THE EFFECTS OF HITCHHIKING ON A GENE FOR RECOMBINATION

C. STROBECK1, J. MAYNARD SMITH AND B. CHARLESWORTH

School of Biological Sciences, University of Sussex, Falmer, Brighton, Sussex, England

Manuscript received April 8, 1975
Revised copy received October 8, 1975

ABSTRACT

This paper proposes that alleles increasing recombination rates may be selected for as a result of the perturbing effects of the spread of selectively favored alleles on neighboring loci maintained polymorphic by selection. The recombination genes are favored since their presence increases the production of selectively advantageous types of gametes with which they tend to remain associated. Numerical examples are presented, and some consequences of this model discussed. One such consequence is the widespread existence of polymorphism for genes affecting recombination values.

Numerous theoretical studies of the population genetics of multi-locus systems at equilibrium under the effects of recombination and selection have led to the conclusion (Fišer 1930; Nei 1967, 1969; Turner 1967; Lewontin 1971; Feldman 1972; Karlin and McGregor 1974) that there is a selection pressure in favor of reducing recombination rates between interacting, polymorphic genes to zero. Phenomena such as inversion polymorphisms (Dobzhansky 1951) are thought to reflect this selection pressure. It is not easy to think of situations in which selection will favor alleles that increase recombination rates.

This paper describes a process which leads to an increase in the frequency of an allele for higher recombination in a random-mating population in a uniform environment. The process is analogous to that by which a “mutator” gene increases in frequency because of the hitchhiking effect of a favorable mutation caused by it (Cox and Gibson 1974). In a similar way, a recombination modifier can increase in frequency because of the hitchhiking effect of a favorable recombinant.

AN EXAMPLE

Suppose that there are three linked loci. At the first locus, two alleles A and a are maintained in the population by the superior fitness of the heterozygote Aa. The population is initially fixed for an allele b at the second locus. A favorable mutant B occurs at this locus and spreads through the population. At the third locus there is a pair of alleles, c for lower recombination between A and B, and

1 Present address: Division of Biological Sciences, State University of New York, Stony Brook, New York 11790. Reprint requests should be mailed to this address.

C for higher recombination between A and B (note that upper and lower case letters do not necessarily indicate dominance). As the mutant B spreads through the population, it will alter the allele frequencies at the other loci.

Consider first a numerical example. The selective values of AA, Aa and aa are as 0.5, 1, and 0.5 respectively; the selective values of BB, Bb and bb are as 1.2, 1 and 0.8. Fitness interactions between loci are assumed to be multiplicative. The recombination values for A and B are $r_{cc} = 0.01$ (CC homozygotes), $r_{cc} = 0$ (Cc heterozygotes) and $r_{cc} = 0$ (cc homozygotes). The recombination value for B and C is 0.01.

The gene order is ABC; there is no interference. Thus it has been assumed that the recombination allele is recessive, and affects only region AB. The low recombination allele c permits no recombination in region AB.

It is assumed that the A locus is initially at equilibrium, with allelic frequencies of 0.5. The initial frequency of C is 0.2, and there is linkage equilibrium between the A and C loci. The favorable mutant B is introduced into the population with a frequency of 0.0005 corresponding to a single mutant in one individual in a population of 1000. The favorable mutation is assumed to be a unique event, and to occur in coupling with allele a. We have to consider separately the cases in which B occurs in coupling with c and with C.

Subsequent changes in the genetic make-up of the population have been followed by computer calculation, using the deterministic equations for a randomly mating population given by Feldman (1972). The results are summarized in Figure 1. If B occurs in coupling with the zero-recombination allele c, the recombination allele C initially falls in frequency because of hitchhiking. In time the population comes to contain mainly the gametes aBc, aBC, Abc and AbC. The allele B cannot increase further in frequency until an AB gamete arises by recombination. This occurs in a CC homozygote, so that the new favorable gamete is ABC. Once this gamete has occurred, B resumes its increase in fre-

![Figure 1](image-url)

**Figure 1.**—Curve (a) shows the trajectories of the gene frequencies at the three loci when B is initially in coupling with c; curve (b) shows the case when B is in coupling with C. Selection and recombination parameters are given in the text.
quency, and carries allele $C$ up with it. Thus $C$ experiences two contrary hitchhiking affects: an initial drop caused by the favorable mutation, and a subsequent rise caused by the favorable recombination. The final frequency of $C$ is 0.383.

When $B$ occurs in coupling with the recombination allele $C$, both hitchhiking effects raise the frequency of $C$ to a final frequency of 0.787.

If the initial frequency of $C$ is 0.2, then 20% of favorable mutations will occur in coupling with it, and 80% in repulsion. Hence the expected frequency of $C$ after the fixation of $B$ is

$$0.2 \times 0.787 + 0.8 \times 0.383 = 0.464.$$  

Thus, on average, with these particular fitness and recombination values, the allele for recombination increases in frequency from 20% to 46.4%.

**VARIATION IN FITNESS AND RECOMBINATION VALUES**

We have carried out rather extensive investigations of the effect of changing the recombination values, fitnesses, dominance relationships and initial frequencies on the magnitude of the gene frequency change.

Some results are given in Tables 1 and 2, and Figures 2–4. The further notation introduced is as follows. The expected change in the frequency of $C$, the allele for increased recombination, is represented by $\Delta P$. (This is the difference between the expected final frequency of $C$, calculated by the method described above, and the initial frequency.) The fitnesses at the heterotic locus are $(1 - t)$: $1 : (1 - t)$ for $AA$, $Aa$ and $aa$ respectively. The fitnesses at the substituting locus are $(1 + s)$: $1 : (1 - s)$ for $BB$, $Bb$ and $bb$ respectively. The recombination values between $A$ and $B$ are represented by $r_{cc}$, $r_{cc}$ and $r_{cc}$ for $CC$, $Cc$ and $cc$, respectively. The recombination value for loci $B$ and $C$ is $R$.

The picture which emerges is not a simple one. Some conclusions can, however, be drawn.

1) $\Delta P$ is always positive, i.e. there is always selection for increased recombination on average. However, it should be noted that $\Delta P$ becomes close to zero for low initial frequencies of $C$ (see Figure 2). This implies that this mechanism can only raise the frequency of a mutant allele for increased recombination to a high level as a result of successive gene substitutions taking place in the region of the heterotic locus, each of which hitches up the $C$ allele a little further.

This selection pressure in favor of increased recombination is not merely a reflection of hitchhiking effects on $C$ of the replacement of $b$ by $B$; it depends on the existence of a stable equilibrium at the $A$ locus. This is proved analytically in Appendix A, where it is shown that $\Delta P$ is zero in the absence of selection at the $A$ locus.

Appendix B gives an analytic proof that this mechanism is capable of increasing recombination from zero by the introduction of a $C$ allele, causing a low level of recombination in $Cc$ heterozygotes, into a $cc$ population with zero recombination. The reinforces the conclusion that $\Delta P$ is always positive, derived from our numerical results.

2) There is no simple relation between $\Delta P$ and the measures of selection
C. STROBECK, J. MAYNARD SMITH AND B. CHARLESWORTH

550

In Figure 2—This shows the expected change in frequency of C (the allele increasing recombination) as a function of its initial frequency. The selection parameters are as in the text; $r_{cc} = 0.01$, $r_{CC} = r_{cc}$ and the recombination value for B and C is 0.01.

intensity at the A and B loci, s and t. Tables 1 and 2 show that the effect on $\Delta P$ of increasing one of these variables can be positive or negative, depending on the values of the other variable and the recombination parameters. This conclusion is reinforced by the results of Appendix B, where it is shown that the intensity of selection for a modifier inducing non-zero recombination is not an increasing function of t, except when t is small.

3) $\Delta P$ is small unless there is reasonably close linkage between the three loci in the absence of allele C. This is particularly so for linkage between A and B; $\Delta P$ is negligibly small unless $r_{cc} \leq 0.01$ (Table 1). Linkage between B and C is less critical; $\Delta P$ may still be appreciable when $R = 0.1$, although it is near zero for $R = 0.5$ (Table 2).

4) The effect on $\Delta P$ of dominance of the C locus is illustrated in Figure 3. If $r_{cc}$ is small, then $\Delta P$ is larger when C is recessive; if $r_{cc} > 0.001$, then $\Delta P$ is larger when C is dominant.

5) It is not necessary for the polymorphism at the A locus to be maintained by heterozygote advantage. The same type of effect is obtained if there is frequency-dependent selection, or if the polymorphism is maintained as a result of different genotype fitnesses in different niches (Levene 1953). The latter point is illustrated in Figure 4.

Dr. D. Charlesworth has pointed out to us that a similar selection pressure in favor of increased recombination may be expected to operate as a result of deleterious genes maintained by mutation pressure interfering with the substitution of favorable alleles at other loci. If, for example, a favorable mutant occurs in a chromosome containing a recessive lethal mutation, then in the absence of recombination the fixation of the favored allele will be prevented, and a two-locus equilibrium with two gamete types (the wild-type allele of the lethal in coupling with the disfavored allele at the other locus, and the lethal in coupling with the favored allele) will be set up. There will thus be selection in favor of a modifier...
TABLE 1

Expected changes ($\Delta P$) in the frequency of the allele C, for various values of s and t

<table>
<thead>
<tr>
<th>$r_{cc} = .000$</th>
<th>t</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>.5</td>
<td>2089</td>
<td>2850</td>
<td>3143</td>
<td>3275</td>
<td>3342</td>
<td>3379</td>
<td>3402</td>
<td>3416</td>
<td>3424</td>
<td>3429</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>2014</td>
<td>2828</td>
<td>3159</td>
<td>3317</td>
<td>3398</td>
<td>3444</td>
<td>3471</td>
<td>3487</td>
<td>3497</td>
<td>3502</td>
</tr>
<tr>
<td></td>
<td>.7</td>
<td>1938</td>
<td>2794</td>
<td>3156</td>
<td>3338</td>
<td>3434</td>
<td>3487</td>
<td>3519</td>
<td>3539</td>
<td>3550</td>
<td>3555</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>1856</td>
<td>2751</td>
<td>3140</td>
<td>3444</td>
<td>3455</td>
<td>3517</td>
<td>3554</td>
<td>3576</td>
<td>3589</td>
<td>3595</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>1760</td>
<td>2694</td>
<td>3111</td>
<td>3336</td>
<td>3464</td>
<td>3536</td>
<td>3579</td>
<td>3605</td>
<td>3620</td>
<td>3626</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r_{cc} = .001$</th>
<th>t</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>.5</td>
<td>0097</td>
<td>0501</td>
<td>0540</td>
<td>0557</td>
<td>0569</td>
<td>0572</td>
<td>0576</td>
<td>0580</td>
<td>0584</td>
<td>0589</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>0017</td>
<td>0407</td>
<td>0442</td>
<td>0460</td>
<td>0471</td>
<td>0478</td>
<td>0484</td>
<td>0490</td>
<td>0497</td>
<td>0505</td>
</tr>
<tr>
<td></td>
<td>.7</td>
<td>0025</td>
<td>0327</td>
<td>0358</td>
<td>0375</td>
<td>0386</td>
<td>0394</td>
<td>0401</td>
<td>0409</td>
<td>0418</td>
<td>0430</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>0019</td>
<td>0257</td>
<td>0283</td>
<td>0298</td>
<td>0309</td>
<td>0317</td>
<td>0326</td>
<td>0335</td>
<td>0346</td>
<td>0360</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>0014</td>
<td>0190</td>
<td>0211</td>
<td>0224</td>
<td>0234</td>
<td>0243</td>
<td>0252</td>
<td>0267</td>
<td>0275</td>
<td>0292</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r_{cc} = .005$</th>
<th>t</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>.5</td>
<td>0100</td>
<td>0124</td>
<td>0133</td>
<td>0137</td>
<td>0139</td>
<td>0140</td>
<td>0140</td>
<td>0141</td>
<td>0141</td>
<td>0141</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>0079</td>
<td>0100</td>
<td>0108</td>
<td>0112</td>
<td>0115</td>
<td>0117</td>
<td>0118</td>
<td>0119</td>
<td>0121</td>
<td>0122</td>
</tr>
<tr>
<td></td>
<td>.7</td>
<td>0062</td>
<td>0080</td>
<td>0087</td>
<td>0091</td>
<td>0094</td>
<td>0097</td>
<td>0099</td>
<td>0101</td>
<td>0103</td>
<td>0106</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>0048</td>
<td>0063</td>
<td>0069</td>
<td>0073</td>
<td>0076</td>
<td>0078</td>
<td>0081</td>
<td>0083</td>
<td>0086</td>
<td>0090</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>0035</td>
<td>0047</td>
<td>0052</td>
<td>0055</td>
<td>0058</td>
<td>0061</td>
<td>0064</td>
<td>0067</td>
<td>0070</td>
<td>0075</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r_{cc} = .010$</th>
<th>t</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>.5</td>
<td>0049</td>
<td>0062</td>
<td>0066</td>
<td>0069</td>
<td>0069</td>
<td>0070</td>
<td>0069</td>
<td>0069</td>
<td>0068</td>
<td>0067</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>0040</td>
<td>0051</td>
<td>0055</td>
<td>0057</td>
<td>0059</td>
<td>0059</td>
<td>0060</td>
<td>0060</td>
<td>0060</td>
<td>0061</td>
</tr>
<tr>
<td></td>
<td>.7</td>
<td>0032</td>
<td>0041</td>
<td>0045</td>
<td>0048</td>
<td>0049</td>
<td>0050</td>
<td>0051</td>
<td>0052</td>
<td>0053</td>
<td>0054</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>0025</td>
<td>0033</td>
<td>0036</td>
<td>0039</td>
<td>0040</td>
<td>0041</td>
<td>0043</td>
<td>0044</td>
<td>0045</td>
<td>0047</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>0018</td>
<td>0025</td>
<td>0028</td>
<td>0030</td>
<td>0031</td>
<td>0033</td>
<td>0034</td>
<td>0036</td>
<td>0038</td>
<td>0040</td>
</tr>
</tbody>
</table>

The initial frequency of C is 0.2; $R = 0.31$, $r_{cc} = 0.1$; the values of $r_{cc} = r_{cc}$ are given in the table.
TABLE 2

Expected changes ($\Delta P$) in the frequency of allele C for various values of $s$ and of the recombination value, $R$, between B and C

<table>
<thead>
<tr>
<th>$r_{cc}$ = .000</th>
<th>$R$</th>
<th>.001</th>
<th>.005</th>
<th>.010</th>
<th>.050</th>
<th>.100</th>
<th>.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.3669</td>
<td>.2794</td>
<td>.2079</td>
<td>.0362</td>
<td>.0097</td>
<td>.0003</td>
<td></td>
</tr>
<tr>
<td>.2</td>
<td>.3819</td>
<td>.3297</td>
<td>.2804</td>
<td>.0978</td>
<td>.0350</td>
<td>.0111</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>.3870</td>
<td>.3482</td>
<td>.3098</td>
<td>.1429</td>
<td>.0637</td>
<td>.0021</td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>.3895</td>
<td>.3575</td>
<td>.3251</td>
<td>.1731</td>
<td>.0886</td>
<td>.0034</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.3909</td>
<td>.3629</td>
<td>.3342</td>
<td>.1933</td>
<td>.1080</td>
<td>.0047</td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>.3918</td>
<td>.3662</td>
<td>.3398</td>
<td>.2068</td>
<td>.1221</td>
<td>.0062</td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>.3923</td>
<td>.3683</td>
<td>.3434</td>
<td>.2155</td>
<td>.1318</td>
<td>.0076</td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>.3927</td>
<td>.3696</td>
<td>.3455</td>
<td>.2205</td>
<td>.1375</td>
<td>.0090</td>
<td></td>
</tr>
<tr>
<td>.9</td>
<td>.3929</td>
<td>.3702</td>
<td>.3464</td>
<td>.2221</td>
<td>.1392</td>
<td>.0101</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$r_{cc}$ = .001</th>
<th>$R$</th>
<th>.001</th>
<th>.0733</th>
<th>.0830</th>
<th>.0830</th>
<th>.0830</th>
<th>.0830</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.0255</td>
<td>.0733</td>
<td>.0830</td>
<td>.0830</td>
<td>.0830</td>
<td>.0830</td>
<td>.0830</td>
</tr>
<tr>
<td>.2</td>
<td>.0180</td>
<td>.0646</td>
<td>.0903</td>
<td>.0643</td>
<td>.0263</td>
<td>.0009</td>
<td>.0009</td>
</tr>
<tr>
<td>.3</td>
<td>.0131</td>
<td>.0519</td>
<td>.0798</td>
<td>.0872</td>
<td>.0467</td>
<td>.0018</td>
<td>.0018</td>
</tr>
<tr>
<td>.4</td>
<td>.0100</td>
<td>.0416</td>
<td>.0677</td>
<td>.0964</td>
<td>.0622</td>
<td>.0029</td>
<td>.0029</td>
</tr>
<tr>
<td>.5</td>
<td>.0078</td>
<td>.0335</td>
<td>.0567</td>
<td>.0973</td>
<td>.0718</td>
<td>.0041</td>
<td>.0041</td>
</tr>
<tr>
<td>.6</td>
<td>.0061</td>
<td>.0271</td>
<td>.0471</td>
<td>.0931</td>
<td>.0763</td>
<td>.0054</td>
<td>.0054</td>
</tr>
<tr>
<td>.7</td>
<td>.0048</td>
<td>.0217</td>
<td>.0386</td>
<td>.0856</td>
<td>.0763</td>
<td>.0067</td>
<td>.0067</td>
</tr>
<tr>
<td>.8</td>
<td>.0037</td>
<td>.0171</td>
<td>.0309</td>
<td>.0755</td>
<td>.0723</td>
<td>.0080</td>
<td>.0080</td>
</tr>
<tr>
<td>.9</td>
<td>.0027</td>
<td>.0127</td>
<td>.0234</td>
<td>.0623</td>
<td>.0643</td>
<td>.0089</td>
<td>.0089</td>
</tr>
</tbody>
</table>

The initial frequency of C is 0.2; $t = 0.5$; $r_{cc} = 0.1$; values of $r_{cc} = r_{cc}$ are given in the table.

Figure 3.—The effect of dominance on $\Delta P$. The initial frequency of C is 0.2; $s = 0.2$; $t = 0.5$; $R = 0.01$; $r_{cc} = 0.01$. Full line, $r_{cc} = r_{cc}$; broken line, $r_{cc} = r_{cc}$.
It is assumed that there are two niches of equal size; in the first niche genotypes AA, Aa and aa have fitnesses 1.5, 1.0 and 0.5 respectively, and in the second niche, 0.5, 1.0 and 1.5 respectively. It is also assumed that C is recessive, with \( r_{cc} = 0.1 \); \( R = 0.01 \); the initial frequency of \( C \) is 0.2.

which allows the production of recombinant gametes which have the nonlethal allele in coupling with the favored allele at the other locus. (This can be demonstrated analytically with the method of APPENDIX B). Since in outbreeding species such as Drosophila 30% or more of the copies of a chromosome in a population may contain a recessive lethal mutation (DOBZHANSKY 1951), this mechanism could be of significance in promoting non-zero recombination levels.

6) We have also simulated a stochastic version of the model, allowing for sampling from a finite population. This gave results which were similar to the deterministic ones presented above, but excessive amounts of computer time are required to get accurate estimates of the average value of \( \Delta P \), on account of the great variation between individual runs.

CONCLUSIONS

It is clear that the hitchhiking effect of favorable recombinants does provide a mechanism which will increase the frequency of alleles increasing recombination, and so prevent the "congealing of the genotype" (TURNER 1967). It is not clear whether this is the main process maintaining recombination. If, however, we assume that it is the main process, certain conclusions can be drawn:

i) Populations will tend to be polymorphic for recombination alleles. This is because the fixation of a favorable allele will shift the frequency of high recombination alleles upwards, but will rarely fix them, except in small populations. This upward tendency will be counterbalanced by the selective disadvantage of producing recombinants between genes with epistatic effects on fitness.

ii) A population that has recently been exposed to strong directional selection will tend to have higher total levels of recombination than a population that has been mainly under stabilizing selection. It should be noted, however, that a high
chiasma frequency would be a consequence of rapid evolution, not a pre-adaptation to it.

iii) Populations will contain recombination modifiers that are local in their effects, or generalized, but they will not contain modifiers that do not alter recombination in their own locality. Several recent studies (Chinnici 1971a, b; Kidwell 1972a, b; Abdullah and Charlesworth 1974) have demonstrated the existence of genetic variability in recombination values in Drosophila melanogaster. No definite evidence is available, however, concerning the frequencies or specificities of action of the genes concerned.

APPENDIX A

In this appendix it is shown that if the A locus is neutral, then the expected frequency of C (as defined above) is unaltered by the fixation of B.

If the A locus is neutral, the three-locus model which we have been using reduces to a two-locus model with a constant recombination fraction, R, for the two loci concerned, B and C. The genotypes BB, Bb and bb are assumed to have the fitness values $W_1$, 1 and $W_2$, respectively. The C locus, which affects the recombination value for A and B, is neutral, since A is neutral. Let the frequencies of the gametes BC, Bc, bC and bc (disregarding the A locus) be $x_1$, $x_2$, $x_3$ and $x_4$ respectively. It is convenient to use the three independent variables:

\[ p_1 = x_1 + x_2 \]  \hspace{1cm} (A.1a)
\[ p_2 = \frac{x_1}{x_1 + x_2} \]  \hspace{1cm} (A.1b)
\[ p_3 = \frac{x_3}{x_3 + x_4} \]  \hspace{1cm} (A.1c)

These give:

\[ x_1 = p_1 p_2 \]
\[ x_2 = p_1 (1-p_2) \]
\[ x_3 = (1-p_3) p_2 \]
\[ x_4 = (1-p_2)(1-p_3) \]

Using these variables, and the selection parameters defined above, the recurrence relations for the two-locus system become:

\[ p_1' = \frac{W_1 p_1^2 + p_1 (1-p_2)}{W_1 p_1^2 + 2 p_1 (1-p_1) + W_2 (1-p_1)^2} \]  \hspace{1cm} (A.2a)
\[ p_2' = p_2 - \frac{R p_1 (1-p_2) (p_2-p_3)}{W_1 p_1^2 + p_1 (1-p_1)} \]  \hspace{1cm} (A.2b)
\[ p_3' = p_3 + \frac{R p_1 (1-p_2) (p_2-p_3)}{W_2 (1-p_1)^2 + p_1 (1-p_1)} \]  \hspace{1cm} (A.2c)

It is assumed that the population is initially fixed for the b allele and that a favorable mutant, B, arises which goes to fixation (see text). Let the initial frequencies of C and c in the population be P and Q, respectively. The probability that B is initially in coupling with C is therefore P, and the probability that it is in coupling with c is Q. If F and f are the frequencies of C at an arbitrary time t when B is initially in coupling with C and c respectively, then we must show that at every such time t, the expected frequency of C, p, is equal to its initial frequency, i.e.

\[ p = PF + Qf = P \]  \hspace{1cm} (A.3)

This is proved below by induction.

Since equation (A.2a) is a function only of $p_1$ (the frequency of B), the frequency of B at
time $t$ is independent of whether $B$ is initially in coupling with $C$ or $c$. This is not true for $p_2$ and $p_3$, however. Let $\Pi_2$ and $\pi_2$ be the values of $p_2$ at time $t$, and $\Pi_3$ and $\pi_3$ be the corresponding values of $p_3$, when $B$ is initially in coupling with $C$ and $c$ respectively. If at time $t$ we have:

$$P = \Pi_2 + Q\pi_2$$

and simultaneously:

$$P = \Pi_3 + Q\pi_3$$

then from equations (A.2) we have:

$$\Pi_2' = \Pi_2 + Q\pi_2 + \frac{R p_2 (1-p_1) (\Pi_2 + Q\pi_2 - \Pi_3 - Q\pi_3)}{W_1 p_1^2 + p_1 (1-p_1)}$$

$$= \Pi_2 + Q\pi_2 = P$$

and

$$\Pi_3' = \Pi_3 + Q\pi_3 + \frac{R p_3 (1-p_1) (\Pi_2 + Q\pi_2 - \Pi_3 - Q\pi_3)}{W_2 (1-p_1)^2 + p_1 (1-p_1)}$$

$$= \Pi_3 + Q\pi_3 = P$$

At time $t = 0$, when the frequency of $B$ takes its initial value $p_0$, we have:

$$x_1 = p_0, \quad x_2 = 0, \quad x_3 = P - p_0, \quad \text{and} \quad x_4 = Q,$$

when $B$ is initially in coupling with $C$.

When $B$ is initially in coupling with $c$, we have:

$$x_1 = 0, \quad x_2 = p_0, \quad x_3 = P, \quad \text{and} \quad x_4 = Q - p_0.$$  

Hence, at time $t = 0$:

$$\Pi_2 = 1, \quad \Pi_3 = (P - p_0)/(1-p_0), \quad \pi_2 = 0, \quad \text{and} \quad \pi_3 = P/(1-p_0),$$

so that

$$\Pi_2 + Q\pi_2 = P + Q \times 0 = P$$

and

$$\Pi_3 + Q\pi_3 = P(P - p_0) + QP/(1-p_0) = P.$$  

Hence, by induction, equations (A.4) are true for all $t$. Since the frequency of the $C$ allele is equal to $p_1 p_2 + (1 - p_1) p_3$, we have for any $t$:

$$P = p_1 P + (1-p_1) P = p_1(\Pi_2 + Q\pi_2) + (1-p_1)(\Pi_3 + Q\pi_3) = PF + Qf$$

which is the desired result (equation (A.3)).

**APPENDIX B**

In this appendix we demonstrate that a population with zero recombination can, under suitable conditions, be unstable to the introduction of a genic modifier which induces recombination. The selective model involves two loci, $A$ and $B$, as described in the text. The population is assumed initially to be segregating for the heterotic locus $A$; a selectively advantageous mutation from $b$ to $B$ is then established. There is no recombination between the two loci, so that $B$ remains associated with the heterotic allele ($A$, arbitrarily) present in the gamete in which it mutated. It is easy to show, using the selection parameters defined in the text, that the gamete $AB$ will spread to fixation if $(1+s)(1-t) > 1$, and that a stable polymorphism with $AB$ and $ab$ present will be established if $(1+s)(1-t) < 1$. The latter situation is the one examined here. It simply requires the selection intensity at the $A$ locus to be about the same or greater than at the $B$ locus. The equilibrium frequencies of $AB$ and $ab$, $p$ and $q$, respectively, are given by

$$p = \frac{s + t - st}{2}, \quad q = \frac{t + st - s}{2}.$$
We may now consider the response of the system to the introduction of a modifier, C, which causes recombination between A and B, thus generating a set of gametes unrepresented in the original population. It will be assumed that the modifier is initially so rare that it is effectively represented in Cc heterozygotes only. Let the frequency in Cc individuals of single crossovers between A and B be $R_1$, the frequency of single crossovers between B and C be $R_2$, and the frequency of double crossovers be $R_3$. (The order of the loci is immaterial in this context). If we neglect the squares and cross products of the frequencies of gametes containing C (ABC, AbC, etc.) and of the C gametes generated by recombination in Cc individuals (Abc and abc), and linearize with respect to the perturbations of Abc and abc from their original equilibrium frequencies, we obtain a set of linear equations for the eight gamete frequencies in terms of their frequencies in the previous generation. The stability analysis of this set is equivalent to that for the following four independent systems of linear equations (c.f. Feldman 1972):

i) A pair of equations describing the response of the original system (Abc and abc) to a perturbation. This system has already been shown to be stable.

ii) An equation for the frequency of Abc. (This gamete is eliminated).

iii) An equation for the frequency of aBc. (This gamete increases in frequency).

iv) A set of four linear equations for the C gametes. The leading eigenvalue of the matrix $M$ for this set determines whether or not C tends to increase in frequency (Feldman 1972); if greater than 1, we can conclude that C will eventually increase; if less than 1, C will be eliminated.

Let $\tilde{w}$ be the mean fitness of the population before the introduction of C ($\tilde{w} = p (1 + s) (1 - t) + q = p + q (1 - s) (1 - t)$); let $w_1$ and $w_2$ be the marginal fitnesses of Ab and aB gametes when introduced into this population. ($w_1 = p (1 - t) + q (1 - s) < \tilde{w}$, and $w_2 = p (1 + s) + q (1 - t) > \tilde{w}$). The matrix $M$ has the form:

$$
\begin{bmatrix}
\tilde{w} - q(R_1 + R_2 + R_3) & p(R_1 + R_2)(1-t) & p(R_1 + R_2)(1+s) & pR_2 \\
qR_3 & w_1 - p(R_1 + R_2)(1-t) & 0 & pR_1 \\
-qR_1 & 0 & w_2 - p(R_1 + R_2)(1+s) & pR_3 \\
qR_2 & q(R_1 + R_2)(1-s) & q(R_2 + R_3)(1-t) & \tilde{w} - p(R_1 + R_2 + R_3)
\end{bmatrix}
$$

When there is no recombination ($R_1 = R_2 = R_3 = 0$), it is easily seen that $M$ has two eigenvalues of unity, one equal to $w_1/\tilde{w}$ ($< 1$), and a leading eigenvalue equal to $w_2/\tilde{w}$ ($> 1$). With low values of recombination induced by C, the method of small parameters (Karlin and McGregor 1972) assures us that the leading eigenvalue of $M$, $\lambda_n$, will be generated from the value of $\lambda_n$ for zero recombination, and hence will be greater than one for sufficiently small values of $R_1$, $R_2$ and $R_3$. Hence a modifier with a small effect on recombination will asymptotically increase in frequency when introduced into the population at a low initial frequency.

An approximate expression for $\lambda_n$ for non-zero recombination can be obtained as follows. Assume that $R_1$, $R_2$ and $R_3$ are functions of a single variable, $R$, which measures the effect of C on the frequency of recombination. Let the derivatives of $R_i$ with respect to $R$ at $R = 0$ be $k_i$ ($i = 1$ to 3). $R_i$ can thus be approximated by $R k_i$. An approximate expression for $\lambda_n$ for small $R$ can, therefore, be obtained by using Taylor's theorem and evaluating $d \lambda_n/dR$ at $R = 0$, by applying the rule for the differentiation of an implicit function to the characteristic equation for $M$. We obtain:

$$
\tilde{w} \lambda_n = w_2 - R[p(k_1 + k_2)(1+s) + q(k_2 + k_3)(1-t)]
$$

(B.2)

This shows that the strength of selection for a rare allele inducing recombination is a decreasing function of the size of its effect.
The effects of variation in $s$ and $t$ on the strength of selection can be understood in terms of the dependence of $\lambda_o$ on $s$ and $t$. If we neglect the terms involving $R$ in equation (B.2), as is reasonable for a modifier of small effect, we obtain:

$$\frac{\partial \lambda_o}{\partial s} = 1 + \frac{s(1-t)}{2t\bar{w}^2} (\bar{w} [3(t+s+st-s) - t] - s^2(1-t)(t+s-st-s)), \quad (B.3)$$

where $\bar{w} = (t + (1-t) [t + s^2 (1-t)]) / 2t$.

It follows that

$$\frac{\partial \lambda_o}{\partial s} > 1 - \frac{s(1-t)}{t + (1-t) [t + s^2 (1-t)]} > 0,$$

noting that $t + st - s > 0$ for the initial equilibrium without $C$ to exist.

Hence, under the conditions assumed here, the strength of selection for increasing recombination increases with $s$. This is consistent with the results for $r_{eo} = 0$ in the parts of Table 1 which satisfy $t + st - s > 0$.

We also have:

$$2t^2\bar{w}^2 \frac{\partial \lambda_o}{\partial t} = (t + (1-t) [t + s^2 (1-t)]) (s - t^2 [1+s])$$

$$+ (1-t) (t + st - s) (t^2 [1-s^2] + s^2) \quad (B.4)$$

When $t = 1$, the R.H.S. of equation (B.4) reduces to $-1$; hence for sufficiently large values of $t$, $\lambda_o$ must be a decreasing function of $t$. For small values of $t$, the R.H.S. reduces to $s (2 + s)$, so that $\lambda_o$ then increases with $t$.

**LITERATURE CITED**


**LEVENE, H.**, 1953 Genetic equilibrium when more than one ecological niche is available. Am. Naturalist 87: 331–333.


Corresponding editor: R. C. LEWONTIN