A NOTE ON THE FIXATION OF A FAVOURED GENE

W. J. EWENS
La Trobe University, Victoria

1. INTRODUCTION

IN this note we consider the fixation behaviour of a favoured gene in the case where the locus " A " at which it occurs is linked to a second locus " B " at which the heterozygote is at a selective advantage. Two cases are considered, firstly that where fitnesses are " additive ", and secondly that where they are " multiplicative ". It might be thought that if the linkage between the two loci is sufficiently close, the population will evolve under natural selection to a state where, at equilibrium, all four possible gametes formed by the gene at the two loci have positive frequency, so that the favoured gene does not become fixed. It is shown in this note that such behaviour cannot occur so long as the recombination fraction between the loci has any positive value; in other words, fixation of the favoured gene must eventually take place. It is not known to what extent the same result holds when fitnesses are neither " additive " nor " multiplicative ".

2. ADDITIVE FITNESSES

We suppose that alleles A_1 and A_2 can occur at a locus " A " and that alleles B_1 and B_2 can occur at a locus " B ". It is assumed that coupling and repulsion double heterozygotes are equally fit and that the fitnesses of the nine possible genotypes are given by the following array:

$$
\begin{array}{cccc}
 & A_1A_1 & A_1A_2 & A_2A_2 \\
B_1B_1 & 1+v+u & 1+s+u & 1+u \\
B_1B_2 & 1+v+t & 1+s+t & 1+t \\
B_2B_2 & 1+v & 1+s & 1
\end{array} \qquad (2.1)
$$

With these values, we would conventionally assume the following values for fitness at the loci separately:

$$
\begin{array}{ccc}
A_1A_1 & A_1A_2 & A_2A_2 \\
1+v & 1+s & 1 \\
B_1B_1 & B_1B_2 & B_2B_2 \\
1+u & 1+t & 1
\end{array}
$$

If, as we assume, $v > s > 0$, then we ascribe a selective advantage to A_1 over A_2, and if further we assume $t > u > 0$, we can say that the heterozygote is selectively favoured at the " B " locus. In such a situation there are only two sorts of equilibrium points possible, firstly where the frequency of A_2 is zero, and secondly where because of tight linkage between " A " and " B " loci, all four genes have positive frequency at equilibrium.

To examine these possibilities quantitatively we suppose that the recombination fraction between " A " and " B " loci is R, and consider initially the case where $R = 0$ If $t+s > u+v$, it is found that at equilibrium

the frequencies c_1, c_2, c_3 and c_4 of the gametes A_1B_1, A_2B_1, A_1B_2 and A_2B_2 are

$$c_1 = (t+s)/(2t+2s-u-v), \quad c_2 = c_3 = 0,$$
$$c_4 = (t+s-u-v)/(2t+2s-u-v), \qquad (2.2)$$

so that the favoured gene A_1 does not become fixed. It might therefore be suspected that, if R is positive but very small, there is an equilibrium point for the gamete frequencies close to (2.2), so that in this case too fixation of the favoured gene would not occur. This view is reinforced by the following argument, due originally to Bodmer and Parsons (1962). Suppose that the frequency of $A_1A_1B_1B_1$ is close to 1, so that the frequency of A_1B_1 is also close to 1 and the frequencies of the other gametes are small. If c'_1, c'_2, c'_3 and c'_4 are the frequencies of the gametes A_1B_1, A_2B_1, A_1B_2, A_2B_2 going to make up the following generation then the recurrence relations for gamete frequencies (Lewontin and Kojima, 1960) show that, in particular,

$$Wc'_4 = (1+s+t)c_1c_4+(1+t)c_2c_4+(1+s)c_3c_4+c_4^2 \qquad (2.3)$$
$$+R(1+s+t)(c_2c_3-c_1c_4)$$

Here W is the mean fitness of the population and may be taken as $1+v+u$. Putting $c_1 \approx 1$ we get

$$c'_4 = (1-R)(1+s+t)(1+v+u)^{-1}c_4 \qquad (2.4)$$

Thus if

$$R < (s+t-v-u)/(1+s+t) \qquad (2.5)$$

the frequency of A_2B_2 will increase; in other words, if linkage is sufficiently tight, and if $s+t>u+v$, there is a tendency away from fixation of the favoured gene.

Both the above analyses suggest the view that with sufficiently tight linkage and with sufficient selective advantage for the heterozygote at the " B " locus, the favoured gene A_1 will not become fixed. This view is now shown not to be correct.

To do this, we suppose that R is small and that c_1, c_2, c_3 and c_4 differ by very small amounts from their equilibrium values (2.2). Specifically, we put

$$c_1 = (t+s)/(2t+2s-u-v)+\delta_1, \quad c_2 = \delta_2, \quad c_3 = \delta_3,$$
$$c_4 = (t+s-u-v)/(2t+2s-u-v)+\delta_4,$$

where the δ_i are small. Ignoring terms of order δ_i^2 we find that the frequency δ'_3 of the gamete A_1B_2 for the following generation is given by the equation $[1+\{(s+t)^2/(2t+2s-u-v)\}]\delta'_3$

$$= [1+\{(t^2+s^2+2ts+tv-su)/(2t+2s-u-v)\}]\delta_3$$
$$+ R(1+s+t)(t+s)(t+s-u-v)/(2t+2s-u-v)^2. \qquad (2.6)$$

If there were a (small) equilibrium value for the frequency of A_2B_1, it would be found by putting $\delta'_3 = \delta_3$ and solving the resulting equation for δ_3. This procedure gives a negative equilibrium frequency, implying that no such equilibrium exists. More directly, (2.4) shows that if $\delta_3 = 0$ and $R > 0$, then $\delta'_3 > 0$, and further that $\delta''_3 > \delta'_3$, etc. Thus the frequency of A_1B_2 monotonically increases and A_1 eventually becomes fixed.

The following numerical example concerns the case where initially $c_1 \approx 1$. Here an initial increase in the frequency of A_2B_2 is noted (as predicted

above).   After a time, however, the frequency of A_2B_2 begins to decrease and eventually A_1 reaches fixation.

TABLE 1

Values of the frequencies of the games A_1, B_1, A_2B_1, A_1, B_2, A_2, B_2 *in various generations for a process with fitness values (2.1) with* s = 0·1, t = 0·3, u = 0, v = 0·2, R = 0·02

Generation	A_1A_1	A_2B_1	A_1B_2	A_2B_2	A_1
0	0·9500	0·0200	0·0200	0·0100	0·9700
2	0·9446	0·0147	0·0298	0·0109	0·9744
4	0·9340	0·0109	0·0436	0·0116	0·9775
6	0·9175	0·0081	0·0623	0·0121	0·9798
8	0·8949	0·0061	0·0867	0·0124	0·9815
10	0·8662	0·0047	0·1169	0·0122	0·9831
20	0·6840	0·0014	0·3077	0·0068	0·9917
30	0·5681	0·0005	0·4294	0·0020	0·9975
100	0·500	0·0000	0·5000	0·0000	1·0000

It is of some interest to note that in the above process, despite the fact that the frequency of A_1B_1 steadily decreases, that of A_1 steadily increases. While this is expected intuitively, it should be noted that, with a fitness matrix of the form (2.1), it *is* possible for the frequency of A_1 to decrease, at least for a small number of generations.   To see this, we note that the recurrence relations defining c_1', c_2', c_3' and c_4' (of which (2.3) is a particular example) show that

$$W(c_1'+c' -c_1 -c_3) = s\{1 -2(c_1+c_3)\}(c_1+c_3)(c_2+c_4) +v(c_1+c_3)^2(c_2+c_4)$$
$$+ t\{(c_1+c_2)(c_3+c_4)(1 -2(c_1+c_3)) +(c_1c_3 -c_2c_4)\}$$
$$+ u(c_1+c_2)(c_1c_4 -c_2c_3). \qquad (2.7)$$

From this it follows that $c_1'+c_3' < c_1 +c_3$ provided c_1 and t are sufficiently large and c_1c_3 is small.   Thus if we put $c_1 = 0\cdot8$, $c_2 = c_4 = 0\cdot1$, $c_3 = 0$, $s = 0\cdot1$, $t = 0\cdot4$, $u = 0\cdot1$, $v = 0\cdot15$, $R = 0\cdot5$, the following gamete frequencies are found:

Generation	A_1B_1	A_2B_1	A_1B_2	A_2B_2	A_1
0	0·8000	0·1000	0·0000	0·1000	0·8000
1	0·7463	0·1413	0·0468	0·0656	0·7931
2	0·7144	0·1580	0·0791	0·0485	0·7935
3	0·6939	0·1617	0·1040	0·0407	0·7976

For values of R smaller than 0·5, the phenomenon of an initial decrease in the frequency of A_1 still occurs; indeed the effect of tighter linkage is to delay the time until the frequency of A_1 starts to increase.

3. MULTIPLICATIVE FITNESSES

When the matrix of fitness values assumes the form

$$\begin{matrix} (1+u)(1+v) & (1+u)(1+s) & 1+u \\ (1+t)(1+v) & (1+t)(1+s) & 1+t \\ 1+v & 1+s & 1 \end{matrix} \qquad (3.1)$$

it is found that an analysis parallel to that of the previous section applies. In particular, it is found again that there is no stable " internal " equilibrium

point: A_1 must again reach fixation. Once more, it is possible that the frequency of A_1 does not increase monotonically; thus starting at the frequencies given at the end of the previous section, the frequency of A_1 decreases from $0 \cdot 8000$ to $0 \cdot 79285$, after one generation, to increase again in subsequent generations. Since no further points of principle emerge from the case of multiplicative fitnesses, we will consider them no further here.

4. RATE OF CHANGE OF GENE FREQUENCY

A further question concerns the effect of the " B " locus on the rate of change of frequency of the favoured gene A_1. Equation (2.7) exhibits this effect: when $t = u = 0$ (*i.e.* no selective differences at the " B " locus) this equation reduces to the classical one for gene frequency changes. While it has been shown above that non-zero values for t and u can have unexpected effects for a small number of generations, it appears difficult to make general statements about the effect of the numerical values of t and u on the rate of change of the frequency of A_1. One effect that is apparent is that if $u = 0$ and initially $c_1 = c_3$, $c_2 = c_4$, then these equalities hold throughout the process and the coefficient of t in (2.7) is always zero. In other words, the " B " locus has no effect in this case on the rate of increase of the frequency of A_1. If we still suppose $u = 0$ but no longer require $c_1 = c_3$, $c_2 = c_4$, then the coefficient of t in (2.7) is positive except when the frequency of B_1 is close to $\frac{1}{2}$: specifically, if we write $\alpha = c_1/(c_1+c_3)$, $\beta = c_3/(c_1+c_3)$, then the coefficient of t in (2.7) is positive unless

$$
\left.
\begin{aligned}
\alpha < \text{frequency of } B_1 < \tfrac{1}{2} & \qquad \text{if } \alpha < \beta \\
\tfrac{1}{2} < \text{frequency of } B_1 < \alpha & \qquad \text{if } \alpha > \beta
\end{aligned}
\right\} \tag{4.1}
$$

The inequalities (4.1) indicate the range of values for which the rate of increase of A_1 is slowed down because of the heterozygote selective advantage at the B locus.

When $u > 0$ it is rather more difficult to make any general statement, and particular cases should be treated on their own merits.

5. DISCUSSION

The above considerations have some bearing on recent papers on the behaviour of rare genes. Thus Ewens (1967) in considering the probability of fixation of a new mutant concluded that in some cases (typified by (2.5)) a new mutant had positive probability of fixation only if occurring at a locus sufficiently closely linked to a second interacting locus. With the above analysis in mind, it is clear that this result should be interpreted to mean that the survival probabilities so calculated refer only to the probability of surviving the comparatively small number of early generations where chance events are important. That the new mutant should survive these early generations, only to be lost subsequently as a result of natural selection, is a contingency not covered by the analysis of Ewens (1967).

A parallel comment applies to the papers of Parsons on Bodmer (1962) on the condition that a rare gene can increase in frequency. Here the contingency that the rare gene should subsequently be lost through natural selection is again not covered.

6. SUMMARY

1. It is shown that in both the " additive " and " multiplicative " case it is impossible that linkage should have the effect that a favoured gene at one locus should not reach fixation because this locus is closely linked to a second locus where the heterozygote is at a selective advantage.

2. However, it is possible for linkage to have the effect that, for a small number of generations, the frequency of such a favoured gene will decrease under natural selection.

7. REFERENCES

BODMER, W. F., AND PARSONS, P. A. 1962. Linkage and recombination in evolution. *Advan. Genet., 11,* 1-100.

EWENS, W. J. 1967. The probability of fixation of a mutant: the two-locus case. *Evolution, 21,* 532-540.

LEWONTIN, R. C., AND KOJIMA, K. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution, 14,* 458-472.