THE EXPECTED DEGREE OF HETEROZYGOSITY IN A DOUBLE-CROSS HYBRID POPULATION

D. G. PEDERSON

Botany Department, Australian National University, Canberra, A.C.T., Australia

Received October 27, 1965

There have been several recent studies of the relationship between the heterozygosity of plant material and the magnitude of some quantitative character, usually yield or stability over a range of environments (Yermanos and Allard 1961; Finlay 1963; Rowe and Andrew 1964). In each of these studies, single crosses and their derivatives have been used as the experimental material so that estimation of the average degree of heterozygosity, for each generation of plants, presents no difficulty. It may be desired, however, to include double-cross hybrid material in such an investigation. The present study gives prediction formulae for the average degree of heterozygosity in double-cross hybrid populations, for a range of genetic and sampling situations.

Genetic and Sampling Models

Consider a set of \( m \) inbred lines, the set being either “fixed” or chosen at random from a conceptually infinite set of such lines derived from a random-mating population by inbreeding. It will be assumed that there are \( n \) alleles \( A_1, A_2, \ldots, A_n \) occurring at a single locus in the set of lines, with frequencies \( p_1, p_2, \ldots, p_n \) respectively. The maximum value of \( n \) is therefore \( m \).

In producing the \( F_1 \) generation, reciprocal crossing may or may not be carried out. For a cross between two \( F_1 \) genotypes to give the double-cross hybrid, selection of the \( F_1 \) parents may be with or without replacement. There are therefore four distinct ways of constructing the \( F_1 \times F_1 \) set of plants:

- **Case 1**: reciprocal \( F_1 \) plants produced; selection of \( F_1 \) parents with replacement.
- **Case 2**: reciprocal \( F_1 \) plants produced; selection of \( F_1 \) parents without replacement.
- **Case 3**: no reciprocal \( F_1 \) plants produced; selection of \( F_1 \) parents with replacement.
- **Case 4**: no reciprocal \( F_1 \) plants produced; selection of \( F_1 \) parents without replacement.

Expected Heterozygosity

For each of the four cases the situation will be considered where all possible crosses are made between lines and between the \( F_1 \) hybrids, within the limits imposed by the particular model. The expected degree of heterozygosity in each case is then the average over the whole set of double-cross hybrids produced.

Case 1 will be considered in detail as a type example, but only the relevant results will be given for the other three cases.

(a) **Case 1**: Of the \( m \) homozygous lines, a total of \( mp_i \) have been assumed to
possess allele \( A_i \) at the locus under consideration. If the lines are sampled in pairs, without replacement, the number of pairs with each member possessing allele \( A_i \) is \((mp_i) [(mp_i)-1]\). The total number of pairs is \(m(m-1)\).

Therefore the probability that a single cross is homozygous at the locus is given by

\[
Pr(\text{homozygote}) = \sum_i \frac{(mp_i)[(mp_i)-1]}{m(m-1)} = \sum_i \left[ p_i^2 - \frac{p_i}{m} \right] \frac{m}{m-1} = \frac{mp - 1}{m-1}, \text{ where } p = \sum_i p_i^2.
\]

It follows that

\[
Pr(\text{heterozygote}) = 1 - Pr(\text{homozygote}) = \frac{m(1-p)}{(m-1)}.
\]

This result is the same for each of the four cases being considered regardless of whether reciprocal \( F_1 \) plants are produced or not.

The two individuals of the \( F_1 \) generation chosen with replacement may be each homozygous, each heterozygous, or one may be homozygous and the other heterozygous for the locus under consideration. The probabilities of these events may be found as successive terms of the expanded binomial

\[
\left[ \frac{mp - 1}{m-1} + \frac{m(1-p)}{m-1} \right]^2.
\]

The three possible outcomes will be considered separately:

1. **Probability of two homozygotes:**

   \[
   Pr(\text{homozygote, homozygote}) = \left[ \frac{mp - 1}{m-1} \right]^2
   \]

   The homozygotes chosen may be the same or different genotypes, these two outcomes being represented as \((A_iA_i, A_iA_i)\) and \((A_iA_i, A_jA_j)\) respectively. In this type of representation, no significance is attached to the order of the genotypes, or to the order of the alleles within the genotypes. Then

   \[
   Pr(A_iA_i,A_iA_i) = \left[ (p_i^2 - \frac{p_i}{m}) \frac{m}{m-1} \right]^2 + \ldots + \left[ (p_n^2 - \frac{p_n}{m}) \frac{m}{m-1} \right]^2 = \left[ \frac{m}{m-1} \right]^2 \left[ (p_i^2 + \ldots + p_n^2) - \frac{2}{m} (p_i^2 + \ldots + p_n^2) \right] + \frac{1}{m^2} (p_i^2 + \ldots + p_n^2) \]

   \[
   = \frac{1}{(m-1)^2} \left[ m^2 \sum_i p_i^2 - 2m \sum_i p_i^2 + p \right] \quad (1)
   \]

   and

   \[
   Pr(A_iA_i,A_jA_j) = \left[ \frac{mp - 1}{m-1} \right]^2 - (1) \quad (2)
   \]
(2) Probability of one homozygote and one heterozygote:

\[
\Pr(\text{homozygote, heterozygote}) = \frac{2(mp-1) \cdot m(1-p)}{(m-1)^2}
\]

\[
\Pr(A_iA_i,A_iA_j) = 2 \left[ \left( \frac{p_i^2-p_i/m}{m-1} \right) \left[ \frac{2p_i(1-p_i) \cdot m}{m-1} \right] + \ldots \right.
\]
\[
\left. + 2 \left[ \left( \frac{p_i^2-p_n/m}{m-1} \right) \left[ \frac{2p_n(1-p_n) \cdot m}{m-1} \right] \right] \right)
\]
\[
= \frac{1}{(m-1)^2} \left[ 4m \sum_i p_i^2 (1-p_i)(mp_i - 1) \right]
\] (3)

and \( \Pr(A_iA_i,A_jA_k) = \frac{2m(1-p)(mp-1)}{(m-1)^2} - (3) \) (4)

(3) Probability of two heterozygotes:

\[
\Pr(\text{heterozygote, heterozygote}) = \left[ \frac{m(1-p)}{m-1} \right]^2
\]

\[
\Pr(A_iA_j,A_iA_j) = \left[ \frac{mp_is_p_i}{m-1} \right] \left[ \frac{2mp_is_p_j}{m-1} \right] + \ldots + \left[ \frac{mp_{n-1}s_{p_n}}{m-1} \right] \left[ \frac{2mp_{n-1}s_{p_n}}{m-1} \right]
\]
\[
= \frac{2m^2}{(m-1)^2} \sum_{i \neq j} (p_i p_j)^2
\] (5)

\[
\Pr(A_iA_j,A_iA_k) = \left[ \frac{mp_is_p_j}{m-1} \right] \left[ \frac{2m}{m-1} \cdot p_i(1-(p_i+p_j)) + \frac{2m}{m-1} \cdot p_j(1-(p_i+p_j)) \right]
\]
\[
+ \ldots + \left[ \frac{mp_{n-1}s_{p_k}}{m-1} \right] \left[ \frac{2m}{m-1} \cdot p_{n-1}(1-(p_{n-1}+p_k)) \right]
\]
\[
= \frac{2m^2}{(m-1)^2} \sum_{i \neq j} p_ip_j(p_i + p_j) \cdot [1-(p_i + p_j)]
\] (6)

and \( \Pr(A_iA_j,A_kA_l) = \left[ \frac{m(1-p)}{m-1} \right]^2 - (5) - (6) \) (7)

Thus for the set of \( F_1 \times F_1 \) material, the average heterozygosity is given by

\[
H = (2) + \frac{1}{2} (3) + (4) + \frac{1}{2} (5) + \frac{3}{4} (6) + (7)
\]
\[
= (1-p).
\]

The corresponding expressions for the remaining three cases may be similarly determined.

(b) Case 2: \( H = \frac{m(1-p)(2m^2-4m+1)}{2(m-1)(m^2-m-1)} \).

(c) Case 3: \( H = (1-p) \).
### TABLE 1

Average degree of heterozygosity of the $F_1 \times F_1$ generation for the particular case of $m$ homozygous parental lines with $m$ alleles at a locus. Values for the $F_1$ and $F_2$ generations are also given.

<table>
<thead>
<tr>
<th>Generation or Case No.</th>
<th>Number of parental lines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>$F_1$</td>
<td>1.0</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.5</td>
</tr>
<tr>
<td>Case 1</td>
<td>0.5</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.5</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.5</td>
</tr>
<tr>
<td>Case 4</td>
<td>...</td>
</tr>
</tbody>
</table>

(d) Case 4:

$$H = \frac{m^2(1-p)}{m^2-1}, \quad m > 2.$$

Of interest is the situation where each of the $m$ lines possesses a different allele at the locus being considered, so that there are $m$ lines and $m$ alleles. In this case, $p_i = 1/m$ and $\sum p_i^2 = 1/m$, so that the expected degrees of heterozygosity are:

- **Case 1:** $(m-1)/m$
- **Case 2:** $\frac{2m^2 - 4m + 1}{2(m^2 - m - 1)}$
- **Case 3:** $(m-1)/m$
- **Case 4:** $m/(m+1)$

Values of these expressions for a range of values of $m$ are given in Table 1, and shown graphically in Figure 1.

### Ratio of Double-cross to Single-cross Heterozygosity

The expected degree of heterozygosity in the $F_1$ generation has been given as $m(1-p)/(m-1)$. The ratio of the degree of heterozygosity in the $F_1 \times F_1$ generation to that in the $F_1$ generation is, for a segregating locus,

- **Case 1:** $(1-p) \div \frac{m(1-p)}{m-1} = \frac{m-1}{m}$
- **Case 2:** $\frac{m(1-p)(2m^2 - 4m + 1)}{2(m-1)(m^2 - m - 1)} \div \frac{m(1-p)}{m-1} = \frac{2m^2 - 4m + 1}{2(m^2 - m - 1)}$
- **Case 3:** $(1-p) \div \frac{m(1-p)}{m-1} = \frac{m-1}{m}$
- **Case 4:** $\frac{m^2(1-p)}{m^2 - 1} \div \frac{m(1-p)}{m-1} = \frac{m}{m+1}, \quad m > 2.$
DISCUSSION

The results given in the paragraph above are of interest for two reasons. First, it can be seen that the ratios are independent of \( p_i \), being functions only of \( m \). The ratio of double-cross to \( F_1 \) heterozygosity can therefore be specified regardless of the distribution of alleles within the parental set. Secondly, the ratios are algebraically identical with those given in the section on Expected Heterozygosity, for the degree of heterozygosity of the double-cross material in the situation where each of the lines has a different allele. It follows that the \( y \)-axis of Figure 1 may be equally as well in units of

\[
R = \frac{\text{degree of heterozygosity of } F_1 \times F_1}{\text{degree of heterozygosity of } F_1}.
\]

It should be emphasized that the results given are for the \textit{average} degree of heterozygosity over a systematic set of double-cross material, or equivalently, the \textit{expected} heterozygosity of one individual in the set of double-cross plants. However, although average values can be specified, there may well be a considerable amount of variation with respect to heterozygosity within the double-cross hybrid material. For example, in the case of four parent lines the average \( F_1 \times F_1 \) degree of heterozygosity is at a maximum when each parent possesses a different allele, and decreases to zero as the parents are considered increasingly alike for the locus. However, at the same time the variation in absolute degree of heterozygosity, between individual members of the \( F_1 \times F_1 \) generation, increases to a maximum and then becomes zero.

Maximum double-cross heterozygosity for a particular parental set is obtained
when no reciprocal single-cross material is used, and selection of hybrid parents in constructing the double-crosses is without replacement. This result is perhaps intuitively obvious, although it can be seen that the advantage of this crossing scheme over the alternative methods diminishes rapidly as the number of parental lines increases much beyond four.

This work was carried out while the author was in receipt of a studentship from the Australian Wheat Industry Research Council.

SUMMARY

Four different procedures are considered for construction of a complete set of double-crosses from \( m \) inbred parents. For a single locus, the average degrees of heterozygosity of each of the four types of double-cross population are determined in terms of the distribution of alleles in the parental set, as defined by frequency parameters \( p_i \) corresponding to alleles \( A_i \) (\( i = 1, \ldots, n \)). Specific expressions are given for the degrees of heterozygosity when each parent has a different allele at the locus. Also determined are the ratios of \( F_1 \times F_1 \) to \( F_1 \) average heterozygosity for each crossing procedure, and these ratios are found to depend only on \( m \), being independent of \( p_i \).

LITERATURE CITED