The possibility remains open that with appropriate PCR conditions, microsatellite primers designed for certain species may be informative in more divergent taxa than previously believed or demonstrated. With the recent increase in the use of microsatellite techniques for analyzing population genetic structure and evolution there is a need for further studies of this type. Certainly, from the thousands of microsatellites so far isolated from the human genome, a significant number will be applicable for use in a range of other primates, though these experiments demonstrate a variability too high for direct comparisons between species in phylogenetic analysis.

From the Conservation Genetics Group, Institute of Zoology, Regents Park, London NW1 4RY, UK. This work was supported by NSF grant BNS-9223355 to J. Altman, University of Chicago, and by the Institute of Zoology. We are grateful to Research Genetics Inc. for permission to publish results using these primers and for help and information about them. Helen Stanley provided information on other primer sequences. We would like to acknowledge the help of Jeanne Altman and Robert Wayne for helpful discussions and for critically reviewing earlier versions of this manuscript. All samples were from the Institute of Zoology blood and tissue bank, except the savannah baboon samples which were taken from a wild population in Amboseli National Park, Kenya. Valuable technical assistance was provided by Susan Haines and Harriet Green, and photography was carried out by Terry Bennett at the Institute of Zoology. Lastly, we would like to thank those people who gave blood for human analysis: Sheila Boddy, Tamara Bell, and Alkis Saccheri.

The Journal of Heredity 1996:87(5)

Estimating and Testing Hypotheses About the Number of Genes Using Inbred-Backcross Data
K. M. Eskridge and D. P. Coyne

We describe an approach to estimate and test hypotheses about the number of genes that differ between the donor parent and the recurrent parent when using the inbred-backcross breeding procedure. 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.

The backcross breeding procedure has long been used to successfully transfer a major gene(s) controlling a desired trait from a donor or nonrecurrent parent to a superior cultivar or line (recurrent parent). Genes are transferred by repeated backcrossing to the recurrent parent and selecting in each generation for the desired trait (Briggs 1935; Briggs and Knowles 1937). This breeding procedure has generally been used to transfer major genes controlling such traits as disease resistance to superior cultivars of self-pollinated crops, inbred lines of cross-pollinated crops, or cultivars of cross-pollinated crops. Generally, two or more backcrosses are made to the recurrent parent in order to reconstitute most of the genotype of the desired parent along with the designated genes from the donor parent, followed by several generations of selfing to attain a homozygous line. It is useful to plant breeders using the inbred-backcross procedure to know the approximate number of genes (k) that differ between the recurrent and nonrecurrent parent controlling a desired quantitative trait.
tion, 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 (\( \text{Phaseolus vulgaris} \)) 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 \( BC_n \) 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 \( q = (1/2)^{p+1} \) (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 \( 1/2 \) 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 \( 1 - (1 - q)^k \) (Mulitze and Baker 1985; Wehrhahn and Allard 1965). If there are \( m \) lines and \( r \) differ from the recurrent parent, the estimated proportion \( \hat{p}_r = r/m \).

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

\[
\ln [(m!(r!)(m - r)!)] / p_r(k)(1 - p_r(k))^{m-r},
\]

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 \( \hat{k} \). This yields

\[
\hat{k} = \ln(1 - \hat{p}_r)/\ln(1 - q).
\]

The estimator \( \hat{k} \) can also be shown to be the minimum modified chi-square estimator (Agresti 1990). Mulitze 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 Mulitze 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

\[
\hat{p}_r = (\hat{p}_1, \ldots, \hat{p}_r) \text{ be the vector of proportions},
\]

\[
\hat{k} = (k_1, \ldots, k_r) \text{ be the vector of estimates, and}
\]

\[
m_i = \text{the number of lines in the } i \text{th group}.
\]

Let \( p_r \) be the true proportion and \( k_r \) be the function of \( p_r \). The covariance matrix of \( \hat{k} \) is defined as

\[
S = \begin{pmatrix}
S_0 & 0 & 0 & \cdots & 0 \\
0 & S_0 & 0 & \cdots & 0 \\
0 & 0 & S_0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & S_0
\end{pmatrix}
\]

where

\[
S_i = H_i V H_i',
\]

\[
V = \text{cov}(\hat{p}_r),
\]

\[
= (1/m) \begin{pmatrix}
\hat{p}_r(1 - \hat{p}_r) & -\hat{p}_r(1 - \hat{p}_r) \\
-\hat{p}_r(1 - \hat{p}_r) & \hat{p}_r(1 - \hat{p}_r)
\end{pmatrix},
\]

\[
H_i = (d_k/\hat{d}_k, -d_k/\hat{d}_k),
\]

evaluated at \( p_r = \hat{p}_r \).

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

\[
\hat{B} = (X'S^{-1}X)^{-1}X'S^{-1}k
\]

which has the covariance matrix

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

Estimates of various quantities (e.g., \( \hat{k}, k_1 - \hat{k}, \hat{k}, \ldots, \hat{k} \)) can be obtained by appropriately defining a \( 1 \times u \) vector of constants, \( l_i \), and computing \( l_i'B \). The standard error of the estimate \( l_i'B \) is \( l_i'V^{-1}l_i \). Any linear hypothesis that may be stated as \( H_0 : LB = 0 \), where \( L \) is a \( c \times u \) matrix of rank \( c \) with \( c \leq u \), may be tested with

\[
(\hat{B}'(X'S^{-1}X)^{-1}L)^{-1}(\hat{B}-LB)
\]

which is asymptotically \( \chi^2 \) 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-Santana et al. 1994). Sixty-four \( BC_1 \) F, RI lines were obtained from each of two crosses, \( PC-50 \times XAN-159 \) and \( PC-50 \times BAC-6 \), 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 TYPE$ 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</td>
<td>2.6</td>
<td>0.51</td>
</tr>
<tr>
<td>Leaf</td>
<td>XAN159</td>
<td>5.2</td>
<td>0.94</td>
</tr>
<tr>
<td>Pod</td>
<td>BAC6</td>
<td>3.7</td>
<td>0.75</td>
</tr>
<tr>
<td>Pod</td>
<td>XAN159</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>

*Crosses with different letters differ \( (P < .05) \).
BAC6 = (CROSS = 'PC50XBB6'); XAN159 =
(CROSS = 'PC50XXAN');**
** COMPUTES ESTIMATES AND STAN-
DARD ERRORS;
PROC CATMOD; BY TYPE; WEIGHT WT;
DIRECT BAC6 XAN159;
MODEL PHENO = BAC6 XAN159 /
NOINT COV;
RESPONSE -7.488875 • 1 LOG 0 1;
** TESTS EQUALITY OF CROSSES;
PROC CATMOD; BY TYPE; WEIGHT WT;
MODEL PHENO = CROSS;
RESPONSE -7.488875 • 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 Ne-
braska, Lincoln, Nebraska.

The Journal of Heredity 1996:87(5)

References
Agresti A. 1990. Categorical data analysis. New York:
Wiley.
Arnaud-Santana E, Coyne DP, Eskridge KM, and Vidaver
AK, 1994. Inheritance; low correlations of leaf, pod and
seed reactions to common blight disease in common
beans; and implications for selection. J Am Soc Hort
Briggs FN, 1935. The backcross method in plant breed-

Briggs FN and Knowles PF, 1967. Introduction to plant
breeding. New York: Reinhold.
Grizzle JE, Starmer CF, and Koch GG, 1969. Analysis of
categorical data by linear models. Biometrics 25:489–
504.

Mulizke DK and Baker RJ, 1985. Evaluation of biomet-

Wehrhahn C and Allard RW, 1965. The detection and
measurement of the effects of individual genes in-
volved in the inheritance of a quantitative character in

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