A Study of Genetic Linkage Relative to Success in Backcross Breeding Programs*

Hiroshi FUJIMAKI** and Ralph E. COMSTOCK

Department of Genetics and Cell Biology, University of Minnesota, St. Paul, Minnesota 55101, U.S.A.

The problem of genetic linkage between an allele to be transferred from a donor parent and alleles with negative effects on any quantitative trait is investigated theoretically in relation to success in backcross breeding programs. A method is outlined for detecting the effect of such linkage. The amount of data required is discussed in terms of the power of the statistical test for the linkage detection. Useful procedure when linkage is found to be present is also discussed.

Introduction

Backcross breeding often aims at the incorporation of a single allele into an otherwise-unchanged genotype. The purpose then is to alter a specific attribute; e.g., disease-resistance, shattering resistance, or grain quality, of a superior variety without changing any other aspect of performance. We will refer to the allele to be incorporated as the target allele and to its locus as the target locus. Assuming that plants heterozygous for the target allele can always be distinguished and that the target allele, when homozygous in the superior variety, will have the desired effect; the only barriers should be unfortunate pleiotropic effects or close genetic linkage between the target locus and alleles with favorable effects, on one or another quantitative trait, that are present in the recurrent but not in the donor parent.

The problem of linkage in backcross programs has been discussed by Harlan and Pope (1922), Briggs (1938), Briggs and Allard (1963), Hanson (1959), and Allard (1960). Various workers, including Sax (1923), Rasmusson (1935), Thoday (1961), Jayakar (1970), Mather and Jinks (1971) and Soller and Brody (1976), have investigated linkage between genes with visible effects and genes on quantitative traits.

This paper is devoted to the problems of learning (1) whether unfavorable effects on any quantitative trait are associated with a specific target allele derived from a specific donor parent, (2) the total magnitude of such effects, and (3) whether all or part of this total is due to pleiotropy or to linkages that for practical purposes are too tight to be broken. Sound information concerning these issues, obtained early enough, would provide the basis for constructive decisions relative to procedures in backcross breeding programs.

---

* Minnesota Agricultural Experiment Station journal series No. 9654.
** The present address of the senior author: Crop Research Division, Central Agricultural Experiment Station, Konosu, Saitama 365.

Received October 22, 1976
Genetic Model

Assumptions made throughout are

1. that both varieties, donor and recurrent, are homozygous,
2. that normal diploid meiosis occurs in the F₁ and all subsequent generations, and
3. that the target allele has a distinct phenotypic effect by which heterozygotes can be identified.

At some points absence of epistasis, in the pleiotropic effects of the target allele and the effects of alleles linked with the target locus, is implied. The possible impact of epistasis will be considered in a later section of the manuscript.

Detection of Unfavorable Effects

Standard procedure in the type of backcross breeding program being considered is as follows. The F₁ of the two varieties is backcrossed to the recurrent parent to obtain the B₁F₁ generation. If the target allele is symbolized as B and its counterpart from the recurrent parent as b, genotypes appearing in B₁F₁ will obviously be Bb and bb in frequencies expected to be equal. Heterozygous (Bb) individuals from B₁F₁, identified by their phenotype, are crossed to the recurrent (bb) parent to obtain the second backcross, the B₂F₁ generation. The process is repeated for as many generations as seem required for an adequate approach to recovery of the recurrent parent genotype except at the the target locus where in each generation the expected frequency of the Bb (and of the bb) genotype is one-half. The final step involves self-fertilization of heterozygous plants to obtain material homozygous for the target allele.

Let X represent the allele of the recurrent parent and x that of the donor parent at any locus other than the target locus and let c be the recombination frequency between the X and B genes. Considering only the B gene and any other, the genotype of backcross individuals used as parents will be either Bx/bX or BX/bX. The gametes produced by random Bb individuals, the corresponding genotypes in the following backcross generation and the expected frequencies of both are shown in Table 1 where \( q_t \) symbolizes the expected frequency of Bx/bX among the Bb genotypes of the B₁F₁ population (obtained by \( t \)-time backcrosses). It is apparent from the first two frequencies of Table 1 that

\[ q_{t+1} = q_t (1-c) \]  

(1)

Because Bx/bX is the only F₁ genotype, \( q_0 = 1 \) and it follows, therefore,
Table 2. Expected frequencies of the X-locus genotypes in the two offspring groups of the B$_4$F$_2$

<table>
<thead>
<tr>
<th>X-locus Genotypes</th>
<th>Offspring From Bb Parents</th>
<th>Offspring From bb Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>$1 - \frac{3}{4} (1-c)^t$</td>
<td>$1 - \frac{3}{4} c(1-c)^{t-1}$</td>
</tr>
<tr>
<td>Xx</td>
<td>$\frac{1}{2} (1-c)^t$</td>
<td>$\frac{1}{2} c(1-c)^{t-1}$</td>
</tr>
<tr>
<td>xx</td>
<td>$\frac{1}{4} (1-c)^t$</td>
<td>$\frac{1}{4} c(1-c)^{t-1}$</td>
</tr>
</tbody>
</table>

respectively.

It is intuitively obvious that in the B$_4$F$_2$ generation produced by self-fertilization of B$_4$F$_1$ individuals effects of alleles linked with the target allele plus any pleiotropic effect of B itself would be sources of difference between averages, of any quantitative trait, for offspring derived from Bb and bb parents. Using expected B$_4$F$_1$ frequencies provided above, the expected frequencies of X-locus genotypes are easily determined. For example, that of the XX genotype in the offspring of Bb parents is

$$\frac{1}{4} (1-c)^t + \left[1 - (1-c)^t\right] = 1 - \frac{3}{4} (1-c)^t$$

The first term on the left is obtained as a result of self-fertilization of Bx/bX and the second term from BX/bX. The complete set of frequencies is shown in Table 2. Note that for loci not linked with the target locus, c=0.5, expected frequencies are the same for the two progeny groups. Thus only genes linked with the B-locus (or the B gene itself) will contribute to a difference between groups in the average value of a quantitative trait.

Let the contributions of genotype at the X-locus to total genotypic value for a quantitative trait be as follows

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>2u</td>
</tr>
<tr>
<td>Xx</td>
<td>u+au</td>
</tr>
<tr>
<td>xx</td>
<td>0</td>
</tr>
</tbody>
</table>

Using these and the frequencies of Table 2 we find that the expected contributions of the X-locus to the averages for the two progeny groups in the B$_4$F$_2$ are

$$\bar{X}(Bb)_t = [2 - (1-c)^t]u + \frac{1}{2} (1-c)^{t-1}au$$

and

$$\bar{X}(bb)_t = [2 - c(1-c)^{t-1}]u + \frac{1}{2} c(1-c)^{t-1}au$$

and that the difference between them is

$$d\bar{X}_t = \bar{X}(bb)_t - \bar{X}(Bb)_t = (1-2c)(1-c)^{t-1}\left(u - \frac{1}{2} au\right)$$

If pleiotropic effects of B-locus genotypes are 2ub, ub+abub, and zero for the bb, Bb, and
BB genotypes, respectively, the contributions of the B-locus to the progeny group averages will be

\[ \bar{B}(Bb)_t = u_b + a_b u_b / 2 \]

and

\[ \bar{B}(bb)_t = 2u_b \]

so that the expected difference between them is

\[ d \bar{B}_t = \bar{B}(bb)_t - \bar{B}(Bb)_t = u_b - \frac{1}{2} a_b u_b \quad (5) \]

which is the same for all values of \( t \); i.e., unaffected by number of generations of backcrossing. Finally, summing over all loci, we have

\[ d_t = d \bar{B}_t + \sum_i d \bar{X}_{ti} \]

\[ = \left( u_b - \frac{1}{2} a_b u_b \right) + \sum_i \left( 1 - 2c_i \right) \left( 1 - c_i \right)^{t-1} \left( u_i - \frac{1}{2} a_i u_i \right) \quad (6) \]

Because the breeder needs early information concerning genetic effects that are unfavorable relative to his objectives, \( d_t \) for \( t=1 \) has special significance. From (6) we obtain

\[ d_1 = \left( u_b - \frac{1}{2} a_b u_b \right) + \sum_i \left( 1 - 2c_i \right) \left( u_i - \frac{1}{2} a_i u_i \right) \quad (7) \]

If \( d_1 \) were shown to be very small (or favorable in its deviation from zero) the breeder could be quite sure that achievement of his objective would not be blocked by unfavorable effects of linked alleles or unfavorable pleiotropic effects of the target allele. On the other hand, if \( d_1 \) were shown to be significantly large in the unfavorable direction such unfavorable effects would be indicated. Thus estimation of \( d_1 \) is worthwhile though, obviously, only if precision is adequate. Planning for the realization of adequate precision calls for information concerning (1) the smallest value of \( d_1 \) that would be significant relative to the breeding objective and breeding procedures and (2) the approximate standard error of the estimate, as a function on nature and amount of data collected, to be anticipated. More explicitly it is the ratio of these quantities that is critical.

**Smallest significant value of \( d_1 \):** Let \( D \) be the reduction in any quantitative trait (e.g., yield) resulting from incorporation, in homozygous state, of the target allele plus others, if any, closely linked with the target allele into the recurrent parent genotype. Clearly

\[ D = 2u_b + \sum_i 2u_i \quad (8) \]

where the final term summation is over loci linked so tightly with the target allele that recombination could not be expected even in a backcross program designed to maximize the occurrence and selection of recombinants. It is apparent from equations (7) and (8) that if the dominance deviations (the \( a u \)'s) are ignored, the smallest significant value of \( d_1 \) may reasonably be defined as \( (1/2) D \) where \( D \) is the value \( D \) such that the advantage of the target attribute, gained by incorporation of the target allele, is exactly cancelled by the disadvantage resulting from change by the amount \( D \) in the quantitative trait. Symbolically,

\[ d_{1s} = \frac{1}{2} D \]
Since the existence of heterosis indicates more positive than negative a's, it may be more reasonable in the case of heterotic traits, to write

\[ d_{1s} \leq \frac{1}{2} D_s \]

**Standard error of the estimate of \( d_1 \):** Suppose that \( B_1F_2 \) families are obtained by self-fertilization of \( n \ B_1F_1 \) plants having the \( Bb \) genotype and an equal number having the \( bb \) genotype. Assume further that the \( 2n \) families are evaluated in an appropriate field design with \( r \) replication. Then the variance of \( \hat{d} \) (the estimate of \( d_1 \) available from the data) would be

\[ \sigma_{\hat{d}}^2 = 2\frac{\sigma^2 + \sigma^2_{\hat{d}}}{n} \]  

(9)

where \( \sigma_{\hat{d}}^2 \) is the variance of families within groups arising from genetic difference and \( \sigma^2 \) is the error variance.

With heritability defined as

\[ H = \frac{\sigma^2}{\sigma^2 + \sigma_{\hat{d}}^2} \]

it is easily shown that

\[ \sigma_{\hat{d}}^2 = \frac{2\sigma^2}{nr} \left[ \frac{1 + (r-1)H}{1-H} \right] \]  

(10)

The ratio of interest to us may now be written as follows

\[ \frac{d_{1s}}{\sigma_{\hat{d}}} = \frac{d_{1s}}{\sigma} \sqrt{\frac{nr(1-H)}{2[1+(r-1)H]}} \]  

(11)

Solving for \( n \) we obtain

\[ n = \left( \frac{d_{1s}}{\sigma_{\hat{d}}} \right)^2 \left( \frac{\sigma^2}{d_{1s}} \right)^2 \left[ \frac{2[1+(r-1)H]}{r(1-H)} \right] \]  

(12)

which can be used to find the number of families required to achieve any specified magnitude of \( d_{1s}/\sigma_{\hat{d}}^2 \).

**Estimation precision:** Almost certainly the real value of \( d_1 \) will not, in any specific instance, be exactly equal to \( d_{1s} \) but the difference between them might be small in which case a great deal of data would be required to determine whether \( d_1 > d_{1s} \) or \( d_1 < d_{1s} \). For practical purposes, we have chosen to investigate the amounts of data required to achieve specified magnitudes of \( d_{1s}/\sigma_{\hat{d}}^2 \). The larger this ratio, the higher the probability of concluding that \( d_1 > d_{1s} \) if it really is larger or that \( d_1 < d_{1s} \) if it really is smaller.

We show in Table 3 the values of \( n \) required to make \( d_{1s}/\sigma_{\hat{d}}^2 = 1.0 \). These were obtained using

<table>
<thead>
<tr>
<th>( r )</th>
<th>( H )</th>
<th>1.0</th>
<th>0.5</th>
<th>0.25</th>
<th>0.125</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>3</td>
<td>9</td>
<td>36</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>3</td>
<td>11</td>
<td>46</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>7</td>
<td>27</td>
<td>107</td>
<td>427</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>2</td>
<td>8</td>
<td>30</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>3</td>
<td>12</td>
<td>48</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>6</td>
<td>23</td>
<td>91</td>
<td>363</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>1</td>
<td>3</td>
<td>12</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>.3</td>
<td>2</td>
<td>6</td>
<td>22</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>3</td>
<td>10</td>
<td>40</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>.7</td>
<td>6</td>
<td>21</td>
<td>83</td>
<td>330</td>
<td></td>
</tr>
</tbody>
</table>

--- 19 ---
equation (12).

It is apparent from equation (12) that if \( d_{1s}/\sigma^2 = 2 \) is desired, all \( n \) values would be 4 times those shown in Table 3. In like manner, for \( d_{1s}/\sigma^2 = 3 \), the \( n \)'s would be 9 times those of Table 3.

We see from Table 3 that increased replication has little value in terms of numbers of families required unless \( H \) is very small. Unless cross pollinations are very difficult it will obviously be better to increase \( n \) than to increase replication. In fact, where \( H \geq 0.3 \) and cross pollinations are relatively easy, there appears no reason for making \( r \geq 1 \).

The broad range of numbers in Table 3 raises immediate questions concerning values of \( d_{1s}/\sigma \) and \( H \) that will most often be pertinent. Most specifically, since the required \( n \) becomes very large when \( H \) is large and \( d_{1s}/\sigma \) is small, the most critical issues are how small \( d_{1s}/\sigma \) may be and how large \( H \) is likely to be. Note first that

\[
\frac{d_{1s}}{\sigma} = \frac{d_{1s}}{\mu} \frac{\mu}{\sigma}
\]

If \( \mu \) represents the population mean of the recurrent parent, the first term on the right is the smallest significant value of \( d_{1s} \) as a fraction of the mean for the recurrent parent and the second is 100 times the reciprocal of the coefficient of variation (of plot values). Since 20% is high for the coefficient of variation of traits such as yield that are most subject to random variation, we suggest that 5 may tentatively be considered a practical lower limit for \( \mu/\sigma \). Turning to \( d_{1s}/\mu \), we argue that the genetic target of a backcross program should always have potential value equal to at least a 5 percent change in any quantitative trait; otherwise the program would hardly be worth its cost when compared with other possible breeding projects. This argument sets 0.05 as a lower limit of \( D_1/\mu \) and 0.025 as the lower limit of \( d_{1s}/\mu \). Using the limiting values proposed above we had (0.025)5 = 0.125 as our lower limit for \( d_{1s}/\sigma \).

In the case of \( H \), it should first be noted that \( H \) will be smaller for \( B_3F_2 \) families than for \( F_3 \) families obtained by self-fertilization of \( F_2 \) plants. If \( \sigma^2 \) is assumed the same for \( B_3F_2 \) and \( F_3 \) families and dominance is assumed absent or trivial (which appears justified for traits having highest heritabilities) the heritability in \( B_3F_2 \) should be \( H_3/(2-H_2) \) where \( H_3 \) is heritability of \( F_3 \) family plot values. The upper end (0.7) for our range of \( H \) was chosen because it is the value expected when \( H_2 = 0.824 \) which is close to the upper end of the range for heritabilities of \( F_3 \) family plot values.

Also worth noting is the fact that for various traits coefficients of variation near 10 are much more probable than ones near 20. In these cases 0.25 becomes a more likely lower limit for \( d_{1s}/\sigma \) than 0.125. In the case of traits, like yield, with higher coefficients of variation, \( H \) is not likely to be higher than 0.5 in \( B_3F_2 \). In extreme cases where \( d_{1s}/\sigma \) is judged to be as low as 0.125 and \( H \) as high as 0.5, it is tempting to think it may be worthwhile to wait for a generation in order to use \( B_2F_2 \) material for which heritability would be less. With assumptions as before, \( H = 0.5 \) in \( B_3F_2 \) would be reduced to 0.41 in \( B_2F_2 \) and the required \( n \) would be 153 instead of 192. However, a modest increase in plot size along with a field design in which blocks contain less plots would decrease the coefficient

--- 20 ---
Table 4. Approximate power $t$-test when numbers of families are those required to achieve indicated values of $d_{1s}/\sigma_2$.

<table>
<thead>
<tr>
<th>$Z^*$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{1s}$</td>
<td>0.26</td>
<td>0.64</td>
<td>&gt;0.98</td>
</tr>
<tr>
<td>$2d_{1s}$</td>
<td>0.64</td>
<td>&gt;0.98</td>
<td>&gt;0.98</td>
</tr>
<tr>
<td>$3d_{1s}$</td>
<td>0.91</td>
<td>&gt;0.98</td>
<td>&gt;0.98</td>
</tr>
</tbody>
</table>

$Z^*$ is the real value of the deviation to be tested. It would be $d_i$ or $|d_i-d_{1s}|$ depending on the test being made.

The question of how large $d_{1s}/\sigma_2$ should be made is illuminated by the probabilities of finding $d_i$ significantly greater than zero or significantly greater (or smaller) than $d_{1s}$. Power tables provided by COHEN (1969) were used for this purpose. One of the arguments for entering the table is a standardized deviation. In our case the standard deviation used to obtain this parameter was

$$\sigma \sqrt{\frac{1+(r-1)H}{r(1-H)}}$$

Power (the probability of a difference significant at the 0.05 level) was obtained for one-tailed $t$-tests. Findings are summarized in Table 4. It indicates, for example, that if $d_i = d_{1s}$, the probability of showing it to be significantly greater than zero is relatively low unless enough material has been observed to make $d_{1s}/\sigma_2$ greater than 2. The same holds for finding $d_i$ significantly greater (or less) than $d_{1s}$ when $|d_i-d_{1s}|=d_{1s}$. If the number of families required for $d_{1s}/\sigma_2 = 1$ when $H$ is high (0.7) and $d_{1s}/\sigma$ is low (0.25) were employed the power would be relatively good for many of the other cases ($H<0.7$ and/or $d_{1s}/\sigma>0.25$).

Table 4 and 3 considered together emphasize the point that unless $H$ is very low, number of families is much more important than replications (plots per family). For example, when $H=0.5$ and $d_{1s}/\sigma=0.25$, power is the same whether number of families is 64 with 1 plot per family, 48 with 2 plots, or 40 with 4 plots per family. Total number of field plots would be 64, 96, and 160, respectively. When $H=0.7$, this kind of comparison is still more favorable to increasing families rather than replications.

**The problem when $d_i > d_{1s}$**

When $d_i$ is large, knowledge of the fraction due to pleiotropy or alleles linked too tightly to be separated from the target allele and of the portion of the observed negative effect that could be eliminated by procedure designed to maximize recombination and the selection of favorable recombinant genotypes, would provide the basis for useful decisions concerning future procedure. As an example at one extreme, work aimed at the transfer of a specific donor allele from a specific donor parent might be abandoned.

It is clear from equation (6) that to the extent that $d_i$ is due to the effects of linked alleles it should be smaller in more advanced generations. This suggests the possibility of testing...
for the role of linkage by comparing the estimate of \( d_1 \) with an estimate of \( d_2, d_3, \) or \( d_4. \) Working from equation (6) we find that the expected difference between \( d_1 \) and \( d_{v+1} \) is

\[
d_1 - d_{v+1} = \sum [1 - (1 - c_i)^v] (1 - 2c_i)\left(u_i - \frac{1}{2}a_i u_i\right)
\]

(13)

It is readily apparent (1) that \( v \) cannot exceed 2 (or 3 at the most) if an estimate of this difference is to be obtained soon enough to be of much use to the breeder and (2) that if \( v = 2 \) or 3, the difference will be small relative to \( d_1 \) itself. A few values of \([1 - (1 - c)^v]\) are shown below.

<table>
<thead>
<tr>
<th>( c )</th>
<th>( 1 - (1 - c)^2 )</th>
<th>( 1 - (1 - c)^3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.19</td>
<td>.271</td>
</tr>
<tr>
<td>.2</td>
<td>.36</td>
<td>.488</td>
</tr>
<tr>
<td>.3</td>
<td>.51</td>
<td>.643</td>
</tr>
<tr>
<td>.4</td>
<td>.64</td>
<td>.884</td>
</tr>
</tbody>
</table>

Comparing (13) with (7) in the light of these values of \([1 - (1 - c)^v]\) it appears likely that even \( d_1 - d_4 \) would never be as large as \((1/2)d_1\) and would often be less. Relevant also is the fact that the variance of an estimate of \( d_1 - d_{v+1} \) would be the sum of the variances of the separate estimates of \( d_1 \) and \( d_{v+1} \). Taken together these facts indicate that to achieve the same relative precision, the amount of data required in estimation of \( d_1 - d_{v+1} \) would be 8 or more times as much as in estimation of \( d_1. \)

If much of \( d_i \) were the consequence of alleles with \( c \)-values less than 0.2, even massive amounts data would hardly suffice to provide the basis for showing \( d_1 - d_3 \) or \( d_1 - d_4 \) significantly different from zero. At the same time, the possibilities for recombination could easily be sufficient to allow success in a well designed backcross program.

Consider an allele for which \( c = 0.05 \) and suppose it to be responsible for a large portion of \( d_1. \) An estimate of \( d_1 - d_4 \) could not distinguish the effect of this allele from a pleiotropic effect of the target allele since \([1 - (1 - 0.05)^v] = 0.143 \) is so small. Now consider the probability of obtaining a recombination and therefore the BX/bX genotype. The probability of an individual of that type is given by (3) and, with \( c = 0.05, \) would be 0.23 in the \( B_1 F_1. \) If \( m \) \( B_1 F_1 \) individuals have independent lines of descent from the \( F_1 \) generation, it is apparent from (2) that the probability that they will all be non-recombinant (BX/bX) types is \((1-c)^tm\) and hence that \([1 - (1 - c)^tmc] = 0.004 \) is the probability that one or more will have the BX/bX genotype. With \( c = 0.05, t = 5, \) and a very small sample (say \( m = 20, \)) we find the probability of one or more recombinants to be

\[
[1 - (1 - 0.05)^{10m}] = 0.994
\]

Thus it is almost certain that at least one of the \( B_1 F_1 \) plants would have the BX/bX genotype and that if homozygous (BB) lines are developed from all 20 plants that those from at least 1 of the 20 will also be of the desired BX/BX genotype.
Discussion

Our investigation has assumed that the \( u \)-value for any particular locus would be the same in the \( B_1F_2 \) where about one-quarter of the background alleles are from the donor parent as in the final derivatives of a backcross program where less than 2 percent of the background alleles are from the donor parent. We have, in effect, assumed no epistasis of which there can actually be no guarantee. It must be considered possible for an allele to have an average effect in the background genotypes of the \( B_1F_2 \) that would not be evidenced in the background genotypes of the final products of a backcross program. Conversely, it is possible that an allele from the donor would have no effect in \( B_1F_2 \) but would have an effect against the background genotype in final products. However, the evidence relative to epistasis, though it may be inadequate, does not suggest such extremes. Without arguing that all \( u \)-values will be exactly constant over the range of background genotypes anticipated from \( B_1F_2 \) to the final derivatives, we do believe that changes will not be great enough to invalidate application of the logic that has been presented.

Appropriate conclusions from our study appear to us to be as follows. First, we believe that estimation of \( d_1 \) (the mean phenotypic difference between \( B_1F_2 \) offspring of plants heterozygous for the target allele and of plants lacking the target allele) is worthwhile. Until now we have indicated that the estimate can be concerned with any quantitative trait. In practice, however, it seems obvious that once the required experimental material has been produced data should be collected on all important traits to provide the basis for an estimate of \( d_1 \) with respect to overall phenotypic merit of the material in addition to estimates for each important trait. The amounts of experimental material should be adequate. Amounts required for specific levels of precision can be approximated using logic that we have outlined. The basic requirements are (1) judgment concerning the value of the primary effect of the target allele in terms of increments in averages for quantitative traits, most specifically overall value of phenotype excluding the primary effect of the target allele, and (2) reasonable estimates of the coefficient of variation of plot values for each quantitative trait to be measured.

Second, if the estimate of \( d_1 \) is large for one or more important traits, it will not be practical to attempt to assess the role of linked alleles by estimation of the decrease in \( d \) from the \( B_1F_2 \) to the \( B_2F_2 \) or \( B_2F_2 \). Instead the more practical step will be to make the number of lines of descent in the backcross program large enough to make the probability of one or more recombinations between the target allele and alleles tightly linked with it (\( c \) as small as 0.02) very high.

It is possible that a useful test, to discriminate effects of alleles for which \( c \) is large enough to allow recombination to be expected in an adequate program, can be based on response to selection among the \( B_2F_2 \) families (of plants heterozygous for the target allele) grown for estimation of \( d_1 \). While this would be complicated by the segregation still occurring in the background genotype, the potential of such a test, or a variation of it,
merits study.

Acknowledgement

Sincere appreciation is expressed to the Japanese Agency of Science and Technology and the Japanese Ministry of Agriculture and Forestry for supporting the senior author with a grant to study at the University of Minnesota in the U.S.A.

Literature Cited

Hanson, W. D. 1959. Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics 44: 833-837.
戻し交雑法による育種における連鎖の研究

藤巻 宏・R.E. コムストック
(米国ミネソタ大学遺伝学細胞生物学部)

戻し交雑法による育種では、一回親から導入する有用遺伝子（目標遺伝子）と農業上重要な量的形質を支配する遺伝子との間に不利な連鎖があると、育種を進める上での大きな障害となる。この研究では、戻し交雑集団におけるこのような不利な連鎖を評価するための統計遺伝学的方法を開発した。さらに、その方法の精度と頻度、特に依存と結果数の効果を論識した。

両親となる自然系を交雑し、雑種世代では異常な2倍性分離が起こるとする。遺伝子座間のエピスタスはなく、目標遺伝子に関するヘテロ接合体が発現率により選別できるとする。目標遺伝子を A (その対立遺伝子を a) とし、量的形質を支配し B と組換え帯 c で連鎖する遺伝子を x (その対立遺伝子を X)とする。

1 回の戻し交雑によって得られる B1F2 世代の遺伝子型分布は第 1 表のとおりになる。これらの個体を自殖して B1F2 系統とし、B-b 座の分離の有無に関して系統を分割する。各群における X-x 座の遺伝子型分布は第 2 表に示すとおりになる。これら 2 つの系統群間の遺伝的差異が B-b 座と X-x 座との間の連鎖効果を表わすことになる。したがって、B1F2 系統を適当な圃場試験計画で栽培して、農業上重要な量的形質に関して測定を行い、B-b 座の分離の有無に関して 2 分した系統群間の遺伝的差異を統計的に検定すれば、目標形質と量的形質との間の連鎖効果を評価することができる。

系統群間の遺伝的差異を統計的に検出するのに必要なデータの量を求めるには第 3 表のとおりになる。d2 は目標遺伝子の導入に伴う有利性を丁度相殺する量的形質の変化量 (D2) に対応する B1F1 世代の系統群間差。H は配置誤差、H は系統群内系統間の遺伝的変異に基づく遺伝力を表わす。この値の値から反復数よりも系統数を増す方が効率がよいことがわかる。

系統群間差の統計的検定の精度を求めたのが第 4 表である。この表から、充分な精度で検定を行うには、真の差 (Z) が d2 程度ではきわめて多くのデータが必要であるが、Z = 3d2 と大きな差がある場合にはわずかなデータで足りることがわかる。

初期世代（たとえば B1F1 世代）において不利な連鎖が検出される場合、さらに戻し交雑を進めてこの連鎖の破れる程度を統計的に明かにすることはきわめて困難である。そこで、不利な連鎖が存在する組換えを使って育種を行う場合には、何回か戻し交雑を重ねて望ましい組換え型の頻度をある程度高めた上で必要な選抜を行い、目的とする遺伝子型を得るのが最もよいと考えられる。