = 1{\cdot}19$. At equilibrium with overdominant selection we have $P = t/(s+t)$, $Q = s/(s+t)$, and $Z = (s+t)/(s+t-st)$. Therefore, in the case of $s = 0{\cdot}1$ and $t = 0{\cdot}2$, $Z = 1{\cdot}07$. Wright & Dobzhansky (1946) obtained the estimates of $s = 0{\cdot}3$ and $t = 0{\cdot}7$ in a laboratory population of *D. pseudoobscura*, though these estimates are almost certainly inflated by the associative overdominance generated at the time of sampling chromosomes for the initial population (Nei, 1975). In this case we have $Z = 1{\cdot}27$ at equilibrium.

The above examples show that Z is close to 1 in most cases. Therefore, for practical purposes (6) may be written as $d' = d(1-r)$ approximately. The d value at the tth generation (d_t) is then given by

$$d_t = d_0(1-r)^t \approx d_0 e^{-rt}, \quad (7)$$

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

since r is generally very small. This is equivalent to Ishii and Charlesworth's formula (10), though the meaning is somewhat different. Our formula is applicable to a wide variety of selection whether the inversion polymorphism is transient or stable or whether W_{II}, W_{IN}, and W_{NN} are constant or not.

In this connexion it should be noted that although the behaviour of d is simple, that of x or y is not. The recurrence equations for x and y can be obtained from (4a)–(4d) and become

$$\begin{aligned} x' &= x_1'/(x_1'+x_2') \\ &= x - rQ(x-y)\,W_{IN}/W_I, \end{aligned} \tag{8a}$$

$$y' = y + rP(x-y)\,W_{IN}/W_N. \tag{8b}$$

Therefore, the amounts of changes in x and y per generation depend not only on the selection for gene arrangement but also on the frequencies of gene arrangements. It is not easy to derive general formulae for the values of x and y in the tth generation (x_t and y_t) when the gene arrangements are subject to selection. However, when there is no selection or when P and Q stay constant and W_{IN}/W_I and W_{IN}/W_N are close to 1, we have the following recurrence equations.

$$x_{t+1} = x_t - rQ(x_t - y_t), \tag{9a}$$

$$y_{t+1} = y_t + rP(x_t - y_t). \tag{9b}$$

Solution of the above equations gives

$$x_t = x_0 - Q(x_0 - y_0)\,[1-(1-r)^t], \tag{10a}$$

$$y_t = y_0 + P(x_0 - y_0)\,[1-(1-r)^t]. \tag{10b}$$

At $t = \infty$, $x_\infty = y_\infty = Px_0 + Qy_0$, as expected.

3. STOCHASTIC TREATMENT

In finite populations gene frequency changes are probabilistic and d_t takes various values with a certain probability distribution. It is, therefore, important to know the mean and variance of d_t. Furthermore, when d_t takes various values, positive and negative, a more appropriate measure of non-random association than the mean would be the expectation of d_t^2. There are two more important properties of non-random association in finite populations. One is the number of generations required for x to become equal to y for the first time. In the deterministic theory this time is infinitely large, but in a finite population x may become equal to y in a finite number of generations. It is then interesting to know the average of this number of generations, i.e. the average first arrival time. Another quantity of interest is the probability that the inversion chromosome remains monomorphic for the initial allele A_1 at generation t. This probability is essentially the same as the probability of monomorphism of inversion chromosome Nei & Li (1975) studied under the assumption of mutation pressure. In this section we shall study these three problems.

(i) *Mean and variance of non-random association*

The stochastic treatment of non-random association is far more complicated than the deterministic treatment, and to make the mathematical computation manageable we assume constancy of P and Q and approximate the mean changes in x and y by (9a) and (9b). This assumption seems to be satisfactory when the inversion frequency is kept constant by selection but W_{IN}/W_I and W_{IN}/W_N are close to 1. In practice, however, even when the frequency of inversion chromosome is changing, it does not seem to affect the final result drastically, as will be seen from the computer simulation given later. At any rate, under this assumption we evaluate the first and second moments of d_t. To do this we use the Kolmogorov backward equation in probability theory. From (9a) and (9b) the mean changes of x and y per generation are given by $M_{\delta x} = -rQ(x-y)$ and $M_{\delta y} = rP(x-y)$, respectively. On the other hand, the variances of the changes of x and y are $V_{\delta x} = x(1-x)/N_I$ and $V_{\delta y} = y(1-y)/N_N$, where N_I and N_N are the 'effective numbers' of inversion and non-inversion chromosomes, respectively. If the effective size of the population is N, $N_I = 2PN$ and $N_N = 2QN$. The above variances are obtained under the assumption that P and Q are not subject to sampling error.

Let $\phi(x, y, p, q, t)$ be the probability density that the frequencies of A_1 in the inversion and non-inversion groups become x and y in generation t respectively, given that the initial frequencies are $p = x_0$ and $q = y_0$, respectively. It can then be shown that ϕ satisfies the following Kolmogorov backward equation when N_I and N_N are sufficiently large.

$$\frac{\partial \phi}{\partial t} = \frac{p(1-p)}{2N_I}\frac{\partial^2 \phi}{\partial p^2} + \frac{q(1-q)}{2N_N}\frac{\partial^2 \phi}{\partial q^2} - rQ(p-q)\frac{\partial \phi}{\partial p} + rP(p-q)\frac{\partial \phi}{\partial q}. \tag{11}$$

The nth moment of d at time t is given by

$$E(d_t^n) = \int_0^1 \int_0^1 (x-y)^n \phi(x, y, p, q, t)\, dx\, dy. \tag{12}$$

Using (11) and (12) and the relations $N_I = 2NP$ and $N_N = 2NQ$, we can show that

$$\frac{\partial}{\partial T}E(d_T^n) = \frac{p(1-p)}{2P}\frac{\partial^2}{\partial p^2}E(d_T^n) + \frac{q(1-q)}{2Q}\frac{\partial^2}{\partial q^2}E(d_T^n) - RQ(p-q)\frac{\partial}{\partial p}E(d_T^n) + RP(p-q)\frac{\partial}{\partial q}E(d_T^n), \tag{13}$$

where $T = t/2N$ and $R = 2Nr$.

It can easily be shown that the first moment of d is given by

$$\bar{d}_T \equiv E(d_T) = (p-q)\, e^{-RT} \tag{14a}$$

$$= (p-q)\, e^{-rt}. \tag{14b}$$

To obtain the second moment, we try the following solution.

$$E(d_T^2) = (Ap^2 + Bq^2 + Cpq + Dp + Eq)\, e^{\lambda T}.$$

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

Putting this into equation (13), we obtain the following equations.

$$-AP^{-1}-2ARQ+CRP=A\lambda, \tag{15a}$$
$$-BQ^{-1}-2BRP+CRQ=B\lambda, \tag{15b}$$
$$2ARQ+2BRP-CR=C\lambda, \tag{15c}$$
$$AP^{-1}-DRQ+ERP=D\lambda, \tag{15d}$$
$$BQ^{-1}+DRQ-ERP=E\lambda. \tag{15e}$$

It can be shown, that the eigenvalues are

$$\lambda_1=-R-\eta/3+2\sqrt{\alpha}\cos\sigma/3,$$
$$\lambda_2=-R-\eta/3-2\sqrt{\alpha}\cos\frac{\pi-\sigma}{3},$$
$$\lambda_3=-R-\eta/3-2\sqrt{\alpha}\cos\frac{\pi+\sigma}{3},$$
$$\lambda_4=-R,$$
$$\lambda_5=0,$$

where

$$\alpha=(3R^2-12R+3\eta R-3\eta+\eta^2)/q,$$
$$\beta=(54R^2-9\eta R^2+36\eta R-9\eta^2R+9\eta^2-2\eta^3)/27,$$
$$\eta=1/(PQ),$$
$$\sigma=\cos^{-1}(\beta/2\alpha\sqrt{\alpha}).$$

Thus, the solution is given by

$$E(d_T^2)=\sum_{i=1}^{5}(A_ip^2+B_iq^2+C_ipq+D_ip+E_iq)\,e^{\lambda_iT}. \tag{16}$$

However, since $\lambda_5=0$ and $E(d^2)$ approaches zero as T becomes infinity, all coefficients of e^{λ_5T} must be zero. That is, $A_5=B_5=C_5=D_5=E_5=0$. Also, from (15*c*, *d*, *e*), we can show that $A_4=B_4=C_4=0$ and $D_4=-E_4$. Equation (16) now reduces to

$$E(d_T^2)=\sum_{i=1}^{3}(A_ip^2+B_iq^2+C_ipq+D_ip+E_iq)\,e^{\lambda_iT}+(D_4p+E_4q)\,e^{-RT}. \tag{17}$$

Putting the initial condition $E(d_0^2)=(p-q)^2=p^2-2pq+q^2$ into (17), we derive the following relations:

$$A_1+A_2+A_3=1, \tag{18a}$$
$$B_1+B_2+B_3=1, \tag{18b}$$
$$C_1+C_2+C_3=-2, \tag{18c}$$
$$D_1+D_2+D_3+D_4=0, \tag{18d}$$
$$E_1+E_2+E_3+E_4=0. \tag{18e}$$

Using these relations and equations (15*a–e*), we can show that $D_4=E_4=0$ and

$$A_1=(2b_2-2b_3+c_2-c_3+b_2c_3-c_2b_3)/\Delta, \tag{19a}$$

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

$$A_2 = (2b_3 - 2b_1 + c_3 - c_1 + b_3 c_1 - c_3 b_1)/\Delta, \tag{19b}$$

$$A_3 = (2b_1 - 2b_2 + c_1 - c_2 + b_1 c_2 - c_1 b_2)/\Delta, \tag{19c}$$

$$B_i = b_i A_i, \tag{19d}$$

$$C_i = c_i A_i, \tag{19e}$$

$$D_i = A_i(\lambda_i^2 + R\lambda_i)^{-1}[P^{-1}(\lambda_i + RP) + b_i RPQ^{-1}], \tag{19f}$$

$$E_i = A_i(\lambda_i^2 + R\lambda_i)^{-1}[RQP^{-1} + (\lambda_i + RQ)b_i Q^{-1}], \tag{19g}$$

where

$$i = 1, 2 \text{ or } 3 \quad \text{and} \quad \Delta = b_1(c_2 - c_3) + b_2(c_3 - c_1) + b_3(c_1 - c_2),$$

$$b_i = P^{-1}Q(\lambda_i + 2RQ + P^{-1})/(\lambda_i + 2RP + Q^{-1}),$$

$$c_i = (\lambda_i + 2RQ + P^{-1})/(RP).$$

We can now evaluate $E(d_T^2)$ by using these coefficients and equation (17). In particular, when $P = Q = \frac{1}{2}$, we have the following simple formula.

$$E(d_T^2) = \sum_{i=1}^{2} [A_i(p^2 + q^2) + C_i pq + D_i(p + q)]e^{\lambda_i T}, \tag{20}$$

where $\lambda_1 = -R - 1 + (R^2 + 1)^{\frac{1}{2}}$, $\lambda_2 = -R - 1 - (R^2 + 1)^{\frac{1}{2}}$, $A_1 = -\lambda_1/(\lambda_2 - \lambda_1)$, $A_2 = \lambda_2/(\lambda_2 - \lambda_1)$, $C_1 = -2\lambda_1(2 + R + \lambda_1)R^{-1}(\lambda_2 - \lambda_1)^{-1}$, $C_2 = 2\lambda_2(2 + R + \lambda_2)R^{-1}(\lambda_2 - \lambda_1)^{-1}$, $D_1 = -2(\lambda_2 - \lambda_1)^{-1}$ and $D_2 = 2/(\lambda_2 - \lambda_1)$.

Formula (14*b*) is identical with the equivalent formula (7) for the deterministic model, and shows that the change in mean of d is independent of N, P, and Q. The second moment of d is, however, a complicated function of these parameters. Table 2 shows the values of $E(d)$ and $E(d^2)$ for various values of N, P, Q, and t. In all cases $p = 1$, $q = 0{\cdot}1$, and $d_0 = 0{\cdot}9$ are assumed. It is clear that $E(d)$ declines very slowly when the recombination value is small. For example, in Case 1, where $P = Q = 0{\cdot}5$, $E(d)$ for $R = 0{\cdot}1$ is 0·546 even at the $10N$th generation. $E(d^2)$ declines at a slower rate than that for $E(d)$ and is often larger than $E(d)$ at later generations. For example, in Case 1 $E(d)$ and $E(d^2)$ for $R = 1$ and $t = 10N$ are 0·006 and 0·025, respectively. Thus, the square root of $E(d^2)$ is 0·158. Therefore, in later generations even if the mean of d is small, a considerable amount of non-random association is expected to exist. In general, $s_d = [E(d^2)]^{\frac{1}{2}}$ gives a better idea of the magnitude of non-random association than $E(d)$ in finite populations. Needless to say, the larger value of s_d than $E(d)$ is caused by random genetic drift (Hill & Robertson, 1968).

In Table 2 the standard deviation of d, i.e. $\sigma = [E(d^2) - E^2(d)]^{\frac{1}{2}}$ is also presented. As expected, this value increases with increasing generation. In later generations σ can be larger than the mean of d. This indicates that d can be negative even if the initial value is a large positive value. It is also noted that s_d is of the same order of magnitude as that of $E(d)$ in early generations, but as $E(d)$ decreases, it tends to be close to σ.

When the frequency of inversion chromosome is small compared with that of non-inversion (Case 2), the change in $E(d^2)$ with time is somewhat complicated but the magnitude of $E(d^2)$ is more or less the same as that for Case 1. A detailed comparison of Case 1 and Case 2 shows that when $2Nr = 0{\cdot}1$, $E(d^2)$ for Case 1 is greater than that for Case 2 in the early generations as well as in the very late generations,

whereas for $t = 2N \sim 20N$ generations it is smaller. This complicated pattern of the change of $E(d^2)$ can be explained if we examine the eigenvalues for these cases. As will be seen from Table 3, there are three eigenvalues for Case 2 but there are two for Case 1, and it is noted that $|\lambda_3|$ is much larger than $|\lambda_1|$ and $|\lambda_2|$. Therefore, in the early generations $E(d^2)$ is expected to decline faster in Case 2 than in Case 1. In

Table 2. *Means* $[E(d)]$, *second moments* $[E(d^2)]$ *and standard deviations* (σ) *of non-random association*

($p = 1$, $q = 0{\cdot}1$, $d_0 = 0{\cdot}9$ and $E(d_0^2) = 0{\cdot}81$ are assumed.)

		Generation						
$2Nr$		$0{\cdot}02N$	$0{\cdot}2N$	N	$2N$	$10N$	$20N$	$100N$
		Case 1: $P = Q = 0{\cdot}5$						
0·1	$E(d)$	0·899	0·891	0·856	0·814	0·546	0·331	0·0061
	$E(d^2)$	0·810	0·811	0·801	0·774	0·533	0·331	0·0074
	σ	0·042	0·130	0·261	0·333	0·485	0·471	0·0859
1	$E(d)$	0·891	0·814	0·546	0·331	0·006	4×10^{-5}	2×10^{-22}
	$E(d^2)$	0·795	0·685	0·412	0·272	0·025	0·001	9×10^{-14}
	σ	0·043	0·146	0·337	0·403	0·158	0·037	3×10^{-7}
10	$E(d)$	0·814	0·331	0·006	4×10^{-5}	2×10^{-22}	3×10^{-44}	—
	$E(d^2)$	0·665	0·139	0·032	0·020	4×10^{-4}	4×10^{-6}	—
	σ	0·048	0·172	0·179	0·141	0·021	0·002	—
		Case 2: $P = 0{\cdot}1$, $Q = 0{\cdot}9$						
0·1	$E(d)$	0·899	0·891	0·856	0·814	0·546	0·331	0·0061
	$E(d^2)$	0·810	0·806	0·800	0·783	0·549	0·336	0·0066
	σ	0·035	0·112	0·260	0·346	0·501	0·476	0·0813
1	$E(d)$	0·891	0·814	0·546	0·331	0·006	4×10^{-5}	2×10^{-22}
	$E(d^2)$	0·795	0·698	0·485	0·327	0·019	6×10^{-4}	5×10^{-16}
	σ	0·038	0·186	0·432	0·466	0·137	0·024	2×10^{-7}
10	$E(d)$	0·814	0·331	0·006	4×10^{-5}	2×10^{-22}	3×10^{-44}	—
	$E(d^2)$	0·667	0·186	0·043	0·026	5×10^{-4}	4×10^{-6}	—
	σ	0·066	0·277	0·208	0·161	0·024	0·002	—

Table 3. *Eigenvalues for various values of P and 2Nr*

$2Nr$	P	λ_1	λ_2	λ_3
0·1	0·5	−0·095	−2·10	
	0·1	−0·098	−1·13	−10·18
1	0·5	−0·586	−3·41	
	0·1	−0·693	−1·60	−11·82
10	0·5	−0·950	−21·05	
	0·1	−0·961	−11·17	−28·98

the intermediate generations, however, the rate of decrease in $E(d^2)$ is largely determined by λ_2 of which the absolute value is greater in Case 1 than in Case 2. Thus, $E(d^2)$ becomes smaller in Case 1 than in Case 2. On the other hand, in the very late generations the rate of decrease in $E(d^2)$ is dictated by λ_1. Since $|\lambda_1|$ is smaller in Case 1 than in Case 2, $E(d^2)$ is expected to be larger in the former than in the latter.

A similar trend has been observed in the cases of $2Nr = 1$ and $2Nr = 10$ though in Table 2 the pattern of the change in $E(d^2)$ in the early generations is not clear because of the rapid decrease of $E(d^2)$ for these cases.

(ii) *Computer simulation*

In the above formulation we have implicitly assumed that the frequency of inversion chromosome reaches the equilibrium value immediately after the occurrence by mutation. In practice, of course, the frequency gradually increases. To see the effect of this gradual increase, we have conducted a computer simulation. In this simulation we assumed that the inversion chromosome is subject to overdominant selection and the fitnesses of II, IN, and NN are $1-s$, 1, and $1-s$, respectively, s being 0·01. The number of inversion chromosomes (N_I) in the population was initially one and increased to $2N/2 = N$. It took about 1500 generations for N_I to become close to N when $N = 10^4$ and about 2000 generations when $N = 10^5$. In each generation N_I was rounded to an integral number which was the closest to the N_I value obtained after deterministic selection. The initial frequency of the A_1 allele was $x = 1$ in the inversion chromosomes and $y = 0{\cdot}1$ in the non-inversion chromosomes. Selection, recombination, and sampling of gametes were assumed to occur in this order. Selection and recombination were treated deterministically in the scheme mentioned earlier. In practice, of course, the frequency of new inversions is subject to stochastic changes and the majority of them will be eliminated (Nei & Roychoudhury, 1973), but since we are interested in only those inversions which have become polymorphic in the population, we can neglect this factor. Our scheme of random sampling was briefly as follows: The number of A_1 genes in the non-inversion chromosomes was assumed to follow the binomial distribution with parameters y and N_N. This binomial distribution was approximated by the Poisson distribution if $N_N y \leqslant 15$ and by the normal distribution if $N_N y > 15$. Similarly, the number of A_1 genes in the inversion chromosomes was assumed to follow the binomial distribution with parameters x and N_I. When $N_I \leqslant 50$, the binomial sampling was directly conducted to determine the value of x for the next generation. When $50 < N_I \leqslant 200$, the binomial distribution was approximated either by the normal distribution or by the Poisson distribution. The normal approximation was used when $5 < N_I x < N_I - 5$; otherwise the Poisson approximation was used with parameter $N_I x$ or $N_I(1-x)$. When $N_I > 200$, the normal approximation was used if $10 < N_I x < N_I - 10$; otherwise the Poisson approximation was used. We studied two different cases, i.e. $N = 10^4$, $r = 10^{-5}$ (Case 1) and $N = 10^5$, $r = 10^{-4}$ (Case 2). In both cases 1000 replications were conducted, and the means of d and d^2 were computed for several different generations.

The results obtained are presented in Table 4 together with the theoretical values obtained by (14*b*) and (20). These theoretical values were obtained under the assumption of $P = Q = 0{\cdot}5$. It is clear that both the mean and second moment of non-random association (d) are close to the theoretical values. There is some tendency for the second moment to be a little larger than the theoretical value. This is apparently caused by the small value of P in the early generations where the effect of

random genetic drift is large. Nevertheless, for the practical purpose our formulae (14*b*) and (20) give sufficiently accurate results. The reason for this seems to be that in the early generations, where the frequency of inversion chromosomes increases rapidly, the change in x is almost deterministic and by the time the variance of x becomes significantly large the frequency of inversion chromosomes has increased to a value close to the equilibrium frequency.

Fig. 1 shows the distributions of d obtained by simulation for $t = 2000$, 5000 and

Table 4. *Means* $[E(d)]$ *and second moments* $[E(d^2)]$ *of non-random association*

(The results from computer simulation are based on 1000 replications. The initial values of x and y are 1 and 0·1, respectively.)

Generation		1000	2000	5000	10000
	Case 1: $N = 10^4$, $r = 10^{-5}$				
$E(d)$	Formula (14*b*)	0·891	0·882	0·856	0·814
	Simulation:	0·895	0·890	0·864	0·833
$E(d^2)$	Formula (20)	0·803	0·796	0·775	0·740
	Simulation:	0·809	0·807	0·791	0·774
	Case 2: $N = 10^5$, $r = 10^{-4}$				
$E(d)$	Formula (14*b*)	0·814	0·737	0·546	0·331
	Simulation:	0·813	0·734	0·544	0·327
$E(d^2)$	Formula (20)	0·664	0·546	0·306	0·124
	Simulation:	0·700	0·574	0·321	0·129

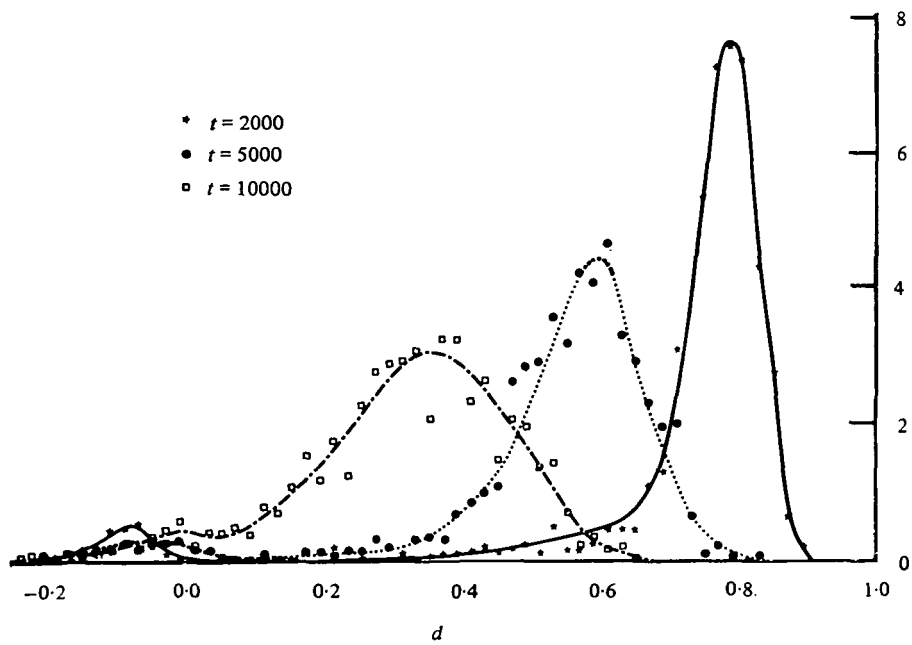

Fig. 1. Distributions of d for $t = 2000$, 5000, and 10000 for the case of $N = 10^5$ and $r = 10^{-4}$. These distributions were obtained by computer simulation. The curves representing the distributions are visually fitted to the simulation results. The number of replications used is 1000 for each generation.

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

10000 for the case of $N = 10^5$ and $r = 10^{-4}$. The curves representing the distributions were visually fitted to the simulation results just to show the general pattern. Since the number of replications is only 1000, they may not be very accurate. It is seen that in the case of $t = 2000$ the d value is generally close to the mean (0·737) but may become much smaller. There is a secondary peak around $-0{\cdot}08$ in the distribution. It is not clear why the secondary peak arises at this value of d. It is possible that the actual location of the secondary peak is at $d = 0$ and by chance it has moved to $d = -0{\cdot}08$. At $t = 5000$ and 10000 the distribution of d becomes flatter than that at $t = 2000$ and has a long tail toward the negative side. In both generations there is a secondary peak around $d = 0$.

(iii) *Mean first arrival time*

Let us now study the number of generations required for the frequency (x) of A_1 in the inversion chromosomes to become equal to the frequency (y) of A_1 in the non-inversions, i.e. the first arrival time. We note that since the initial frequency of non-inversion chromosomes is much larger than that of inversion chromosomes, y generally does not change as much as x does in the evolutionary process. Therefore, we assume that y remains constant. Although this assumption is not strictly correct, it simplifies the mathematical treatment considerably and yet gives a rough idea about the first arrival time. At any rate, under this assumption the mean first arrival time can be computed by Maruyama's (1977) formula (see also Nagylaki, 1974). In our case the mean change in x per generation is $M_{\delta x} = -rQ(x-y)$, whereas the variance is $V_{\delta x} = x(1-x)/(2N_I)$. We have a reflecting boundary at $x = 1$ and a absorbing boundary at $x = y$. If we put these conditions in formula (4·58) of Maruyama (1977), we obtain the mean sojourn time at each gene frequency class. The mean first arrival time is given by the integration of this mean sojourn time from y to 1. It becomes

$$t(y) = 4N_I \int_y^1 (1-\zeta)^{A(1-y)-1} \zeta^{Ay-1} \int_y^\zeta (1-\xi)^{-A(1-y)} \xi^{-Ay} \, d\zeta \, d\xi, \tag{21}$$

where $A = 4N_I rQ$.

The mean first arrival times for two different values of y, i.e. 0·1 and 0·5, are presented in Table 5. The mean first arrival time depends on the values of $4N_I rQ$ and y. When $4N_I rQ$ is small, the time is very long. For example, $N_I = 10^4$, $Q = 0{\cdot}5$, and $r = 5 \times 10^{-5}$, it is 60000 generations for $y = 0{\cdot}1$. When $y = 0{\cdot}5$, the time is a little shorter than that for $y = 0{\cdot}1$.

(iv) *Probability of monomorphism*

The probability that the inversion chromosome remains monomorphic for the initial allele A_1 at generation t can be obtained by the same method as that of Nei & Li (1975). We again assume that the frequency (y) of A_1 among non-inversion chromosomes remains constant. This assumption would lead to an underestimation of the probability of monomorphism, since in practice y increases with time. At any rate, under this assumption we have $M_{\delta x} = -rQ(x-y)$ and $V_{\delta x} = x(1-x)/(2N_I)$.

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

The process is essentially the same as that for the change of gene frequency under linear pressure, and the distribution of x at time t is given by

$$\phi(p, x, t) = \sum_{i=0}^{\infty} X_i(x) \exp\left[-i\left(rQ + \frac{i-1}{4N_I}\right)t\right], \tag{22}$$

$$X_i(x) = x^{B-1}(1-x)^{A-B-1} F(A+i-1, -i, A-B, 1-x) \times F(A+i-1, -i, A-B, 1-p) \times \frac{\Gamma(A-B+i)\,\Gamma(A+2i)\,\Gamma(A+i-1)}{i!\,\Gamma^2(A-B)\,\Gamma(B+i)\,\Gamma(A+2i-1)},$$

where $A = 4N_I rQ$, $B = Ay$ and p is the initial frequency of x (Crow & Kimura, 1970). If $p = 1$, then $F(A+i-1, -i, A-B, 1-p) = 1$. The probability that x is equal to or greater than $1-\alpha$ is

$$P(x \geqslant 1-\alpha; t) = \int_{1-\alpha}^{1} \phi(p, x, t)\, dx. \tag{23}$$

Table 5. *Mean first arrival time of x to y*

(Time is measured in the unit of $4N_I$ generations and the initial value of x is 1.)

Frequency of A_1 among non-inversions	$4N_I rQ$ = 0·1	1	10
$y = 0{\cdot}1$	15·76	1·50	0·267
$y = 0{\cdot}5$	12·29	1·29	0·196

Table 6. *Probability that the frequency (x) of A_1 in the inversion group is equal to or greater than $1-\alpha$*

(N_I, r, and Q are the number of inversion chromosomes, recombination value, and the frequency of non-inversions, respectively.)

		Generation				
$4N_I rQ$		$0{\cdot}04N_I$	$0{\cdot}4N_I$	$2N_I$	$4N_I$	$20N_I$
0·1	$\alpha = 0{\cdot}01$	0·978	0·841	0·724	0·670	0·463
	$\alpha = 0{\cdot}05$	0·999	0·948	0·834	0·774	0·536
	$\alpha = 0{\cdot}10$	1·000	0·985	0·888	0·827	0·574
1	$\alpha = 0{\cdot}01$	0·676	0·124	0·029	0·014	0·002
	$\alpha = 0{\cdot}05$	0·995	0·445	0·119	0·060	0·008
	$\alpha = 0{\cdot}10$	0·999	0·682	0·215	0·111	0·016
10	$\alpha = 0{\cdot}01$	10^{-6}	$1{\cdot}1 \times 10^{-14}$	$1{\cdot}7 \times 10^{-18}$	$1{\cdot}0 \times 10^{-18}$	$1{\cdot}0 \times 10^{-18}$
	$\alpha = 0{\cdot}05$	0·080	$1{\cdot}6 \times 10^{-8}$	$3{\cdot}3 \times 10^{-12}$	$2{\cdot}0 \times 10^{-12}$	$2{\cdot}0 \times 10^{-12}$
	$\alpha = 0{\cdot}10$	0·723	$5{\cdot}6 \times 10^{-6}$	$1{\cdot}6 \times 10^{-9}$	$1{\cdot}0 \times 10^{-9}$	$1{\cdot}0 \times 10^{-9}$

Table 6 shows the probabilities of monomorphism for three different values of α, i.e. 0·01, 0·05, and 0·10. In this table $p = 1$ and $y = 0{\cdot}1$ are assumed. It is clear that the probability of monomorphism remains high for a long time if the recombination value is low. However, it declines very rapidly if the recombination value is high. Since the values in Table 5 are obtained under the assumption that y is

https://doi.org/10.1017/S001667230001394X Published online by Cambridge University Press

constant, they are expected to be underestimates, particularly when t is large, as mentioned earlier.

4. DISCUSSION

We have seen that the degree of non-random association depends on the recombination value, the effective numbers of inversion and non-inversion chromosomes, and the time after the inversion chromosome was introduced. The last two factors are generally unknown, but some information is available about the recombination value particularly in *Drosophila melanogaster*. Ishii & Charlesworth (1977) recently reviewed the works on recombination values inside inverted segments in Drosophila and concluded that it is of the order of 10^{-4}. According to Chovnick (1973), however, the majority of recombinants from inversion heterozygotes are the products of gene conversion rather than the classical double crossovers. Examining about 5×10^6 zygotes by using a selective system, Chovnick estimated that in inversion heterozygotes of *D. melanogaster* the rate of gene conversion in the rosy region is of the order of 10^{-5}, whereas the frequency of double crossovers is much less. For our purpose, of course, double crossing-over and gene conversion have the same effect. We can therefore conclude that the recombination value is of the order of $10^{-5} \sim 10^{-4}$. Of course, the recombination value would depend on the location of the locus under consideration. If the locus is located close to the breakpoint, the recombination value could be even smaller.

At any rate, with this magnitude of recombination value, the non-random association introduced by a new inversion mutation or migration is expected to last for a long time, particularly in small populations. For example, if $r = 5 \times 10^{-5}$, $N = 10^4$, $P = 0{\cdot}1$, $p = 1$, $q = 0{\cdot}1$, and $d_0 = 0{\cdot}9$, s_d is 0·580 even after 20000 generations. If there are 10 generations in a year in Drosophila, this corresponds to 2000 years.

In the present paper we have neglected the effect of mutations. This can be justified as long as the mutation rate is lower than the recombination value and the evolutionary time considered is not extremely large. However, when the evolutionary time is long, our formulation is no longer realistic, since in our formation the population will eventually become monomorphic for an electromorph. When a pair of gene arrangements are selectively maintained in the population and new mutation occurs repeatedly at the protein locus, $E(d^2)$ is expected to reach an equilibrium value as in the case of ordinary linkage disequilibrium (Ohta & Kimura, 1969). This equilibrium value of $E(d^2)$ is expected to be considerable when $4N_I r$ is about 1 or less. Therefore, as long as $4N_I r$ remains small, protein loci and inversion chromosomes would generally be associated non-randomly.

In practice, it is not easy to determine the magnitudes of N_I and N_N in natural populations. However, the size of natural populations generally fluctuates considerably from generation to generation. Since the long-term effective population size is the harmonic mean of the effective sizes for individual generations, N_I and N_N could be much smaller than the present actual sizes. This is particularly so when a population starts from a small number of founders. Ishii & Charlesworth (1977) pointed out the possibility that the United States and Japanese populations

of *Drosophila melanogaster* are only about a few hundred years old, since these populations apparently immigrated from West Africa through human migration and travel. If we consider this type of population fluctuation, most of the association between inversions and electromorphs so far observed seems to be explainable by historical relics and genetic drift. Of course, this does not mean that the electromorphs on these chromosomes are selectively neutral. Rather it indicates that this type of information is not very useful for the study of the maintenance of protein polymorphism.

However, a special comment seems to be necessary on the strong association of gene arrangements and electromorphs at the *Pt*-10 locus in *Drosophila pseudoobscura* and *D. persimilis*. As mentioned earlier, Nei & Li (1975) have shown that the identical monomorphism of the *ST* phylad for the 1·04 allele in the two species can be explained by historical accidents and genetic drift if the effect of recombination is neglected. The present study indicates that the recombination in inversion heterozygotes decreases the probability of monomorphism considerably. In the case of the *Pt*-10 locus, however, this does not pose any serious problem for the hypothesis of historical accident. This is because *D. persimilis* is monomorphic for the electromorph 1·04, whereas the polymorphism at the *Pt* – 10 locus in *D. pseudoobscura* may be of recent origin. It is known that gene arrangement *ST* exists in both species but the gene arrangements of the *SC* phylad are present only in *D. pseudoobscura* with a relatively low frequency (Dobzhansky, 1970; Prakash & Lewontin, 1971). It is thus possible that the *SC* phylad was derived from the *ST* gene arrangement through the so-called Hypothetical gene arrangement some time after the two species diverged (Dobzhansky & Sturtevant, 1938). If this is the case, the first gene arrangement in the *SC* phylad as well as the parental gene arrangement Hypothetical seems to have had allele 1·06 at the *Pt*-10 locus. This can happen if the population of *ST* gene arrangements was polymorphic for alleles 1·04 and 1·06 when the Hypothetical was formed but later allele 1·06 was eliminated by genetic drift. Therefore, if the time of occurrence of the *SC* gene arrangement is rather recent, Prakash & Lewontin's (1968) observation can easily be accommodated with the neutral mutation hypothesis. We also note that there is no information about the rate of gene exchange (recombination frequency) at this locus in inversion heterozygotes.

At this point some readers might recall Epling's (1944) conjecture that the *SC* gene arrangement has existed from the time of Miocene (13 million years ago). From this conjecture and the present distribution of *D. pseudoobscura*, Prakash & Lewontin (1968) speculated that the *ST* and *SC* phylads diverged about $3 \sim 5$ million years ago. In our view, however, Epling's conjecture is illogical and cannot be accepted. (We will be happy to send our detailed comments on Epling's paper if any one is interested.) Indeed, Epling himself was not sure about his conjecture and stated: 'It is possible, as originally implied by Dobzhansky & Sturtevant (1938), that the present distribution of gene arrangements took place in relatively recent time, perhaps Pleistocene or post-Pleistocene.' In this connexion it should be mentioned that the genetic distance between *D. pseudoobscura* and *D. persimilis*

for protein loci corresponds to a divergence time of about 250000 years (Nei, 1975), though this estimate is also subject to a large standard error.

Recently, Olvera *et al.* (1979) indicated the possibility that the *SC* is older than the *ST*, the latter being derived from the former through the Hypothetical. In this case the non-random association of the electromorphs at the *Pt*-10 locus and inversions can still be accommodated with the neutral mutation hypothesis if we assume that the *SC* or Hypothetical was polymorphic for 1·04 and 1·06 and the first *ST* gene arrangement happened to have 1·04. It is also possible that the Hypothetical is oldest and both the *ST* and *SC* were derived from this gene arrangement. In this case the neutral explanation is simpler, since we can assume that the Hypothetical was polymorphic for 1·04 and 1·06 and the *ST* was derived from a gamete carrying allele 1·04, whereas the *SC* was derived from a gamete carrying 1·06.

Table 7. *Non-random associations* (d) *and coefficients of associations* (A_s) *for the pairs of electromorphs and inversions in the same arms of chromosomes II and III in* Drosophila melanogaster

(x and y are the frequencies of a given allele in the inversion and non-inversion chromosomes. B = Brownsville, U.S.A.; Katsunuma, Japan. Data are taken from Langley *et al.* (1974).)

Locus and inversion	Population	d	A_s	x	y
αGpd–In(2L)*t*	*B*	0·136*	1·000*	1·000	0·864
	K	0·079	0·317	0·896	0·817
Adh–In(2L)*t*	*B*	0·325**	1·000**	1·000	0·675
	K	0·746**	1·000**	1·000	0·254
Amy–In(2R)*NS*	*B*	0·112*	0·791*	0·983	0·871
	K	0·081	0·627	0·973	0·892
Est-6–*In*(3L)*P*	*B*	0·449**	0·830**	0·897	0·448
	K	0·256**	0·583**	0·412	0·156
Pgm–In(3L)*P*	*B*	0·087	0·593	0·966	0·879
	K	0·226	0·593	0·882	0·656

* Significant from 0 at the 5% level.
** Significant from 0 at the 1% level.

In the past the degree of non-random association between inversions and electromorphs has been measured either by the usual linkage disequilibrium (D) or by the correlation coefficient (r). However, these measures are not very appropriate from the theoretical point of view. As is clear from (1), the linkage disequilibrium is a product of PQ and d. Therefore, when P or Q is small, D is also small even if d is large. We note that P is necessarily small when a new inversion is introduced and thus D is very small. In this case r is also expected to be generally small. However, as P increases due to selection or genetic drift, D or r may increase even if d declines steadily in every generation. On the other hand, the coefficient of association (A_s) gives either 1 or -1 when a new inversion is introduced as mentioned earlier. We believe that this is a good property for studying non-random association of electromorphs and inversions. Of course, the relationship between A_s and evolutionary time is not as simple as that of d. However, a high value of A_s indicates that the

inversion chromosome was recently introduced either by mutation or by migration or the recombination frequency is very low.

Langley *et al.* (1974) published gametic frequencies for many different pairs of inversions and electromorphs in two populations of *Drosophila melanogaster*. Table 7 gives the values of d and A_s for five pairs of enzyme loci and inversion in the same arms of chromosomes II and III. The data for the inversions in the right arm of chromosome III are not included, since there are two polymorphic inversions and x and y cannot be obtained from their paper. Table 7 shows that the A_s value is generally high but only six of the ten estimates are significantly different from 0. (The statistical test was done by the usual $2 \times 2\chi^2$ test.) In the case of *Adh–In*(2*L*)*t* the A_s value is 1·0 in both Brownsville and Katsunuma, i.e. one allele (A_1) at the *Adh* locus is always associated with inversion *In*(2*L*)*t*. According to Langley *et al.* (1974), the linkage disequilibrium for *Adh–In*(2*L*)*t* is 0·047 in Brownsville and 0·111 in Katsunuma, whereas the correlation coefficient is 0·280 in Brownsville and 0·589 in Katsunuma. Therefore, it is difficult to recognize the complete association of A_1 and *In*(2*L*)*t* from these quantities. It is also noted that in each enzyme locus and inversion pair there is a considerable difference in the value of d between the Brownsville and Katsunuma populations but the difference in A_s is relatively small except in *XGpd–In*(2*L*)*t*. The difference in d could be due to either natural selection or the bottleneck effect at the time of formation of these populations.

This study was supported by research grants from the U.S. National Institute of Health and the U.S. National Science Foundation.

REFERENCES

CHOVNICK, A. (1973). Gene coversion and transfer of genetic information within the inverted region of inversion heterozygote. *Genetics* **75**, 123–131.

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

DOBZHANSKY, TH. (1970). *Genetics of the Evolutionary Process*. New York: Columbia University Press.

DOBZHANSKY, TH. & EPLING, C. (1948). The suppression of crossing over in inversion heterozygotes of *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences, U.S.A.* **34**, 137–141.

DOBZHANSKY, TH. & STURTEVANT, A. H. (1938). Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23**, 28–64.

EPLING, C. (1944). *Contributions to the Genetics, Taxonomy, and Ecology of* Drosophila pseudoobscura *and Its Relatives*. III. The historical background. *Carnegie Institute Washington Publication*, no. 554, pp. 147–183.

HALDANE, J. B. S. (1957). The conditions for coadaptation in polymorphism for inversions. *Journal of Genetics* **55**, 218–225.

HILL, W. G. & ROBERTSON, A. (1968). Linkage disequilibrium in finite populations. *Theoretical & Applied Genetics* **38**, 226–231.

ISHII, K. & CHARLESWORTH, B. (1977). Associations between allozyme loci and gene arrangements due to hitch-hiking effects of new inversions. *Genetical Research* **30**, 93–106.

KOJIMA, K., GILLESPIE, J. & TOBARI, Y. N. (1970). A profile of *Drosophila* species' enzymes assayed by electrophoresis. I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. *Biochemical Genetics* **4**, 627–637.

LANGLEY, C. H., TOBARI, Y. N. & KOJIMA, K. (1974). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**, 921–936.

LEVINE, R. P. (1956). Crossing over the inversions in coadapted systems. *American Naturalist* **90**, 4–45.

MARUYAMA, T. (1977). *Stochastic Problems in Population Genetics*. Berlin: Springer.

MUKAI, T., METTLER, L. E. & CHIGUSA, S. I. (1971). Linkage disequilibrium in a local population of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **68**, 1065–1069.

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). The moments of stochastic integrals and the distribution of sojourn times. *Proceedings of the National Academy of Sciences, U.S.A.* **71**, 746–749.

NEI, M. (1975). *Molecular Population Genetics and Evolution*. Amsterdam: North-Holland.

NEI, M., KOJIMA, K. & SCHAFFER, H. F. (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**, 741–750.

NEI, M. & LI, W.-H. (1975). Probability of identical monomorphism in related species. *Genetical Research* **26**, 31–43.

NEI, M. & ROYCHOUDHURY, A. K. (1973). Probability of fixation and fixation time of an overdominant mutation. *Genetics* **74**, 371–380.

OHTA, T. & KIMURA, M. (1969). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* **68**, 571–580.

OLVERA, O., POWELL, J. R., DE LA ROSA, M. E., SALCEDA, V. M., GASO, M. I., GUZMAN, J., ANDERSON, W. W. & LEVINE, L. (1979). Population genetics of Mexican *Drosophila*. VI. Cytogenetic aspects of the inversion polymorphism in *Drosophila pseudoobscura*. *Evolution* **33**, 381–395.

PRAKASH, S. & LEWONTIN, R. C. (1968). A molecular approach to the study of genic heterozygosity in natural populations. III. Direct evidence of coadaptation in gene arrangements of *Drosophila*. *Proceedings of the National Academy of Sciences, U.S.A.* **59**, 398–405.

PRAKASH, S. & LEWONTIN, R. C. (1971). A molecular approach to the study of genic heterozygosity in natural populations. V. Further direct evidence of coadaptation in inversions of *Drosophila*. *Genetics* **69**, 405–408.

WRIGHT, S. & DOBZHANSKY, TH. (1946). Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics* **31**, 125–156.