

Selection of new inversions in multi-locus genetic systems

By BRIAN CHARLESWORTH AND DEBORAH CHARLESWORTH

Department of Genetics, University of Liverpool, Liverpool, L69 3BX

(Received 23 October 1972)

SUMMARY

An approximate expression is derived for the rate of change in frequency of an inversion introduced at a low initial frequency into a multi-locus system at equilibrium under recombination and selection. It is shown that this expression gives accurate predictions of the rate of progress of the inversion, even if the initial population is perturbed somewhat from equilibrium. Extensions to the cases where there are sex differences in recombination and selection are considered. An implication of the results is that selection pressure for newly arisen inversions depends on the existence of a stable equilibrium with linkage disequilibrium. The expected chance of survival of a new inversion in a large population is shown to be approximately one half the square root of the loss in fitness due to recombination.

1. INTRODUCTION

Cytological and genetic observations of many species of plants and animals have revealed polymorphisms for chromosome rearrangements which reduce crossing-over when heterozygous. The best-studied case of such polymorphism is that of the paracentric inversions of the third chromosome of *Drosophila pseudoobscura*, which has been exhaustively investigated by Dobzhansky and his associates. Several theories have been proposed to account for the maintenance of these polymorphisms by natural selection. Perhaps the simplest is that the chromosome rearrangements are associated with position effects which confer an advantage on the heterozygotes. Some evidence for this view has been presented by Sperlich (1959), for the case of inversion polymorphisms in *D. subobscura*. Another theory was proposed by Sturtevant & Mather (1938), and elaborated by Kojima (1967), Nei, Kojima & Schaffer (1967) and, most recently, by Ohta (1971). It states that the superior fitness of heterozygotes for chromosome rearrangements suppressing crossing-over (for brevity, we will refer to such rearrangements as inversions in what follows) is due to the accumulation by chance of different recessive, partially recessive or over-dominant mutations in different rearrangements. Homozygotes for inversions will, on average, thus be homozygous for more genes reducing fitness when homozygous than inversion heterozygotes.

These two theories fail to account for the observations of Dobzhansky's school on the behaviour of artificial populations of *D. pseudoobscura* formed from mixtures of races of different geographic origin. Although, for example, the arrangements *AR* and *CH* come to a stable and repeatable polymorphic equilibrium when the two

sequences are derived from the same locality, there may be no stable equilibrium (and the results of replicate experiments are usually highly unpredictable) when *AR* is derived from a different locality from *CH* (Dobzhansky, 1951). These observations led Dobzhansky (1951) to propose that inversion polymorphisms are maintained because the suppression of crossing-over in inversion heterozygotes reduces the average rate at which recombination breaks up favourable complexes of genes on the same chromosome. On this theory, different populations may tend to develop different sets of interacting (coadapted) genes associated with the same inversion, and this explains the findings in the inter-population crosses.

This theory is related to the idea, put forward by Fisher (1930), that genetic factors reducing crossing-over will be favoured in populations polymorphic for two loci, whose alleles interact in such a way that the combinations *AB* and *ab* have superior fitness to the combinations *Ab* and *aB*. The lower the amount of crossing-over between the two loci, the lower the rate at which the unfavourable combinations of alleles are produced in the double heterozygote *AB/ab*. This theory has been made more precise by the mathematical investigations of, among others, Kimura (1956), Haldane (1957), Lewontin & Kojima (1960), Bodmer & Felsenstein (1967), Nei (1967), Turner (1970), Lewontin (1971), Deakin (1972) and Feldman (1972). It does not appear, however, that a general deterministic theory of the conditions for establishment of a new inversion has been developed. In this paper, we hope to contribute to the development of such a theory.

We will derive a simple expression for the rate of change in frequency of an inversion introduced at a low frequency into an infinitely large population at equilibrium under selection for an arbitrarily large number of loci. It turns out that the inversion will spread at a significant rate only if introduced into a gamete whose marginal fitness is greater than the mean fitness of the population; the rate of change of the inversion frequency is proportional to the deviation of this marginal fitness from the population mean fitness. This implies that there is effectively no selection for an inversion if the initial population is in linkage equilibrium. Gene interaction is thus, in itself, insufficient to favour inversion polymorphism, although it is a necessary condition.

2. THEORY

Let us consider a population segregating for an arbitrary number of autosomal loci, such that there are n types of gametes with frequencies x_1, x_2, \dots, x_n . Let the relative fitness of the genotype formed by the union of the i th and j th gametic types be w_{ij} (we assume that generations are distinct and that there are no fitness differences between the sexes). Provided that the parameters of recombination are the same in both sexes (this restriction will be removed later), we can write the following general equation for the change in frequency from one generation to the next of the i th gametic type in a random mating population (cf. Lewontin, 1964),

$$\Delta x_i = \frac{x_i(w_{i.} - \bar{w})}{\bar{w}} - \rho_i \quad (1)$$

where $w_{i.}$ is the marginal fitness of the i th gametic type ($w_{i.} = \sum_j x_j w_{ij}$), \bar{w} is the population mean fitness ($\bar{w} = \sum_i x_i w_{i.}$) and ρ_i is a parameter measuring the contribution of cross-over events to the change in x_i .

At equilibrium, we have

$$\frac{\hat{x}_i(\hat{w}_{i.} - \hat{w})}{\hat{w}} = \hat{\rho}_i. \tag{2}$$

If there is complete linkage equilibrium at equilibrium, then $\hat{w}_{i.} = \hat{w}$ and $\hat{\rho}_i = 0$ for all i . If there is linkage disequilibrium at equilibrium, then some of the $\hat{\rho}_i$ must be negative and some positive, although (for systems of greater complexity than two loci each with two alleles) some may be zero. We may speak of gametes for which $\hat{\rho}_i > 0$ as being in excess and those for which $\hat{\rho}_i < 0$ as in deficiency, although (except for the two-locus, two allele case) it does not follow that their frequencies are necessarily respectively greater or smaller than those expected on the basis of random combination of alleles at different loci.

We shall now consider the fate of a newly arisen inversion covering the region containing the loci in question. The inversion is assumed to be introduced into a large, equilibrium population, and its initial frequency is assumed to be small, as would be the case if it arose by a mutation. Since such a mutation is effectively a unique event and since all but very large inversions, when heterozygous, completely suppress crossing-over in and around the chromosome segment between their breakpoints, we can assume that the inversion is introduced into only one type of gamete, with which it remains associated indefinitely. We will also assume that the fitnesses of the carriers of the inversion are exactly the same as non-inversion individuals of the same genotypes.

Let the inversion be introduced into the i th gametic type. Let the frequency of the inversion among the i th gametic type be q , the overall frequency of the inversion be y , and the overall frequency of the i th type of gamete be x_i . Since the inversion, when heterozygous, completely suppresses crossing-over among the genes we are considering, we have in any generation

$$\Delta y = \frac{y(w_{i.} - \bar{w})}{\bar{w}} \tag{3}$$

where $w_{i.} = \sum_j x_j w_{ij}$, $\bar{w} = \sum_i x_i w_{i.}$, as before.

The frequencies of the other gametic types in the population, including the non-inverted i th gamete, will in general be perturbed from their equilibrium values by quantities of the order of y , owing to their encounters with the inversion in heterozygotes. We can therefore write:

$$\begin{aligned} w_{i.} &= \hat{w}_{i.} + O(y), \\ \bar{w} &= \hat{w} + O(y) \end{aligned}$$

so that equation (3) becomes

$$\Delta y = ys + O(y^2) \tag{4}$$

where

$$s = \frac{\hat{w}_i - \hat{w}}{\hat{w}} = \frac{\hat{\rho}_i}{\hat{x}_i}$$

Since $y = qx_i$, we have

$$\begin{aligned}\Delta qx_i &= q\Delta x_i + x_i\Delta q + O(qy) \\ &= \hat{x}_i\Delta q + O(qy)\end{aligned}$$

so that

$$\Delta q = qt + O(q^2) \quad (5)$$

where

$$t = \frac{\hat{w}_i - \hat{w}}{\hat{x}_i \hat{w}} = \frac{\hat{\rho}_i}{\hat{x}_i^2}$$

An equation similar to equation (4) is given by Kimura (1956) for a special case of a two-locus system.

Equations (4) and (5) demonstrate that both the overall frequency of the inversion and its frequency among gametes of the i th type will change initially at rates dependent on $\hat{\rho}_i$. If $\hat{\rho}_i > 0$, y and q will increase; if $\hat{\rho}_i < 0$, the inversion will be eliminated; if $\hat{\rho}_i = 0$, the change in the inversion frequency will be at most of the order of the square of its frequency. For an inversion introduced by mutation into a large population, such changes are negligible.

These results can be summed up by saying that a newly arisen inversion will be selected for in a large population only if it is introduced into a gamete present in excess at equilibrium, and that the selection coefficient in favour of the inversion is equal to the deviation of the equilibrium marginal fitness of the gamete concerned from the mean fitness of the equilibrium population, relative to the latter.

In the special case of two loci with alleles A , a and B , b respectively, there are four gamete types AB , Ab , aB and ab , with frequencies x_1 , x_2 , x_3 and x_4 . The coefficient of linkage disequilibrium, D , is defined as $x_1x_4 - x_2x_3$, and in this case we have

$$\rho_i = \pm \frac{w_H RD}{\bar{w}}$$

where R is the recombination fraction, w_H is the fitness of the double heterozygote $A/a B/b$, and the sign is positive for x_1 and x_4 and negative for x_2 and x_3 (Lewontin & Kojima, 1960).

Equations (4) and (5) thus become

$$\Delta y = \pm y \frac{w_H R \hat{D}}{\hat{x}_i \hat{w}} + O(y^2) \quad (6)$$

and

$$\Delta q = \pm q \frac{w_H R \hat{D}}{\hat{x}_i^2 \hat{w}} + O(q^2). \quad (7)$$

In the next section of this paper, we will illustrate the adequacy of the approximations made in deriving the above formulae by comparing the predictions of equations (4) and (5) with the results of exact calculations of population trajectories for systems of two, three and five loci.

3. NUMERICAL TESTING OF FORMULAE

The results derived above were tested by first determining the equilibrium for a given set of fitnesses, recombination parameters and initial gamete frequencies, by iteration of the appropriate recurrence relations for the system without the inversion, using the formulae and methods of Lewontin & Kojima (1960) for two-locus systems, Kojima & Klekar (1969) for three-locus systems, and Lewontin (1964) for the five-locus case. Only cases with two alleles at each locus were studied. Using the equilibrium gamete frequencies obtained in this way, the selection coefficients s and t of equations (4) and (5) can be calculated for an inversion introduced into a given type of gamete. These coefficients can be used to calculate the expected values of Δy and Δq (given the current values of y and q), which can be compared with the values obtained by direct calculation of the course of change of the system, incorporating the presence of the inversion.

Some results of these computations are given in Tables 1, 2 and 3. Table 1 gives the results for two two-locus systems, with the inversion introduced at an initial q of 0.005 into a gametic type (AB) present in excess. In both of the cases shown, agreement between the theoretical values of Δy and Δq and those obtained by direct calculation was good, up to frequencies of the inversion of 5% or so. In both cases, the inversion increases steadily in frequency, at the expense of the non-inverted gamete with the same genotype. By the 200th generation, the inversion has risen to a frequency of 50% in case (a), and represents nearly 100% of all the AB gametes; in case (b), the corresponding values are 25 and 50%.

Table 1. Selection for inversions in two-locus systems

(a) Symmetric fitness matrix

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	0.6	0.8	0.6
<i>Bb</i>	0.2	1.0	0.2
<i>bb</i>	0.6	0.8	0.6

(b) Asymmetric fitness matrix

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	0.606	0.800	0.602
<i>Bb</i>	0.200	1.000	0.180
<i>bb</i>	0.560	0.750	0.600

$R = 0.10, D = 0.144, s = 0.052, t = 0.132$

$R = 0.05, D = 0.205, s = 0.026, t = 0.050$

All values $\times 10^3$

Genera- tion	(a)					(b)				
	<i>y</i>	Δy		Δq		<i>y</i>	Δy		Δq	
		Found	Exp.	Found	Exp.		Found	Exp.	Found	Exp.
10	3.11	0.16	0.16	0.41	0.41	3.30	0.09	0.09	0.17	0.17
20	5.18	0.27	0.27	0.68	0.68	4.30	0.11	0.11	0.21	0.21
30	8.59	0.44	0.45	1.11	1.13	5.50	0.14	0.14	0.27	0.28
40	14.21	0.73	0.74	1.82	1.86	7.10	0.18	0.18	0.35	0.36
50	23.43	1.19	1.22	2.92	3.05	9.20	0.23	0.24	0.45	0.46
60	38.36	1.92	2.00	4.60	4.94	11.80	0.30	0.30	0.58	0.59
70	62.02	2.99	3.23	6.98	7.86	15.00	0.38	0.39	0.74	0.75
80	98.19	4.44	5.12	9.94	12.18	19.40	0.49	0.50	0.94	0.97
90	149.97	6.07	7.82	12.88	18.07	24.80	0.62	0.64	1.18	1.23
100	216.78	7.28	11.30	14.58	25.57	31.70	0.78	0.82	1.49	1.57

Table 2. Selection for inversions in three- and five-locus systems

(a) Three-locus system

(b) Five-locus system

		Fitness matrix			Number of loci heterozygous	Fitness
		AA	Aa	aa		
BB	CC	0.500	0.181	0.500	0	0.074
	Cc	0.138	0.241	0.138	1	0.111
	cc	0.500	0.181	0.500	2	0.222
Bb	CC	0.655	0.759	0.655	3	0.407
	Cc	0.793	1.000	0.793	4	0.667
	cc	0.655	0.759	0.655	5	1.000
bb	CC	0.500	0.181	0.500	$s = 0.037, t = 0.085$	
	Cc	0.138	0.241	0.138		
	cc	0.500	0.181	0.500		

$s = 0.025, t = 0.054$

All values $\times 10^3$

Generation	(a)					(b)				
	y	Δy		Δq		y	Δy		Δq	
		Found	Exp.	Found	Exp.		Found	Exp.	Found	Exp.
10	5.93	0.15	0.15	0.32	0.32	6.32	0.23	0.24	0.53	0.54
20	7.57	0.19	0.19	0.40	0.41	9.06	0.33	0.34	0.75	0.77
30	9.65	0.24	0.24	0.51	0.52	12.97	0.47	0.48	1.06	1.10
40	12.23	0.30	0.31	0.64	0.66	18.50	0.64	0.69	1.45	1.51
50	15.64	0.38	0.39	0.81	0.84	26.26	0.93	0.98	2.07	2.21
60	19.87	0.48	0.50	1.02	1.07	37.05	1.28	1.38	2.84	3.11
70	25.68	0.60	0.64	1.28	1.35	50.16	1.68	1.87	3.72	4.19
80	31.79	0.75	0.79	1.59	1.70	71.55	2.23	2.67	4.99	5.94
90	40.02	0.92	1.00	1.96	2.14	97.26	2.93	3.63	6.29	8.02
100	50.16	1.13	1.25	2.40	2.70	129.51	3.59	4.83	7.56	10.58

Table 2 shows some similar results for three and five-locus systems. Part (a) shows a three-locus case, with loci in the order *ABC*, with the map distance between adjacent loci 1.9 cM and no interference. The inversion was introduced with a *q* of 0.01 into the gamete *aBC*, which made up 46.4% of the initial population. Part (b) shows a system with five loci. Here, the map distance between adjacent loci was 1 cM, and interference was complete. The fitness of a genotype was determined solely by the number of loci for which it was heterozygous, the effect of increased heterozygosity increasing (on a linear scale) with the level of heterozygosity. In this case, the gamete into which the inversion was introduced (again with *q* = 0.01) was present at a frequency of 43.9% in the initial population. In both these cases, and in many others studied, agreement between the expected and observed values of Δy and Δq was very good, provided that the values of *y* and *q* respectively were below about 5%. Above this level, the fit becomes progressively worse as *y* and *q* increase.

When the inversion is introduced into a gamete in deficiency at equilibrium, it

is rapidly eliminated in accordance with equations (4) and (5). For example, with the fitness matrix of Table 1(a) and $R = 0.10$, when an inversion is introduced into a deficient gamete with an initial q of 0.005 and y of 0.0005, y is reduced to less than 0.000001 and q to 0.000009 by generation 30, at rates which agree with those predicted by equations (4) and (5).

With certain types of fitness interactions, notably with the two-locus systems studied by Lewontin & Kojima (1960) or with multiplicative heterotic interactions, it is possible, with suitable recombination parameters, to have equilibria with no linkage disequilibrium. In such cases, the change in y and q should be, at most, of order y^2 and q^2 respectively. This is found to be the case in practice, so that with values of y less than 1% or so, there is virtually no change in frequency of the inversion, even after several hundred generations. For example, with the fitness matrix of Table 1(a) and $R = 0.20$, there is a stable equilibrium with $D = 0$ in the absence of inversions. An inversion introduced with an initial q of 0.005 and y of 0.00125 reached $q = 0.00648$ and $y = 0.00163$ after 200 generations. This represents a very slight selection pressure, corresponding to effective s and t of 0.00013 and 0.0013 respectively, compared with their theoretical value of zero.

If, however, an inversion is introduced at a high enough initial frequency, selection can increase its frequency quite rapidly, even if the initial population is in linkage equilibrium. An example of this is shown in Table 3. With this fitness matrix, an inversion introduced into an AB gamete, with an initial y of 0.0025, increased its frequency to only 0.0029 after 200 generations. In part (a) of Table 3, it can be seen that the inversion increases slowly when introduced at a y of 0.0125 (after 200 generations, it reaches $y = 0.035$). With higher initial frequencies, as

Table 3. *Effect of initial frequency on the progress of an inversion introduced into a two-locus system in linkage equilibrium*

Fitness matrix	AA	Aa	aa
BB	0.250	0.375	0.250
Bb	0.625	1.000	0.625
bb	0.250	0.375	0.250

$R = 0.10, D = 0, s = 0, t = 0$

All values $\times 10^3$

Generation	(a) Initial $y = 0.0125$			(b) Initial $y = 0.025$		
	y	Δy	Δq	y	Δy	Δq
10	12.8	0.03	0.10	29.5	0.26	0.96
20	13.1	0.04	0.12	32.2	0.28	1.01
30	13.5	0.05	0.16	35.2	0.33	1.12
40	14.0	0.05	0.19	38.7	0.35	1.20
50	14.6	0.06	0.21	42.9	0.46	1.51
60	15.2	0.06	0.23	48.0	0.56	1.81
70	15.8	0.07	0.26	54.2	0.70	2.19
80	16.6	0.08	0.28	62.0	0.89	2.70
90	17.3	0.08	0.30	72.1	1.16	3.34
100	18.2	0.09	0.33	85.3	1.55	4.30

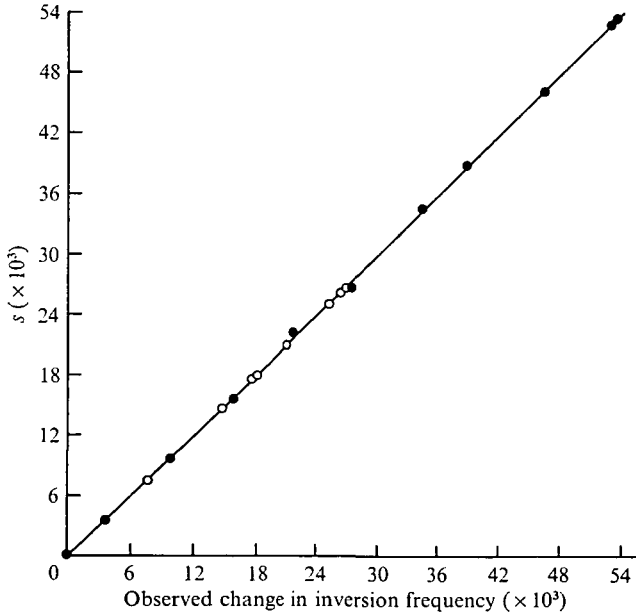


Fig. 1. Exact values of $\Delta y/y$ five generations after the introduction of the inversion, at an initial q of 0.01, plotted against the predicted values. The fitness matrices of Table 1(a) and Table 3 were used.

in part (b) of Table 3, the inversion increases more rapidly (to $y = 0.15$ after 200 generations, with initial $y = 0.025$). With an initial y of 0.05, the inversion reaches its equilibrium frequency of 0.50 after only 120 generations.

These results demonstrate that an inversion is effectively neutral when introduced at a low frequency into an equilibrium population in linkage equilibrium, but may be selected for strongly in such a population when its initial frequency is high enough.

The accuracy of equation (4) in predicting the rate of progress of a low frequency inversion is illustrated graphically in Fig. 1, in which the exact values of $\Delta y/y$ five generations after the introduction of the inversion, with an initial q of 0.01, are plotted against the values of s obtained from the appropriate equilibrium formulae. The fitness matrices of Tables 1(a) and Table 3 were used.

4. FACTORS AFFECTING THE SELECTION COEFFICIENT OF AN INVERSION

It is of some interest to consider how the selection coefficient of a newly arisen inversion in a gamete in excess at equilibrium is affected by the factors determining the nature of the equilibrium population: the level of epistatic selection and the recombination parameters for the genes involved. It is fairly obvious that, other things being equal, the greater the amount of epistasis, the greater the selection coefficients s and t are likely to be. In the case of the symmetric fitness, two-locus model of Lewontin & Kojima (1960), the amount of epistasis can be summed

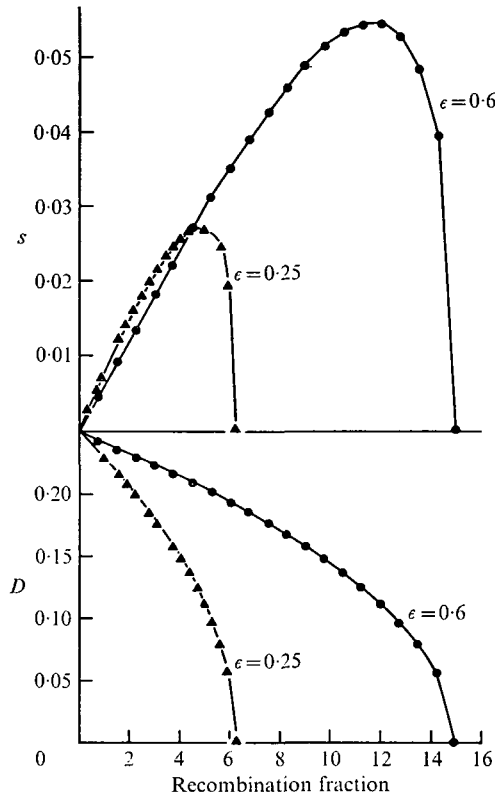


Fig. 2. Upper graph: Values of s plotted against R for two different fitness matrices. Lower graph: D plotted against R , for the same fitness matrices.

up in a single parameter ϵ , which measures the difference between the effects of making locus A heterozygous in heterozygotes and homozygotes for locus B , relative to the fitness of $A/a B/b$. The fitness matrices of Table 1(a) and Table 3 are of the symmetric type: ϵ is 0.6 for the former, and 0.25 for the latter. In Fig. 2, the values of s and D for these fitness matrices are plotted against R . It will be seen that the curve of s for the system with higher ϵ is mostly the higher of the two.

Recombination affects the selection coefficient in a more complex way. As can be seen from Fig. 2, there may be an optimum value of R . Interestingly enough, the selection coefficient is not an increasing function of the amount of linkage disequilibrium. This is obvious from the fact that the \hat{p}_i must be zero when linkage is complete, and also when it is loose enough for no disequilibrium to exist. In the two-locus examples just discussed, D decreases steadily with increasing R , whereas the maximum in s occurs quite near the point above which $D = 0$. This also seems to be true for systems involving more than two loci, where the advantage of heterozygosity increases with the number of loci already heterozygous. Quite high selection coefficients can be obtained with only moderate linkage disequilibrium; for example, with the curve of Fig. 2 corresponding to the fitness matrix of

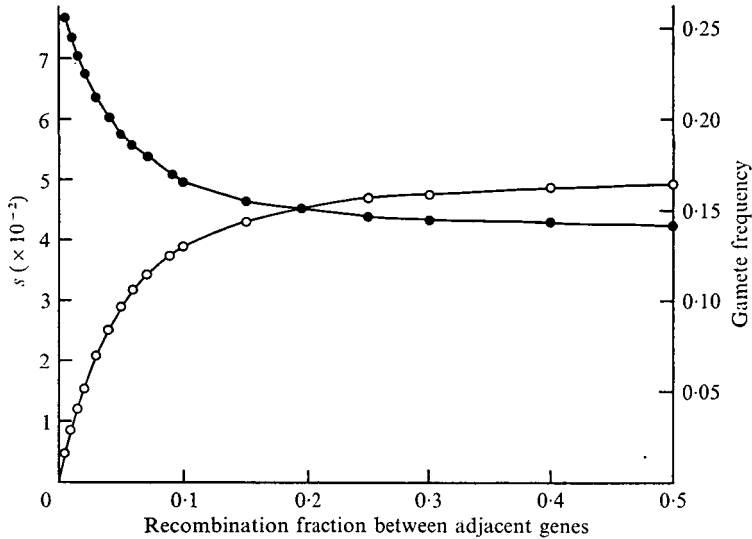


Fig. 3. Values of s and gamete frequency, for the gamete Abc , plotted against recombination fraction.

Table 1(a), an s of 0.04 is obtained when $D = 0.056$, i.e. when the gametes AB and ab each have frequencies of 0.306, compared with frequencies of 0.194 for the deficient gametes Ab and aB .

With more complex fitness interactions, the pattern of relationship of s with the amount of recombination may be more complex. With the general symmetric two-locus fitness matrix of Bodmer & Felsenstein (1967), for example, it is impossible for equilibria with $D = 0$ to exist, so that there is always some selection pressure for an inversion regardless of R . A similar three-locus case, with an asymmetric fitness matrix, is illustrated in Fig. 3. The map distances $A - B$ and $B - C$ were the same and there was no interference. It will be seen that s , for the gamete Abc , increases continuously with map distance, while the frequency of Abc in the initial population falls from about 26% (with no recombination) to 14% (with independent assortment).

5. SEX DIFFERENCES IN RECOMBINATION AND FITNESS

Up to now, we have assumed that the recombination and fitness parameters are the same for both sexes. Providing that the fitnesses and gamete frequencies do not differ too much between the sexes, the arguments of §2 of this paper can be extended to the more general case, as follows. Consider first the case when only the cross-over values differ in males and females. We can write the frequency of the i th gametic type among gametes produced by females as x_{if} , and the corresponding frequency for paternally-derived gametes as x_{im} . If x_i is the average of x_{if} and x_{im} , we can write $x_{if} = x_i + \delta_i$ and $x_{im} = x_i - \delta_i$. It will be assumed that terms $O(\delta_i^2)$ can be neglected. Write w_{if} for the marginal fitness of maternal gametes

of the i th type, and w_{im} for the corresponding function for paternal gametes. We have

$$\begin{aligned}
 w_{if} &= \sum_j x_{jm} w_{ij}, \\
 w_{im} &= \sum_j x_{jf} w_{ij}, \\
 \bar{w} &= \sum_i x_{if} w_{if} = \sum_i x_{im} w_{im}
 \end{aligned}$$

and so:

$$\Delta x_{if} = \frac{\frac{1}{2}x_{if}w_{if} + \frac{1}{2}x_{im}w_{im} - x_{if}\bar{w}}{\bar{w}} - \rho_{if} \tag{8a}$$

$$\Delta x_{im} = \frac{\frac{1}{2}x_{if}w_{if} + \frac{1}{2}x_{im}w_{im} - x_{im}\bar{w}}{\bar{w}} - \rho_{im}. \tag{8b}$$

ρ_{if} and ρ_{im} are parameters measuring the effects of cross-over events in multiple heterozygotes on the frequencies of the i th gametic type in males and females respectively. ρ_{if} is a function of the form

$$\frac{1}{2} \sum_{k,l} (x_{kf}x_{im} + x_{km}x_{if}) R_{kl}^{(f)} \tag{9a}$$

where $R_{kl}^{(f)}$ measures the contribution to Δx_{if} of cross-over events in females formed from the union of the k th and l th gametic types.

If we neglect terms $O(\delta^2)$, equation (9a) can be written as

$$\rho_{if} = \sum_{kl} x_k x_l R_{kl}^{(f)}. \tag{9b}$$

Similarly, we can write

$$\rho_{im} = \sum_{kl} x_k x_l R_{kl}^{(m)} \tag{10}$$

where $R_{kl}^{(m)}$ has an analogous meaning to $R_{kl}^{(f)}$.

Substituting equations (9b) and (10) into equations (8), and neglecting terms $O(\delta^2)$, we get

$$\Delta x_{if} = \frac{x_i(w_i - \bar{w})}{\bar{w}} - \sum_{kl} x_k x_l R_{kl}^{(f)}$$

$$\Delta x_{im} = \frac{x_i(w_i - \bar{w})}{\bar{w}} - \sum_{kl} x_k x_l R_{kl}^{(m)}$$

where $w_i = \sum_j x_j w_{ij}$.

If we write $\rho_i = \frac{1}{2} \sum_{kl} R_{kl}^{(f)} + \frac{1}{2} \sum_{kl} R_{kl}^{(m)}$, we get

$$\Delta x_i = \frac{x_i(w_i - \bar{w})}{\bar{w}} - \rho_i. \tag{11}$$

This is of identical form to equation (1), and shows that, to an accuracy of the order of the square of the differences between maternal and paternal gamete frequencies, the mean of the recombination parameters in males and females can be used as the measure of recombination in predicting the change in gamete frequencies under the joint action of recombination and selection. This may be compared with the similar finding of Geiringer (1948) for the case of no selection.

Table 4. Selection for an inversion in a two-locus system with no crossing-over in males

				♀♀	♂♂
Initial equilibrium gamete frequencies		<i>AB</i>		0.375	0.416
		<i>Ab</i>		0.125	0.084
		<i>aB</i>		0.125	0.084
		<i>ab</i>		0.375	0.416
$s = 0.053, \quad t = 0.133$					
All values $\times 10^3$					
Generation	<i>y</i>	Δy		Δq	
		Found	Exp.	Found	Exp.
10	1.68	0.09	0.09	0.23	0.23
20	2.86	0.16	0.15	0.39	0.38
30	4.83	0.26	0.25	0.65	0.64
40	8.15	0.44	0.43	1.09	1.08
50	13.38	0.72	0.70	1.80	1.80
60	22.44	1.20	1.18	2.93	3.00
70	37.41	1.94	1.97	4.65	4.91
80	62.00	3.05	3.26	7.11	7.91
90	100.40	4.56	5.28	10.18	12.35
100	152.18	6.22	8.00	13.18	18.40

When there are differences in fitness between the sexes, an analogous argument holds good, provided that the squares of the sex differences in gamete frequency and fitness, and also of the changes in gamete frequencies per generation, can be neglected. The measure of fitness is the average of the male and female fitnesses.

Combining these results, it is obvious that expressions similar to equations (4) and (5) can be derived for these more general cases with the above assumptions. Gamete frequencies, fitnesses and recombination parameters are all to be measured as averages for the two sexes.

This result is of particular importance in the case of organisms, such as many *Drosophila* species, where recombination is restricted to one sex. An example of the process of selection for an inversion in such a system is given in Table 4. The fitness matrix of Table 1(a) was used, with an *R* of 0.20 in females and 0 in males. The inversion was initially introduced into maternal gametes with a *q* of 0.005. As may be seen from the table, the selection coefficients *s* and *t*, as calculated from the composition of the equilibrium population lacking the inversion (which was found by iteration of the appropriate equations), are very close to their corresponding values for the case of equal recombination (*R* = 0.10) in both sexes. The agreement between the exact and approximate theoretical values of Δy and Δq is not significantly worse than in the case of equal recombination.

6. EFFECTS OF PERTURBATIONS FROM EQUILIBRIUM

So far, the results we have derived have been based on the assumption that the initial population, into which the inversion is introduced, is in equilibrium under the joint action of selection and recombination. It is clearly of interest to investigate how far our conclusions must be modified when the initial population is perturbed from such an equilibrium state. We studied this question with various two-locus systems. In general, it appears that when there is strong enough epistatic selection to generate appreciable potential selection for inversions, the system returns back towards the equilibrium so quickly that there is no serious disagreement between the observed and expected rates of change.

Table 5 illustrates this conclusion for the fitness matrix of Table 1(a), and $R = 0.10$. In this case, the equilibrium gamete frequencies are $x_1 = x_4 = 0.394$ and $x_2 = x_3 = 0.106$. In case (a) of the table, the inversion was introduced with initial $q = 0.005$ into an AB gamete whose frequency was above its equilibrium value, and which therefore tended to decrease in frequency in the initial generations (the initial gamete frequencies were chosen so that the decrease was due to selection as well as recombination). In case (b) of Table 5, the inversion was introduced into a population with AB in lower frequency than at equilibrium.

It will be seen in both these examples that the fit of the approximate equations is somewhat worse than in the equilibrium case, but not greatly so. When the inversion is introduced into a gamete which is deficient at the nearest equilibrium, it is eliminated, even with considerable perturbations from equilibrium. For ex-

Table 5. Effect of perturbations from equilibrium of initial population on the progress of inversions in two-locus systems

(a) Initial gamete frequencies.						(b) Initial gamete frequencies.				
	<i>AB</i>	0.45		<i>AB</i>	0.30		<i>AB</i>	0.30		
	<i>Ab</i>	0.05		<i>Ab</i>	0.20		<i>Ab</i>	0.20		
	<i>aB</i>	0.10		<i>aB</i>	0.12		<i>aB</i>	0.12		
	<i>ab</i>	0.40		<i>ab</i>	0.38		<i>ab</i>	0.38		
$s = 0.052$						$t = 0.132$				
All values $\times 10^3$										
	(a)					(b)				
		Δy		Δq			Δy		Δq	
Genera- tion	<i>y</i>	Found	Exp.	Found	Exp.	<i>y</i>	Found	Exp.	Found	Exp.
10	3.21	0.17	0.17	0.44	0.36	2.80	0.15	0.15	0.34	0.53
20	5.35	0.28	0.28	0.72	0.60	4.67	0.24	0.24	0.58	0.85
30	8.93	0.47	0.47	1.17	1.00	7.69	0.39	0.40	0.97	1.39
40	14.85	0.77	0.77	1.91	1.66	12.63	0.64	0.66	1.59	2.27
50	24.54	1.25	1.28	3.07	2.72	20.70	1.05	1.08	2.57	3.71
60	40.22	2.01	2.10	4.82	4.40	33.81	1.69	1.76	4.08	6.02
70	65.02	3.13	3.39	7.26	7.00	54.72	2.66	2.85	6.25	9.63
80	100.27	4.61	5.23	10.27	10.80	87.06	4.01	4.53	9.10	15.04
90	156.20	6.23	81.4	13.14	15.94	134.37	5.63	7.00	12.14	22.61
100	224.29	7.35	11.69	14.62	22.13	197.46	7.03	10.29	14.31	32.18

ample, with the same fitness matrix and recombination fraction as above, when the inversion is introduced with initial $q = 0.005$ into an Ab gamete in a population with $x_1 = 0.20$, $x_2 = 0.30$, $x_3 = 0.12$ and $x_4 = 0.38$, y is reduced to 0.0001 after twenty generations. The quantitative agreement between the approximate and exact values of Δy and Δq is slightly worse than in the corresponding equilibrium case.

Several cases of an inversion introduced into a system perturbed from a $D = 0$ equilibrium were studied. Even when the inversion is introduced into a gamete which is increasing temporarily in frequency, any marked increase in y and q is confined to the first few generations (provided that the initial inversion frequency is low), after which there is at most a very slow increase. For example, with the fitness matrix and R of Table 3, an inversion introduced, with initial $y = 0.0005$, into a population with gamete frequencies $x_1 = 0.10$, $x_2 = 0.20$, $x_3 = 0.40$ and $x_4 = 0.30$ increased to $y = 0.00074$ after five generations, fell to 0.00060 after eighty generations, and reached 0.00072 after a thousand generations.

These results suggest that the effects of perturbations of the initial population from its equilibrium state are relatively small. In particular, the qualitative conclusion that selection for an inversion is dependent on linkage disequilibrium, in the equilibrium state from which the population has been perturbed, seems to hold good.

7. DISCUSSION

One of the main conclusions of this work is that there will be appreciable selection for newly arising inversions in multi-locus systems only if the population is at or near an equilibrium with linkage disequilibrium. Feldman (1972) has reached the same conclusion for the case of a modifier of recombination of arbitrary effect, with the symmetric two-locus fitness matrix of Bodmer & Felsenstein (1967). It has been widely assumed that epistasis with respect to fitness between the genes located in the chromosome segment covered by an inversion is sufficient for a new inversion to be favoured by selection. For example, Dobzhansky (1970) writes: 'Inversions binding together coadapted gene complexes will be favoured by natural selection'.

Since epistasis of fitness effects (on a linear scale) is necessary for the existence of linkage disequilibrium at equilibrium (Kojima & Kelleher, 1961), it is also a necessary condition for the selection of a newly-arisen inversion. But there are several types of fitness interaction which can generate stable equilibria with no linkage disequilibrium for certain ranges of the recombination parameters, e.g. multiplicative fitness interactions (Lewontin, 1964; Bodmer & Felsenstein, 1967). In such cases, as we have seen, there is virtually no selection in favour of a new inversion, unless it is introduced at an appreciable frequency ($y = 5\%$ or more). This finding explains the multi-locus simulation results of Fraser & Burnell (1967), who failed to obtain any selection for an inversion unless it was introduced above a certain critical frequency. The system they were using generated unstable equilibria with no linkage disequilibrium.

It is of course possible that an inversion introduced into a multi-locus system with epistatic selection but no linkage disequilibrium may drift up to a high enough frequency to be selected. In a large population of this sort, the chance of incorporation of an inversion is likely to be of the order of the reciprocal of the total population size, so that a very large number of mutations to inversions would have to occur for a single polymorphism to become established. It seems somewhat unlikely that the more than forty known inversion polymorphisms in a species such as *D. subobscura*, for example, could be explained by such a mechanism, in view of the extreme rarity of spontaneous mutation to new inversions in *Drosophila*.

On the other hand, if an inversion appears in a gamete in excess in a population at equilibrium, its chance of survival in a large population will be, from equation (4), approximately $2s$. This follows from the standard result of Fisher (1930). As has been shown (Figs. 2 and 3), given large enough selection pressures and suitable linkage relations, values of s as high as 2% can be produced with quite slight linkage disequilibrium.

It is of some interest to derive an expression for the expected chance of survival of an inversion introduced into a multi-locus system. As we have seen from equation (4), the chance of survival of a new inversion occurring in a representative of the i th type of gamete is either approximately $2(w_i - \bar{w})/\bar{w}$ ($w_i > \bar{w}$) or 0 ($w_i \leq \bar{w}$). The probability that a newly-arisen inversion occurs in the i th gametic type is x_i . The expected probability of survival of a new inversion is thus $2 \sum_i \hat{x}_i (\hat{w}_i - \hat{\bar{w}}) / \hat{\bar{w}}$ (for $\hat{w}_i > \hat{\bar{w}}$). This can be approximated as follows, if we assume a normal distribution for $(\hat{w}_i - \hat{\bar{w}}) / \hat{\bar{w}}$ with zero mean and variance $\hat{\sigma}^2$. On this basis, there is a probability of $\frac{1}{2}$ that the inversion has a non-zero chance of survival. Using the usual formula for the mean of a truncated normal distribution, the expected chance of survival of an inversion is approximately

$$P = 2 \times \frac{1}{2} \times 2 \times 0.4\hat{\sigma} = 0.8\hat{\sigma}. \tag{12}$$

It can be shown as follows that $\hat{\sigma}$ is directly related to the loss in fitness produced by recombination every generation in the equilibrium population. We can write the change in mean fitness of a multi-locus system in an arbitrary generation as

$$\Delta\bar{w} = \Delta_s - \Delta_r$$

where Δ_s is the change in fitness due to selection alone (i.e. not involving cross-over probabilities), and Δ_r is the change due to recombination.

It follows from the multi-locus version of the fundamental theorem of natural selection developed by Kojima & Kelleher (1961) that we have

$$\Delta_s \approx 2\bar{w}\sigma^2. \tag{13}$$

At equilibrium, $\Delta_s = \Delta_r$. The loss in fitness of the equilibrium population due to recombination is therefore equal to $2\bar{w}\hat{\sigma}^2$, approximately. The mean fitness of a hypothetical population formed by random union of gametes drawn from the

equilibrium population after selection, but before recombination has taken place, is therefore $\hat{w}(1 + 2\hat{\sigma}^2)$. We can thus define the genetic load due to recombination as

$$L_r = \frac{2\hat{w}\hat{\sigma}^2}{\hat{w}(1 + 2\hat{\sigma}^2)} \approx 2\hat{\sigma}^2. \quad (14)$$

From equation (12), we therefore have

$$P = 0.56\sqrt{L_r}. \quad (15)$$

The expected chance of survival of a new inversion therefore depends on the contribution of the segment of chromosome covered by it to the marginal variance in fitness, or, equivalently, to the loss in fitness due to recombination in the original population. Clearly, the same values of this quantity can be produced by a virtually infinite set of combinations of possible interaction systems and numbers of loci, so there is little point in speculating on the details of the genetic systems underlying inversion polymorphisms.

Finally, it should be emphasized that the selection coefficient s with which we have chiefly been concerned here is not necessarily related to the selection coefficients for inversions which are observed in naturally occurring inversion systems. These relate to the genetic system after the inversion has spread from its low initial frequency. In many cases (but by no means all), the inversion will oust the gamete with the same genotype and the final population will be dominated by two types of gamete – the inversion, and the complementary type which gives the maximum heterozygosity in combination with the inversion. The selective differences between the hetero- and homo-karyotypes in such cases reflect the differences between relatively homozygous genotypes and genotypes heterozygous for many loci. For example, with the fitness matrix of Table 1(a), the equilibrium population with an inversion arising in an AB gamete can be shown to consist of AB (inversion) and ab in equal frequencies. This gives a selection coefficient of 40% against homokaryotypes, which is nearly 8 times as large as the largest s of Fig. 2.

REFERENCES

- BODMER, W. F. & FELSENSTEIN, J. (1967). Linkage and selection: theoretical analysis of the deterministic two locus random mating model. *Genetics* **57**, 237–265.
- DEAKIN, M. A. B. (1972). A model for inversion polymorphism. *Journal of Theoretical Biology* **35**, 191–212.
- DOBZHANSKY, T. (1951). *Genetics and the Origin of Species*. New York: Columbia University Press.
- DOBZHANSKY, T. (1970). *The Genetics of the Evolutionary Process*. New York: Columbia University Press.
- FELDMAN, M. W. (1972). Selection for linkage modification: 1. Random mating populations. *Theoretical Population Biology* **3**, 324–346.
- FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- FRASER, A. S. & BURNELL, D. (1967). Simulation of genetic systems. XI. Inversion polymorphism. *American Journal of Human Genetics* **19**, 270–286.
- GEIRINGER, H. (1948). On the mathematics of random mating in case of different recombination values for males and females. *Genetics* **33**, 548–554.
- HALDANE, J. B. S. (1957). The conditions for co-adaptation in polymorphism for inversions. *Journal of Genetics* **55**, 218–225.

- KIMURA, M. (1956). A model of a system which leads to closer linkage by natural selection. *Evolution* **9**, 419–435.
- KOJIMA, K. (1967). Likelihood of establishing newly induced inversion chromosomes in small populations. *Ciencia e Cultura* **19**, 67–77.
- KOJIMA, K. & KELLEHER, T. M. (1961). Changes of mean fitness in random mating populations when epistasis and linkage are present. *Genetics* **46**, 527–540.
- KOJIMA, K. & KLEKAR, A. (1969). Deterministic simulation of evolutionary changes in three-locus genetic systems. In *Computer Applications in Genetics*, ed. N. E. Morton, pp. 147–160. Honolulu: University of Hawaii Press.
- LEWONTIN, R. C. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67.
- LEWONTIN, R. C. (1971). The effect of genetic linkage on the mean fitness of a population. *Proceedings of the National Academy of Sciences, U.S.A.* **68**, 984–986.
- LEWONTIN, R. C. & KOJIMA, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.
- NEI, M. (1967). Modification of linkage intensity by natural selection. *Genetics* **57**, 625–641.
- NEI, M., KOJIMA, K. & SCHAFFER, H. (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**, 741–750.
- OHTA, T. (1971). Associative overdominance caused by linked detrimental mutations. *Genetical Research* **18**, 227–286.
- SPERLICH, D. (1959). Experimentelle Beiträge zum Problem des positiven Heterosis-effekts bei der struktur-polymorphen Art *Drosophila subobscura*. *Zeitschrift für Vererbungslehre* **90**, 273–287.
- STURTEVANT, A. H. & MATHER, K. (1938). The interrelations of inversions, heterosis and recombination. *American Naturalist* **72**, 447–452.
- TURNER, J. R. G. (1970). Some properties of two locus systems with epistatic selection. *Genetics* **64**, 147–155.