Variance of Marker Estimates of Parental Contribution to F_2 and BC_1-Derived Inbreds

Jiankang Wang and Rex Bernardo*

ABSTRACT

An Essentially Derived Variety is a cultivar or inbred that largely retains the characteristics of an ancestral cultivar or inbred. The parental contribution to F_2-derived inbreds (p_{F2}) and BC_1-derived inbreds (p_{BC1}) can be estimated with molecular markers. A recombinant inbred (RI) with p_{F2} or p_{BC1} greater than a specified threshold is then considered essentially derived. Our objectives were (i) to derive the variance of p_{F2} and p_{BC1}, and (ii) to determine the probability of obtaining an essentially derived RI for different numbers of marker loci in different species. The variances of p_{F2} and p_{BC1} are a function of the number of chromosomes, length of each chromosome, and number of marker loci on each chromosome. The standard errors (SE) of p_{F2} and p_{BC1} were smallest when the two marker loci closest to the ends of each chromosome were included. The minimum values of $SE(p_{F2})$ and $SE(p_{BC1})$ are useful for setting minimum values of thresholds for declaring essential derivation. Suppose selfing from the BC_1 is permissible and the maximum error rate for falsely declaring an RI is essentially derived is set at 2.5%. The minimum value of the threshold for these conditions is 0.881 in maize (*Zea mays* L.). For a threshold of 0.90, the probabilities of an essentially derived RI from the BC_1 generation were >6% in rye (*Secale cereale* L.), >3% in barley (*Hordeum vulgare* L.), <3% in tomato (*Lycopersicon* spp.), rice (*Oryza* spp.), and maize, and <1% wheat (*Triticum aestivum* L.). These results suggest that the thresholds used to declare essential derivation should differ among species.

Parental contribution is the proportion of the genome contributed by a parent to its inbred progeny. Expected parental contributions with Mendelian inheritance are 0.5 for either parent of an F_2-derived inbred, 0.75 for the recurrent parent of a BC_1-derived inbred, and 0.25 for the donor parent of a BC_1-derived inbred. Selection and genetic drift during selfing may cause differences between observed and expected parental contributions to inbred progeny (Lorenzen et al., 1995; Bernardo et al., 1997). Parental contribution determined from pedigree records may therefore be inaccurate.

In 1991, the Union Internationale pour la Protection des Obtentions Végétales established the concept of an Essentially Derived Variety, i.e., when ''the essential part of the genome of an initial variety has been included in the new variety'' (Smith et al., 1995). Molecular markers are useful for estimating parental contribution, and the use of molecular markers for assessing essential derivation has gained widespread acceptance (Dillmann et al., 1995). Implicit in the concept of an Essentially Derived Variety is that thresholds will be established beyond which an inbred will be declared essentially derived from an initial (i.e., parental or ancestral) inbred. A consensus has not been reached regarding appropriate thresholds, although a threshold of 0.90 has been proposed for maize (Smith et al., 1995).

The variance of marker estimates of parental contribution, among a set of RIs, is crucial in determining appropriate thresholds for essential derivation in different crop species. Such information is needed to calculate the probability of obtaining an essentially derived RI from an F_2 or BC_1 population in the absence of selection. The variance of marker estimates of parental contribution would vary among species because of differences in the number of chromosomes and length of each chromosome. Because of linkage among marker loci, a simple binomial distribution is not applicable for the variance of marker estimates of parental contribution.

Assessing the amount of variation among random RIs, in terms of their marker estimate of parental contribution, is conceptually different from determining the minimum number of marker loci for estimating the parental contribution for a specific RI. We do not make any recommendations regarding the number of marker loci needed for assessing essential derivation. Specifically, our first objective was to derive the variance of marker estimates of parental contribution for different lengths of chromosomes and numbers of marker loci on each chromosome. Our second objective was to determine the probability of obtaining a random F_2- or BC_1-derived RI that is essentially derived, given different thresholds and numbers of marker loci in different species.

THEORY

Genetic Model and Notation

We considered an RI derived from a $(P_1 \times P_2)F_2$ or $[(P_1 \times P_2) \times P_1]BC_1$ population. The species has n pairs of chromosomes. On the kth chromosome, P_1 and P_2 are polymorphic at l_k mapped, single-locus markers. Across all chromosomes, P_1 and P_2 are polymorphic at $l = \sum_{k=1}^{n} l_k$ marker loci. The marker genotypes are $M_1M_1M_2M_2 \ldots M_lM_l$ in P_1, and $m_1m_1m_2m_2 \ldots m_lm_l$ in P_2.

The l_k marker loci are evenly distributed along a chromosome that is D centimorgans (cM) long. The distance (d) between adjacent loci depends on the placement of markers on the chromosome. We considered two models for the placement of marker loci (Fig. 1). In the Terminal Marker Model, the distance between adjacent markers is $d = D/(l_k - 1)$, and the marker loci are at the 0, d, $2d$, ..., D cM positions on the chromosome. The two marker loci that map closest to the ends of each chromosome are chosen first, then marker loci

J. Wang, Laboratory Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002, P.R. China; R. Bernardo, Dep. of Agronomy, Purdue Univ., 1150 Lilly Hall of Life Sciences, West Lafayette, IN 47905-1150. Purdue Agric. Res. Programs Journal Paper 16015. Received 4 June 1999. *Corresponding author (bernardo@purdue.edu).

Published in Crop Sci. 40:659–665 (2000).

Abbreviations: cM, centimorgans; RI, recombinant inbred.

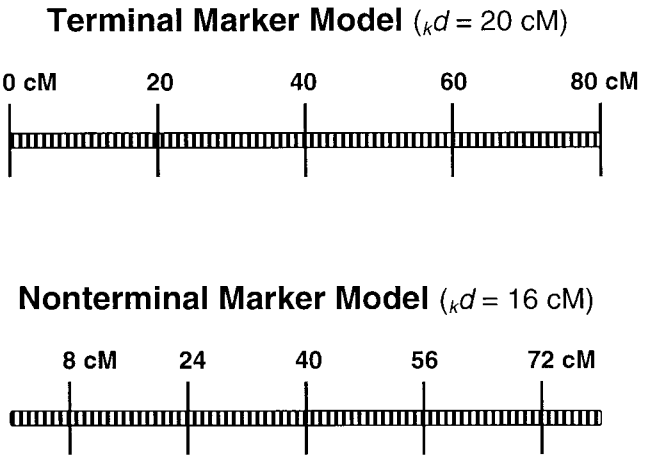

Fig. 1. Locations of five marker loci on an 80 centimorgan (cM) chromosome with the Terminal Marker Model and Nonterminal Marker Model. The $_kd$ values are the distances between adjacent markers on the *k*th chromosome.

located between these two terminal markers are chosen next. In the Nonterminal Marker Model, the distance between adjacent marker loci is $d = D/l_k$, and the marker loci are at the $(0.5)d$, $(1.5)d$, $(2.5)d$, ..., $(l_k - 0.5)d$ cM positions.

For the *k*th chromosome, marker estimates of parental contribution are denoted as $_kp_{F2}$ for an F_2-derived inbred and $_kp_{BC1}$ for a BC_1-derived inbred. Across all chromosomes, marker estimates of parental contribution are denoted as p_{F2} for an F_2-derived inbred and p_{BC1} for a BC_1-derived inbred.

Individual Chromosome

F_2-Derived RI

At the *k*th chromosome, $_kX_i$ ($i = 1, ..., l_k$) is an indicator variable that is equal to 1 if the RI has the same marker genotype as P_1, and 0 if the RI has the same marker genotype as P_2. The number of marker loci at which the RI and P_1 are homozygous for the same allele is $_kX = {}_kX_1 + {}_kX_2 + ... + {}_kX_{l_k}$. The parental contribution at Chromosome *k* of P_1 to the F_2-derived RI is estimated as $_kp_{F2} = {}_kX/l_k$, with a mean and variance of

$$E(_kp_{F2}) = E(_kX)/l_k$$
$$V(_kp_{F2}) = V(_kX)/l_k^2$$

The mean and variance of $_kX$ are

$$E(_kX) = \sum_{i=1}^{l_k} E(_kX_i)$$

$$V(_kX) = \sum_{i=1}^{l_k} V(_kX_i) + 2 \sum_{1\le i<j\le l_k} Cov(_kX_i, {}_kX_j)$$

Consider two linked marker loci, *i* and *j* ($i < j$). With a single meiosis, the recombination frequency between *i* and *j* is r_{ij}. The frequency of recombinant genotypes among RIs is $R_{ij} = 2r_{ij}/(1 + 2r_{ij})$(Haldane and Waddington, 1931). On the basis of the expected frequencies of marker genotypes among RIs (Table 1), the means, variances, and covariances of $_kX_i$ and $_kX_j$ are

$$E(_kX_i) = E(_kX_j) = 1/2$$
$$V(_kX_i) = V(_kX_j) = 1/4$$
$$Cov(_kX_i, {}_kX_j) = (1 - 2R_{ij})/4 = C_{ij}/4,$$

where $C_{ij} = (1 - 2R_{ij}) = (1 - 2r_{ij})/(1 + 2r_{ij})$. Let d_{ij} be the distance in cM between *i* and *j*. We used the Kosambi (1944) mapping function because, unlike the Haldane mapping function, it allows for modest crossover interference in adjacent marker intervals. With the Kosambi mapping function, $d_{ij} = (25)\ \ln[(1 + 2r_{ij})/(1 - 2r_{ij})]$. Consequently, $C_{ij} = \exp(-d_{ij}/25)$.

Consider three linked marker loci, *i*, *j*, and *h*. If Locus *h* is between *i* and *j*, then $d_{ij} = d_{ih} + d_{hj}$, and $C_{ij} = \exp[-(d_{ih} + d_{hj})/25] = C_{ih}C_{hj}$. Therefore, the mean and variance of $_kX$ are

$$E(_kX) = l_k/2$$
$$V(_kX) = \frac{l_k}{4} + \frac{1}{2} \sum_{1\le i<j\le l_k} \prod_{k=0}^{j-i-1} C_{(i+k),(i+k+1)} \qquad [1]$$

If the l_k loci are evenly distributed along the chromosome, the distance between any two adjacent markers is $d_{ij} = d$. For both the Terminal Marker and Nonterminal Marker Models, the frequency of recombinants between any two adjacent loci is

$$r = \frac{1 - \exp(-d/25)}{2[1 + \exp(-d/25)]}$$

and therefore,

$$C = [(1 - 2r)/(1 + 2r)] = \exp(-d/25)$$

From Eq. [1], the variance of $_kX$ is

$$V(_kX) = \frac{l_k}{4} + \frac{1}{2} \sum_{1\le i<j\le l_k} C^{j-i}$$

We found that the last part of this previous equation reduces to

$$\sum_{1\le i<j\le l_k} C^{j-i} = \frac{(l_k - 1)C}{1 - C} - \frac{C^2(1 - C^{l_k-1})}{(1 - C)^2}$$

Hence, the variance of $_kX$ is equal to

$$V(_kX) = \frac{l_k}{4} + \frac{1}{2}\left[\frac{(l_k - 1)C}{1 - C} - \frac{C^2(1 - C^{l_k-1})}{(1 - C)^2}\right]$$

For the *k*th chromosome, the mean and variance of the parental contribution of P_1 to an F_2-derived RI is

$$E(_kp_{F2}) = 1/2$$
$$V(_kp_{F2}) = \frac{1}{4l_k} + \frac{1}{2l_k^2}\left[\frac{(l_k - 1)C}{1 - C} - \frac{C^2(1 - C^{l_k-1})}{(1 - C)^2}\right] \qquad [2]$$

For a chromosome that is *D* cM long, the limit of $V(_kp_{F2})$ as l_k approaches infinity is

Table 1. Frequency of marker genotypes, at Loci *i* and *j* on Chromosome *k*, among recombinant inbreds (RI) derived from a ($P_1 \times P_2$)F_2 population.

Marker genotype†	Frequency‡	$_kX_i$§	$_kX_j$
$M_iM_iM_jM_j$	$(1 - R_{ij})/2$	1	1
$M_iM_im_jm_j$	$R_{ij}/2$	1	0
$m_im_iM_jM_j$	$R_{ij}/2$	0	1
$m_im_im_jm_j$	$(1 - R_{ij})/2$	0	0

† Inbred P_1 has the $M_iM_iM_jM_j$ genotype whereas P_2 has the $m_im_im_jm_j$ genotype.
‡ R_{ij} is the frequency of recombinant genotypes between loci *i* and *j* among RI's.
§ $_kX_i$ and $_kX_j$ are indicator variables equal to 1 if the RI and P_1 are homozygous for the same allele, and 0 if the RI and P_2 are homozygous for the same allele.

Table 2. Frequency of marker genotypes, at Loci *i* and *j* on Chromosome *k*, among recombinant inbreds (RI) derived from a [($P_1 \times P_2$) $\times$ P_1]BC_1 population.

BC_1 individuals		Frequency of marker genotypes among BC_1-derived RI's			
Marker genotype†	Frequency	$M_iM_iM_jM_j$	$M_iM_im_jm_j$	$m_im_iM_jM_j$	$m_im_im_jm_j$
$M_iM_iM_jM_j$	$(1 - r_{ij})/2$‡	1	0	0	0
$M_iM_iM_jm_j$	$r_{ij}/2$	1/2	1/2	0	0
$M_im_iM_jM_j$	$r_{ij}/2$	1/2	0	1/2	0
$M_im_iM_jm_j$	$(1 - r_{ij})/2$	$(1 - R_{ij})/2$§	$R_{ij}/2$	$R_{ij}/2$	$(1 - R_{ij})/2$
Total frequency		$1/2 + (1 - r_{ij})(1 - R_{ij})/4$	$1/4 - (1 - r_{ij})(1 - R_{ij})/4$	$1/4 - (1 - r_{ij})(1 - R_{ij})/4$	$(1 - r_{ij})(1 - R_{ij})/4$
${}_kX_i^{BC1}$¶		1	1	0	0
${}_kX_j^{BC1}$		1	0	1	0

† Inbred P_1 has the $M_iM_iM_jM_j$ genotype whereas P_2 has the $m_im_im_jm_j$ genotype.
‡ r_{ij} is the frequency of recombination with a single meiosis.
§ R_{ij} is the frequency of recombinant genotypes between loci *i* and *j* among RI's.
¶ ${}_kX_i^{BC1}$ and ${}_kX_j^{BC1}$ are indicator variables equal to 1 if the BC_1-derived RI and P_1 are homozygous for the same allele, and 0 if the RI and P_2 are homozygous for the same allele.

$$\lim_{l_k \to \infty} V({}_kp_{F2}) = \frac{25[D - 25(1 - e^{-D/25})]}{2D^2}$$

BC_1-Derived RI

At the *k*th chromosome, ${}_kX_i^{BC1}(i = 1, ..., l_k)$ is an indicator variable that is equal to 1 if the BC_1-derived RI has the same marker genotype as P_1, and 0 if the RI has the same marker genotype as P_2. The number of marker loci at which the RI and P_1 are homozygous for the same allele is ${}_kX^{BC1} = {}_kX_1^{BC1} + {}_kX_2^{BC1} + ... + {}_kX_{l_k}^{BC1}$. On the basis of the expected frequencies of marker genotypes in the BC_1 population (Table 2), the means, variances, and covariances of ${}_kX_i^{BC1}$ and ${}_kX_j^{BC1}$ are

$$E({}_kX_i^{BC1}) = E({}_kX_j^{BC1}) = 3/4$$

$$V({}_kX_i^{BC1}) = V({}_kX_j^{BC1}) = \frac{3}{16} = \frac{3}{4}V({}_kX_i) = \frac{3}{4}V({}_k\overline{X}_j)$$

$$Cov({}_kX_i^{BC1}, {}_kX_j^{BC1}) = \frac{3}{16}C_{ij} = \frac{3}{4}Cov({}_kX_i, {}_kX_j)$$

Therefore, the means and variances of ${}_kX^{BC1}$ and ${}_kp_{BC1}$ are

$$E({}_kX^{BC1}) = (3/4)l_k$$
$$V({}_kX^{BC1}) = (3/4)V({}_kX)$$
$$E({}_kp_{BC1}) = 3/4$$
$$V({}_kp_{BC1}) = (3/4)V({}_kp_{F2})$$

Several Chromosomes

Across all *n* chromosomes, the total number of marker loci at which an F_2-derived RI and P_1 are homozygous for the same allele is $X_T = \sum_{k=1}^{n} {}_kX$. The ${}_kX$ values are independently distributed, and the mean and variance of X_T are

$$E(X_T) = l/2$$
$$V(X_T) = \sum_{k=1}^{n} V({}_kX),$$

where *l* is the total number of marker loci across all chromosomes. Therefore, the mean and variance of p_{F2} are

$$E(p_{F2}) = 1/2$$
$$V(p_{F2}) = \sum_{k=1}^{n} V({}_kX)/l^2$$

For a BC_1-derived RI, the mean and variance of p_{BC1} are

$$E(p_{BC1}) = 3/4$$
$$V(p_{BC1}) = (3/4)V(p_{F2}) \quad [3]$$

The estimates of parental contribution are expected to approach a normal distribution as the number of marker loci increases. The standard errors (SE) of p_{F2} and p_{BC1} are equal to the square root of their respective variances. We transformed p_{F2} and p_{BC1} into *z*-scores, i.e., $z = (p_{F2} - 0.5)/SE(p_{F2})$ for F_2-derived RIs and $z = (p_{BC1} - 0.75)/SE(p_{BC1})$ for BC_1-derived RIs. On the basis of *z*-scores, we determined the probability of obtaining an essentially derived RI for thresholds of 0.65, 0.70, and 0.75 for F_2-derived RIs, and thresholds of 0.85, 0.90, and 0.95 for BC_1-derived inbreds.

RESULTS AND DISCUSSION

Terminal versus Nonterminal Marker Models for a Single Chromosome

The location of marker loci on a chromosome affected the nature of the relationship between the number of marker loci and $SE({}_kp_{F2})$. In the Terminal Marker Model (Fig. 1), there was a specific number of marker loci that minimized $SE({}_kp_{F2})$ for each chromosome length (Fig. 2). In contrast, $SE({}_kp_{F2})$ in the Nonterminal Marker Model, decreased asymptotically as the number of marker loci increased. These results indicated that $V({}_kp_{F2})$ and $V({}_kp_{BC1})$ cannot be infinitely small. The difference in $SE({}_kp_{F2})$ between the two models was greater with shorter chromosomes than with longer chromosomes. But regardless of chromosome length, the Terminal Marker Model always had a lower minimum $SE({}_kp_{F2})$ than the Nonterminal Marker Model. For example, the minimum $SE(p_{F2})$ in the Terminal Marker Model was 0.3624 for a 50 cM chromosome and 0.2101 for a 250 cM chromosome. In contrast, the $SE({}_kp_{F2})$ in the Nonterminal Marker Model as the number of marker loci approached infinity was 0.3767 for a 50 cM chromosome and 0.2121 for a 250 cM chromosome.

As indicated in Eq. [2], $V({}_kp_{F2})$ has a variance component (i.e., $1/4l_k$) and a covariance component. We are unable to provide a simple, intuitive explanation for the existence of a specific number of markers that minimized $V({}_kp_{F2})$ in the Terminal Marker Model but not

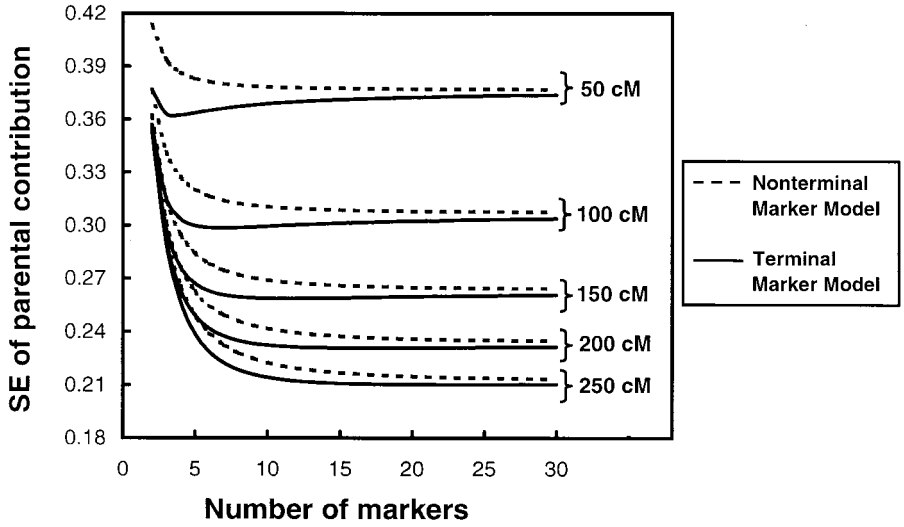

Fig. 2. Standard error (SE) of the parental contribution to F_2-derived recombinant inbreds for a single chromosome for (i) different numbers of marker loci on the chromosome, (ii) different lengths (in centimorgans, cM) of the chromosome, and (iii) terminal versus nonterminal locations of markers.

in the Nonterminal Marker Model. Instead, we can only describe the behavior of the variance and covariance components. As the number of marker loci on a chromosome of fixed length increases, the variance component decreases whereas the covariance component increases. The variance component is equal for both models. But due to a smaller distance between adjacent marker loci, the Nonterminal Marker Model has a larger covariance component than the Terminal Marker Model. Lower thresholds will help minimize unjustified disputes regarding essential derivation. The remainder of our discussion will focus only on the results for the Terminal Marker Model because its standard error, being always smaller than that for the Nonterminal Marker Model, leads to lower thresholds for declaring essential derivation.

Probability of Obtaining Essentially Derived Inbreds

Crop species differ in genome size (in centimorgans), number of chromosomes, and length of each chromosome. Estimated genome sizes and the numbers of linkage groups, corresponding to the haploid number of chromosomes (n), are 724 cM and $n = 7$ in rye (Wanous et al., 1998); 1088 cM and $n = 7$ in barley (Langridge et al., 1995); 1294 cM and $n = 12$ in tomato (Tanksley et al., 1992); 1491 cM and $n = 12$ in rice (Causse et al., 1994); 1749 cM and $n = 10$ in maize (Senior et al., 1996); and 2828 cM and $n = 21$ in wheat (Gale et al., 1995). Rye, which has the fewest chromosomes and smallest genome, had the largest values of $SE(p_{F2})$ (Fig. 3). Wheat, which has the most chromosomes and largest genome, had the smallest values of $SE(p_{F2})$. The numbers of polymorphic marker loci that minimized $SE(p_{F2})$ were 56 in rye, 76 in barley, 97 in tomato, 111 in rice, 128 in maize, and 206 in wheat. The minimum $SE(p_{F2})$ with these numbers of marker loci were 0.1117 in rye, 0.0965 in barley, 0.0842 in tomato, 0.0801 in rice, 0.0771 in maize, and 0.0588 in wheat. These numbers of marker loci corresponded to distances between adjacent marker loci of 15 cM in rye, tomato, rice, maize, and wheat, and 16 cM in barley. As indicated by Eq. [3], the minimum $SE(p_{BC1})$ was equal to $(3/4)^{1/2}\ SE(p_{F2})$.

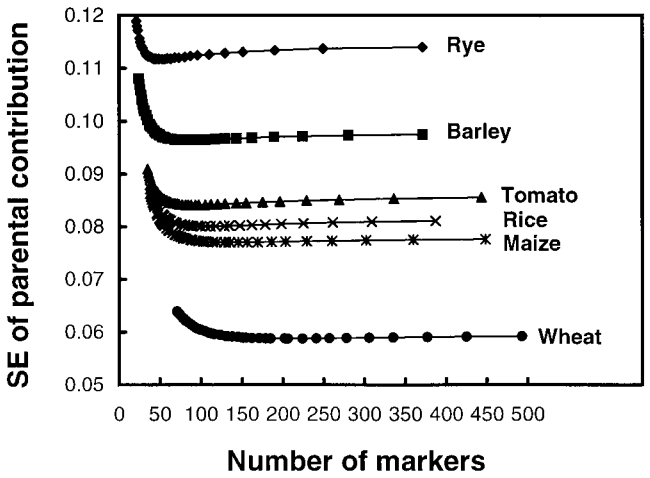

Fig. 3. Standard error (SE) of the parental contribution to F_2-derived recombinant inbreds with different numbers of marker loci in different species.

In a study of essential derivation in maize, the American Seed Trade Association has proposed estimating $SE(p_{F2})$ and $SE(p_{BC1})$ empirically with at least 80 marker loci (A.R. Hallauer, 1997, personal communication; Bernardo, 1999, unpublished). The observed variation in p_{F2} and p_{BC1} in a diverse set of populations would provide benchmarks for setting thresholds. An alternative approach is to use the theoretical values of $SE(p_{F2})$ and $SE(p_{BC1})$ we have derived. The $SE(p_{F2})$ and $SE(p_{BC1})$ indicate the variation in p_{F2} and p_{BC1} among a series of random RIs, but do not indicate the error of the estimate of p_{F2} and p_{BC1} for a single RI. In other words, our results are consistent with the American Seed Trade Association approach for determining appropriate thresholds, but not for determining whether a specific RI is essentially derived or not. Further research is needed on the variance of the error of the estimate of p_{F2} and p_{BC1} for an individual RI.

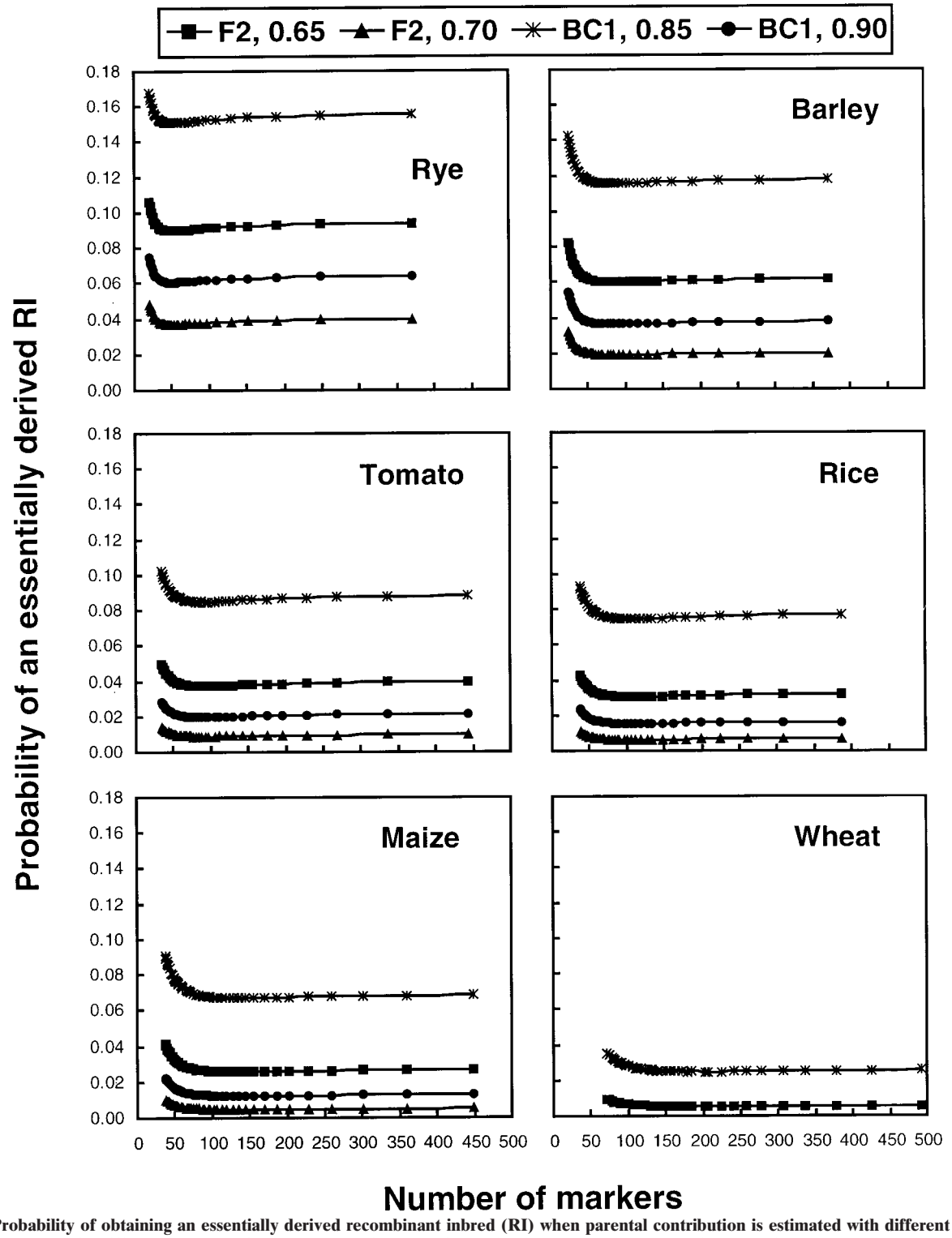

Fig. 4. Probability of obtaining an essentially derived recombinant inbred (RI) when parental contribution is estimated with different numbers of marker loci in different crop species. Thresholds for essential derivation are 0.65 and 0.70 for F_2-derived RIs, and 0.85 and 0.90 for BC_1-derived RIs.

We propose that the minimum values of $SE(p_{F2})$ and $SE(p_{BC1})$ be used to establish minimum values of thresholds for declaring essential derivation. Suppose selfing is permissible from a BC_1 population with a third party inbred as the recurrent parent. Assume the maximum error rate for falsely declaring an RI is essentially derived is set at 2.5%. The upper limit of a 95% confidence interval on p_{BC1}, equal to $0.75 + z_{0.975} \times SE(p_{BC1})$, would then serve as the minimum value of the threshold for declaring essential derivation. This upper limit is equal to 0.881 in maize. This result implies that, given the specified rate of 2.5% for false positives, any BC_1-

derived RI with $p_{BC1} \leq 0.881$ should not be in any danger of being declared essentially derived. A BC_1-derived RI with $p_{BC1} > 0.881$ may or may not be declared essentially derived, depending on the final threshold used for assessing essential derivation.

Averaged across the six crop species we considered, the probability of obtaining an essentially derived RI was greatest when the threshold was 0.85 among BC_1-derived RIs (Fig. 4). For the threshold of 0.75 among F_2-derived RIs, the probabilities of an essentially derived RI were <2% in rye, <1% in barley, and <0.3% in tomato, rice, maize, and wheat (results not shown in Fig. 4). For the extreme threshold of 0.95 among BC_1-derived RIs, the probabilities were <3% in rye, <2% in barley, and <0.6% in tomato, rice, maize, and wheat (results not shown in Fig. 4). Rye, which has the smallest genome, had the highest probability of an essentially derived RI across all thresholds. The probability of an essentially derived RI from the BC_1 was >15% when the threshold was 0.85, and >6% when the threshold was 0.90. The probabilities were lowest in wheat, which has the largest genome. When the threshold was 0.85, the probability of an essentially derived RI from the BC_1 was about five times lower in wheat than in rye. The probabilities in wheat were <0.1% when the threshold was 0.70 among F_2-derived RIs, and <0.4% when the threshold was 0.90 among BC_1-derived RIs (results not shown in Fig. 4). The probabilities of essentially derived RIs in barley, tomato, rice, and maize were intermediate to those in rye and wheat. In maize, the probability in the BC_1 was 6 to 9% when the threshold was 0.85, and 1 to 2% when the threshold was 0.90. These results suggest that the thresholds used to declare essential derivation should differ among species.

Robustness of the Model

The probabilities we calculated depended on four assumptions: (i) the linkage maps we used are representative of the crop species; (ii) the marker loci are evenly spaced along a chromosome; (iii) p_{F2} and p_{BC1} among RIs follow a normal distribution, and (iv) selection is absent. The probabilities of obtaining essentially derived RIs are only as good as the linkage maps from which they were determined. Different linkage maps for the same species vary in the estimated size of each chromosome. If a linkage map differs substantially from the map we used, then the probabilities of obtaining essentially derived RIs may need to be recalculated.

The assumption of evenly spaced markers may not be met in practice. The number of marker loci that minimized $SE(p_{F2})$ was 11 for a 150 cM chromosome, with a corresponding distance of 15 cM between adjacent marker loci and $SE({}_kp_{F2})$ of 0.2586. Suppose the distance between marker loci is 10, 15, or 20 cM instead of a constant 15 cM. If the marker loci are at the 0, 10, 25, 45, 55, 70, 90, 100, 115, 135, and 150 cM positions, the $SE({}_kp_{F2})$ increases only slightly from 0.2586 to 0.2587. When the distance between marker loci is 5, 15, or 25 cM, $SE({}_kp_{F2})$ increases to 0.2609. Hence, the formulas for $SE({}_kp_{F2})$ and $SE({}_kp_{BC1})$ are robust with regards to variation in the distance between marker loci.

In simulation studies, we found that p_{F2} among 100 000 F_2-derived RIs closely followed a normal distribution (results not shown). For BC_1-derived RIs, the p_{BC1} values approached a normal distribution as the number of chromosomes and marker loci increased. Suppose the genome comprises only $n = 4$ chromosomes, each 80 cM long and each with five marker loci. When the threshold was 0.85, the probability of obtaining an essentially derived RI in the simulation study, compared with the probability for a normal distribution (in parentheses), was 0.213 (0.166). The corresponding probability was 0.114 (0.085) when the threshold was 0.90. The probabilities in this paper may therefore underestimate the actual probability of obtaining an essentially derived RI if n and the number of marker loci are small. This discrepancy largely disappeared when the genome comprised $n = 8$ chromosomes, each 140 cM long and each with eight marker loci. The probability of an essentially derived RI was 0.127 (0.126) when the threshold was 0.85, and 0.037 (0.047) when the threshold was 0.90.

The assumption of no selection implies that the mean parental contribution among a set of RIs is 0.50 for F_2-derived RIs and 0.75 for BC_1-derived RIs. We speculate that selection could change the probability of obtaining an RI with p_{F2} or p_{BC1} exceeding the threshold. A comparison of our theoretical results with empirical data on p_{F2} and p_{BC1} among sets of RIs, developed with selection during inbreeding, would be useful.

ACKNOWLEDGMENTS

The China Scholarship Council supported Dr. Jiankang Wang as a visiting scholar at Purdue University. We thank an anonymous reviewer for pointing out a conceptual error in a previous version of this manuscript.

REFERENCES

Bernardo, R., A. Murigneux, J.P. Maisonneuve, C. Johnsson, and Z. Karaman. 1997. RFLP-based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor. Appl. Genet. 94: 652–656.

Causse, M.A., T.M. Fulton, Y.G. Cho, S.N. Ahn, J. Chunwongse, K. Wu, J. Xiao, Z. Yu, P.C. Ronald, S.E. Harrington, G. Second, S.R. McCouch, and S.D. Tanksley. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274.

Dillmann, C., A. Charcosset, A. Bar-Hen, B. Goffinet, J.S.C. Smith, Y. Dattee, and J. Guiard. 1995. The estimation of molecular genetic distances in maize for DUS and ED protocols: Optimisation of the information and new approaches of kinship. *In* 3rd session, Working group on biochemical and molecular techniques and DNA profiling in particular, Wageningen, Netherlands. 19–21 Sept. 1995. Union Internationale pour la Protection des Obtentions Végétales, Geneva, Switzerland.

Gale, M.D., M.D. Atkinson, C.N. Chinoy, R.L. Harcourt, J. Jia, Q.Y. Li, and K.M. Devos. 1995. Genetic maps of hexaploid wheat. p. 29–40. *In* Z.S. Li and Z.Y. Xin (ed.) Proc. Intl. Wheat Genet. Symp., 8th. Beijing, China. 20–25 July 1993. China Agric. Scientech Press, Beijing, China.

Haldane, J.B.S., C.H. Waddington. 1931. Inbreeding and linkage. Genetics 16:357–374.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. Ann. Eugen. 12:172–175.

Langridge, P., A. Karakousis, N. Collins, J. Kretchmer, and S. Manning. 1995. A consensus linkage map of barley. Mol. Breed. 1:389–395.

Lorenzen, L.L., S. Boutin, N. Young, J.E. Specht, and R.C. Shoe-

maker. 1995. Soybean pedigree analysis using map-based molecular markers. 1. Tracking RFLP markers in cultivars. Crop Sci. 35: 1326–1336.
Senior, M.L., E.C.L. Chin, M. Lee, J.S.C. Smith, and C.W. Stuber. 1996. Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. Crop Sci. 36:1676–1683.
Smith, J.S.C., D.S. Ertl, and B.A. Orman. 1995 Identification of maize varieties. p. 253–264 *In* C.W. Wrigley (ed.) Identification of food-grain varieties. Am. Assoc. Cereal Chemists, St. Paul, MN.
Tanksley, S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, et al. 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160.
Wanous, M., O. Heredia-Diaz, X. Ma, P. Goicoechea, G. Wricke, E. Ferrer, and J.P. Gustafson. 1997. Progress of molecular and cytological mapping in rye (*Secale cereale* L.). p. 121–122 *In* P.E. McGuire and C.O. Qualset (ed.) Progress in genome mapping of wheat and related species: Joint Proc. 5th and 6th Public Workshops of the Intl. Triticeae Mapping Initiative, Norwich, UK. 1–3 Sept. 1995, and Sydney, Australia, 30–31 Aug. 1996. Univ. of California Genet. Resources Conservation Program, Report No. 18. Davis, CA.

Reaction of Soybean Cultivars to Sclerotinia Stem Rot in Field, Greenhouse, and Laboratory Evaluations

H. S. Kim, G. L. Hartman, J. B. Manandhar, G. L. Graef, J. R. Steadman, and B. W. Diers*

ABSTRACT

Sclerotinia stem rot of soybean [*Glycine max* (L.) Merr.], caused by the fungal pathogen *Sclerotinia sclerotiorum* (Lib.) de Bary, recently has increased in importance in the northern U.S. soybean production area. The objective of our study was to determine the effectiveness of three different inoculation techniques in predicting the field reactions of cultivars to sclerotinia stem rot. Eighteen soybean cultivars were field tested in six Michigan environments from 1994 to 1996 and tested in the greenhouse or laboratory with three inoculation methods. The cultivars were inoculated by placing infested oat (*Avena sativa* L.) seed or mycelial plugs on cotyledons or by placing mycelial plugs on detached leaves. There were significant ($P < 0.05$) differences in resistance to sclerotinia stem rot among cultivars at all but one field environment and for all inoculation methods. The disease severity ratings based on the inoculations were significantly correlated with the field results, with the exception of one method. Disease severity ratings for the three inoculation methods were significantly correlated with only two exceptions. Cultivars such as Novartis S19-90 and Corsoy 79 consistently had the lowest disease severity ratings in the field tests and for the inoculation methods. Similarly, a number of cultivars were rated as susceptible in all tests. Ratings for cultivars with intermediate reactions were not consistent across tests. The inoculation methods tested can provide some useful information on the resistance of soybean genotypes to sclerotinia stem rot. However, resistance identified by inoculation methods should be confirmed with field tests, since these methods can misclassify the resistance of some cultivars.

Sclerotinia stem rot (syn. white mold) of soybean is caused by the fungal pathogen *Sclerotinia sclerotiorum* (Lib.) de Bary (Grau and Hartman, 1999). This disease has recently increased in importance in the northern USA, and breeding for resistance has become an objective for many soybean cultivar development programs. Soybean cultivars have been evaluated for resistance to sclerotinia stem rot under field conditions and some with partial resistance to the disease have been identified (Grau et al., 1982; Boland and Hall, 1987; Nelson et al., 1991; Kim et al., 1999). Although researchers have been successful in identifying partial resistance using field evaluations, these evaluations are difficult because of the need for a cool, wet environment for disease development and the high spatial variability of disease foci across fields. For these reasons, researchers would benefit from having a controlled-environment screening method that accurately predicts the reaction of soybean germplasm in field environments.

Both physiological resistance and escape mechanisms contribute to differences in the reaction of cultivars to sclerotinia stem rot in field trials. Escape mechanisms include early flowering and maturity, less lodging, and an upright, open canopy. One or more of these mechanisms have been shown to be significantly associated with reduced levels of sclerotinia stem rot in several studies (Boland and Hall, 1987; Nelson et al., 1991; Kim et al., 1999; Kim and Diers, 2000). Kim and Diers (2000) found genetic evidence of both escape mechanisms and physiological resistance. In a population derived from a cross between Novartis S19-90 by 'Williams 82', they mapped three quantitative trait loci (QTL) controlling sclerotinia stem rot resistance. Two of these loci were significantly associated with flowering date or plant height and lodging, indicating these loci contribute to resistance through disease escape. The third QTL was not associated with escape mechanisms, indicating it may be a gene contributing to physiological resistance to the disease.

Several research groups have developed inoculation techniques for evaluating soybeans for resistance to sclerotinia stem rot. Chun et al. (1987) and Nelson et al. (1991) inoculated excised stems with *S. sclerotiorum* mycelium and measured the length of the lesions that developed. Although both groups observed significant differences among cultivars for lesion length, these re-

H.S. Kim, Wheat and Barley Research Division, National Crop Exp. Stn., Suwon 441-100, Korea; G.L. Hartman, USDA-ARS and Dep. of Crop Sciences, University of Illinois, Urbana, IL 61801; J.B. Manandhar and B.W. Diers, Dep. of Crop Sciences, Univ. of Illinois, Urbana, IL 61801; G.L. Graef, Dep. of Agronomy, Univ. of Nebraska, Lincoln, NE 68583; and J.R. Steadman, Dep. of Plant Pathology, Univ. of Nebraska, Lincoln, NE 68583. Research is a joint contribution from the Michigan Agricultural Experiment Station, the Nebraska Agric. Exp. Stn., Nebraska Journal Series No. 12824. The research was supported in part by grants from the Michigan Soybean Promotion Committee and the Nebraska Soybean Board. Received 19 July 1999. *Corresponding author (bdiers@uiuc.edu).

Published in Crop Sci. 40:665–669 (2000).

Abbreviations: AUDPC, area under the disease progress curve; DSI, disease severity index; MSU, Michigan State University; PDA, potato-dextrose agar; QTL, quantitative trait loci; UIUC, University of Illinois at Urbana-Champaign.