


nition, it would be useful to be able to statistically test if different nonrecurrent parents differ regarding $k$. This information can be used to determine the number of recombinant inbred (RI) lines needed for replicated trials to find RI lines that have an adequate number of genes from the donor parent to ensure an acceptable level of the introduced trait.

This article describes an approach to estimate and test hypotheses about the number of loci with genes that differ between the donor parent and the recurrent parent. The method is used to estimate the number of genes determining resistance of common beans (Phaseolus vulgaris L.) to common bacterial blight incited by Xanthomonas campestris pv. phaseoli (Smith) Dye. beans.

An $F_1$ is produced from the cross of two pure lines. The progeny are successively backcrossed $b$ (BC$_b$) times to one of the parent lines, resulting in BC$_b$ plants. These plants and their progeny are selfed until near homozygosity and propagated until enough seed is available for replicated trials. These trials are used for determining the number of inbred-backcross lines that differ in response from the recurrent parent. At any particular locus, the probability that an inbred-backcross line differs from the recurrent parent is $q = (1/2)^{b+1}$ (Wehrhahn and Allard 1965). This follows since the cross of two pure lines (AA and aa where aa is the recurrent parent) backcrossed $b$ times results in progeny, before inbreeding, that have a $(1/2)^b$ chance of being Aa. Selfing these Aa progeny until homozygosity gives lines of which 1/2 are AA. Thus, the proportion of all lines that are AA, that is, that differ from the recurrent parent, is $q = (1/2)^{b+1}$. The expected proportion of lines that differ from the recurrent parent at one or more of the $k$ unlinked loci is then $1 - (1 - q)^k$ (Mullite and Baker 1985; Wehrhahn and Allard 1965). If there are $m$ lines and $r$ differ from the recurrent parent, the estimated proportion is $p_r = r/m$.

The number of unlinked loci with genes affecting the difference between two parental lines for the trait of interest may be estimated using maximum likelihood (Agresti 1990). Given $m$ lines, the number of lines that differ from the recurrent parent ($r$) at one or more of the $k$ unlinked loci has a binomial distribution expressed as

$$\text{Pr}(k) = \binom{m}{r}(1 - q)^k q^r,$$

where $p_r(k) = 1 - (1 - q)^k$. The maximum likelihood estimator, $\hat{k}$, is the value that maximizes the above expression. Specifically, $\hat{k}$ is obtained by differentiating with respect to $k$, setting the result equal to zero and solving for $k$. This yields

$$\hat{k} = \ln(1 - p_r)/\ln(1 - q).$$

The estimator $\hat{k}$ can also be shown to be the minimum modified chi-square estimator (Agresti 1990). Mullite and Baker (1985) obtained the same estimator using a method of moments estimation approach. The estimator $\hat{k}$ is based on the assumptions of no epistasis, no linkage, and normal diploid meiosis. Any deviation from these assumptions will likely decrease the precision of the estimates as stated in Mullite and Baker (1985).

Standard errors and hypothesis tests may be based on weighted least squares as described by Grizzle et al. (1969). Assume there are $s$ groups to be compared regarding $k$. For the $i$th group, let

$$\hat{r}_i = (\hat{k}_i, 1 - \hat{k}_i)$$

be the vector of proportions, $\hat{k}_i = (k_i, \ldots, k_i)'$ be the vector of estimates, and $m_i$ be the number of lines in the $i$th group.

Let $p_i$ be the true proportion and $k_i$ be the function of $p_i$. The covariance matrix of $\hat{k}_i$ is defined as

$$S = \begin{pmatrix}
S_1 & 0 & 0 & \ldots & 0 \\
0 & S_2 & 0 & \ldots & 0 \\
0 & 0 & \ldots & \ldots & S_s \\
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0
\end{pmatrix}
$$

where

$$S_i = k_i^2 V_i,$$

$$V_i = \text{cov}(\hat{r}),$$

$$= \frac{1}{m_i} \begin{pmatrix}
\hat{r}_i(1 - \hat{r}_i) & \hat{r}_i(1 - \hat{r}_i) \\
\hat{r}_i(1 - \hat{r}_i) & \hat{r}_i(1 - \hat{r}_i)
\end{pmatrix},$$

and $H_i = \text{cov}(\hat{r}, \hat{r})$ evaluated at $p_i = \hat{p}_i$.

Now define the model $k = XB$ where $X$ is an $s \times u$ design matrix and $B$ is a $u \times 1$ vector of coefficients. Estimated weighted least squares may be used to estimate $B$:

$$\hat{B} = (X'X)^{-1}X'Y,$$

which has the covariance matrix

$$\text{cov}(\hat{B}) = (X'X)^{-1}.$$

Estimates of various quantities (e.g., $\hat{k}_i$, $\hat{k}_j$, $\hat{k}_j - \hat{k}_i$, etc.) may be obtained by appropriately defining a $1 \times u$ vector of constants, $t_i$, and computing $t_i' \hat{B}$. The standard error of the estimate $t_i' \hat{B}$ is $t_i' (X'X)^{-1} t_i'$. Any linear hypothesis that may be stated as $H_0: L \hat{B} = 0$, where $L$ is a $c \times u$ matrix of rank $c$ with $c \leq u$, may be tested with

$$\left(L' \hat{B}ight) (L'X^{-1}L)^{-1} (L' \hat{B}),$$

which is asymptotically chi square with $c$ degrees of freedom when $H_0$ is true.

SAS PROC CATMOD may be used to obtain results using an appropriately defined RESPONSE statement and design matrix.

To demonstrate the approach, data were used from a study on the inheritance of resistance of common bean to the common bacterial blight (Arnaud-Santaana et al. 1994). Sixty-four BC$_1$, F$_1$ RI lines were obtained from each of two crosses, PC-50 × XAN-159 and PC-50 × BAC-6, where PC-50 was the susceptible recurrent parent. The resistant lines XAN-159 and BAC-6 were derived from different germplasm. For each cross, the number of RI lines that showed significantly improved disease resistance over PC-50 were obtained for seeds, leaves, and pods. The maximum likelihood/weighted least squares approach was used to estimate the number of unlinked loci affecting disease resistance for the three traits and to test if the number of loci differed across the two crosses (Table 1). Results indicated that BAC-6 had significantly fewer genes for leaf resistance than XAN-159, but more genes for pod resistance. The SAS statements necessary to conduct the analysis are given in the Appendix.

### Appendix
SAS statements for estimating and testing the number of genes affecting a trait

DATA A; INPUT TYPES CROSS$ @@;
DO I = 1 TO 2; INPUT PHENO$ WT @@;
OUTPUT; END;
CARDS;

<table>
<thead>
<tr>
<th>Trait</th>
<th>Nonrecurrent parent</th>
<th>Estimated number of genes</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf</td>
<td>BAC6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.6</td>
<td>0.61</td>
</tr>
<tr>
<td>Leaf</td>
<td>XAN159&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.2</td>
<td>0.94</td>
</tr>
<tr>
<td>Pod</td>
<td>BAC6&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.7</td>
<td>0.75</td>
</tr>
<tr>
<td>Pod</td>
<td>XAN159&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.1</td>
<td>0.38</td>
</tr>
<tr>
<td>Seed</td>
<td>BAC6</td>
<td>2.5</td>
<td>0.59</td>
</tr>
<tr>
<td>Seed</td>
<td>XAN159</td>
<td>3.3</td>
<td>0.70</td>
</tr>
</tbody>
</table>

<sup>a</sup> Crosses with different letters differ ($p < .05$).

Data from Table 1. Estimates of the number of genes affecting common bacterial blight resistance in leaves, pods, and seeds in BC, F<sub>i</sub> progeny of two common bean crosses PC-50 × XAN-159 and PC-50 × BAC-6 where PC-50 is susceptible and XAN-159 and BAC-6 are resistant.
BAC6 = (CROSS = 'PC50xB6'); XAN159 = (CROSS = 'PC50xXAN');
**COMPUTES ESTIMATES AND STANDARD ERRORS;
PROC CATMOD; BY TYPE; WEIGHT WT;
DIRECT BAC6 XAN159;
MODEL PHENO = BAC6 XAN159 / NOINT COV;
RESPONSE -7.48875*1 LOG 0 1;
**TESTS EQUALITY OF CROSSES;
PROC CATMOD; BY TYPE; WEIGHT WT;
MODEL PHENO = CROSS;
RESPONSE -7.48875*1 LOG 0 1;
From the Departments of Biometry (Eskridge) and Horticulture (Coyne), University of Nebraska, Lincoln, NE 68583-0712. Published as journal series paper no. 11107, Agricultural Research Division, University of Nebraska, Lincoln, Nebraska.

The Journal of Heredity 1996;87(5)

References


Received May 18, 1995
Accepted February 1, 1996
Corresponding Editor: James L. Hamrick