SELECTION FOR RECOMBINATION IN PARTIALLY SELF-FERTILIZING POPULATIONS

D. CHARLESWORTH*, B. CHARLESWORTH* AND C. STROBECK†

*National Institute of EnvironmentalHealth Sciences, P. O. Box 12233,
†Department of Genetics, University of Alberta, Edmonton, Alberta, Canada

Manuscript received November 7, 1978
Revised copy received March 9, 1979

ABSTRACT

This paper confirms HOLDEN's (1979) suggestion that certain types of fitness interactions between a pair of loci in partially self-fertilizing populations may promote selection for increased recombination between them. Our results are based on both algebraic and computer calculations of the fate of alleles at a third locus, which control the level of recombination between the selected pair. We also show that the behavior of the population mean fitness as a function of recombination fraction is not necessarily an indicator of the direction of selection on recombination in partially selfing populations.

IN a recent detailed examination of a two-locus system with a highly symmetrical fitness matrix, HOLDEN (1979) discovered that, in some cases, population mean fitness in partially selfing populations was an increasing function of the recombination fraction. He suggested that modifiers increasing recombination would be selected for in such cases. This contradicts the result of our previous study (CHARLESWORTH, CHARLESWORTH and STROBECK 1977) of the spread of modifiers of recombination in partially selfing populations. However, only three different fitness matrices were studied by us, and it remained possible that parameter values leading to selection for increased recombination were not included in our study. It seems important to re-examine this question, bearing HOLDEN's new results in mind. In particular, HOLDEN suggested by analogy with the random-mating case (and confirmed by computer studies) that, with the form of fitness matrix which he studied, the condition for mean fitness to increase with recombination fraction (assumed also to determine whether increased recombination will be favored) is that the fitness matrix and the selfing rate must be such that the equilibrium with zero linkage disequilibrium is stable for all recombination fractions. HOLDEN finds that this occurs when a quantity $\Delta(s)$ is negative, where $\Delta(s)$ is given by

$$\Delta(s) = \hat{W}(0,s) - W^*(0,s) .$$

Here, $\hat{W}(0,s)$ is the mean fitness of the equilibrium population with linkage disequilibrium when there is no recombination and the selfing rate is s; $W^*(0,s)$

* Present address: School of Biological Sciences, University of Sussex, Brighton BN1 9QG, England.

is the mean fitness for the equilibrium population with no linkage disequilibrium, in the same circumstances. We have therefore done deterministic computer calculations of the spread of genes modifying recombination introduced into populations at equilibrium under HOLDEN's (1979) form of fitness matrix. We have compared the results with those predicted by the mean fitnesses and $\Delta(s)$, and find that $\Delta(s)$ does, in all cases tested, predict the changes in the modifier. We have also studied an asymmetrical fitness matrix related to HOLDEN's form of fitness matrix; with this type of matrix, the difference in mean fitnesses did not always predict the changes in the modifier.

Before describing these results, we will present some analytical results relating to the spread of inversions in selfing populations with the form of fitness matrix assumed by HOLDEN (1979). The fitness matrix is given in Table 1.

FATE OF AN INVERSION

We assume that a population is at a stable equilibrium under a fitness matrix of the form studied by HOLDEN (1979) and that there is no linkage disequilibrium. We then introduce an inversion (or, equivalently, a dominant recombination suppressor that is completely linked to the selected loci). It is possible to derive an expression for the rate of change of such an element when introduced at low frequency into a population with a low selfing rate (s) and with an initial recombination fraction R between the selected loci. This calculation is outlined in the APPENDIX. The asymptotic relative rate of change in frequency of an inversion proves to be given by the following expression:

$$\frac{\Delta p}{p} = \frac{4sR(1-R)(1+a-2b)}{(1+a+2b)^2}, \qquad (1)$$

where p is the inversion frequency, and a and b are the two parameters of HOLDEN's fitness matrix (see Table 1). Equation (1) shows that the sign of the change does not depend on R (provided that R is above the critical value that guarantees no linkage disequilibrium).

It is convenient to specify the fitness matrix in the more general form of LANGLEY and CROW (1974), as shown in Table 1. With this notation, the two selected

TABLE 1

Comparison of HOLDEN's *fitness matrix (above) with a more general form of fitness matrix (below)*

	AA	*Aa*	*aa*
BB	a	b	a
	$1-a_1-b_1-k_1$	$1-b_1$	$1-a_2-b_1-k_3$
Bb	b	1	b
	$1-a_1$	1	$1-a_2$
bb	a	b	a
	$1-a_1-b_2-k_2$	$1-b_2$	$1-a_2-b_2-k_4$

loci interact additively if all the k_i values are zero. If the k_i are all equal, and if $a_1 = b_1 = a_2 = b_2$, we have the form of fitness matrix used by HOLDEN (1979) with $a_1 = b_1 = a_2 = b_2 = 1 - b$, and $k_1 = k_2 = k_3 = k_4 = 2b - a - 1 = k$, say. Thus, the term $1 + a - 2b$ in the numerator of Equation (1) is equal to $-k$. Equation (1) therefore shows that, when there is no epistasis, the inversion is not expected to change in frequency. When the double homozygotes have lower fitness than predicted from the fitness of the single homozygotes with additivity across loci, Equation (1) predicts that an inversion will be eliminated, whereas it should spread if the epistasis is in the opposite direction (k negative).

The value of these results depends on the relations between the conditions under which an inversion will spread and those governing the spread or elimination of genic modifiers of recombination, which might have lesser effects and which might be linked or unlinked, dominant or recessive, or of intermediate dominance. In previous studies (*e.g.*, CHARLESWORTH 1976), the direction of change of an inversion has been found to predict that of genic modifiers in all cases. The computer results described below gave the same result, unless the selfing rate was very high, in which case the approximations involved in obtaining Equation (1) are not valid.

RESULTS OF COMPUTER CALCULATIONS

We have used a three-locus model of the type described by CHARLESWORTH, CHARLESWORTH and STROBECK (1977). The alleles at one locus (the modifier locus, C) control the recombination fraction between the other two (the selected loci, A and B). The modifier locus was assumed to be to the right of the selected loci and linked to it with recombination fraction R_{BC}. We assumed no interference. The population was run to equilibrium before adding a low frequency of the modifier allele in one gamete type. The rate of increase ($\Delta p/p$) of the modifier was followed for some generations until it became constant; this rate was taken as an estimate of the asymptotic rate of increase of the modifier when rare. The rate did not depend on the initial modifier frequency in any of a number of cases tested; a value of 0.00025 for the initial frequency was used in most of the runs.

We first made some comparisons between rates of increase of inversions predicted by Equation (1) and those observed in the computer runs. The initial recombination fraction between the selected loci was assumed to be 0.5. The agreement is very good with both positive and negative k values (which lead to elimination and incorporation of inversions, respectively), when s is small. Quite good agreement is maintained even when s is not very low, but serious discrepancies are seen when s is very high and k is positive, when selection for increased recombination may be incorrectly predicted. When k is zero, increased s values lead to increased selection for an inversion. Similar results were obtained with other initial R values.

Runs with genic modifiers of recombination, rather than inversions, were also done. Table 2 shows results with some examples of HOLDEN's (1979) form of fitness matrix. As in HOLDEN's (1979) results, the sign of $\Delta(s)$ correctly predicts

TABLE 2

Results of computer runs of unlinked modifiers with intermediate dominance

k	a	s	Difference in mean fitness between equilibrium with R=0 and that with R=0.5	Rate of spread of modifier Increasing R from 0 to 0.5	Decreasing R from 0.5 to 0
		$b = 0.9$			
0	0.8	0.1	0.00025	—0.00013	0
		0.5	0.00153	—0.00108	0.00008
		0.8	0.00203	—0.00176	0.00019
0.02	0.78	0.01	—0.00001	0.00001	—0.00004
		0.1	—0.00013	0.00010	—0.00004
		0.5	—0.00031	0.00027	—0.00011
		0.8	0.00007	—0.00007	0.00001
0.04	0.76	0.01	—0.00003	0.00002	0
		0.1	—0.00026	0.00023	—0.00008
		0.5	—0.00084	0.00080	—0.00029
		0.8	—0.00046	0.00049	—0.00019
		$b = 0.8$			
0	0.6	0.1	0.00120	—0.00083	0.00001
		0.5	0 00759	—0.00684	0.00054
		0.8	0.01270	—0.01482	0.00205
0.04	0.56	0.1	—0.00026	0.00025	—0.00009
		0.5	0.00115	—0.00108	0.00009
		0.8	0.01071	—0.01326	0.00189
0.2	0.4	0.1	—0.00155	0.00154	—0.00056
		0.5	—0.00439	0.00522	—0.00205
		0.8	0.00643	—0.00966	0.00164

the sign of the difference in mean fitness between the equilibrium with $R = 0$ (which may have linkage disequilibrium if k is negative, or if k is small and s is high enough) and that with $R = 0.5$ (where there is no linkage disequilibrium with these fitness matrices). In those cases in which increased recombination was selected against, the mean fitness was highest when the recombination fraction between the selected loci was zero; at the initial equilibrium with $R = 0$, there was linkage disequilibrium ($D = 0.25$). When the mean fitness increased with R, modifiers increasing recombination spread, and modifiers decreasing it from an initial value of 0.5 were eliminated.

The effect of changes in the selfing rate may be summarized as follows (see Table 2). With a value of $k(> 0)$ giving a fitness matrix that can generate selection for increased recombination, there is no lower limit to the value of s that produces such selection. Selection is weak when s is very small, and increases to a maximum when s reaches some value less than 1. For higher values of s, selection for recombination weakens again, and there is often a critical s value beyond which the same fitness matrix now produces selection for reduced, rather than increased recombination. This s value tends to be lower, the lower the value of k. Thus, with this form of fitness matrix, selection for increased recombination occurs only with $k > 0$, and even then only when s is not too high.

In all cases tested, we found that when a modifier increasing recombination from zero increased in frequency, a modifier reducing recombination was eliminated, and *vice versa*. This suggests that the equilibrium recombination fraction that is reached under selection is always 0 or 0.5 with this form of selection. This has been confirmed for a selected set of parameter values, by starting the runs with a range of initial recombination fractions between the selected loci and introducing modifiers either increasing or decreasing that value. Furthermore, the equilibrium mean fitness was always found to be either a monotonically increasing or decreasing function of recombination fraction, for this form of fitness matrix.

We have also studied the effect of recessivity of the modifier gene, and of its linkage to the selected loci. With a recessive modifier, the changes were always the same direction as with intermediate dominance, but were slower, often considerably so. As would be expected, this effect was especially noticeable when s was low. A linked modifier also increased more slowly than one unlinked to the selected loci; with $R_{BC} = 0.05$, the rate of increase of the modifier when the starting recombination fraction was zero was, in all cases tested, slightly less than for the case of $R_{BC} = 0.5$.

Finally, runs with a variant of the basic fitness matrix were done in order to see the effects, if any, of making the matrix asymmetrical. To preserve a single parameter k as an epistasis measure, the asymmetry was introduced by setting $a_1 = b_1 = 0.1$, $a_2 = b_2 = 0.2$.At equilibrium, these systems always had linkage disequilibrium, which was still present (though small) when the selected loci recombined freely. With these parameter values, we nevertheless found that with all positive values of k (0.02 and higher), increased recombination was favored. We could not use high selfing rates, because fixation of one gamete type occurred. The maximum value of s used was therefore 0.5, in most cases. It is interesting to note that, with some parameter sets (*e.g.*, $k = 0.04$, $s = 0.5$ or $k = 0.1$, $s = 0.8$), the population mean fitness was lower when the recombination fraction was 0.5 than when it was zero. In these cases, the mean fitness decreased as R increased from zero, then increased again, but never rose to as high a value as with complete linkage. Nevertheless, when a modifier increasing the recombination fraction from zero to 0.5 was introduced, it spread, the mean fitness decreasing as it did so.

DISCUSSION

The results described above clearly show that our earlier generalization (CHARLESWORTH, CHARLESWORTH and STROBECK 1977) was unjustified, and that a constant fitness matrix can, in a partially selfing population, give selection for increased recombination. Therefore, in order to account for the observation that recombination rates tend to be higher in selfing than in outcrossing species [see CHARLESWORTH, CHARLESWORTH and STROBECK (1977) for references], it may be unnecessary to seek for models of selection for increased recombination that become stronger under selfing; we now have a mechanism that promotes recombination when there is some selfing, but not when there is random mating,

all other population parameters being unaltered. To accept this as the explanation of the findings in selfing species of plants, we would have to know that fitness matrices that generate selection for increased recombination when there is selfing are likely to be common. HOLDEN's (1979) form of fitness matrix is, of course, a very special case and unlikely to be generally applicable, but this does not imply that other more realistic types of fitness matrix would not have the same property. We have studied one form of an asymmetrical case, and found that the selection for increased recombination is preserved. The suggestion that the loci must interact in such a way that homozygosity for both loci gives lower fitness than would be predicted from the single-locus effects seems to be helpful, but is hard to generalize; in the general case with four epistasis parameters, there seems little alternative to obtaining empirical results by computer runs.

Furthermore, this explanation for increased recombination in selfing species encounters difficulties in cases with high selfing rates. Even when variation is preserved, high selfing rates tend to give selection for decreased recombination, or to weaken any selection for increased recombination that exists with less selfing. This model is therefore more attractive as an explanation of differences between populations with low to moderate selfing rates than for very high rates of selfing.

The result that with selfing the mean fitness may be a nonmonotonic function of the recombination fraction is new, but a similar relation has previously been discovered by KARLIN and CARMELLI (1975) in random-mating populations; in some of the examples studied by these authors, mean fitness was highest when there was no recombination and declined as recombination increased, but then increased again. FELDMAN (personal communication), has shown that a modifier decreasing recombination will always spread in a random-mating population that is at equilibrium with linkage disequilibrium, regardless of the relation between the recombination fraction and the equilibrium mean fitness. A modifier reducing recombination from some high rate to that rate which gives the minimum population mean fitness at equilibrium would therefore presumably spread, resulting in a decrease of mean fitness. We have also found cases where mean fitness decreased during the spread of a modifier, in this case one that increases recombination; this occurred with certain asymmetric fitness matrices that gave the maximum mean fitness when there was no recombination.

This work was stimulated by our reading L. R. HOLDEN's paper before its publication. We thank him for his comments on the first draft of this paper.

LITERATURE CITED

CHARLESWORTH, B. and D. CHARLESWORTH, 1973 Selection of new inversions in multi-locus genetic systems. Genet. Res. **21**: 167–183.

CHARLESWORTH, D., B. CHARLESWORTH and C. STROBECK, 1977 Effects of selfing on selection for recombination. Genetics **86**: 213–226.

HOLDEN, L. R., 1979 New properties of the two-locus partial selfing model with selection. Genetics **93**: 217–236.

KARLIN, S. and D. CARMELLI, 1975 Numerical studies on two-loci selection models with general viabilities. Theoret. Pop. Biol. **7**: 399–421.

LANGLEY, C. H. and J. F. CROW, 1974 The direction of linkage disequilibrium. Genetics **78**: 937–941.

Corresponding editor: B. S. WEIR

APPENDIX

We shall use the notation of HOLDEN (1979) throughout this APPENDIX. Let the frequencies of the gametes AB, Ab, aB and ab in a given generation be x_1, x_2, x_3 and x_4, respectively, at the stage before selection has taken place. Let the frequency before selection of the genotype formed from gametes i and j be g_{ij}. If w_{ij} is the fitness of this genotype and W is the population mean fitness, $\Sigma g_{ij}w_{ij}$, the frequency after selection but before recombination is $g'_{ij} = g_{ij}w_{ij}/W$. The corresponding frequency of gamete i is $x'_i = g'_{ii} + \frac{1}{2}\sum_{j\neq i} g'_{ij}$.

An inversion in gamete type 1 is introduced into the population. It can be treated as an additional gamete whose frequency is x_5 and has the same fitness effects as gamete 1 (*i.e.*, $w_{1i} = w_{5i}$) but fails to recombine. Due to the assumption of self fertilization, it is necessary to consider the frequencies of all five possible genotypes that contain the inversion, g_{5i} $(i = 1, \ldots, 4)$. It is convenient to use these as the components of a 5-dimensional vector, $\mathbf{y}$, which describes the state of the population with respect to the inversion.

In order to simplify the calculations, we assume that the inversion is introduced at a low frequency into a population at a stable equilibrium with zero linkage disequilibrium and with fitnesses of the symmetrical form assumed by HOLDEN (1979). In such a population, the gamete frequencies x_i $(i = 1, \ldots, 4)$ are all equal to $\frac{1}{4}$, and the mean fitness takes the value $\hat{W}$, given by the cubic equation,

$$(2\hat{W}-sb)\{2(\hat{W}-a)[2\hat{W}-s(1-2z)] - (1-a)(1-s)\hat{W}\} - 2(b-a)(1-s)\hat{W}[2\hat{W}-s(1-4z)] = 0\,, \tag{A.1}$$

where $z = R(1-R)$ (HOLDEN 1979).

If second-order terms in the inversion frequencies are neglected, we obtain the following matrix equation for $\tilde{\mathbf{y}}$, the value of $\mathbf{y}$ in the next generation

$$\tilde{\mathbf{y}} = \mathbf{A}\mathbf{y}\,, \tag{A.2a}$$

where

$$\mathbf{A} = \begin{pmatrix} \frac{(1+s)a}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1-s)}{4\hat{W}} & \frac{(1-s)a}{2\hat{W}} \\ \frac{(1-s)a}{4\hat{W}} & \frac{(1+s)b}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1-s)}{4\hat{W}} & \frac{(1-s)a}{2\hat{W}} \\ \frac{(1-s)a}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1+s)b}{4\hat{W}} & \frac{(1-s)}{4\hat{W}} & \frac{(1-s)a}{2\hat{W}} \\ \frac{(1-s)a}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1-s)b}{4\hat{W}} & \frac{(1+s)}{4\hat{W}} & \frac{(1-s)a}{2\hat{W}} \\ \frac{sa}{4\hat{W}} & \frac{sb}{4\hat{W}} & \frac{sb}{4\hat{W}} & \frac{s}{4\hat{W}} & \frac{sa}{\hat{W}} \end{pmatrix} \tag{A.2b}$$

If λ_0 is the leading eigenvalue of **A**, the asymptotic rate of increase in the frequency of the inversion while still rare is given by the equation

$$\Delta x_5 = (x_5 - x_5) = x_5(\lambda_0 - 1)\,. \tag{A.3}$$

The Perron-Frobenius theorem applied to **A** proves that λ_0 is real, positive and unique. Unfortunately, Equation (A.1) can be solved explicitly only when $s = 0$, when $\hat{W} = (1+a+2b)/4$. It is therefore impossible to obtain λ_0 directly when $s > 0$. We can, however, apply the following method; if we know λ_0 and $\delta\lambda_0/\delta s$ at $s = 0$, we can approximate λ_0 for small s by

$$\lambda_0(s) \approx \lambda_0(0) + s\left(\frac{\delta\lambda_0}{\delta_0}\right)_{s=0} \tag{A.4}$$

It is not difficult to see that with $s = 0$, **A** has one eigenvalue of unity, and four of zero. (This corresponds with the fact that there is no selection for or against a rare inversion in an equilibrium random-mating population with zero linkage disequilibrium [CHARLESWORTH and CHARLESWORTH 1973].) For small s, λ_0 is therefore generated from this eigenvalue of unity; we can apply the rule for differentiating an implicit function to the characteristic equation for A, and obtain the relation

$$\left(\frac{\delta\lambda_0}{\delta s}\right)_{s=0} = -\frac{(\delta|\mathbf{A} - \lambda_0\mathbf{I}|/\delta s)\ s=0}{(\delta|\mathbf{A} - \lambda\mathbf{I}|/\delta\lambda)\ s=0, \lambda=1} \tag{A.5}$$

where **I** is the 5×5 unit matrix.

After considerable algebra, we obtain

$$\left(\frac{\delta\lambda_0}{\delta s}\right)_{s=0} = \frac{4R(1-R)(1+a-2b)}{(1+a+2b)^2} \tag{A.6}$$

This can be substituted into Equations (A.3) and (A.4) to yield Equation (1) of the text. From the symmetry of HOLDEN's (1979) fitness model, this result does not depend on the gamete into which the inversion is introduced.