Genet. Res., Camb. (1977), **30**, *pp.* 93–106
With 3 text-figures
Printed in Great Britain

Associations between allozyme loci and gene arrangements due to hitch-hiking effects of new inversions

By K. ISHII* and B. CHARLESWORTH
School of Biological Sciences, University of Sussex, Brighton BN1 9QG, Sussex

(*Received* 9 *November* 1976)

SUMMARY

This paper examines the hypothesis that the associations between alleles at allozyme loci and gene arrangements, which are observed in many Drosophila populations, result from inversions remaining associated with the alleles contained in the gametes in which they originally occurred. The effects of double crossing over in heterokaryotypes and of selection at loci linked to the allozyme loci, which are themselves assumed to be neutral, are studied theoretically. It is concluded that selection at linked loci is unlikely to have an important effect in retarding the decay of this type of association. The literature on associations between allozymes and gene arrangements is surveyed, and it is concluded that a 'neutralist' explanation of this sort cannot be excluded, except possibly in some cases where the allozyme locus is outside the inversion breakpoints and crosses over with measurable frequency in heterokaryotypes.

1. INTRODUCTION

Prakash & Lewontin (1968) reported the existence of strong non-random associations (linkage disequilibrium) between alleles at two allozyme loci and third chromosome gene arrangements of *Drosophila pseudoobscura*. Similar results have since been described by many workers on several species of Drosophila, as reviewed in the discussion section of this paper. Prakash & Lewontin suggested that these associations provide evidence for selection acting on the allozyme loci themselves. According to them, only interactions in fitness effects between the allozyme alleles and loci contained in the inversions can explain the maintenance of consistent associations over large geographical areas, which have in some cases probably persisted for a very long time. As Lewontin (1974, p. 314) has emphasized, it is very difficult to account for such a pattern on the hypothesis of genetic drift in a finite population, unless the rate of exchange of alleles between gene arrangements is of the order of the reciprocal of the total species effective population size.

A purely neutral alternative to the drift hypothesis is that a new inversion arises in a gamete with a unique allelic content. Since crossing over is virtually absent

* This work was done while K. Ishii was on leave from Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Japan. Reprint requests should be sent to this address.

in heterokaryotypes, the inversion remains associated with these alleles even when it has reached equilibrium. The associations would be expected to decay gradually as alleles are exchanged between gene arrangements in heterokaryotypes, but this process might take a very long time to complete. (This is an example of a hitch-hiking effect (Maynard Smith & Haigh, 1974).) This explanation of associations between allozymes and inversions has independently been proposed by Mukai, Watanabe & Yamaguchi (1974) and Sheppard (1975, p. 150). Nei & Li (1975) have made a mathematical model of it which they believe fits the facts on *D. pseudo-obscura*. This model does not, however, take into account the exchange of alleles which must occur, at a low rate, between mutually inverted gene arrangements in heterokaryotypes (Sturtevant & Beadle, 1936).

The purpose of this paper is to explore this hitch-hiking hypothesis, taking into account the exchange of alleles between different gene arrangements. In particular, we attempt to evaluate the possibility that the allozyme loci (assumed to be neutral) are surrounded by loci under epistatic selection, so that crossovers in heterokaryotypes are likely to carry an unfavourable gene combination and are therefore usually quickly lost from the population. This possibility was suggested to us by Dr C. Strobeck; its effect would be to retard the effects of crossing over. We also review the evidence from Drosophila species on the existence of associations between allozymes and inversions, and attempt to assess the plausibility of hitch-hiking versus selectionist interpretations.

2. THE MODEL

Three loci are considered. The first has alleles A and a and is selectively neutral. The other two, with alleles B, b and C, c repectively, are under selection. The selection pressures are such that an inversion, introduced into a BC gamete in

Table 1. *The symmetric fitness model*

	AA	Aa	aa
BB	$1-\delta$	$1-\beta$	$1-\alpha$
Bb	$1-\gamma$	1	$1-\gamma$
bb	$1-\alpha$	$1-\beta$	$1-\delta$

a population segregating for the selected loci, would reach an equilibrium where the inverted sequence (*In*) gametes would all be BC and the standard sequence (*St*) gametes would be bc, in the absence of double crossing over in heterokaryotypes. This type of selection is exemplified by the symmetric fitness model of Table 1 when $\gamma+\beta > \alpha, \delta$ (Charlesworth, 1974). If the original population were also polymorphic at the neutral locus, and the inversion occurred in an a gamete, the *In* gametes would all be a in the absence of double crossing over in heterokaryotypes, whereas the *St* gametes would retain the polymorphism at this locus. As a result of double crossing over in heterokaryotypes, alleles will be exchanged between *In* and *St* gametes. It is known, however, that rates of double crossing over are usually

extremely low for naturally occurring inversions (Payne, 1924; Sturtevant, 1931; Spurway & Philip, 1952; Levine, 1956), so that with sufficiently strong selection the *In* and *St* gametes will have much the same composition as with no double crossing over, with respect to the selected loci (Charlesworth, 1974). The *In* and *St* gametes will gradually converge as far as the neutral locus is concerned, and the purpose of this paper is to examine the rate of this convergence, starting from the type of equilibrium described above, and following the perturbing effects of double crossing over in heterokaryotypes. With sufficiently strong selection, this should approximate closely the real situation.

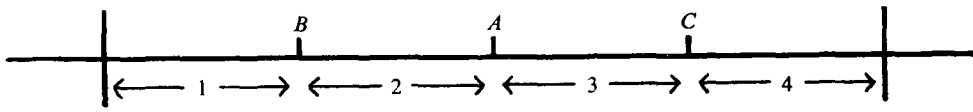

Fig. 1. This shows the intervals defined by the loci with respect to each other and the inversion breakpoints, which are represented by the pair of heavy vertical lines. *B* and *C* are selected loci, such that *St* gametes are nearly all *bc* and *In* gametes are *BC*. *A* is a neutral locus, such that *St* is initially polymorphic for alleles *A* and *a*, and *In* is initially monomorphic for *a*.

We assume an infinite, random-mating population. The neutral locus *A* is assumed to be located between loci *B* and *C*, and all the loci fall within the inversion, as shown in Fig. 1. The effects of crossing over can be represented as follows. In homokaryotypes we let r_μ be the frequency of a crossover in region μ. For mathematical convenience we assume here that rate of double crossing over is of similar small order in both homo- and heterokaryotypes. We therefore write $\epsilon R_{\mu\nu}$ and $\epsilon R^*_{\mu\nu}$ for the frequencies of simultaneous crossovers in regions μ and ν in homo- and heterokaryotypes respectively, where ϵ^2 is negligible compared with ϵ. In general we expect $R^*_{\mu\nu} < R_{\mu\nu}$. Our assumption about the order of magnitude of double crossover frequencies may be inaccurate for species such as *D. subobscura*, where crossover rates are rather high (Spurway & Philip, 1952), but should be a good approximation for species such as *D. melanogaster* where interference is nearly complete for regions less than 15 map units long. It probably does not affect our conclusions greatly, since the homogeneity of the gene arrangements with respect to the *B* and *C* loci means that double crossovers within homokaryotypes involve only pairs of gamete types, one of which is kept consistently rare by selection, and so they have a relatively small impact on the decay of the association between *a* and *In*.

The effects of selection at the *B* and *C* loci can be represented as follows. Gametes containing *BC*, *Bc*, *bC* and *bc* are labelled 1–4; w_{ij} is the fitness of an individual carrying gametes i and j. The equilibrium frequencies of *In* and *St* are written as p and q respectively. Given sufficiently strong selection of the type mentioned above, *In* and *St* will be virtually all *BC* and *bc* respectively. This requires $w_{11}, w_{44} < w_{14}$ and $w_2, w_3 < w_1, w_4$, where $w_i = pw_{i1} + qw_{i4}$ (Charlesworth, 1974). We have

$$p = (w_{14} - w_{44})/(2w_{14} - w_{11} - w_{44}).$$

It is convenient in what follows to work with normalized fitness $\tilde{w}_{ij} = w_{ij}/\overline{w}$, where $\overline{w}$ is the mean fitness of the population ($\overline{w} \simeq w_1 = w_4$).

This model is developed in the next section to provide a formula (equation [10]) for the rate of convergence of the frequency of A in the *In* and *St* gametes. Only an outline of the mathematical analysis will be given, owing to the large amount of tedious algebra involved in the calculations.

3. MATHEMATICAL DEVELOPMENT AND RESULTS

We will write $x(n)$ and $X(n)$ for the frequencies of A and a among *St* gametes in an arbitrary generation, and $y(n)$ and $Y(n)$ for the corresponding frequencies among *In* gametes. The initial values for generation 0 will be written simply as x, X, etc. (x takes some value between 0 and 1, determined by the frequency of A before the introduction of the inversion, whereas $y = 0$ because of the hitch-hiking effect of the inversion). Our purpose is to derive an expression for $x(n)$ and $y(n)$, in terms of x and y, and the recombination and selection parameters.

There is a total of 16 different gamete types in this system, 8 being *St* and 8 *In* in karyotype. (The genotypes of *In* gametes will be distinguished from those of *St* gametes by an asterisk.) It is possible to write down recurrence relations for all 16 gamete types. These can be simplified as follows. Let $\mathbf{u}(n)$ be the 16-dimensional row vector of deviations of gamete frequencies in generation n from their values in generation 0. The components of $\mathbf{u}(n)$ describe, in order, the following gamete types *Abc*, *abc*, *ABC**, *aBC**, *ABC*, *aBC*, *ABc*, *aBc*, *AbC*, *abC*, *ABc**, *aBc* AbC**, *Abc**, *abc**. (The first 4 components of $\mathbf{u}$ have been separated from the rest since they are the only ones which selection does not tend to eliminate.) Ignoring second-order terms in the components of $\mathbf{u}$, we obtain the following linearized expression

$$\mathbf{u}(n+1) = \mathbf{u}(n)\,\mathbf{M} + \epsilon\mathbf{a}, \tag{1}$$

where $\mathbf{M}$ is a 16×16 matrix, ϵ is the order of the total frequency of double crossovers, and $\mathbf{a}$ is a 16-dimensional row vector. $\mathbf{M}$ describes the relations between the perturbations in gamete frequencies in successive generations while $\mathbf{a}$ describes the generation of new gametes by double crossing over involving the gametes present in the initial generation, using their unperturbed frequencies. The components of $\mathbf{a}$ are given in the Appendix.

This equation can be solved using the initial condition $\mathbf{u}(0) = 0$,

$$\mathbf{u}(n) = \epsilon\mathbf{a}(\mathbf{I} - \mathbf{M})^{-1}(\mathbf{I} - \mathbf{M}^n), \tag{2}$$

where $\mathbf{I}$ is the 16×16 unit matrix.

Now $\mathbf{M}^n$ can be written as the spectral expansion

$$\mathbf{M}^n = \sum_{i=1}^{16} \mathbf{V}_i \lambda_i^n \mathbf{U}_i, \tag{3}$$

where λ_i, $\mathbf{U}_i$ and $\mathbf{V}_i$ are respectively the i eigenvalue, row eigenvector and column eigenvector of $\mathbf{M}$. We assume that the $\mathbf{U}_i$ and $\mathbf{V}_i$ constitute an orthonormal system of eigenvectors.

Substituting from (3) into (2), we obtain

$$u(n) = \epsilon \sum_{i=1}^{16} \frac{\mathbf{a}\mathbf{V}_i}{(1-\lambda_i)}(1-\lambda_i^n)\,\mathbf{U}_i. \tag{4}$$

When the amount of double crossing over is small, we can write

$$\mathbf{M} = \mathbf{M}_0 + \epsilon \mathbf{M}', \tag{5}$$

where $\mathbf{M}_0$ is the matrix for the case of zero double crossing over, and $\mathbf{M}'$ is the derivative of $\mathbf{M}$ with respect to ϵ at $\epsilon = 0$.

We can similarly write

$$\begin{aligned} \lambda_i &= \lambda_{i0} + \epsilon\lambda_{i1}, \\ \mathbf{U}_i &= \mathbf{U}_{i0} + \epsilon\mathbf{U}_{i1}, \\ \mathbf{V}_i &= \mathbf{V}_{i0} + \epsilon\mathbf{V}_{i1}. \end{aligned} \tag{6}$$

The component of $\mathbf{M}_0$, and its eigenvectors corresponding to the repeated leading eigenvalue of one, are given in the Appendix. It is unnecessary to give the components and eigenvectors of $\mathbf{M}'$ here, since they are not used in the detailed calculation. Substituting from (6) into (4) and using the fact that $\mathbf{M}_0$ has a repeated leading eigenvalue of 1, we obtain

$$\mathbf{u}(n) = -\sum_{i=1}^{2} \frac{\mathbf{a}\mathbf{V}_{i0}}{\lambda_{i1}}[1-(1+\epsilon\lambda_{i1})^n]\,\mathbf{U}_{i0} + \epsilon \sum_{i=3}^{16} \frac{\mathbf{a}\mathbf{V}_{i0}}{(1-\lambda_{i0})}(1-\lambda_{i0}^n)\,\mathbf{U}_{i0} + o(\epsilon). \tag{7}$$

Since $(1+\epsilon\lambda_{i1})^n = 1 + n\epsilon\lambda_{i1} + O(n^2\epsilon^2)$, and neglecting all but the terms involving the leading eigenvalues of $\mathbf{u}(n)$, we can approximate equation (7) further as

$$\mathbf{u}(n) = n\epsilon \sum_{i=1}^{2} \mathbf{a}\mathbf{V}_{i0}\,\mathbf{U}_{i0}, \tag{8}$$

where terms $O(\epsilon)$ and $O(n^2\epsilon^2)$ have been neglected. For this to be a good approximation, we need to restrict n to values of order $\epsilon^{-\frac{1}{2}}$. Since $\mathbf{u}$ must remain small to justify the linear approximation of equation (1), equation (8) can only be used meaningfully when $n\epsilon = O(\sqrt{\epsilon}) \leqslant O(0{\cdot}01)$, i.e. $\epsilon \leqslant O(10^{-4})$. This is reasonable for the rates of double crossing over in which we are interested.

For n of order $\epsilon^{-\frac{1}{2}}$, therefore, we can use equation (8) to write down expressions for $x(n)$, $y(n)$, etc. by using the eigenvectors of $\mathbf{M}_0$ given in the Appendix and summing over the appropriate components of $\mathbf{u}(n)$. It turns out that the frequencies of *St* and *In* gametes remain, as expected, constant to order ϵ. This gives the result of main interest

$$x(n) - y(n) = (x-y)\,(1-nm), \tag{9a}$$

where

$$\begin{aligned} m = \epsilon\tilde{w}_{14}\Bigg\{ & R^*_{23} + r_2pq\left(R^*_{14}\frac{r_3}{r_2+r_3} + R^*_{13}\right)\left(\frac{\tilde{w}_{24}}{1-\tilde{w}_2+r_2p\tilde{w}_{24}} + \frac{\tilde{w}_{13}}{1-\tilde{w}_3+r_2q\tilde{w}_{13}}\right) \\ & + r_3pq\left(R^*_{14}\frac{r_2}{r_2+r_3} + R^*_{24}\right)\left(\frac{\tilde{w}_{12}}{1-\tilde{w}_2+r_3q\tilde{w}_{12}} + \frac{\tilde{w}_{34}}{1-\tilde{w}_3+r_3p\tilde{w}_{34}}\right)\Bigg\}. \end{aligned} \tag{9b}$$

https://doi.org/10.1017/S0016672300017511 Published online by Cambridge University Press

It is shown in the Appendix that this approach can be extended to generations such that $n \gg \epsilon^{-\frac{1}{2}}$. The final formula is

$$x(n) - y(n) = (x - y)\,\mathrm{e}^{-nm}. \tag{10}$$

When there is no selection formula (9*b*) reduces to

$$m = (R^*_{13} + R^*_{14} + R^*_{23} + R^*_{24})\epsilon \tag{11}$$

i.e. m is equal to the total probability of an exchange involving the A locus. This is identical with the rate of decay obtained directly by treating both the inversion and A as neutral units.

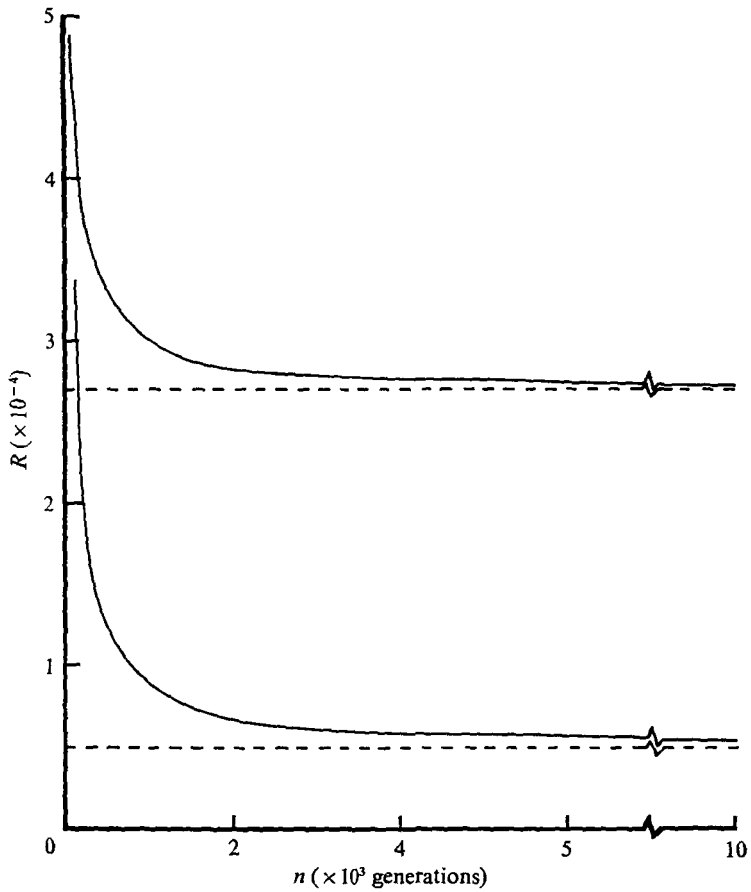

Fig. 2. The solid lines show the true values of $R(n)$ as functions of n; the broken lines are the corresponding values predicted by equation (10). In both cases $\alpha = 0{\cdot}4$, $\beta = \gamma = 0{\cdot}5$ and $\delta = 0{\cdot}8$. The initial frequency of A was 0·5 in the *St* gametes and 0 in the *In* gametes. In the top case the loci were equally spaced, such that the probabilities of the single crossover events in each region were 0·1 (i.e. $r_\mu = 0{\cdot}1$, $\mu = 1$–4). In the bottom case, A was closely linked to B, so that $r_2 = 0{\cdot}01$ and $r_3 = 0{\cdot}19$; the other single crossover parameters were the same as in the top case. In both cases the probabilities of a crossover in both regions μ and ν in homokaryotypes were $0{\cdot}1 \times r_\mu \times r_\nu$; the corresponding probabilities in heterokaryotypes were $0{\cdot}01 \times r_\mu \times r_\nu$.

https://doi.org/10.1017/S0016672300017511 Published online by Cambridge University Press

The accuracy of the approximations involved in deriving equations (9) and (10) was checked by comparing them with population trajectories computed using the full set of recurrence relations for the system. We found that when selection was reasonably intense, the approximate formulae gave satisfactory results for rates

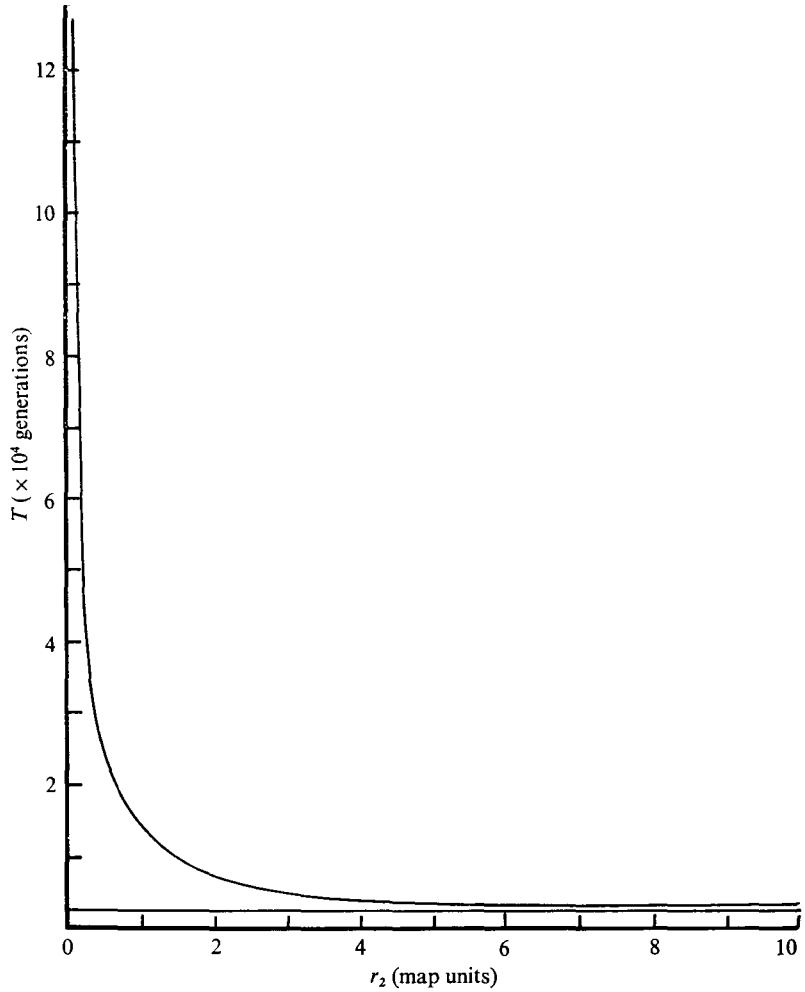

Fig. 3. This shows the value of T, the half-life of the decay process, as a function of r_2, the distance between A and B. The upper line is the case with selection as in Fig. 2; the lower line is for the case of no selection. The other recombination parameters are as in Fig. 2.

of crossing over up to 10^{-3} or so. Fig. 2 shows two examples involving the symmetric fitness model. The function

$$R(n) = (-1/n)\log\{[x(n)-y(n)]/[x-y]\}$$

is plotted against n, the number of generations. Equation (10) predicts that $R(n)$ should follow straight lines parallel to the abscissa (broken lines). The actual curves of $R(n)$ values approach these lines quite closely within 500 generations or so. This

https://doi.org/10.1017/S0016672300017511 Published online by Cambridge University Press

is a short time relative to the half-life of the convergence process, which is of the order of 10^4 generations.

Fig. 3 shows the half-life of the decay process, as given by equations (9*b*) and (10), as a function of the distance of A from the selected locus B. Despite the intense selection in this case, it can be seen that there is a significant retarding effect of selection at the B locus only when A is rather close to B (less than 1 map unit away). This conclusion appears to be rather general. It should be noted that the numerical values in Figs. 2 and 3 are for the case of an inversion with equilibrium frequency 0·5. When the karyotype frequencies are more extreme than this, the overall rate of decay will be lessened since pq will be smaller giving less weight to the terms in m other than that involving R^*_{23}. This effect will be marked only when R^*_{23} is small compared with the other double crossing over frequencies, i.e. when A is closely linked to one of the selected loci.

4. DISCUSSION

The results presented above indicate that, unless a neutral locus is very closely linked to one of the selected loci which are involved in the maintenance of the inversion polymorphism, the half-life of the decay of an association between the neutral locus and the inversion is of the order of the reciprocal of the rate of double crossing over in heterokaryotypes. It therefore seems likely that experimental estimates of the rate of exchange of alleles at an allozyme locus between two mutually inverted gene arrangements will provide a good estimate of the rate of decay on the neutral hypothesis. Unfortunately no reliable direct estimates of this sort have been obtained to date. We therefore have to rely on indirect estimates based on measurements of the rates of crossing over between morphological mutants and inversions. Since in only a few cases, to be discussed below, is it clear that these genes are located close to allozyme loci of interest, our knowledge of crossing over rates is correspondingly limited.

Measurements of crossing over between mutants and inversions have been carried out in a number of Drosophila species. For example, Payne (1924) published extensive data on the naturally occurring inversions *In(3L)P* and *In(3R)P* of *D melanogaster*. He found that the loci *h* and *e*, located respectively near the middle of the *3L* and *3R* inversions, showed a rate of exchange of the order of 2×10^{-4}. (Each inversion covers about 30 map units.) Spurway & Philip (1952) studied the inversion A_2 of *D. subobscura*, and found only 1 double crossover involving a pair of loci near the middle of the inversion, out of some 20000 individuals. This inversion covers 80 units of a chromosome about 250 units long. Levine (1956) found somewhat higher rates (10^{-3}–10^{-4}) for a number of inversions in *D. pseudoobscura*. In his study there was no marked difference between the rate of exchange of alleles between gene arrangements differing in simple and complex, overlapping inversions. This is interesting in view of the work of Chovnick (1973), who found that the overwhelming majority of so-called double crossovers recovered from a simple inversion heterozygote in *D. melanogaster* were in fact the products of gene conversion events, and did not involve the exchange of closely linked markers. He

estimated the rate of conversion as being of order 10^{-5}, however, rather than the 10^{-4} rate of exchange found in the experiments quoted above. The gene conversion interpretation suggests that the rate of exchange should be more or less independent of the position of the gene within the inversion, except for loci sufficiently close to the break points that pairing is badly disturbed. Complex inversions should show similar rates to simple ones. These points need further experimental investigation. Furthermore, since conversion events in Drosophila probably only involve a few hundred bases at a time (Chovnick, Ballantyne & Holm, 1971), the exchange of alleles at a neutral locus would not be impeded by selection, apart from any selection operating on adjacent cistrons.

From these experiments, a rate of 10^{-4} per generation may be assigned as typical for a locus near the middle of a simple inversion, giving 10^4 generations for the order of the time-scale of the decay of an association between an allozyme locus and an inversion which contains it. How does this relate to observations on natural populations? Unfortunately the data cannot be unambiguously interpreted in most cases. Many of the observations reported in the literature concern gene arrangements which differ by more than one inversion, where one might often expect a very low rate of exchange of alleles at a locus within the breakpoints. A hitch-hiking explanation for associations between allozymes and inversions cannot be ruled out in the absence of knowledge concerning crossing over rates. This applies to the original work of Prakash & Lewontin (1968) on the *Pt-10* and *α-Amylase* loci of *D. pseudoobscura* and *D. persimilis*, and to the associations between four loci and the sex-ratio inversion complex of *D. pseudoobscura* described by Prakash & Merritt (1972) and Prakash (1974). It is known that *α-Amylase* is included within the inversion breakpoints separating the two phylads which are genetically distinct with respect to this locus (Yardley, 1974); there is no other information available at present concerning crossing over values in these systems. In *D. subobscura* there is an association between alleles at the *Lap* locus and the inversion complex O_{3+4} (Charlesworth, Charlesworth & Loukas, 1976; Prevosti, personal communication). Nothing is known for certain about the location of *Lap* with respect to the inversion breakpoints, but no crossovers out of 800 backcross individuals have been recovered from double heterozygotes. The results of Prakash & Lewontin (1971) on the *Pt-12* locus of *D. pseudoobscura* cannot be so readily interpreted on a hitch-hiking model, since the association with inversions is such that one allele (*1.20*) is found only in the *St* arrangement and is virtually absent from the others. It is possible, however, that *1.20* has arisen relatively recently within *St* and has not yet had time to diffuse into the other arrangements.

A number of examples of strong associations between allozymes and simple inversions have been reported in *D. melanogaster*. In two of these cases, involving α-glycerophosphate dehydrogenase (*α-Gpd*) and esterase-6 (*Est*-6), the enzyme locus is known to be near the middle of the inversion concerned, and would be expected to cross over at a rate of about 10^{-4}. *α-Gpd* in fact showed no association with the inversion concerned (*In(2L)t*) in two populations (the Katsunuma, Japan population studied by Kojima, Gillespie and Tobari (1970) and by Langley, Tobari

& Kijoma (1974), and the N. Carolina population studied by Mukai, Chigusa & Mettler (1971) and Mukai *et al.* (1974)). In the Brownsville, Texas population studied by Langley *et al.*, there was a significant association. The *Est*-6 locus shows consistent association with *In(3L)P* in the Katsunuma and Brownsville populations. At first sight, these associations suggest natural selection. It should be remembered, however, that Japanese and American populations of *D. melanogaster* have probably only been in existence for 100 years or so. An association between the inversions and these allozyme loci could have been set up as a result of a restricted number of founder individuals, and could still be largely intact today. Since there are probably about 10 generations or so a year, about 10^3 generations would have elapsed since this event, leaving plenty of scope for a significant but not absolute association. It would be interesting to have data from an African population, since West Africa is the supposed place of origin of *D. melanogaster*. With the other examples of associations between inversions and included loci in *D. melanogaster*, the rates of crossing over are probably even lower, since the loci concerned (amylase and phosphoglucomutase) are near to the inversion breakpoints.

Data on allozyme loci which are known to be located outside the inversions with which they show associations are somewhat more encouraging for a selectionist view. Prakash & Levitan (1973), working with *D. robusta*, showed that alleles at the *Est*-1 locus showed geographically consistent associations with the inversions *2L1* and *2L3*. They demonstrated about 10% crossing over between the enzyme locus and either inversion. Unfortunately, the females used in their mapping experiments were heterokaryotypic on 4 other chromosome arms, and this must greatly have increased the amount of crossing over. Nevertheless, there must be an appreciable rate of crossing over in natural populations. It is therefore difficult to explain this case without appealing to selection. Langley *et al.* (1974) report associations between alleles at the aldehyde oxidase (*Aldox*) and octanol dehydrogenase (*Odh*) loci of *D. melanogaster*, and *In(3R)P* in the Katsunuma and Brownsville populations. The association involving *Odh* is found only in the Katsunuma population, but is consistent over two years, whereas the *Aldox* association is found in both populations. Neither of these associations are absolute and they contrast with the other associations in *melanogaster* in that the commonest allele in the standard sequence does not overwhelmingly predominate in the inverted sequence. Furthermore, *Odh* is close to the locus of *p*, which Payne (1924) showed to cross over at rate of about $2{\cdot}5 \times 10^{-3}$ with the inversion. *Aldox* is 1 unit to the left of *ss*, which crosses over at a rate of about $3{\cdot}5 \times 10^{-4}$ with the inversion (Payne, 1924). It seems unlikely that the *Odh* association could be explained by a hitch-hiking or founder effect model, although this is possible for the *Aldox* case. Some more data on associations between these enzymes and the inversions and more extensive data on crossing over seem desirable. There is also evidence for strong associations between *Adh* alleles and 2L inversions in many populations of *D. melanogaster* (e.g. Langley *et al.* 1974; Watanabe & Watanabe, 1977). *Adh* is known to lie between the inversion breakpoints and the centromere in these cases (Watanabe & Watanabe, 1977), and so probably shows a very low rate of crossing over with the inversions.

APPENDIX

(*a*) *Components of* $\mathbf{M}_0$ *and* $\mathbf{a}$.

$$\mathbf{M}_0 = \begin{pmatrix} \mathbf{A} & \mathbf{O} & \mathbf{O} \\ \mathbf{B} & \mathbf{C} & \mathbf{O} \\ \mathbf{D} & \mathbf{O} & \mathbf{E} \end{pmatrix}, \tag{A 1}$$

where the component non-zero matrices are given as follows:

$$\mathbf{A} = \begin{pmatrix} 1+px(\tilde{w}_{44}-2) & pX(\tilde{w}_{44}-2) & qy(\tilde{w}_{14}-2) & qY(\tilde{w}_{14}-2) \\ px(\tilde{w}_{44}-2) & 1+pX(\tilde{w}_{44}-2) & . & . \\ px(\tilde{w}_{14}-2) & pX(\tilde{w}_{14}-2) & 1+qy(\tilde{w}_{11}-2) & qY(\tilde{w}_{11}-2) \\ . & . & qy(\tilde{w}_{11}-2) & 1+qY(\tilde{w}_{11}-2) \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} px[(1-r_2-r_3)\tilde{w}_{14}-2] & pX[(1-r_2-r_3)\tilde{w}_{14}-2] & qy(\tilde{w}_{11}-2) & qY(\tilde{w}_{11}-2) \\ . & . & . & . \\ p[(x+r_2X)\tilde{w}_{24}-2x\tilde{w}_2] & pX[(1-r_2)\tilde{w}_{24}-2\tilde{w}_2] & qy(\tilde{w}_{12}-2\tilde{w}_2) & qY(\tilde{w}_{12}-2\tilde{w}_2) \\ px[(1-r_2)\tilde{w}_{24}-2\tilde{w}_2] & p[(X+r_2x)\tilde{w}_{24}-2X\tilde{w}_2] & . & . \\ p[(x+r_3X)\tilde{w}_{34}-2x\tilde{w}_3] & pX[(1-r_3)\tilde{w}_{34}-2\tilde{w}_3] & qy(\tilde{w}_{13}-2\tilde{w}_3) & qY(\tilde{w}_{13}-2\tilde{w}_3) \\ px[(1-r_3)\tilde{w}_{34}-2\tilde{w}_3] & p[(X+r_3x)\tilde{w}_{34}-2X\tilde{w}_3] & . & . \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1-p(r_2+r_3)\tilde{w}_{14} & 0 & p(r_2x+r_3)\tilde{w}_{14} & pr_2X\tilde{w}_{14} & p(r_3x+r_2)\tilde{w}_{14} & pr_3X\tilde{w}_{14} \\ 0 & 1-p(r_2+r_3)\tilde{w}_{14} & pr_2x\tilde{w}_{14} & p(r_2X+r_3)\tilde{w}_{14} & pr_3x\tilde{w}_{14} & p(r_3X+r_2)\tilde{w}_{14} \\ 0 & 0 & \tilde{w}_2-r_2pX\tilde{w}_{24} & r_2pX\tilde{w}_{24} & 0 & 0 \\ 0 & 0 & r_2px\tilde{w}_{24} & \tilde{w}_2-r_2px\tilde{w}_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{w}_3-r_3pX\tilde{w}_{34} & r_3pX\tilde{w}_{34} \\ 0 & 0 & 0 & 0 & r_3px\tilde{w}_{34} & \tilde{w}_3-r_3px\tilde{w}_{34} \end{pmatrix}.$$

The dots indicate that the relevant element is identical to the one immediately above.

Matrix **D** is obtained from **B** by interchanging terms as follows: column 1 with 3, 2 with 4; row 1 with 5, 2 with 6; p, x and X with q, y and Y respectively; subscript 1 with 4, 2 with 3, in $\tilde{w}_{ij}$ and $\tilde{w}_i$.

Matrix **E** is obtained from **C** in a similar way, except that columns 1 and 5, 2 and 6 are interchanged.

https://doi.org/10.1017/S0016672300017511 Published online by Cambridge University Press

(*b*) *Eigenvalues and eigenvectors of* $\mathbf{M}_0$.

The eigenvalues of $\mathbf{M}_0$ are the total of those of **A**, **C** and **E**. Since **C** and **E** describe the frequency changes of gametes which are ultimately eliminated, in the absence of double crossovers, their largest eigenvalues must be less than one in modulus. It is easy to show that the two largest eigenvalues of **A** are equal to one; these represent the largest eigenvalues of $\mathbf{M}_0$. It can be shown by straightforward calculation that the corresponding row eigenvectors of $\mathbf{M}_0$ are

$$\mathbf{u}_{10} = (1, -1, 0, 0, \mathbf{0}, \mathbf{0}); \quad \mathbf{u}_{20} = (0, 0, -1, -1, \mathbf{0}, \mathbf{0}), \tag{A 2}$$

where **0** denotes a 6-dimensional zero row vector.

Similarly the corresponding column eigenvectors are given by

$$\mathbf{V}'_{10} = (X, -x, 0, 0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \mathbf{0}), \tag{A 3}$$

$$\mathbf{V}'_{20} = (0, 0, Y; -y, \mathbf{0}, \theta_5^*, \theta_6^*, \theta_3^*, \theta_4^*, \theta_1^*, \theta_2^*),$$

where

$$\theta_1 = \frac{pXr_2r_3}{r_2+r_3}\left(\frac{\tilde{w}_{24}}{1-\tilde{w}_2+r_2p\tilde{w}_{24}} + \frac{\tilde{w}_{34}}{1-\tilde{w}_3+r_3p\tilde{w}_{24}}\right),$$

$$\theta_3 = \frac{pXr_2\tilde{w}_{24}}{1-\tilde{w}_2+r_3p\tilde{w}_{24}}, \quad \theta_5 = \frac{pXr_3\tilde{w}_{34}}{1-\tilde{w}_3+r_3p\tilde{w}_{34}}, \quad -\theta_{2i} = \frac{x\theta_{2i-1}}{X},$$

and θ_j^* is derived from θ_j by substituting q, y and Y for p, x and X, and interchanging 1 with 4 and 2 with 3 in the subscripts for $\tilde{w}_{ij}$ and $\tilde{w}_i$.

These eigenvectors constitute an orthonormal system.

(*c*) *Components of row vector* **a**.

We can write

$$\mathbf{a} = pq\tilde{w}_{14} \sum_{\mu=1}^{3} \sum_{\nu=\mu+1}^{4} R_{\mu\nu}^* \mathbf{a}_{\mu\nu}^*, \tag{A 4}$$

where the $\mathbf{a}_{\mu\nu}^*$ are row vectors defined as follows

$$a_{12}^* = (-x, -X, -y, -Y, 0, 0, x, X, 0, 0, 0, 0, y, Y, 0, 0)$$
$$a_{13}^* = (-x, -X, -y, -Y, 0, 0, y, Y, 0, 0, 0, 0, x, X, 0, 0)$$
$$a_{14}^* = (-x, -X, -y, -Y, y, Y, 0, 0, 0, 0, 0, 0, 0, 0, x, X)$$
$$a_{23}^* = (y-x, x-y, x-y, y-x, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$
$$a_{24}^* = (-x, -X, -y, -Y, 0, 0, 0, 0, y, Y, x, X, 0, 0, 0, 0)$$
$$a_{34}^* = (-x, -X, -y, -Y, 0, 0, 0, 0, x, X, y, Y, 0, 0, 0, 0)$$

(*d*) *Extension of equations* (9)

We here show how equations (9) can be generalized to cover generations such that n is of higher order than $\epsilon^{-\frac{1}{2}}$. Consider a generation $N = O(\epsilon^{-\frac{1}{2}})$. From the analysis sketched in the text we find that the frequencies of all gametes except

https://doi.org/10.1017/S0016672300017511 Published online by Cambridge University Press

bc and BC^* are $O(\epsilon)$. We can represent the gamete frequencies in a generation $N+n$ as a vector of deviations from those given by equation (8) for generation N, $\mathbf{u}^*(n)$. Hence we can write

$$\mathbf{u}^*(n+1) = u^*(n)\,\mathbf{M}^* + \epsilon\mathbf{a}^*, \tag{A 5}$$

where the initial condition is $\mathbf{u}^*(0) = O(\epsilon)$, and $\mathbf{M}^*$ and $\mathbf{a}^*$ are obtained in the same way as $\mathbf{M}$ and $\mathbf{a}$, replacing $x(0)$ and $y(0)$ with $x(N)$ and $y(N)$ respectively.

Equation (A 5) has the solution

$$\mathbf{u}^*(n) = \epsilon\mathbf{a}^*(\mathbf{I}-\mathbf{M}^*)^{-1}(\mathbf{I}-\mathbf{M}^{*n}) + \mathbf{u}^*(0)\,\mathbf{M}^{*n}. \tag{A 6}$$

If $\mathbf{M}^*$ is analysed by methods similar to those used for equations (5) and (6), we find that it has leading eigenvalues which are slightly less than one in modulus, provided that ϵ is sufficiently small. Hence $\mathbf{u}^*(0)\,\mathbf{M}^{*n}$ is at most $O(\epsilon)$. Extending this approach to the sequence of generations $n = kN$ $(k = 1, 2, \ldots)$ we obtain the analogue of equations (9)

$$x(kN)-y(kN) = \{x([k-1]N)-y([k-1]N)\}\{1-Nm\}+O(k^{\frac{1}{2}}\epsilon). \tag{A 7}$$

The estimate of the error term is based on the fact that it is a sum of k terms of order ϵ but with varying signs. A sum of a sequence of n random variables which take the values ± 1 with equal frequency has standard deviation $n^{\frac{1}{2}}$, so that the order of magnitude of the error term will be $k^{\frac{1}{2}}\epsilon$.

Equation (A 7) can be iterated k times to give $x(kN)-y(kN)$ in terms of $x(0)-y(0)$. Applying the above argument about the magnitude of the error term to the sum of k terms of order $k^{\frac{1}{2}}\epsilon$, we obtain

$$x(kN)-y(kN) = (x-y)\,(1-Nm)^k + O(k\epsilon)$$

or

$$x(n)-y(n) \approx (x-y)\,e^{-nm} + O(nN^{-1}\epsilon). \tag{A 8}$$

This formula is meaningful for n such that $\mathrm{e}^{-nm} \gg O(nN^{-1}\epsilon) = O(nm\epsilon^{\frac{1}{2}})$. For $\epsilon = 10^{-4}$ this gives $e^{-nm} > nm10^{-2}$. With $nm = 2{\cdot}3$ we have $e^{-2{\cdot}3} = 0{\cdot}10 > 2{\cdot}3 \times 10^{-2}$. so that equation (A 8) can be used until $x(n)-y(n)$ is 10 % of its original value. It therefore provides a satisfactory solution to our problem.

We thank Professor Alan Robertson for suggesting several improvements to the original draft of this paper. K.I. would like to express his sincere gratitude to Professor John Maynard Smith for his hospitality during his stay in Sussex.

REFERENCES

CHARLESWORTH, B. (1974). Inversion polymorphism in a two-locus genetic system. *Genetical Research* **23**, 259–280.

CHARLESWORTH, B., CHARLESWORTH, D. & LOUKAS, M. (1976). Linkage disequilibrium studies on *D. subobscura* populations. *Drosophila Information Service* **52** (In the Press).

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

CHOVNICK, A., BALLANTYNE, G. H. & HOLM, D. G. (1971). Studies on gene conversion and its relationship to linked exchange in *Drosophila melanogaster*. *Genetics* **69**, 179–209.

Kojima, K., Gillespie, J. H. & 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 and inversions in coadapted systems. *American Naturalist* **90**, 41–45.

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

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

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**, 1056–1069.

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.

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

Payne, F. (1924). Crossover modifiers in the third chromosome of *Drosophila melanogaster*. *Genetics* **9**, 327–342.

Prakash, S. (1974). Gene differences between the sex ratio and standard gene arrangements of the X chromosome and linkage disequilibrium between loci in the standard gene arrangement of *Drosophila pseudoobscura*. *Genetics* **77**, 795–804.

Prakash, S. & Levitan, M. (1973). Associations of alleles of the Esterase-1 locus with gene arrangements of the left arm of the second chromosome in *Drosophila robusta*. *Genetics* **75**, 371–379.

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

Prakash, S. & Merritt, R. B. (1972). Direct evidence of genic differentiation between sex ratio and standard gene arrangements of X chromosome in *Drosophila pseudoobscura*. *Genetics* **72**, 169–175.

Sheppard, P. M. (1975). *Natural Selection and Heredity* (4th edn.) London: Hutchinson.

Spurway, H. & Philip, U. (1952). Genetics and cytology of *Drosophila subobscura*. *Journal of Genetics* **51**, 198–215.

Sturtevant, A. H. (1931). Known and probably inverted sections of the autosomes of *Drosophila melanogaster*. In *Contributions to the Genetics of Certain Chromosome Anomalies in Drosophila melanogaster* (A. H. Sturtevant and T. Dobzhansky), pp. 1–27. Washington: Carnegie Institution Publ. No. 421.

Sturtevant, A. H. & Beadle, G. W. (1936). The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* **21**, 544–604.

Watanabe, T. K. & Watanabe, T. (1977). Enzyme and chromosome polymorphisms in Japanese natural populations of *Drosophila melanogaster*. *Genetics* **85**, 319–329.

Yardley, D. (1974). Linkage analysis of α-amylase in *D. pseudoobscura*. *Drosophila Information Service* **51**, 25.